UNIVERSIDADE FEDERAL DE VIÇOSA

Efeitos da filtragem ambiental e de uma seca extrema experimental na diversidade de formigas em florestas tropicais

Isadora Gerheim de Vasconcellos Moreira
Doctor Scientiae

VIÇOSA - MINAS GERAIS
2024

ISADORA GERHEIM DE VASCONCELLOS MOREIRA

Efeitos da filtragem ambiental e de uma seca extrema experimental na diversidade de formigas em florestas tropicais

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ecologia, para obtenção do título de *Doctor Scientiae*.

Orientador: Ricardo R. de Castro Solar

VIÇOSA - MINAS GERAIS
2024

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
M838e
2024

Moreira, Isadora Gerheim de Vasconcellos, 1990-
Efeitos da filtragem ambiental e de uma seca extrema experimental na diversidadede formigas em florestas tropicais / Isadora Gerheim de Vasconcellos Moreira. – Viçosa, MG, 2024.
1 tese eletrônica (361 f.): il. (algumas color.).

Texto em português e inglês.
Inclui apêndices.
Orientador: Ricardo Ribeiro de Castro Solar.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Biologia Geral, 2024.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2025.084
Modo de acesso: World Wide Web.

1. Formigas - Mata Atlântica. 2. Formigas - Amazônia. 3. Formigas - Efeito das mudanças climáticas. 4. Formigas - Efeito da seca. 5. Formigas - Efeito do uso do solo. I. Solar, Ricardo Ribeiro de Castro, 1985-. II. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-Graduação em Ecologia. III. Título.

CDD 22. ed. 595.7961782

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317

ISADORA GERHEIM DE VASCONCELLOS MOREIRA

Efeitos da filtragem ambiental e de uma seca extrema experimental na diversidade de formigas em florestas tropicais

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ecologia, para obtenção do título de *Doctor Scientiae*.

APROVADA: 24 de outubro de 2024.

Assentimento:

_____

Isadora Gerheim de Vasconcellos Moreira
Autora

_____

Ricardo Ribeiro de Castro Solar
Orientador

AGRADECIMENTOS

Gostaria primeiramente de agradecer à minha família, Beatriz, Maria, Dudu, Eduardo & Giovana. Sem vocês nada disso teria sido possível. Obrigada por me incentivarem a seguir meus sonhos e minha realização profissional, mesmo que o caminho tenha parecido tão tortuoso às vezes. Obrigada pelo carinho e zelo ao construírem nosso lar, que sempre será minha referência quando preciso de descanso e de “voltar pra casa”. Obrigada por terem criado tantas tradições afetivas que hoje fazem parte do meu dia a dia e viraram um alento nos dias difíceis.

Gostaria de agradecer à Larissa, Paola, Alicia e Beta pela linda amizade que criamos. Agradeço especialmente Larissa e Paola por terem me inspirado a sair da minha zona de conforto profissional e por causa disso, cresci muito e conheci minha força. Sem vocês essa jornada não teria sido tão divertida. Gostaria também de agradecer aos amigos do CSEC pelas vivências e partilhas. O que tem sido construído no CSEC é muito especial e fico muito feliz de ter feito parte dessa experiência.

Queria agradecer imensamente ao Bob pela orientação durante esses anos. Por ter entendido e me dado liberdade de levar essa jornada do meu jeito. Essa liberdade foi um grande ato de respeito e foi essencial para que eu construísse a confiança que eu tenho hoje no meu trabalho. Agradeço por todas as oportunidades que você me deu ao longo desses anos. Você foi um excelente orientador. Agradeço também por ter me dado um dos maiores aprendizados e que eu tento levar para vida: as coisas às vezes são mais simples do que parecem.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

“Os filósofos não fizeram mais do que interpretar de diversos modos o mundo; trata-se, agora, de transformá-lo.”
(Karl Marx)

RESUMO

MOREIRA, Isadora Gerheim de Vasconcellos, D.Sc., Universidade Federal de Viçosa, outubro de 2024. **Efeitos da filtragem ambiental e de uma seca extrema experimental na diversidade de formigas em florestas tropicais**. Orientador: Ricardo Ribeiro de Castro Solar.

A Amazônia e a Mata Atlântica constituem dois dos mais importantes centros de biodiversidade do mundo. As comunidades de formigas contribuem para a resiliência e estabilidade dessas florestas, e a presença dessas espécies pode ser influenciada por diversos fatores ambientais através do processo de filtragem ecológica. O primeiro capítulo desta tese tem como objetivo investigar a influência de fatores ambientais na diversidade de formigas da Mata Atlântica. Embora a filtragem ecológica constitua um processo natural e esperado, afetando vários aspectos da biodiversidade, ela está sujeita a modificações por atividades antropogênicas, como eventos de seca, que são previstos de se tornarem mais frequentes e prolongados na região amazônica. Portanto, no segundo capítulo desta tese, investigamos se esses eventos de seca extrema levam a mudanças na diversidade filogenética da comunidade de formigas da Floresta Amazônica. Para o primeiro capítulo, usamos um conjunto de dados de 63 locais de amostragem ao longo da Mata Atlântica para explorar a influência de 31 variáveis ambientais na composição da comunidade de formigas. Para o segundo capítulo, extraímos métricas filogenéticas de comunidades de formigas amostradas em uma área de seca contínua induzida na Floresta Amazônica. Nossos resultados demonstram que, enquanto as variáveis climáticas explicam parte da distribuição espacial da diversidade de formigas da Mata Atlântica, são fatores edáficos que emergem como a influência mais dominante, mas o efeito sinérgico supera a influência de qualquer uma isoladamente. A precipitação durante o mês mais seco e o teor de argila são os fatores com maior importância na distribuição da composição de formigas. Portanto, tanto as mudanças climáticas quanto as atividades de uso da terra representam uma ameaça à diversidade de formigas. Nossos resultados demonstram que a seca reduziu a riqueza filogenética de todas as subfamílias, mas não selecionou espécies mais próximas filogeneticamente. Embora tenhamos observado uma diminuição geral na riqueza e abundância de espécies, houve uma ampla variedade de respostas, inclusive dentro de um mesmo gênero. As métricas filogenéticas da comunidade de controle exibiram variação temporal gradual, enquanto a comunidade sob seca apresentou picos, oscilando entre valores baixos e altos, o que pode ser atribuído à amostragem

estocástica como consequência da perda de espécies. Essas mudanças imprevisíveis poderiam levar à diminuição da resiliência e desestabilizar a estrutura e função da comunidade. Devido à distribuição variada de respostas ao longo da filogenia, sugerimos que a hipótese de conservadorismo de nicho (a tendência de espécies filogeneticamente próximas responderem de maneira semelhante às pressões ambientais) não se aplica a essas comunidades de formigas e pode não ser um preditor confiável de respostas a eventos de seca para as formigas amazônicas. Nesta tese nós abordamos os efeitos de fatores ambientais naturais e antropogênicos em escalas macro e locais sobre as formigas da Amazônia e Mata Atlântica, demonstrando como as mudanças climáticas e as atividades de uso da terra representam ameaças a biodiversidade das formigas e sugerimos que o conservadorismo de nicho pode não auxiliar a predição de respostas dessas comunidades a eventos de seca.

Palavras-chave: Comunidades de formigas; Mudanças climáticas ; Beta diversidade ; Diversidade das florestas tropicais; Macroecologia; Diversidade taxonômica; Diversidade Filogenética

ABSTRACT

MOREIRA, Isadora Gerheim de Vasconcellos, D.Sc., Universidade Federal de Viçosa, October, 2024. **Effects of environmental filtering and an experimental extreme drought on ant diversity in tropical forests**. Adviser: Ricardo Ribeiro de Castro Solar.

The Amazon and the Atlantic Forest constitute two of the most important centers of biodiversity in the world. The ant community contributes to the resilience and stability of these forests and the presence of these ant species can be influenced by various environmental factors through the process of ecological filtering. The first chapter of this thesis aims to investigate the influence of environmental factors on the diversity of ants in the Atlantic Forest. While ecological filtering constitutes a natural and expected process, affecting several aspects of biodiversity, it is subject to modification by anthropogenic activities such as drought events, which are expected to be more frequent and long in the Amazon region. Therefore, in the second chapter of this thesis, we analyze data from a drought experiment in the Amazon Forest to investigate if drought events lead to changes in phylogenetic composition of the ant community. For the first chapter we used a dataset of 63 sample sites along the Atlantic Forest to explore the influence of 31 environmental variables on ant community composition. For the second chapter, we extracted phylogenetic metrics from ant communities sampled within the context of a induced drought experiment in Amazonian rainforest. We found that that while climatic variables explain part of the spatial distribution of ant diversity in the Atlantic Forest, it is the soil characteristics that emerge as the more dominant influence, but the synergistic effect surpasses the influence of either one alone. Precipitation during the driest month and the clay content has the most importance in ant composition distribution. Therefore, both climate change and land use activities pose a threat to ant diversity. We found that, as a consequence of the drought, all subfamilies presented a decrease in phylogenetic richness, but it did not select for species with close phylogenetic relationship. Although we observed a general decrease in species richness and abundance, there was also a wide array of responses, including within the same genus. The phylogenetic metrics of the control community exhibited gradual temporal variation, while drought community displayed peaks, fluctuating between low and high values, which can be attributable to stochastic sampling as consequence of species loss. These unpredictable changes could lead to decreased ecosystem resilience and further destabilize the community structure and function. Due to the varied distribution of responses

throughout the phylogeny, we suggest that the hypothesis of niche conservatism (the tendency of closely related species to respond similarly to environmental pressures) does not apply to these ant communities and might not be a reliable predictor of responses to drought events for the Amazonian ants. In this thesis we addressed the effects of natural and human-induced climatic and soil factors at macro and local scales on ant from Amazon and Atlantic Forests, demonstrating how climate change and land use activities pose significant threats to various aspects of ant biodiversity and suggesting that niche conservatism might not help to predict the response of these communities to extreme drought events.

Keywords: Ant communities ; Climate change; Beta diversity ; Rainforest diversity ; Macroecology; Taxonomic diversity; Phylogenetic diversity

SUMÁRIO

1 INTRODUÇÃO GERAL

Durante o Paleogeno, as florestas tropicais das regiões norte, nordeste e leste da América do Sul eram espacialmente contínuas e conectadas, compondo uma única floresta (Costa et al., 2017; Morley, 2000; Sobral-Souza et al., 2015). No entanto, grandes mudanças geológicas e climáticas, incluindo o surgimento de uma região seca, levaram à separação em duas regiões de floresta tropical: a Floresta Amazônia, situada nas regiões norte e noroeste da América do Sul e a Mata Atlântica, predominantemente cobrindo a costa leste do continente (Costa, 2003; Hoorn et al., 2010). Ao longo de milhões de anos de evolução, essas duas florestas se desenvolveram e tornaram-se dois dos mais importantes centros de biodiversidade do mundo (Myers et al., 2000; Ribeiro et al., 2009).

A Floresta Amazônica é a maior floresta tropical do mundo, abrangendo 60% do território brasileiro e 40% das florestas tropicais remanescentes do mundo (Kauano et al., 2017). É considerada a principal fonte de diversidade e um centro de dispersão para organismos terrestres neotropicais, atuando como um repositório global de serviços ecossistêmicos (Bandopadhyay e Sánchez, 2020; Rodrigues et al., 2013). Além disso, a região amazônica é conhecida por suas características físico-químicas, incluindo a qualidade de suas águas superficiais (do Nascimento Monte et al., 2021), desempenhando um papel crucial na manutenção da biodiversidade, dos serviços ecossistêmicos e dos ciclos hídricos do continente sul-americano (Kauano et al., 2017). Apesar de sua importância, ela enfrenta desafios, como o uso ilegal de recursos naturais, devido à expansão da agricultura e pecuária, que causam desmatamento ilegal e incêndios florestais (Barlow et al., 2016; DeFries et al., 2008; Gollnow et al., 2022; Morton et al., 2006; Silva et al., 2021).

A Mata Atlântica é conhecida mundialmente por abrigar espécies únicas, com mais de 8.000 espécies endêmicas registradas (Tabarelli et al., 2005). No entanto, a industrialização maciça e a expansão agrícola fragmentaram sua área e atualmente apenas 11,6% da sua cobertura florestal original permanece, estando distribuída em um mosaico de pequenos fragmentos desconectados (Joly et al., 2014; Ribeiro et al., 2009; Scarano e Ceotto, 2015). Como decorrência, ela é considerada um *hotspot* de biodiversidade de alta prioridade (Jenkins et al., 2015), sendo listada entre os 25 *hotspots* de conservação prioritários globais (Myers et al., 2000).

Um dos fatores que sustentam essas florestas é a diversidade de espécies que as habitam, incluindo as formigas. As formigas são um grupo-chave dominante que ocorre na maior parte da superfície terrestre (Ward et al., 2015), mas dois terços de sua abundância estão concentrados em ecossistemas tropicais (Schultheiss et al., 2022). Elas desempenham vários papéis ecossistêmicos, atuando como predadores, necrófagos, granívoros, herbívoros, onívoros, coletoras de sementes, na bioturbação do solo e ajudando na sequestração de carbono do solo (Del Toro et al., 2012; Dorn, 2014; Lach et al., 2010; Parr et al., 2016). Além disso, as formigas estão envolvidas em relações mutualísticas com espécies de plantas, atuando eventualmente na polinização, mas principalmente na proteção contra herbívoros (Brian e Beattie, 1986; Oliver et al., 2008). No geral, a presença e as atividades das formigas contribuem para a estabilidade e resiliência das florestas tropicais (Del Toro et al., 2012; Dorn, 2014; Lach et al., 2010; Parr et al., 2016).

A presença e distribuição das espécies de formigas nas florestas tropicais são influenciadas por diversos fatores, como condições ambientais e impacto das atividades humanas na região. Durante o processo de montagem da comunidade, certas condições ambientais podem atuar na seleção de espécies com base em características que lhes permitem se estabelecer e permanecer em um habitat específico, enquanto excluem aquelas menos

adaptadas àquelas condições - um processo conhecido como filtragem ecológica (Bazzaz, 1991; Cornwell e Ackerly, 2009; Kraft et al., 2014). A força desse processo varia com a escala espacial, é influenciada por outros processos, como interações bióticas, história evolutiva e processos neutros (Chesson, 2000; Snyder e Chesson, 2004; Cavender-Bares et al., 2006; Biswas et al., 2015), mas pode ser um dos principais fatores que moldam a distribuição das espécies (Engelbrecht et al., 2007; John et al., 2007). Enquanto a filtragem ecológica constitui um processo natural e esperado, ela está cada vez mais sujeita a modificações por atividades antropogênicas e suas consequências, como mudanças climáticas, atividades agrícolas, pecuária e mineração (Hufbauer et al., 2011; Tuomainen e Candolin, 2011). Essas alterações rápidas nas dinâmicas de filtragem ecológica deixam as comunidades mais suscetíveis a mudanças sem precedentes, potencialmente sem a janela temporal para adaptação (Tuomainen e Candolin, 2011; Otto 2018; Wood et al., 2021).

Entender como as condições ambientais impactam as comunidades de formigas é importante para prever como alterações no ambiente, como aquecimento global e mudanças no uso da terra, influenciam as comunidades de formigas e, como consequência, os processos ecossistêmicos nos quais elas estão envolvidas (Wiescher et al., 2012). Por exemplo, fatores como temperatura e precipitação desempenham um papel significativo na formação dessas comunidades (Warren e Chick, 2013). Estudos indicam que em florestas tropicais a abundância e a atividade de vários gêneros de formigas são afetadas por sua tolerância térmica e pela temperatura ambiente (Boyle et al., 2021). Mudanças nos padrões de temperatura e precipitação também podem influenciar o habitat e a disponibilidade de recursos para as formigas, afetando sua abundância, distribuição e comportamento (Oliveira et al., 2017). Alguns estudos têm sugerido cada vez mais que fatores climáticos desempenham um papel importante na formação das comunidades de formigas (Dunn et al., 2009; Holway et al., 2002; Jenkins et al., 2011; Kaspari et al., 2000; Perillo et al., 2021; Vasconcelos et al., 2010).

Além da importância das condições climáticas, dado que as formigas interagem diretamente com o solo e suas propriedades, há um grande potencial para fatores edáficos (como textura, umidade e conteúdo de nutrientes no solo) influenciarem essas comunidades (Campbell e Crist, 2017; da Costa-Milanez et al., 2017), mas essa relação tem sido menos investigada. Para aumentar nosso conhecimento sobre o impacto dos fatores ambientais na diversidade de formigas na Mata Atlântica, investigamos no primeiro capítulo desta tese quais fatores climáticos e edáficos mais influenciam a similaridade na composição de espécies dessas comunidades e como eles afetam sua distribuição espacial. Ao explorar como a diversidade varia ao longo de gradientes espaciais e ambientais, podemos melhorar nossa compreensão dos fatores que moldam os padrões de biodiversidade, auxiliando na identificação das principais ameaças à biodiversidade de formigas na Mata Atlântica.

Além do processo natural de montagem da comunidade, outros processos podem interagir, introduzindo novas variáveis ou interrompendo os ciclos ambientais naturais. Por exemplo, o aumento na queima de combustíveis fósseis nos últimos 60 anos está causando mudanças climáticas e tem sido considerado uma das principais causas das mudanças na biodiversidade, afetando diretamente a sobrevivência e persistência das espécies (Hooper et al., 2012; IPCC/AR6, 2012; Lade, 2020; Rockström et al., 2018; Steffen et al., 2015; Urban et al., 2016). A projeção climática para a América do Sul sugere que um aumento de 1,5°C e 2°C na temperatura aumentará a duração e a frequência das secas em 47% e 79%, respectivamente (Cook e Vizy, 2008; IPCC, 2012; Xu et al., 2019), o que poderia perturbar a dinâmica das florestas tropicais (Flores e Staal, 2022). Esses eventos extremos de seca podem provocar mudanças significativas nos ciclos climáticos das florestas tropicais, levando a alterações na composição da comunidade e impactando vários aspectos da biodiversidade, como as diversidades taxonômica, funcional e filogenética.

A diversidade filogenética reflete a história evolutiva e a relação entre os táxons (Grab et al., 2019; Pio et al., 2011; Webb et al., 2002). Ela captura semelhanças nas respostas ao ambiente entre os clados, o que pode influenciar os padrões de perda de espécies e a dinâmica da comunidade durante eventos de seca (Pio et al., 2011; Webb et al., 2002). Além disso, influencia indiretamente outros aspectos da biodiversidade, como estabilidade da comunidade, diversidade funcional, propensão à invasão e potencial de resistência às mudanças climáticas através da adaptação (Bononi et al., 2018; Cadotte et al., 2012; Carroll e Fox, 2008; El-Barougy et al., 2020; Hoffmann e Sgrò, 2011). Compreender a estrutura filogenética das comunidades pode fornecer novos conhecimentos sobre como a seca afeta a persistência das espécies e clados filogenéticos e suas implicações mais amplas para a biodiversidade. Portanto, no segundo capítulo desta tese, analisamos dados de um experimento de seca contínua induzida por 10 anos na Floresta Amazônica para investigar o impacto desse evento extremo na estrutura filogenética de uma comunidade de formigas. Analisando a estrutura filogenética da comunidade de formigas, podemos investigar se os eventos de seca afetam os clados de maneira diferente, levando a mudanças na composição filogenética da comunidade.

Referências

ALMEIDA, R. P. S. et al. Induced drought strongly affects richness and composition of ground-dwelling ants in the eastern Amazon. **Oecologia**, v. 201, n. 2, 11 ago. 2020.

BANDOPADHYAY, S.; SÁNCHEZ, D. A. C. Amazonian Fire Events Disturbed the Global Carbon Cycle: A Study from 2019 Amazon Wildfire Using Google Earth Engine. **Environmental Sciences Proceedings**, v. 3, n. 1, p. 43, 2020.

BARLOW, J. et al. Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. **Nature**, v. 535, n. 7610, p. 144–147, 29 jun. 2016.

BONONI, L. et al. Higher phylogenetic diversity prevents loss of functional diversity caused by successive drying and rewetting cycles. **Antonie van Leeuwenhoek**, v. 111, n. 7, p. 1033–1045, 29 dez. 2017.

BOYLE, M. J. W. et al. Localised climate change defines ant communities in human-modified tropical landscapes. **Functional Ecology**, v. 35, n. 5, p. 1094–1108, 11 jan. 2021.

BRIAN, M. V.; BEATTIE, A. J. The Evolutionary Ecology of Ant-Plant Mutualisms. **The Journal of Ecology**, v. 74, n. 4, p. 1213, dez. 1986.

CADOTTE, M. W.; DINNAGE, R.; TILMAN, D. Phylogenetic diversity promotes ecosystem stability. **Ecology**, v. 93, n. sp8, p. S223–S233, ago. 2012.

CAMPBELL, K. U.; CRIST, T. O. Ant species assembly in constructed grasslands is structured at patch and landscape levels. **Insect Conservation and Diversity**, v. 10, n. 2, p. 180–191, 5 jan. 2017.

CARROLL, S. P.; FOX, C. W. **Conservation biology : evolution in action**. [s.l.] Oxford University Press, 2008.

COOK, K. H.; VIZY, E. K. Effects of Twenty-First-Century Climate Change on the Amazon Rain Forest. **Journal of Climate**, v. 21, n. 3, p. 542–560, fev. 2008.

COSTA, G. C. et al. Biome stability in South America over the last 30 kyr: Inferences from long-term vegetation dynamics and habitat modelling. **Global Ecology and Biogeography**, v. 27, n. 3, p. 285–297, 1 mar. 2018.

COSTA, L. P. The historical bridge between the Amazon and the Atlantic Forest of Brazil: a study of molecular phylogeography with small mammals. **Journal of Biogeography**, v. 30, n. 1, p. 71–86, jan. 2003.

COSTA-MILANEZ, C. B. DA et al. Influence of soil granulometry on average body size in soil ant assemblages: implications for bioindication. **Perspectives in Ecology and Conservation**, v. 15, n. 2, p. 102–108, 1 abr. 2017.

DEFRIES, R. S. et al. Fire-related carbon emissions from land use transitions in southern Amazonia. **Geophysical Research Letters**, v. 35, n. 22, 25 nov. 2008.

DEL TORO, I. et al. The little things that run the world revisited: a review of ant-mediated ecosystem services and disservices (Hymenoptera: Formicidae). **Myrmecol. News**, v. 17, n. 10, p. 133–146, 2012.

DELSINNE, T.; ARIAS-PENNA, T.; LEPONCE, M. Effect of rainfall exclusion on ant assemblages in montane rainforests of Ecuador. **Basic and Applied Ecology**, v. 14, n. 4, p. 357–365, jun. 2013.

DO NASCIMENTO MONTE, C. et al. The physical-chemical characteristics of surface waters in the management of quality in clearwater rivers in the Brazilian Amazon. **Water Policy**, v. 23, n. 5, p. 1303–1313, 14 set. 2021.

DORN, R. I. Ants as a powerful biotic agent of olivine and plagioclase dissolution. **Geology**, v. 42, n. 9, p. 771–774, set. 2014.

DUNN, R. R. et al. Climatic drivers of hemispheric asymmetry in global patterns of ant species richness. **Ecology Letters**, v. 12, n. 4, p. 324–333, abr. 2009.

EL-BAROUGY, R. et al. Richness, phylogenetic diversity, and abundance all have positive effects on invader performance in an arid ecosystem. **Ecosphere**, v. 11, n. 2, fev. 2020.

FINER, M. et al. Future of oil and gas development in the western Amazon. **Environmental Research Letters**, v. 10, n. 2, p. 024003, 28 jan. 2015.

FLORES, B. M.; STAAL, A. Feedback in tropical forests of the Anthropocene. **Global Change Biology**, v. 28, n. 17, p. 5041–5061, 30 jun. 2022.

GOLLNOW, F. et al. Gaps in adoption and implementation limit the current and potential effectiveness of zero-deforestation supply chain policies for soy. **Environmental Research Letters**, v. 17, n. 11, p. 114003, 28 out. 2022.

GRAB, H. et al. Agriculturally dominated landscapes reduce bee phylogenetic diversity and pollination services. **Science**, v. 363, n. 6424, p. 282–284, 17 jan. 2019.

HOFFMANN, A. A.; SGRÒ, C. M. Climate change and evolutionary adaptation. **Nature**, v. 470, n. 7335, p. 479–485, 23 fev. 2011.

HOLWAY, D. A.; SUAREZ, A. V.; CASE, T. J. Role of Abiotic Factors in Governing Susceptibility to Invasion: A Test with Argentine Ants. **Ecology**, v. 83, n. 6, p. 1610–1610, 1 jun. 2002.

HOOPER, D. U. et al. A Global Synthesis Reveals Biodiversity Loss as a Major Driver of Ecosystem Change. **Nature**, v. 486, n. 7401, p. 105–108, 2 maio 2012.

HOORN, C. et al. Amazonia Through Time: Andean Uplift, Climate Change, Landscape Evolution, and Biodiversity. **Science**, v. 330, n. 6006, p. 927–931, 12 nov. 2010.

IPCC. **Managing the risks of extreme events and disasters to advance climate change adaption : special report of the Intergovernmental Panel on Climate Change.** Cambridge: Cambridge University Press, 2012.

JENKINS, C. N. et al. Global diversity in light of climate change: the case of ants. **Diversity and Distributions**, v. 17, n. 4, p. 652–662, 16 maio 2011.

JENKINS, C. N. et al. Patterns of Vertebrate Diversity and Protection in Brazil. **PLOS ONE**, v. 10, n. 12, p. e0145064, 17 dez. 2015.

JOLY, C. A.; METZGER, J. P.; TABARELLI, M. Experiences from the Brazilian Atlantic Forest: ecological findings and conservation initiatives. **New Phytologist**, v. 204, n. 3, p. 459–473, 10 set. 2014.

KASPARI; O'DONNELL; KERCHER. Energy, Density, and Constraints to Species Richness: Ant Assemblages along a Productivity Gradient. **The American Naturalist**, v. 155, n. 2, p. 280, 2000.

KAUANO, É. E.; SILVA, J. M.; MICHALSKI, F. Illegal use of natural resources in federal protected areas of the Brazilian Amazon. **ProQuest**, v. 5, 10 out. 2017.

LACH, L.; PARR, C. L.; ABBOTT, K. L. **Ant ecology**. Oxford: Oxford University Press, 2010.

LADE, S. J. et al. Human impacts on planetary boundaries amplified by Earth system interactions. **Nature Sustainability**, v. 3, n. 2, p. 119–128, 16 dez. 2019.

LUCAS NEVES PERILLO et al. Disentangling the effects of latitudinal and elevational gradients on bee, wasp, and ant diversity in an ancient neotropical mountain range. **Journal of Biogeography**, v. 48, n. 7, p. 1564–1578, 29 mar. 2021.

MORLEY, R. J. Origin and Evolution of Tropical Rain Forests. **PALAIOS**, v. 15, n. 6, p. 580–580, 1 dez. 2000.

MORTON, D. C. et al. Cropland expansion changes deforestation dynamics in the southern Brazilian Amazon. **Proceedings of the National Academy of Sciences**, v. 103, n. 39, p. 14637–14641, 14 set. 2006.

MYERS, N. et al. Biodiversity hotspots for conservation priorities. **Nature**, v. 403, n. 6772, p. 853–858, fev. 2000.

OLIVEIRA, F. M. P. et al. Chronic anthropogenic disturbance as a secondary driver of ant community structure: interactions with soil type in Brazilian Caatinga. **Environmental Conservation**, v. 44, n. 2, p. 115–123, 19 ago. 2016.

OLIVER, T. H.; LEATHER, S. R.; COOK, J. M. Macroevolutionary patterns in the origin of mutualisms involving ants. **Journal of Evolutionary Biology**, v. 21, n. 6, p. 1597–1608, nov. 2008.

PARR, C. L. et al. Suppression of savanna ants alters invertebrate composition and influences key ecosystem processes. **Ecology**, v. 97, n. 6, p. 1611–1617, jun. 2016.

PIO, D. V. et al. Spatial Predictions of Phylogenetic Diversity in Conservation Decision Making. **Conservation Biology**, v. 25, n. 6, p. 1229–1239, 9 nov. 2011.

RIBEIRO, M. C. et al. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. **Biological Conservation**, v. 142, n. 6, p. 1141–1153, jun. 2009.

ROCKSTRÖM, J.; BAI, X.; DEVRIES, B. Global sustainability: the challenge ahead. **Global Sustainability**, v. 1, 2018.

RODRIGUES, J. L. M. et al. Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. **Proceedings of the National Academy of Sciences**, v. 110, n. 3, p. 988–993, 27 dez. 2012.

SCARANO, F. R.; CEOTTO, P. Brazilian Atlantic forest: impact, vulnerability, and adaptation to climate change. **Biodiversity and Conservation**, v. 24, n. 9, p. 2319–2331, set. 2015.

SCHULTHEISS, P. et al. The abundance, biomass, and distribution of ants on Earth. **Proceedings of the National Academy of Sciences**, v. 119, n. 40, 19 set. 2022.

SILVA, C. A. et al. Fire Occurrences and Greenhouse Gas Emissions from Deforestation in the Brazilian Amazon. **Remote Sensing**, v. 13, n. 3, p. 376, 22 jan. 2021.

SOBRAL-SOUZA, T.; LIMA-RIBEIRO, M. S.; SOLFERINI, V. N. Biogeography of Neotropical Rainforests: past connections between Amazon and Atlantic Forest detected by ecological niche modeling. **Evolutionary Ecology**, v. 29, n. 5, p. 643–655, 3 jul. 2015.

STEFFEN, W. et al. Planetary boundaries: Guiding human development on a changing planet. **Science**, v. 347, n. 6223, 15 jan. 2015.

TABARELLI, M. et al. Challenges and Opportunities for Biodiversity Conservation in the Brazilian Atlantic Forest. **Conservation Biology**, v. 19, n. 3, p. 695–700, jun. 2005.

URBAN, M. C. et al. Improving the forecast for biodiversity under climate change. **Science**, v. 353, n. 6304, p. aad8466–aad8466, 8 set. 2016.

VASCONCELOS, H. L. et al. Patterns of ant species diversity and turnover across 2000 km of Amazonian floodplain forest. **Journal of Biogeography**, v. 37, n. 3, p. 432–440, mar. 2010.

WARD, P. S. et al. The evolution of myrmicine ants: phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera: Formicidae). **Systematic Entomology**, v. 40, n. 1, p. 61–81, 23 jul. 2015.

WARREN, R. J.; CHICK, L. Upward ant distribution shift corresponds with minimum, not maximum, temperature tolerance. **Global Change Biology**, v. 19, n. 7, p. 2082–2088, 21 mar. 2013.

WEBB, C. O. et al. Phylogenies and Community Ecology. **Annual Review of Ecology and Systematics**, v. 33, n. 1, p. 475–505, nov. 2002.

WIESCHER, P. T.; PEARCE-DUVET, J. M. C.; FEENER, D. H. Assembling an ant community: species functional traits reflect environmental filtering. **Oecologia**, v. 169, n. 4, p. 1063–1074, 1 fev. 2012.

XU, L.; CHEN, N.; ZHANG, X. Global drought trends under 1.5 and 2 °C warming. **International Journal of Climatology**, v. 39, n. 4, p. 2375–2385, 25 dez. 2019.

2 CAPÍTULO 1

Soil and climatic effects: How do environmental factors shape the ant community similarity in the Atlantic Forest?

Authors: Isadora Gerheim[1]*, Ricardo R. C. Solar[2]

[1] Programa de Pós-Graduação em Ecologia, Universidade Federal de Viçosa (UFV), Viçosa, Minas Gerais, 36570-900 Brazil

[2] Centro de Síntese Ecológica e Conservação, Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Avenida Presidente Antônio Carlos 6627, Belo Horizonte, Minas Gerais, 31270-910 Brazil

*corresponding author

Options for target journal: Journal of Biogeography; Ecography; Diversity and Distributions

2.1 Abstract

The Atlantic Forest, a biodiversity hotspot in Brazil, faces significant threats from anthropogenic activities and climate change, which impact its ecological communities. Our study explores the effects of climatic and edaphic factors on the diversity of ant communities within this biome. Using a comprehensive dataset of 63 sample sites from 2002 to 2015, we investigated the influence of 31 environmental variables on ant community composition. Our results indicate that soil variables, particularly clay content, have a more pronounced influence on ant community composition compared to climatic variables. The Generalized Dissimilarity Modeling (GDM) explained 32.36% of the variation in ant composition, with soil factors alone contributing with 14.28% and climatic factors with 2.77%. The overlap of these variables accounted for 15.31%, highlighting synergistic effects. Our analysis identified that precipitation during the driest month and soil texture (clay and silt content) significantly affected ant diversity, indicating that both temporal aspects and soil structure are crucial in shaping these communities. Our results also suggest the presence of critical thresholds beyond which the relationship between environmental variables and ant community composition changes significantly, suggesting complex interactions between these factors. Our results highlight the potential vulnerability of these ant communities to climate change, as the Atlantic Forest is expected to experience increased temperatures and decreased precipitation. Additionally, land use activities, such as agriculture and urbanization, which alter soil properties, also pose threats to the ant diversity of the Atlantic Forest.

2.2 Introduction

The Atlantic Forest is a Brazilian biodiversity hotspot (Myers et al., 2000), among the most threatened biomes in the world and known for its high levels of species endemism (Jenkins et al., 2015; Joly et al., 2014). It is characterized by dense, mixed semideciduous, and open forests (Galindo-Leal and Câmara, 2003; Zachos and Habel, 2011), covering a wide range in latitude, longitude, altitude, soil and climate conditions (Galindo-Leal and Câmara, 2003), hence presenting high heterogeneity. However, the Atlantic Forest confronts challenges and increasing threats attributed to a long-standing history of anthropogenic activities such as expansion of agriculture, deforestation, urbanization, and intensive exploitation of natural resources, leaving only around 12% of its initial coverage remained (Galindo-Leal and Câmara, 2003). In addition, the Atlantic Forest faces significant challenges under the current global climate change, which will lead to more alterations in climate variables. These alterations encompass the increase in average temperatures, modifications in rainfall patterns, and occurrences of extreme weather events such as extended periods of drought and intense precipitation (Kirilenko and Sedjo, 2007).

The environmental conditions can affect different ecological processes, and one of them is the selection of species based on their traits that allow them to thrive in a particular habitat while excluding those that are less suited to the conditions - a process known as ecological filtering (Bazzaz, 1991; Cornwell and Ackerly, 2009; Kraft et al., 2015). The strength of this process varies with spatial scale, and it is also influenced by other processes, such as biotic interactions, evolutionary history, and neutral processes (Biswas et al., 2016; Cavender-Bares et al., 2006; Chesson, 2000; Snyder and Chesson, 2004), but it can be one of the major processes shaping species distribution (Engelbrecht et al., 2007; John et al., 2007). While ecological

filtering constitutes a natural and expected process, it is increasingly subject to modification by fast-paced anthropogenic activities and its consequences, such as climate change, agricultural, livestock and mining activities (Hufbauer et al., 2012; Tuomainen and Candolin, 2011). These rapid alterations on ecological filtering dynamics leave the communities more susceptible to unprecedented change, potentially without the temporal window for adaptation (Otto, 2018; Tuomainen and Candolin, 2011; Wood et al., 2021).

Among these affected groups, ants play a critical role in maintaining forest ecosystems. By collecting and transporting seeds from and to different localities, ants enhance the spatial distribution, survival, diversity and the regeneration of plant communities (Arnan et al., 2010; Avgar et al., 2008; Zelikova and Breed, 2008). They are also effective decomposers, breaking down organic matter, recycling nutrients back into the soil, and aerating the soil, improving its structure, and facilitating water infiltration (Cammeraat and Risch, 2008; Cerdà and Jurgensen, 2008; Frouz and Jilková, 1976; Griffiths et al., 2021). Ants are also known for contributing to nutrient redistribution by removing food resources from the forest floor to their nests. In a study ants were responsible for 52% of total bait removal, with vertebrates and other invertebrates together accounted for the remaining 48% (Griffiths et al., 2018). When ants and vertebrates were excluded, there was no compensation in bait removal, indicating low functional redundancy between these groups and reinforcing the unique role of ants for the ecosystem functioning (Griffiths et al., 2018). Furthermore, ants are engaged in mutualistic relationships with plant species, assisting in their pollination and protection from herbivores (Brian and Beattie, 1986; Oliver et al., 2008). Overall, the presence and activities of ants contribute to the stability and resilience of forest ecosystems (Del Toro et al., 2012; Dorn, 2014; Lach et al., 2010; Parr et al., 2016)

Understanding how environmental conditions impact ant communities is important for predicting how alterations in the environment, such as global warming and land use changes,

influence ant communities and the ecosystem processes in which they engage. Studies indicate that in tropical forests, the abundance and activity of various ant genera are affected by their thermal tolerance and ambient temperature (Boyle et al., 2021). Also, research conducted in the Atlantic Forest demonstrated that ant species richness was higher during certain seasons, thereby linking climate to ant distribution (Dunn et al., 2009; Holway et al., 2002; Jenkins et al., 2011; Kaspari et al., 2000; Perillo et al., 2021; Vasconcelos et al., 2010). Shifts in temperature and precipitation patterns can also influence the habitat and resource availability for ants, affecting their abundance, distribution, and behavior (Oliveira et al., 2017).

Recent studies have increasingly suggested that climatic factors play a major role in shaping ant communities (Dunn et al., 2009; Holway et al., 2002; Jenkins et al., 2011; Kaspari et al., 2000; Perillo et al., 2021; Vasconcelos et al., 2010). However, given that ants directly interact with the soil and its properties, there is a great potential of edaphic factors (such as texture, moisture, and nutrient content) to influence these communities as well (Campbell and Crist, 2017; da Costa-Milanez et al., 2017). Ant species richness and community composition have been observed to be associated with soil type, soil compaction, and soil pH (Campbell and Crist, 2017; Oliveira et al., 2017; Rocha-Ortega and García-Martínez, 2018), with a higher species richness of ants being correlated with lower soil pH values (Rocha-Ortega and García-Martínez, 2018).

We aim with this study to explore the impact of environmental factors on the diversity of ants in the Atlantic Forest. We specifically aim to identify which climatic and edaphic factors are most important in influencing the similarity in species composition of these communities and how they affect their spatial distribution. By examining how diversity varies across spatial and environmental gradients, we can improve our understanding of the drives of biodiversity patterns, helping to identify the major threats to ant biodiversity in this biome.

2.3 Methods

We used a dataset published by Silva et al. (2021), with 153,818 ant records from 7,636 study locations along the Atlantic Forest, from 1886 to 2020. To avoid sampling and disturbance biases, we filtered the data to consider only studies carried out using 'pitfall' traps and on sites labelled as 'no disturbance'. These sites represent a wide range of forest classifications, including Forest, Seasonal Semi-deciduous Atlantic Forest, Secondary Forest, and variations such as Seasonal Semi-deciduous Submontane Secondary Atlantic Forest and Secondary Atlantic Forest. Considering that analyses at the genus level have been demonstrated to be efficient in predicting variations in richness and assemblage composition detected at the species level and due to naming differences in morphospecies across the various studies, we restricted our analysis to the genus level (Souza et al., 2018).

Many sampling sites contained only a single record of a species, which could lead to challenges in representing spatial variability and species distribution patterns. To address this issue, we aggregated samples that were within a 10 km radius. This criterion was applied to ensure that the data reflected broader ecological patterns and to reduce the risk of spatial gaps caused by sparse sampling in the dataset. Therefore, any difference in similarity between communities within this 10 km radius would not be explained by the environmental variables used. We characterized the climatic and edaphic conditions of these sample units using data from WorldClim version 2.1 and SoilGrid version 2.0. The WorldClim variables include: annual mean temperature, mean diurnal range, isothermality, temperature seasonality, maximum temperature of the warmest period, minimum temperature of the coldest period, annual temperature range, mean temperature of the wettest quarter, mean temperature of the driest quarter, mean temperature of the warmest quarter, mean temperature of the coldest quarter,

annual precipitation, precipitation of the wettest quarter, precipitation of the driest quarter, precipitation seasonality, precipitation of the warmest quarter, precipitation of the coldest quarter, precipitation of the wettest month, precipitation of the driest month, and elevation. The soil characteristics (0-5 cm beneath the surface) included: soil density, cation exchange capacity, volumetric fraction of coarse fragments, clay percentage, nitrogen, organic carbon density, organic carbon stock, pH in water, sand percentage, silt percentage, and organic carbon stock. Originally, the soil data had a spatial resolution of 5 km; however, it was resampled to match the climatic sample unit of 10 km.

To address multicollinearity and investigate non-linear relations between variables, we employed the transformation into orthogonal polynomials, for both climatic and soil variables (Dormann et al., 2013; Elbers, 2020). It results in two sets of outcomes, the linear and the quadratic transformations for each variable analyzed. We employed Pearson's correlation coefficient, and an iterative method based on the Variance Inflation Factor (VIF) to address multicollinearity across the orthogonal polynomials (Dormann et al., 2013). Subsequent analyses were performed using the set of environmental variables that showed a VIF less than 5 and therefore demonstrated low correlation with each other (James et al., 2013). To gauge variations in genus composition across sample units, we constructed a dissimilarity matrix using the Sorensen index, using 'betapart' package (Baselga and Orme, 2012). We used Generalized Dissimilarity Modeling (GDM) to generate the model, using the 'gdm' package (Mokany et al., 2022). We applied Bray-Curtis taxonomic distance to presence-absence data and incorporated geographical distance as a predictive variable. We ran the Generalized Dissimilarity Modeling (GDM) separately for climatic and soil data, and then for a combination of both. This was done to identify which configuration exhibited the highest efficacy, indicated by the percentage of explanation and to identify the unique contribution of each set of variables to the model. To

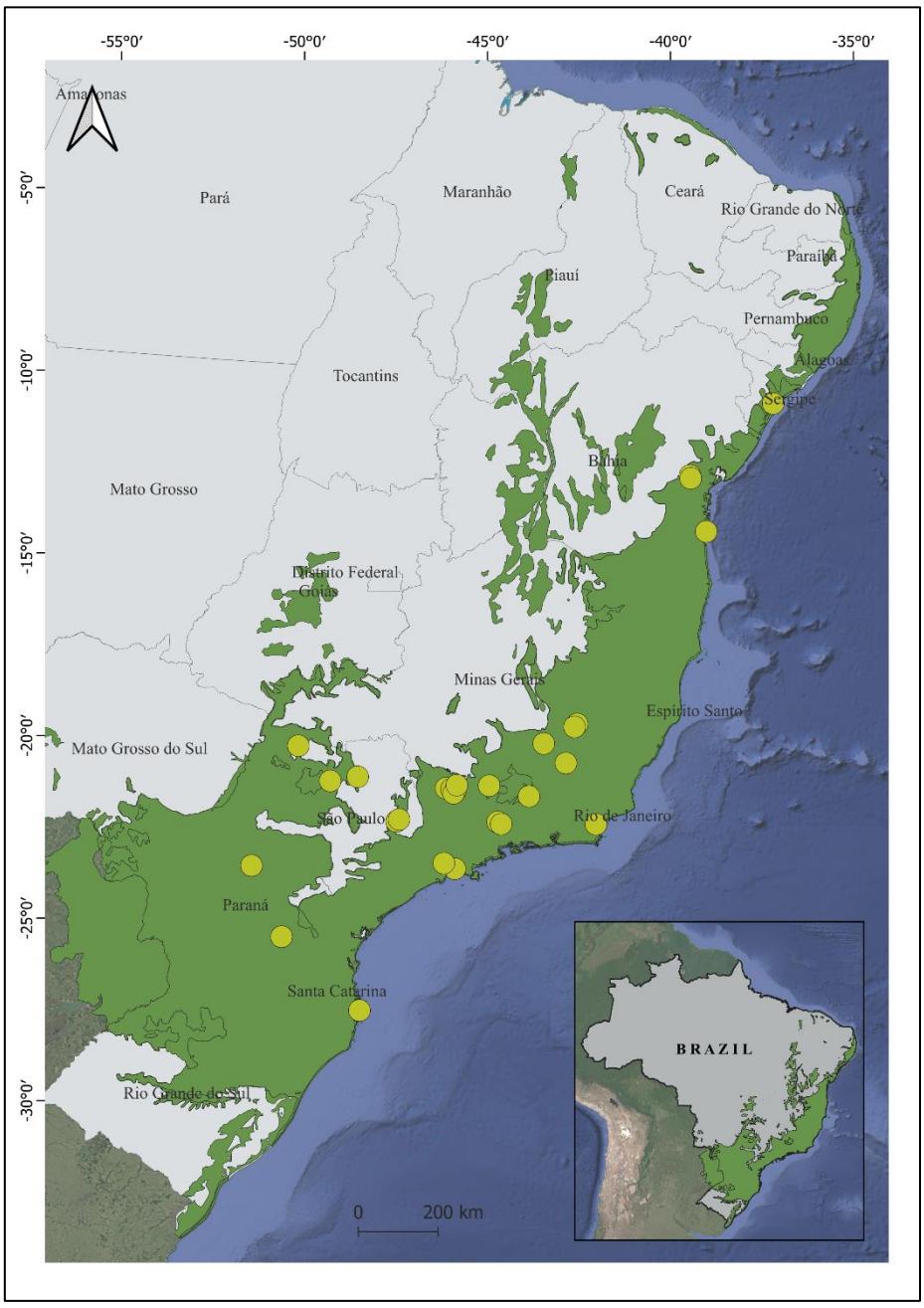
spatially visualize the model, we mapped the first three axes derived from a principal component analysis (PCA) performed on the set of transformed predictor layers.

Once a GDM model has been fitted to the biological data, the model was used to predict the biological uniqueness of each location expected along the Atlantic Forest (Mokany et al., 2022). We transformed geographical and environmental predictors to their respective biological importance, so all these predictors are available as spatial surfaces within a geographical information system (GIS) (Ferrier et al., 2007). Through a subsampling approach, we selected focal and reference pixel, calculating the mean similarity among them (Mokany et al., 2022). Each estimated beta-diversity value was subsequently plotted on a map to visualize the average similarity of that pixel in relation to the entire region, highlight areas of uniqueness diversity (Mokany, 2022). All analyses were performed using R software (R Core Team, 2023).

2.4 Results

The dataset obtained through our filtering process resulted in 63 sample sites, grouped into 28 distinct sampling clusters across Atlantic Forest (Figure 1). Each sampling cluster encompass the sample sites within a 10 km radius, resulting in an average of 2.25 unique study sites per cluster, spanning seven Brazilian states: Sergipe, São Paulo, Santa Catarina, Minas Gerais, Bahia, Paraná, and Rio de Janeiro. Covering a period from 2002 to 2015, the taxa are represented by 11 subfamilies: Formicinae, Myrmicinae, Ponerinae, Pseudomyrmecinae, Dolichoderinae, Dorylinae, Ectatomminae, Paraponerinae, Heteroponerinae, Proceratiinae, and Amblyoponinae, and 74 genera, with the average of 19.54 genera per sampling cluster. For instance, the sampling cluster with the highest site diversity, located in Itatiaia, Rio de Janeiro, has 8 study sites and records 37 unique genera. In contrast, the cluster with the least site diversity, located in São Cristóvão, Sergipe, comprises only 1 site with 10 unique genera.

Figure 1 - Distribution of the 28 ant cluster sites used in this study. The green area represents the Atlantic Forest, and the light green dots are the location of each cluster site.



The Pearson correlation coefficient returned a remarkably wide range in pair correlations, varying from -0.9985 to 0.9952, indicating the risk of multicollinearity. The VIF analysis resulted in the selection of 11 variables with VIF values within the generally accepted critical threshold (5): precipitation of the driest quarter, mean diurnal temperature range, isothermality, organic carbon stock, precipitation of the wettest month, pH in water, maximum temperature of the warmest period, organic carbon density, percentage of silt, percentage of

clay, and volume of fixed organic carbon, all in the quadratic version of the polynomial. After applying the VIF threshold, the range of Pearson correlations among these selected variables was reduced, fluctuating between -0.56 and 0.70, establishing a set of independent variables for subsequent analyses.

The Generalized Dissimilarity Modeling (GDM) models exhibited good performance in explaining the composition variation. When the model was run using only soil variables, the results indicated a null deviance of 28.15, with the model explaining 29.59% of the variation and a GDM deviance of 19.82. The model intercept was 0.50. On the other hand, the model with only climatic variables showed a null deviance of 28.15, with an explained variation of 18.09% and a GDM deviance of 23.05. In terms of the overall assessment of the total (climatic + soil variables) adjusted model, the null deviance explained by the GDM model was 19.02, with 32.05% of the variation explained, using 1000 permutations. Given that the total model (climatic + soil variables) exhibited the highest percentage of explanatory power, henceforth we will use this model to discuss the impacts of environmental variables on the ant communities. On that model, the importance of the predictors varied considerably, with the percentage of clay proportion showing the highest importance (yet not significant), followed by geographical distance (Table 1).

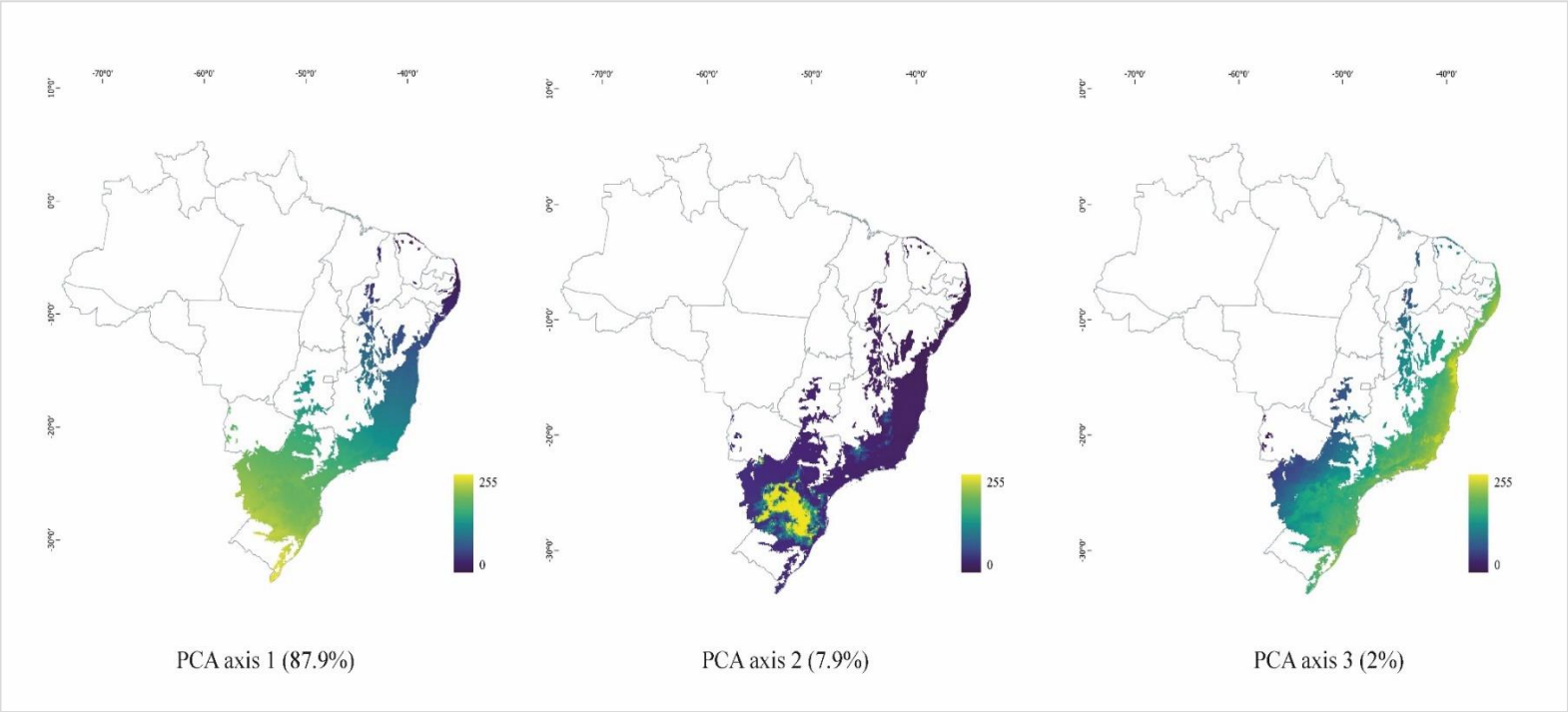
Table 1. Values of importance and p-value for each environmental variable selected by the model.

Variable	Importance
Proportion of clay particles (< 0.002 mm) in the fine earth fraction	25.367
Geographic Distance	8.851
Mean Diurnal Range (Mean of monthly (max temp - min temp))	7.090
Bulk density of the fine earth fraction	3.843

Proportion of silt particles (≥ 0.002 mm and ≤ 0.05 mm) in the fine earth fraction	3.280
Precipitation of Driest Month	2.129
Elevation	1.070
Organic carbon stocks in topsoil	1.113
Volumetric fraction of coarse fragments (> 2 mm)	0.618
Precipitation of Warmest Quarter	0.525

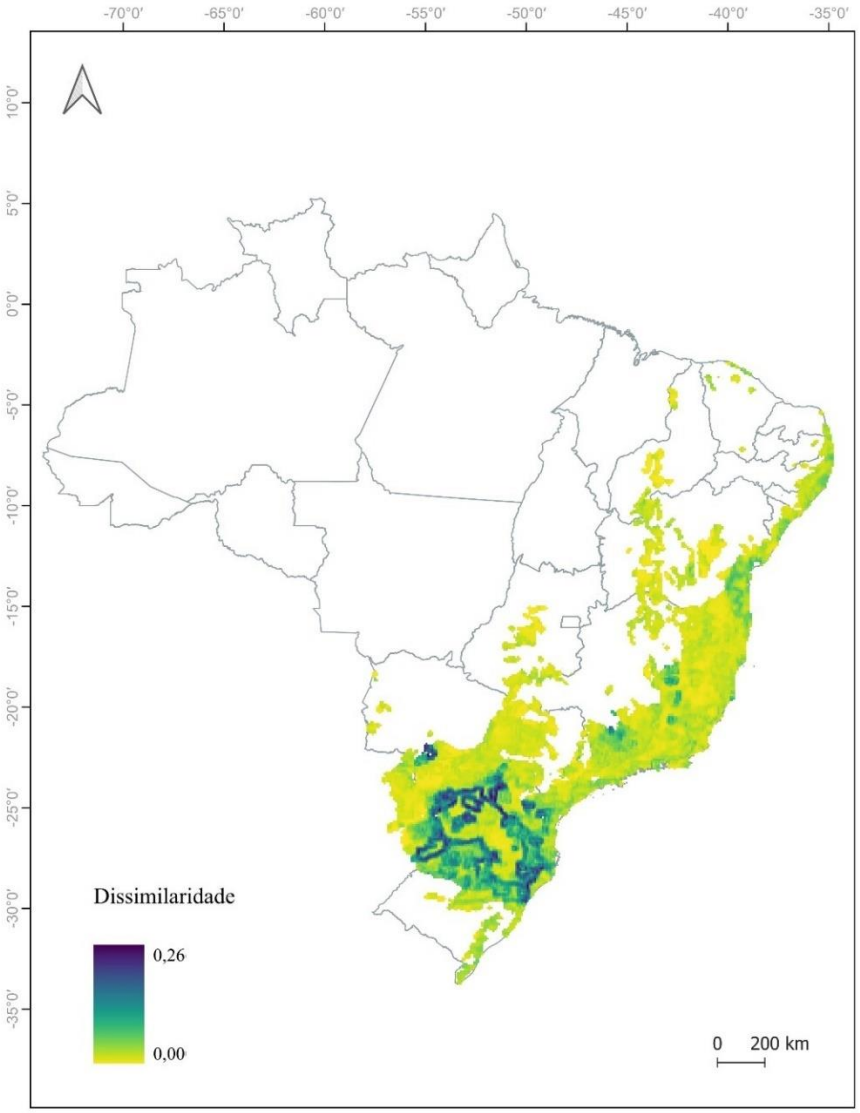
Variables with coefficients of 0 do not contribute significantly to the model and were therefore effectively removed during the modeling process, specifically: isothermality and pH in water. All variables achieved complete convergence with 1000 iterations, which reiterates the stability of our GDM model. As for the PCA analysis, the three main axes combined explain 97.7% of the variation in climate and soil data. The three principal components of the model were simplified into spatial layers to visualize areas predicted to be more or less similar (Figure 2).

Figure 2 - Predicted differences in community composition using Principal Component Analysis (PCA) of GDM-transformed environmental predictors. Similar color shades indicate more similar expected compositions.



The estimated uniqueness diversity is presented in Figure 3. The color range represents the ant diversity, with yellow areas indicating more similarity and dark blue areas indicating more diversity uniqueness. The uniqueness indicated in the south region and small areas in Minas Gerais and coastal Bahia suggests that these areas are characterized by greater species turnover (Figure 3).

Figure 3 - Predicted uniqueness of each location. Dark blue areas are predicted to have higher dissimilarity to the neighboring locations, indicating greater biological turnover.



Analyzes of the Venn diagram shows a significant contribution from both climatic and edaphic variables to ant diversity in the Atlantic Forest. The unique contribution of climatic variables was 2.77%, while that of edaphic variables was 14.28%, demonstrating a greater influence of soil on the diversity of ant communities than climate. However, the overlap of climatic and edaphic variables presented the highest value (15.31%), and the total contribution of climate and soil was 32.36%, indicating that the combination of these variables has the highest influence on the composition of ant communities.

2.5 Discussion

In this study, we aimed at exploring the influence of climatic and soil variables on the similarity of composition of ant communities in the Atlantic Forest. Most of the variables exhibit a pattern where their increase corresponds to an increase in dissimilarity, suggesting a direct relationship between them (geographic distance, proportion of clay particles, mean diurnal temperature range, precipitation of driest month, proportion of silt particles and bulk density). However, the nature of this relationship varies across variables. Some reveal the presence of tipping points, beyond which the relationship between variables and outcomes shifts significantly (geographic distance, clay particles and soil bulk density). Such thresholds delineate zones where the impact on ant diversity undergoes more dramatic changes, deviating from a straightforward linear progression. Moreover, there are cases where little or no discernible correlation was observed (precipitation of warmest quarter, volumetric fraction of coarse fragments, organic carbon stock and elevation).

Despite initial expectations of a stronger climatic influence, as suggested by Jenkins (2011) and supported by previous studies (Kaspari et al., 2000, 2003; Vasconcelos et al., 2010), our findings present a different scenario. We found that while climatic variables alone

contributed a modest 2.77% to ant dissimilarity, soil variables showed a more pronounced influence of 14.28%. However, the overlap between climatic and soil variables explains 15.31%, leading to a total contribution of 32.36% of soil and climatic factors in the ant composition. Considering that models of compositional dissimilarity commonly exhibit percent explained values ranging between 20% and 50% (Mokany, 2022), our model presented a satisfactory representation of the relationship between compositional diversity and environmental factors. When considering the total contribution of soil and climate, of the nine more important factors, five were soil variables, with the most influential individual variable also being related to the soil. This suggests that soil characteristics play an important role in shaping ant communities in terms of composition, surpassing the impact of climatic conditions when considered independently, but a more comprehensive understanding of ant composition emerges when examining the interplay between these two sets of variables.

Regarding the specific effects of each variable, our results show that the spatial distribution of precipitation significantly impacts the compositional distribution of ant communities in the Atlantic Forest. The precipitation pattern during the driest month has a pronounced significance, ranking as the sixth most important factor in explaining ant composition distribution. Its relationship suggests that as precipitation locally increases, it creates new environmental conditions, increasing compositional dissimilarity (uniqueness). The precipitation pattern in the warmest quarter also showed some correlation with ant composition distribution, but its impact was less pronounced, indicating it may not be a primary structuring factor. In line with these findings, previous studies have also underscored the influence of precipitation on ant community composition, with both positive and negative impacts reported across different species and habitats (Dunn et al., 2009; Uhey et al., 2020). Uhey (2020) highlighted the varied response of ant richness and abundance to precipitation, showing negative correlations in forest habitats and positive ones in open habitats, indicating

the importance of the surrounding vegetation regulating the effects of precipitation. Our results introduce an additional dimension to understanding the effects of precipitation on ant communities, emphasizing the temporal aspect of this environmental factor. Particularly, in regions where ant communities' dissimilarity was highest (the southern region and small area in coastal Bahia), the occurrence of higher precipitation during the driest month suggests that not merely the quantity of precipitation but its timing during more hostile periods plays a crucial role. These areas with an increase in precipitation may offer a form of refuge during dry periods, mitigating the severity of climatic conditions, which affects the ant composition and increases the taxonomic difference between the communities. Therefore, climatic change likely poses a threat to ant diversity, once it is expected that the Atlantic Forest will experience temperatures rising by over 1 °C and precipitation decreasing by more than 4% (Graham et al., 2016), increasing the hostility of drier months.

Studies have been exploring the relationship between biodiversity and elevation and it has been found to be common for species richness to decrease with increasing elevation in tropical mountains (Hoiss et al., 2012; Jankowski et al., 2009), primarily due to greater climate variation with elevation (Perillo et al. 2021). In a recent review paper, five broad patterns of ant elevational diversity have been identified: decreasing patterns (highest diversity at lowest elevation), low plateaus, mid-elevation peaks, increasing patterns, and constant, but the most common patterns are mid-elevation peaks and decreasing patterns (Subedi and Budha, 2020). It differs from older findings, which commonly found that the highest ant species richness is at the low and intermediate elevations (Sanders, 2003). In a study on the effects of elevation in ant diversity, they found that temperature was the main variable explaining the diversity variation in different elevational scales (Perillo et al., 2021). Pinpointing the primary drivers of ant diversity across different elevational gradients can be complicated (Szewczyk and McCain, 2016). Their analysis across various mountain ranges indicated that no single factor, such as

temperature, precipitation, or geographical area, uniformly dictates ant diversity. In our study, the elevation does not seem to affect ant similarity distribution in any way. While other studies found discernible patterns in local ant species distribution with elevation, these patterns seem to become less distinct when viewed across the continental scale of the Atlantic Forest. Given the absence of an elevation gradient, our results should not be considered as an argument against the influence of elevation on ant diversity, but that this relationship might be overshadowed by the underlying changes in the climate-elevation relationship over space at such a large scale as the Atlantic Forest (McCoy, 1990). This is in conformity with the discussion presented in Perillo et al. (2021) that, for the scale and location of their study, elevation was more important than latitude due to the greater climatic variation on the tropical mountains studied. In our study, the climatic variation comes primarily from the latitude, as the influence of mountain elevations is reduced compared to the larger-scale latitudinal climate gradients.

The mean diurnal temperature range was identified as the third more significant variable in explaining the differences in ant composition. Our analysis revealed a positive correlation, where an increase in mean diurnal temperature range corresponded to an increase in the dissimilarity between ant communities. However, this relationship is not mirrored spatially, meaning that areas with higher mean diurnal temperature range values did not consistently align with those exhibiting greater community dissimilarity highlighted. The complexity of species responses to temperature, indicating that factors such as colonization, occupancy, and extinction are not solely determined by mean annual temperature (Diamond, 2016). Instead, these responses are influenced by a combination of temperature and interspecies interactions, forming intricate networks of associations. Ewers and Banks-Leite (2013) provide an insight into habitat-specific temperature effects, noting that forest habitats, typically shaded and cooler, offer more stable microclimatic conditions. Also, the buffering effect of forests reduced maximum outside temperatures by one third or more at ground level within a forest, with a

stronger effect below-ground than one meter above-ground (Ewers and Banks-Leite, 2013). However, the temperature buffering effect was reduced near forest edges (Ewers and Banks-Leite, 2013). This aspect of forest microclimates could buffer the direct influence of temperature, and other climatic factors, in the ant composition. The high importance of this variable to ant composition but a lack of a straightforward spatial correlation suggests that the effect of diurnal temperature is modulated by other factors and cannot be fully understood without a detailed study on these variable interactions.

The direct link of organic carbon stock in soil with ant diversity has been little explored in recent years, but some studies demonstrated its importance to general diversity or its indirect effects. In recent research, Schuldt et al. (2023) identified belowground carbon stock as a significant predictor of multitrophic diversity. The relationships they found were nonlinear and showed a stronger correlation with lower trophic levels, whereas it was nonsignificant for higher trophic level diversity. Our results indicate the lack of correlation between carbon stock and ant diversity, with increases in carbon stock not leading to any changes in ant community similarities. In addition, the importance of this variable in explaining ant composition was low, suggesting that it might be an indirect and weak measure of another factor, such as vegetation biomass. Because of this result, our results corroborate the conclusion by Schuldt et al. (2023) that some conservation strategies should be evaluated cautiously. Specifically, the ones that focus on maximizing carbon stock to also mitigate climate change may fall short in meeting biodiversity conservation goals (Ferreira et al., 2018; Schuldt et al., 2023). Although the Atlantic Forest (specifically Serra do Mar and Bahia coastal forests) are hotspots for both carbon (above- and belowground combined) and local biodiversity (Soto-Navarro et al., 2020), in the context of ants, we suggest that not all carbon-focused climate mitigation strategies will effectively preserve their diversity and the lack of correlation of their diversity with soil carbon might decrease the efficiency of such strategies for this group.

Regarding the soil variables, the distribution of the size of soil particles directly impacts its porosity, permeability, and physical processes (such as transport and deposition), potentially impacting soil organisms (da Costa-Milanez et al., 2017). Our research reveals a linear relationship between soil clay content and ant diversity, identifying it as the primary factor influencing ant species composition. Specifically, areas with higher clay content demonstrated a marked increase in community dissimilarity. Additionally, soil bulk density significantly influences ant community structure, ranking as the fourth key variable. Greater dissimilarity in ant communities was observed in areas exhibiting moderate soil bulk density, particularly in zones transitioning from low to high values. This pattern was not as pronounced in areas where either high or low values of soil bulk density predominate. Furthermore, our findings indicate that soil silt content also plays a role in shaping ant communities, emerging as the fifth most influential factor. Regions with lower silt content tended to have more similar ant communities, whereas increasing silt levels were associated with higher community differentiation. Research exploring how soil granulometric distribution influences ant presence (Costa et al., 2010; Costa-Milanez et al., 2014) identified a strong positive association between the prevalence of fine and ultra-fine sand grains and the density of ants below 5 mm in size (Costa et al., 2010). They found an increase in the proportion of small ant species of 5 mm or less body length with an increase in percentage of fine grains in the sediment, explaining the disproportionate occurrence of workers of small species, but the predominance of fine grains did not explain overall ant species richness and abundance (Costa et al., 2010). Larger ants typically nest underground, suggesting that a high proportion of fine grains, coupled with near-surface groundwater, might not promote colonization in these habitats (da Costa-Milanez et al., 2017). In another study, Costa-Milanez et al. (2014) found a significant positive correlation between the density of small ant species in areas where the proportion of fine and very fine grains was higher (Costa-Milanez et al., 2014). It is possible that the observed differences in ant communities across different soil

particle sizes reflect this habitat selection process. This hypothesis implies that ant species size may suggest a preference for areas with specific soil granulometry, specifically small sized species preferring fine grain soil, which is avoided by larger sized species. With increasing the grain size, the habitat selection eases, thereby leading to varied community structures at a broader ecological scale. This would be consistent with our model, which shows an increase in community dissimilarities with an increase in clay and silt content and soil bulk density. In this case, this process of habitat selection through grain sizes would be a factor in shaping the ecological distribution of ants. In line with this, Johnson (2000) conducted a study on two ant species, noting their segregation among microhabitats based on soil texture, with one species found predominantly in soils with higher clay content and moisture retention. Interestingly, in regions where both species coexisted (sympatric areas), the soil texture was of an intermediate composition compared to areas where each species was found alone (allopatric areas). This pattern is consistent with the spatial distribution of clay content and soil bulk density, where regions with intermediate amount of clay and bulk density (south region and coastal Bahia) presented the highest values of community dissimilarity. Costa Milanez et al. (2017) further contributed to this understanding by noting that soils with high clay content exhibit a greater capacity for water retention, which could reduce the risk of desiccation for ant colonies. This aspect of soil composition is particularly crucial in the early stages of ant colony development, where desiccation is a significant factor in colony mortality (Costa Milanez et al., 2017). Therefore, our study reinforces the observations regarding the impact of soil granulometric on ant species and expands upon these findings by demonstrating how soil compositions and structure in general can influence community dissimilarity in a broader ecological context.

2.6 Conclusion

Our study indicates an intricate dynamic between climatic and soil variables and their impact on the composition of ant communities within the Atlantic Forest. Our findings, settled on a 10km spatial scale, explored the relationship between environmental factors and ant diversity, suggesting that while climatic variables explain part of the spatial distribution of ant composition, it is the soil characteristics that emerge as the more dominant influence, when both sets of variables are considered alone. Our analysis also reveals that the synergistic effect of climatic and edaphic factors surpasses the influence of either factor in determining the composition of ant communities. Moreover, the identification of specific variables, such as precipitation at driest month and clay content, as significant predictors of ant diversity emphasize the value of detailed environmental assessments in ecological efforts. Our research also reinforces the presence of thresholds in the relationship between environmental variables and ant community composition. These thresholds mark critical points beyond which the effects of certain variables on ant diversity become markedly different, suggesting that the impact of environmental changes on biodiversity can be complex and interconnected, as pointed by other studies. Furthermore, our results reveal that, in the considered spatial resolution and scale, neither carbon stocks nor elevation affect the similarity between ant communities.

The direct impact of precipitation during driest months on ant community structure reinforces the importance of temporal aspects in understanding biodiversity dynamics. It also hints at the potential vulnerability of ant communities to climate change, which are expected to increase temperatures and decrease precipitation conditions in Atlantic Forest (Graham et al., 2016), altering precipitation regimes and exacerbate the challenges posed by drier conditions. Additionally, the significant influence of soil clay and silt content, and soil bulk density on ant diversity brings to light the implications of land use and land cover changes on biodiversity. The effect of these variables on community uniqueness underscores the sensitivity of ant populations to alterations in soil structure, which can be directly impacted by human activities

such as agriculture, deforestation, and urbanization (Dror et al., 2021). These activities often lead to changes in soil composition and compaction (Dror et al., 2021), thereby altering the habitat suitability for various ant species. In general, our results highlight the potential of land use changes and climate change to disrupt the spatial distribution of ant communities' composition, reinforcing the importance of conservation strategies to mitigate theses anthropogenic impacts when addressing the multifaceted challenges facing biodiversity in tropical forests.

2.7 References

ARNAN, X. et al. Foraging behaviour of harvesting ants determines seed removal and dispersal. **Insectes Sociaux**, v. 57, n. 4, p. 421–430, 9 maio 2010.

AVGAR, T.; GILADI, I.; NATHAN, R. Linking traits of foraging animals to spatial patterns of plants: social and solitary ants generate opposing patterns of surviving seeds. **Ecology Letters**, v. 11, n. 3, p. 224–234, mar. 2008.

BASELGA, A.; ORME, C. D. L. betapart: an R package for the study of beta diversity. **Methods in Ecology and Evolution**, v. 3, n. 5, p. 808–812, 18 jun. 2012.

BAZZAZ, F. A. Habitat Selection in Plants. **The American Naturalist**, v. 137, p. S116–S130, jun. 1991.

BISWAS, S. R. et al. A conceptual framework for the spatial analysis of functional trait diversity. **Oikos**, v. 125, n. 2, p. 192–200, 12 maio 2015.

BOYLE, M. J. W. et al. Localised climate change defines ant communities in human-modified tropical landscapes. **Functional Ecology**, v. 35, n. 5, p. 1094–1108, 11 jan. 2021.

BRIAN, M. V.; BEATTIE, A. J. The Evolutionary Ecology of Ant-Plant Mutualisms. **The Journal of Ecology**, v. 74, n. 4, p. 1213, dez. 1986.

CAMMERAAT, E. L. H.; RISCH, A. C. The impact of ants on mineral soil properties and processes at different spatial scales. **Journal of Applied Entomology**, v. 132, n. 4, p. 285–294, maio 2008.

CAMPBELL, K. U.; CRIST, T. O. Ant species assembly in constructed grasslands is structured at patch and landscape levels. **Insect Conservation and Diversity**, v. 10, n. 2, p. 180–191, 5 jan. 2017.

CARLOS GALINDO LEAL; IBSEN DE GUSMÃO CÂMARA. **The Atlantic Forest of South America : biodiversity status, threats, and outlook**. Washington: Island Press, 2003.

CAVENDER-BARES, J.; KEEN, A.; MILES, B. PHYLOGENETIC STRUCTURE OF FLORIDIAN PLANT COMMUNITIES DEPENDS ON TAXONOMIC AND SPATIAL SCALE. **Ecology**, v. 87, n. sp7, p. S109–S122, jul. 2006.

CERDÀ, A.; JURGENSEN, M. F. The influence of ants on soil and water losses from an orange orchard in eastern Spain. **Journal of Applied Entomology**, v. 132, n. 4, p. 306–314, maio 2008.

CHESSON, P. Mechanisms of Maintenance of Species Diversity. **Annual Review of Ecology and Systematics**, v. 31, n. 1, p. 343–366, nov. 2000.

CORNWELL, W. K.; ACKERLY, D. D. Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. **Ecological Monographs**, v. 79, n. 1, p. 109–126, fev. 2009.

COSTA-MILANEZ, C. B. DA et al. Influence of soil granulometry on average body size in soil ant assemblages: implications for bioindication. **Perspectives in Ecology and Conservation**, v. 15, n. 2, p. 102–108, 1 abr. 2017.

DEL TORO, I. et al. The little things that run the world revisited: a review of ant-mediated ecosystem services and disservices (Hymenoptera: Formicidae). **Myrmecol. News**, v. 17, n. 10, p. 133–146, 2012.

DIAMOND, S. E. et al. Climatic warming destabilizes forest ant communities. **Science Advances**, v. 2, n. 10, 7 out. 2016.

DORMANN, C. F. et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. **Ecography**, v. 36, n. 1, p. 27–46, 18 maio 2012.

DORN, R. I. Ants as a powerful biotic agent of olivine and plagioclase dissolution. **Geology**, v. 42, n. 9, p. 771–774, set. 2014.

DROR, I.; YARON, B.; BERKOWITZ, B. The Human Impact on All Soil-Forming Factors during the Anthropocene. **ACS Environmental Au**, v. 2, n. 1, 13 set. 2021.

DUNN, R. R. et al. Climatic drivers of hemispheric asymmetry in global patterns of ant species richness. **Ecology Letters**, v. 12, n. 4, p. 324–333, abr. 2009.

ELBERS, B. **Orthogonal Polynomial Contrasts and Applications to Age-Period-Cohort Models**. [s.l: s.n.]. Acesso em: 17 jan. 2025.

ENGELBRECHT, B. M. J. et al. Drought sensitivity shapes species distribution patterns in tropical forests. **Nature**, v. 447, n. 7140, p. 80–82, 1 maio 2007.

EWERS, R. M.; BANKS-LEITE, C. Fragmentation Impairs the Microclimate Buffering Effect of Tropical Forests. **PLoS ONE**, v. 8, n. 3, p. e58093, 4 mar. 2013.

FERREIRA, J. et al. Carbon-focused conservation may fail to protect the most biodiverse tropical forests. **Nature Climate Change**, v. 8, n. 8, p. 744–749, 16 jul. 2018.

FERRIER, S. et al. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. **Diversity and Distributions**, v. 13, n. 3, p. 252–264, maio 2007.

FROUZ, J.; JILKOVÁ, V. The effect of ants on soil properties and processes (Hymenoptera: Formicidae). **Myrmecol. News**, v. 11, p. 191–199, 2008.

GRAHAM, T. L.; MATTHEWS, H. D.; TURNER, S. E. A Global-Scale Evaluation of Primate Exposure and Vulnerability to Climate Change. **International Journal of Primatology**, v. 37, n. 2, p. 158–174, 30 jan. 2016.

GRIFFITHS, H. M. et al. Ants are the major agents of resource removal from tropical rainforests. **Journal of Animal Ecology**, v. 87, n. 1, p. 293–300, 8 ago. 2017.

GRIFFITHS, H. M. et al. The impact of invertebrate decomposers on plants and soil. **New Phytologist**, v. 231, n. 6, p. 2142–2149, 16 jul. 2021.

HOISS, B. et al. Altitude acts as an environmental filter on phylogenetic composition, traits and diversity in bee communities. **Proceedings of the Royal Society B: Biological Sciences**, v. 279, n. 1746, p. 4447–4456, 29 ago. 2012.

HOLWAY, D. A.; SUAREZ, A. V.; CASE, T. J. Role of Abiotic Factors in Governing Susceptibility to Invasion: A Test with Argentine Ants. **Ecology**, v. 83, n. 6, p. 1610–1610, 1 jun. 2002.

HUFBAUER, R. A. et al. Anthropogenically induced adaptation to invade (AIAI): contemporary adaptation to human-altered habitats within the native range can promote invasions. **Evolutionary Applications**, v. 5, n. 1, p. 89–101, 2 nov. 2011.

JAMES, G. et al. **An Introduction to Statistical Learning**. New York, NY: Springer New York, 2013. v. 103

JANKOWSKI, J. E. et al. Beta diversity along environmental gradients: implications of habitat specialization in tropical montane landscapes. **Journal of Animal Ecology**, v. 78, n. 2, p. 315–327, mar. 2009.

JENKINS, C. N. et al. Global diversity in light of climate change: the case of ants. **Diversity and Distributions**, v. 17, n. 4, p. 652–662, 16 maio 2011.

JENKINS, C. N. et al. Patterns of Vertebrate Diversity and Protection in Brazil. **PLOS ONE**, v. 10, n. 12, p. e0145064, 17 dez. 2015.

JOHN, R. et al. Soil nutrients influence spatial distributions of tropical tree species. **Proceedings of the National Academy of Sciences**, v. 104, n. 3, p. 864–869, 16 jan. 2007.

JOHNSON, R. A. Habitat segregation based on soil texture and body size in the seed-harvester ants Pogonomyrmex rugosus and P. barbatus. **Ecological Entomology**, v. 25, n. 4, p. 403–412, nov. 2000.

JOLY, C. A.; METZGER, J. P.; TABARELLI, M. Experiences from the Brazilian Atlantic Forest: ecological findings and conservation initiatives. **New Phytologist**, v. 204, n. 3, p. 459–473, 10 set. 2014.

KAREL MOKANY et al. A working guide to harnessing generalized dissimilarity modelling for biodiversity analysis and conservation assessment. **Global Ecology and Biogeography**, v. 31, n. 4, p. 802–821, 21 fev. 2022.

KASPARI, M.; O'DONNELL, S.; KERCHER, J. R. Energy, Density, and Constraints to Species Richness: Ant Assemblages along a Productivity Gradient. **The American Naturalist**, v. 155, n. 2, p. 280–293, fev. 2000.

KASPARI, M.; YUAN, M.; ALONSO, L. Spatial Grain and the Causes of Regional Diversity Gradients in Ants. **The American Naturalist**, v. 161, n. 3, p. 459–477, mar. 2003.

KIRILENKO, A. P.; SEDJO, R. A. Climate change impacts on forestry. **Proceedings of the National Academy of Sciences**, v. 104, n. 50, p. 19697–19702, 11 dez. 2007.

KRAFT, N. J. B. et al. Community assembly, coexistence and the environmental filtering metaphor. **Functional Ecology**, v. 29, n. 5, p. 592–599, 8 out. 2014.

LACH, L.; PARR, C. L.; ABBOTT, K. L. **Ant ecology**. Oxford: Oxford University Press, 2010.

MCCOY, E. D. The Distribution of Insects along Elevational Gradients. **Oikos**, v. 58, n. 3, p. 313, ago. 1990.

MYERS, N. et al. Biodiversity hotspots for conservation priorities. **Nature**, v. 403, n. 6772, p. 853–858, fev. 2000.

OLIVEIRA, F. M. P. et al. Chronic anthropogenic disturbance as a secondary driver of ant community structure: interactions with soil type in Brazilian Caatinga. **Environmental Conservation**, v. 44, n. 2, p. 115–123, 19 ago. 2016.

OLIVER, T. H.; LEATHER, S. R.; COOK, J. M. Macroevolutionary patterns in the origin of mutualisms involving ants. **Journal of Evolutionary Biology**, v. 21, n. 6, p. 1597–1608, nov. 2008.

OTTO, S. P. Adaptation, speciation and extinction in the Anthropocene. **Proceedings of the Royal Society B: Biological Sciences**, v. 285, n. 1891, 21 nov. 2018.

PARR, C. L. et al. Suppression of savanna ants alters invertebrate composition and influences key ecosystem processes. **Ecology**, v. 97, n. 6, p. 1611–1617, jun. 2016.

PERILLO, L. N. et al. Disentangling the effects of latitudinal and elevational gradients on bee, wasp, and ant diversity in an ancient neotropical mountain range. **Journal of Biogeography**, v. 48, n. 7, p. 1564–1578, 29 mar. 2021.

ROCHA-ORTEGA, M.; GARCÍA-MARTÍNEZ, M. Á. Importance of Nesting Resources and Soil Conditions for the Recovery of Ant Diversity During Secondary Succession in a Tropical Rainforest. **Tropical Conservation Science**, v. 11, n. 11, p. 194008291878706-194008291878706, 1 jan. 2018.

SANDERS, N. J.; MOSS, J.; WAGNER, D. Patterns of ant species richness along elevational gradients in an arid ecosystem. **Global Ecology and Biogeography**, v. 12, n. 2, p. 93–102, mar. 2003.

SCHULDT, A. et al. Carbon–biodiversity relationships in a highly diverse subtropical forest. **Global Change Biology**, v. 29, n. 18, p. 5321–5333, 3 abr. 2023.

SILVA, R. R. et al. ATLANTIC ANTS: a data set of ants in Atlantic Forests of South America. **Ecology**, v. 103, n. 2, 22 dez. 2021.

SNYDER, ROBIN E.; CHESSON, P. How the Spatial Scales of Dispersal, Competition, and Environmental Heterogeneity Interact to Affect Coexistence. **The American Naturalist**, v. 164, n. 5, p. 633–650, nov. 2004.

SOTO-NAVARRO, C. et al. Mapping co-benefits for carbon storage and biodiversity to inform conservation policy and action. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 375, n. 1794, p. 20190128, 27 jan. 2020.

SOUZA, J. et al. Effectiveness of genera as a higher-taxon substitute for species in ant biodiversity analyses is not affected by sampling technique. **Biodiversity and conservation**, v. 27, n. 13, p. 3425–3445, 14 ago. 2018.

SUBEDI, I. P.; BUDHA, P. B. Diversity and distribution patterns of ants along elevational gradients. **Nepalese Journal of Zoology**, v. 4, n. 1, p. 44–49, 16 ago. 2020.

SZEWCZYK, T.; MCCAIN, C. M. A Systematic Review of Global Drivers of Ant Elevational Diversity. **PLOS ONE**, v. 11, n. 5, p. e0155404, 13 maio 2016.

TUOMAINEN, U.; CANDOLIN, U. Behavioural responses to human-induced environmental change. **Biological Reviews**, v. 86, n. 3, p. 640–657, 27 out. 2010.

UHEY, D. A. et al. Climate and vegetation structure shape ant communities along elevational gradients on the Colorado Plateau. **Ecology and Evolution**, v. 10, n. 15, p. 8313–8322, 7 jul. 2020.

VASCONCELOS, H. L. et al. Patterns of ant species diversity and turnover across 2000 km of Amazonian floodplain forest. **Journal of Biogeography**, v. 37, n. 3, p. 432–440, mar. 2010.

WOOD, Z. T. et al. The Importance of Eco-evolutionary Potential in the Anthropocene. **BioScience**, v. 71, n. 8, p. 805–819, 31 mar. 2021.

ZACHOS, F. E.; HABEL, J. C. **Biodiversity Hotspots: Distribution and Protection of Conservation Priority Areas**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

ZELIKOVA, T. J.; BREED, M. D. Effects of habitat disturbance on ant community composition and seed dispersal by ants in a tropical dry forest in Costa Rica. **Journal of Tropical Ecology**, v. 24, n. 3, p. 309–316, maio 2008.

3 CAPÍTULO 2

Does an induced drought change the phylogenetic structure of Amazonian ant communities?

Authors: Isadora Gerheim[1*], Rony P. S. Almeida[2], Rogério R. Silva[2], Ricardo R. C. Solar[3]

[1] Programa de Pós-Graduação em Ecologia, Universidade Federal de Viçosa (UFV), Viçosa, Minas Gerais, 36570-900 Brazil

[2] Museu Paraense Emílio Goeldi, Coordenação de Ciências da Terra e Ecologia, Belém, PA, Brazil

[3] Centro de Síntese Ecológica e Conservação, Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Avenida Presidente Antônio Carlos 6627, Belo Horizonte, Minas Gerais, 31270-910 Brazil

[*]corresponding author

Options for target journal: Insect Conservation and Diversity; Journal of Insect Conservation; Oecologia

3.1 Abstract

Understanding the impacts of drought on biodiversity is essential to predict the effects of climate change on ecosystems. In this study, we explored the impacts of a long-term induced drought on the phylogenetic structure of Amazonian ant communities over a ten-month period. The ant sampling is composed of 162 ant species distributed across 44 genera and 7 subfamilies. We found that the impact on species abundance fluctuated across phylogenetic branches with no uniform response within subfamilies and a few within genera. We confirm our prediction that drought conditions led to a decrease in phylogenetic richness due to species loss. However, drought did not select species with close phylogenetic relationships. The control community displayed gradual temporal variation in mean phylogenetic distance, whereas the drought community exhibited low and high peaks, from one month to another. The induced drought led to a decrease in phylogenetic richness, an increase in phylogenetic gaps between species, and destabilization of the phylogenetic structure of the ant community. This demonstrates how some consequences of climate change, such as extreme drought events, can lead to changes in different facets of ant diversity, including phylogenetic diversity.

3.2 Introduction

Human activities are changing the global climate at an unprecedented rate, and in the last 60 years, these activities were heavily accelerated (IPCC/AR6, 2012). Among others, the increased burning of fossil fuel (Rockström et al., 2018), has led to a rising accumulation of greenhouse gas concentrations (Broecker, 1975), causing several changes in ecosystem dynamic climate (Lade, 2020; Steffen et al., 2015, 2018). Empiric data and mathematical models raise our concerns that these global changes will alter water cycles, soil humidity and temperature, influencing extreme drought regimes (Brando et al., 2019). The climate projection for South America suggests that an increase of 1.5°C and 2°C in temperature will increase drought duration and frequency in 47% and 79%, respectively (Cook and Vizy, 2008; IPCC, 2012; Xu et al., 2019), which could disrupt the feedback that stabilize tropical forests (Flores and Staal, 2022), causing generalized and permanent damage (Brando et al., 2019). Some studies have already shown that the Amazon Forest is in fact vulnerable to drought events, with the potential to feedback on climate change by reversing its role as a large long-term carbon sick (Phillips et al., 2009).

Under such circumstances, global climate changes have been considered one of the main causes of biodiversity change (Urban et al., 2016), by directly affecting the species survival and persistence (Hooper et al., 2012). Among affected groups, ants are a dominant key group that occur in most of the land surface of Earth (Ward et al., 2015), but two thirds of its abundance are concentrated in tropical ecosystems (Schultheiss et al., 2022). They perform several ecosystem roles such as predators, scavengers, granivores, herbivores, omnivores, seed harvesters, acting also on bioturbation of soil and aiding to soil carbon sequestration (Del Toro et al., 2012; Dorn, 2014; Lach et al., 2010; Parr et al., 2016). Ants are sensitive to environment

disturbance (Lach et al., 2010), they generally prefer open and hot environment (Andersen 1997), and some genera are well-adapted to tropical climate (Delsinne et al., 2013; Lach et al., 2010), which makes them an interesting group for studying drought effects at the community level.

Drought events can impose severe alterations on tropical forests climate cycles, changing the community composition and affecting all biodiversity dimensions, including taxonomic, functional and phylogenetic diversities. The latter reflects the amount of evolutionary history and the evolutionary relatedness between taxa (Grab et al., 2019; Pio et al., 2011). It represents the accumulation of adaptations, strategies and behaviors, capturing similarities of responses to the environment among clades (Grab et al., 2019; Pio et al., 2011; Tan et al., 2015). Since related species tend to respond more similarly to the environment (Webb et al., 2002) (for ants this is generally true at the genera level, see examples in Lach et al., 2010), sometimes the species loss is not random nor uniform across the phylogeny (Grab et al., 2019). Investigating the phylogenetic structure of a community (occurrence, abundance, and relatedness of species) can elucidate how the drought affects the clades persistence and its implications. The phylogenetic structure indirectly influences other aspects of biodiversity, such as community stability (Cadotte et al., 2012), functional diversity (Bononi et al. 2018), invasion proneness (El-Barougy et al., 2020) and potential for resisting climate change through adaptation (Hoffmann and Sgrò, 2011; Carroll and Fox, 2008).

The effect of drought on phylogenetic diversity appears to be complex and may depend on the group of interest. One study found that under a rainfall reduction treatment, taxonomic and functional plant diversity decreased but phylogenetic diversity increased (López-Rubio et al., 2022), while another study found a loss of phylogenetic diversity after a natural drought event (Li et al., 2019). Regarding the phylogenetic structure, ecological disturbances seem to select for disturbance-adapted traits, which leads to phylogenetic clustering if there is niche

conservatism or a strong phylogenetic signal (Ding et al., 2012; Dinnage, 2009; Helmus et al., 2010; Miazaki et al., 2015; Qi et al., 2015). For ants, a study on aridity gradient did not find any effect neither at specific nor at higher-phylogenetic levels (Delsinne et al., 2010). However, rainfall exclusion experiments found a decrease in ant species richness (Almeida et al., 2023; Delsinne et al., 2013), but the impact of this disturbance on phylogenetic structure remains uncertain given the potential variation in response among ant genera (Delsinne et al., 2013).

Since the consequences of water deficit in tropical forests dynamics are still poorly understood, an induced drought experiment was set at the Amazon rainforest, with the goal to assess its impact on the forest cycle and biodiversity. They found that prolonged drought can have significant impacts on tree mortality, wood production, and above-ground biomass (da Costa et al., 2010). The experiment reveals an initial decline in leaf fall due to induced drought, with a subsequent re-stabilizing at a new functional state after > 10 years of the ongoing experiment, despite high tree mortality (Rowland et al., 2018). The tree mortality led to a 30% decrease in total forest transpiration, but remaining trees exhibited sustained or enhanced per tree transpiration due to decreased water competition and elevated light availability (Costa et al., 2018). Once the impact of the induced drought was set, new studies have been done on this same experimental site to explore the consequences of drought on other aspects of forest biodiversity. In this study, we delve into the data of this ongoing long-term drought experiment in Brazilian Amazon to explore how induced drought affects the phylogenetic structure of an ant community. Based on previous studies on the decline in ant species richness under rainfall exclusion experiments (Almeida et al., 2023; Delsinne et al., 2013) and the tendency of ecological disturbances to select for disturbance-adapted traits, our hypotheses are that the induced drought will: i) decrease the phylogenetic richness through species loss and ii) lead to phylogenetic clustering.

3.3 Methods

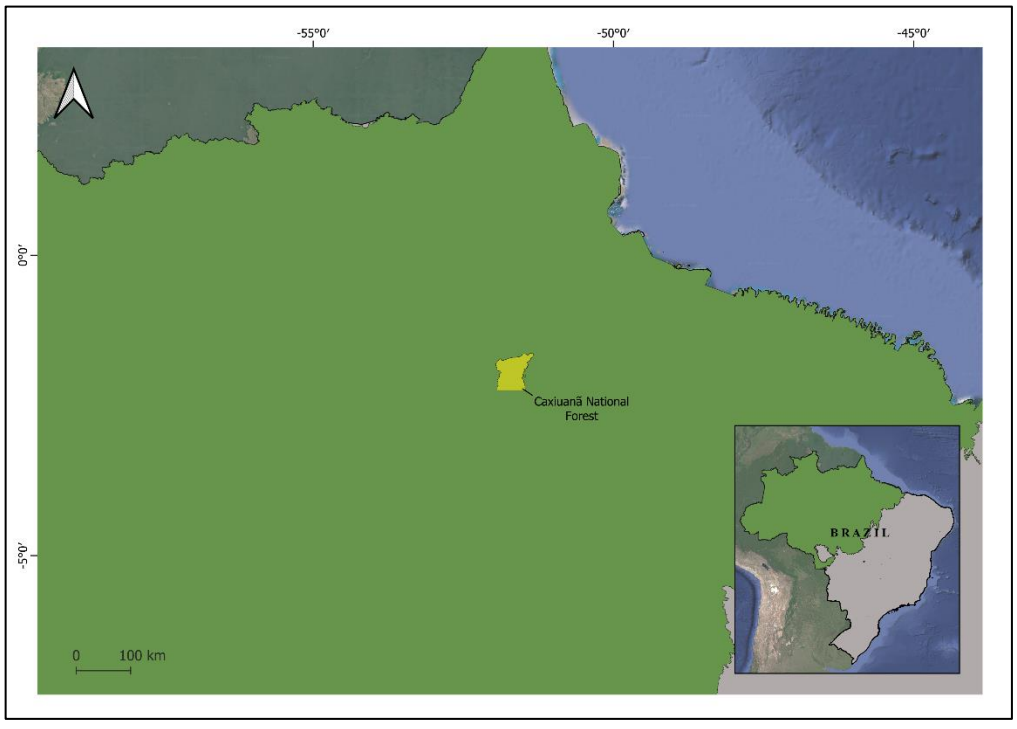
For this study, we used data provided by the ESecaFlor Project, which consisted of ant communities sampled within the context of a long-term (> 10 years) through-fall exclusion experiment (TFE) in an eastern Amazonian rainforest.

The ESecaFlor Experiment

Site

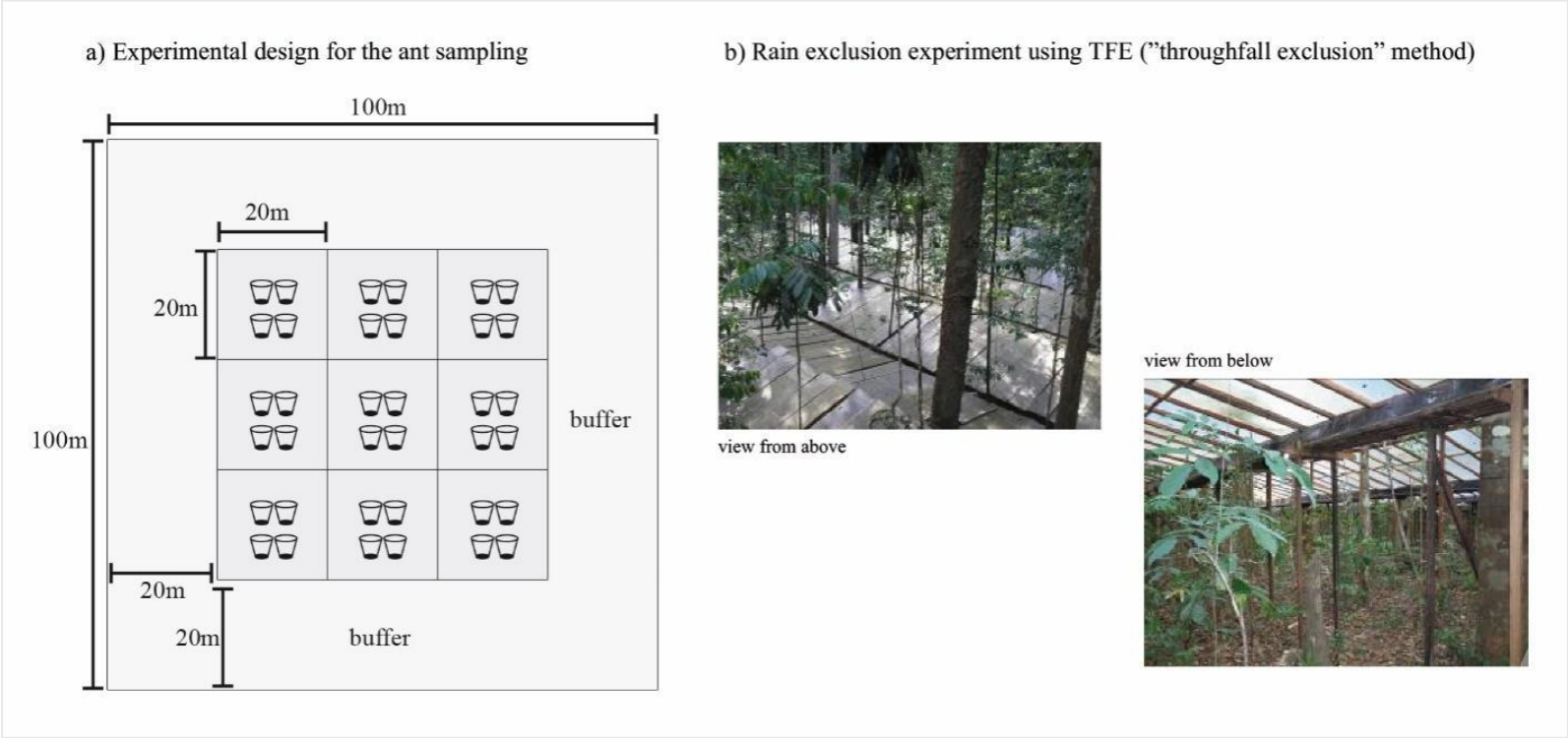
The site where the experiment was conducted is located at a lowland area (*terra firme*) at the Caxiuanã National Forest, near the Estação Científica Ferreira Penn (01º 42' 30" S e 51º 31' 45" W), Pará State, Brazil (Figure 1). The Mean Annual Rainfall is 2272 (±193) mm, with dry season between July-December, with Mean Rainfall of 555 (±116) mm.

Figure 1 - Location of Caxiuanã National Forest (in light green), where the induced drought experiment took place. The green area represents the Amazon biome.



The experiment began in 2001, and it was carried out using TFE ("throughfall exclusion") approach, which consists of two 1ha-plots: 'induced drought plot' and 'a control plot', located 50m apart from each other, at structurally and floristically similar areas. The soil moisture deficit was induced by using 6 thousand transparent plastic panels set at 2m above the soil, intercepting ~50% of the rain and preventing it getting into the soil (more details of the experiment design in Figure 2 and in Fisher et al. (2007)). Due to the immense effort to set this kind of experiment in a dense forest with difficult access such as Amazon, its implementation was limited to two plots with a Control-Impact (CI) design, comparing the impacted site (induced drought as treatment) with a non-randomly allocated control site. For further details on the experimental design and results see: da Costa et al. (2010); Rowland et al. (2015).

Figure 2 - ESECAFLOR experiment design. a) Plot design and ant sampling; b) Plastic panels that intercept the rain, used to induce drought conditions. Photos from the ESECAFLOR website.



Ant Sampling

The ant sampling was conducted by a team, whose work and data are detailed in Almeida et al. (2023). This sampling occurred from October 2011 to September 2012 (ten years after the induced drought experiment started), excluding April and June due to technical issues. Twenty-five plots of 20x20m were set and in the middle of each one 4 pitfall traps of 500ml were placed, containing water, soap, and salt (5%) for 48 hours, for the ant sampling. For this study, we considered the sampling from all subplots as an ant community, summarizing two communities (one from the 'induced drought' and one from the 'control plot') with 10 temporal samplings. Since the area located at the border could be biased by the plot surroundings, we considered only the nine subplots located at the center, hence, apart 20m away from the border (Figure 2a). All species were identified up to the morphospecies level.

Phylogenetic Diversity Metrics

We used the species relatedness found in one of the most complete phylogenetic hypotheses for Formicidae up to date (Moreau and Bell, 2013). Some collected genera were not present in the phylogenetic tree, so we replaced their closest genus with them, using recent phylogenetic studies. Given the lack of species-level phylogeny of the Formicidae family, we built a genus-level phylogenetic tree and used the list of sampled species to simulate 300 species-level phylogenies based on Yule process (pure-birth) (Arnan et al., 2018) using phytools package (Revell, 2012) on R software (R Development Core Team, 2013).

To test our hypotheses, we run the phylogenetic richness, which is the sum of branch lengths of all species in community (Faith, 1992), with "richness" null model. This null model shuffles species occurrences within the sites, while maintaining site species richness (Miller et al., 2017). To assess the relatedness among taxa within each community, two metrics were employed: Mean Pairwise Distances (MPD; Webb et al., 2008) and Mean Nearest Taxon

Distances (MNTD; Webb et al., 2008), with and without abundance weighted. The MPD metric is particularly sensitive to changes at the clade level, while the MNTD metric focuses on changes at more recent branches. We used the "taxa labels" null model, which shuffles tip labels, while maintaining data structure, including tree topology. The null models were run 500 times and *P*-value is calculated through a two-tailed test, thus the $P \leq 0.025$ means phylogenetic clustering (i.e. species are more related than expected by chance) and $P \geq 0.975$ means phylogenetic overdispersion (i.e. species are less related than expected by chance) (Cadotte and Davies, 2016). All analysis were done using R software (*R Core Team*) and the packages: *picante* (Kembel et al., 2010), *MicEco* (Russel, 2021) and *ecoPD* (Regetz et al., 2019).

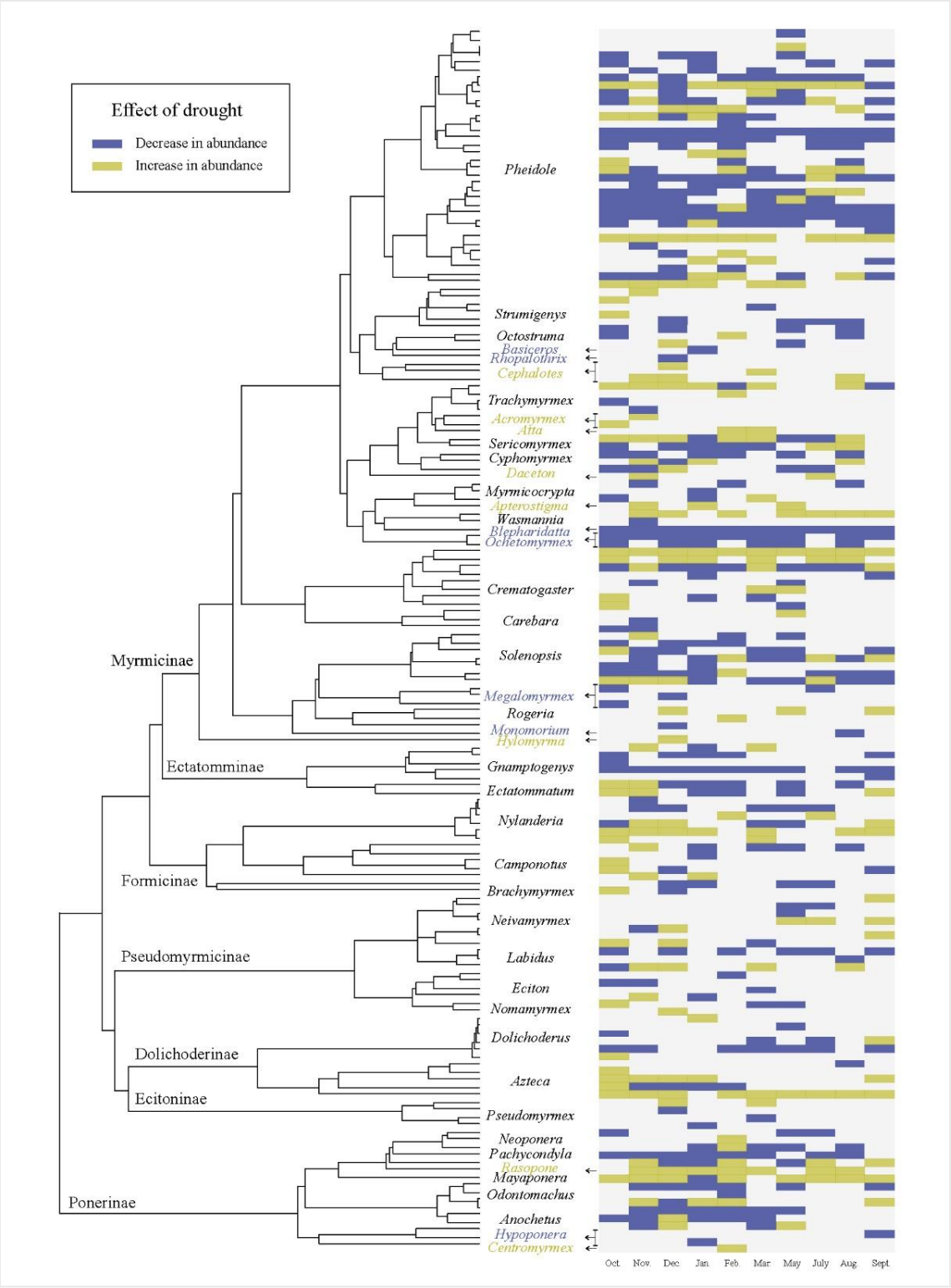
To visualize if there are some phylogenetic patterns in the community response to drought conditions, we plotted if the genera presented increase or decrease in abundance due to drought treatment, for each month. We chose this method of visualization to reduce the bias from natural differences in abundance among the species. Without this procedure, genera with higher abundance could mistakenly exhibit more drastic changes simply due to their larger population size, not necessarily because they were more affected by the drought. This approach was implemented with the purpose of investigating whether certain subfamily or genera were uniformly affected by drought conditions.

3.4 Results

The dataset was composed of 162 ant species, comprising 44 genera and 7 subfamilies. All subfamilies presented an increase and decrease in species abundance, and it does not seem to present a uniform response within each one (Figure 3). At the genus level, only a few showed only one of these responses (increase or decrease) during the experiment and a portion showed some kind of consistency, namely: *Cephalotes* (all 3 species increased in abundance),

Blepharidatta (decreased in all months), *Ochetomyrmex* (decrease in most months) and *Rasopone* (increase in most months) (Figure 3).

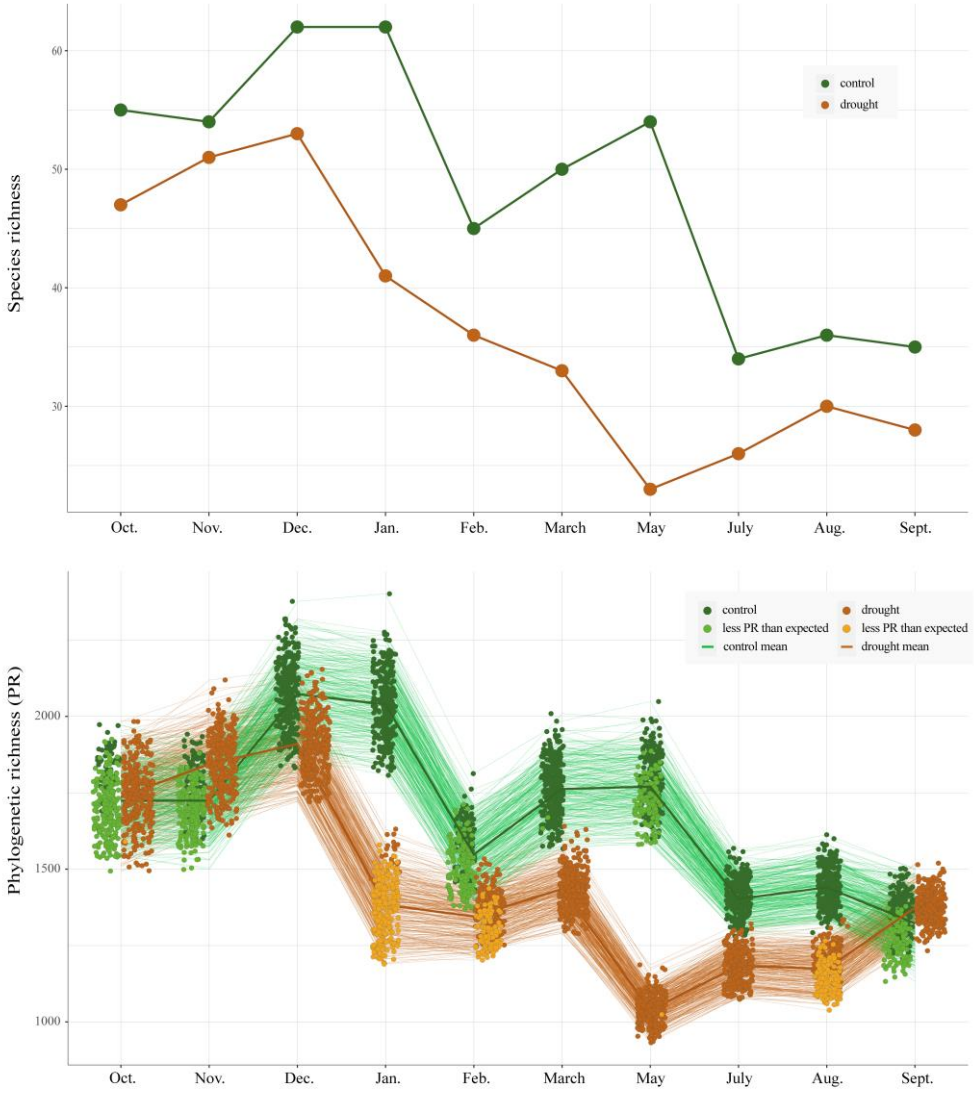
Figure 3. Heatmap displaying the difference in abundance between control and induced drought treatments for each species. Blue coloration indicates a reduction in abundance when comparing the same month in the control and drought conditions. Beige coloration indicates an increase in abundance and grey means it did not change. Genera that exhibited a certain consistency in response (either reduction or increase) during the experiment are highlighted with arrows and color-coded according to the response.



We confirmed our first prediction that induced drought would decrease the phylogenetic richness of the community as consequence of species loss (Figure 4). In the control and drought treatments, we observed variation in phylogenetic richness across the 10-month period (Figure 4). In the first months (October and November), phylogenetic richness showed similar values

for control and drought treatments (Figure 4). In contrast, from December to August, the control community consistently exhibited higher phylogenetic richness than the drought (Figure 4). This trend was most pronounced in January and May, where the richness in the control was considerably higher than in the drought (Figure 4). However, in September, the drought community presented a slight increase in phylogenetic richness relative to the control (Figure 4).

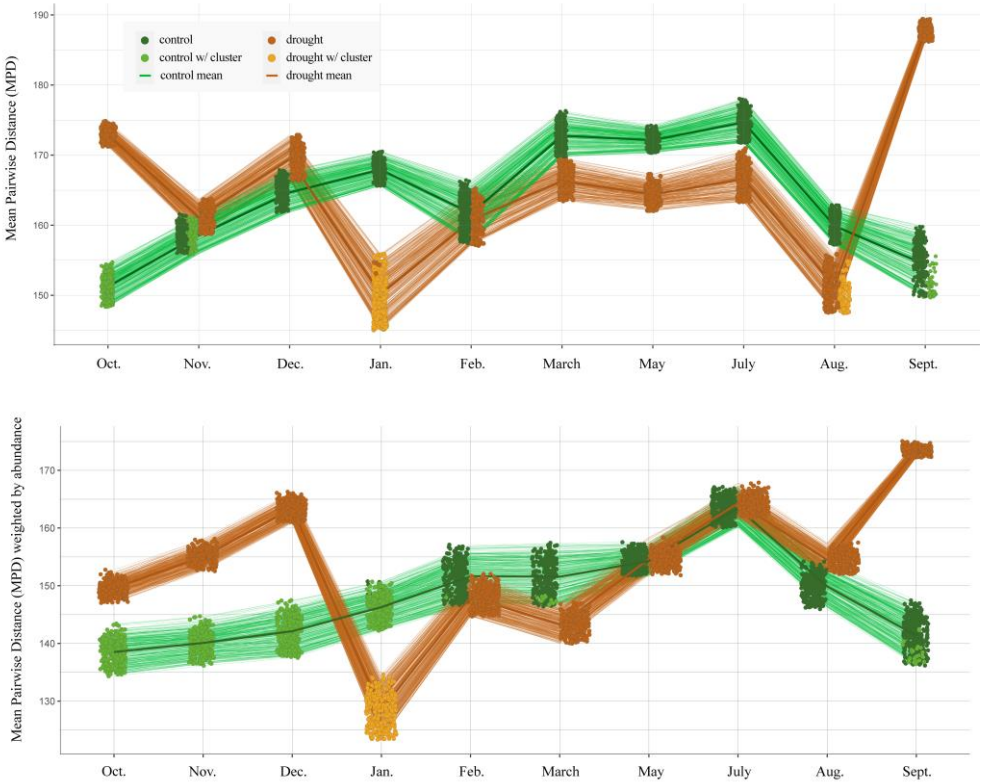
Figure 4. Results of the species richness and phylogenetic richness analyses. Upper graph: Species richness over the course of the 10-month experiment. Lower graph: Phylogenetic richness of the community. Each line and point represent a simulated phylogenetic tree, and the thicker line represents the average phylogenetic richness value for each treatment. The darker colored points represent simulations where the phylogenetic richness was within the expected range given the species richness. Lighter colored points (respectively, light green and light orange), represent simulations where the community exhibited less phylogenetic richness than expected given the species richness.



The Mean Pairwise Distance (MPD) between control and drought treatments also varied across the 10-month period (Figure 5). In October, the drought community showed a higher MPD relative to the control, which presented phylogenetic clustering in all simulated trees (Figure 5). In November, the values equated, but only the control community presented clustering in half simulated trees (Figure 5). In December, the values were similar and none of them presented phylogenetic clustering (Figure 5). The communities displayed a substantial difference between the two treatments in January, with the drought community showing a considerably lower MPD and phylogenetic clustering (Figure 5). From February to August, the MPD values were relatively similar, except that in August the drought community presented

phylogenetic clustering (Figure 5). However, in September, the drought community presented a marked increase in MPD compared to the control, which presented phylogenetic clustering in some simulated trees (Figure 5). When we weighed with the species abundance, the MPD values did not fluctuate in the same way (Figure 5). From October to December, the control community showed a lower MPD and phylogenetic clustering, which were not observed in drought community (Figure 5). In January, the control community shows a higher MPD than the drought group and both display phylogenetic clustering (Figure 5). From February to August, the MPD values between the control and drought communities were similar (Figure 5). However, in September, the drought community displayed a marked increase in MPD compared to the control (Figure 5).

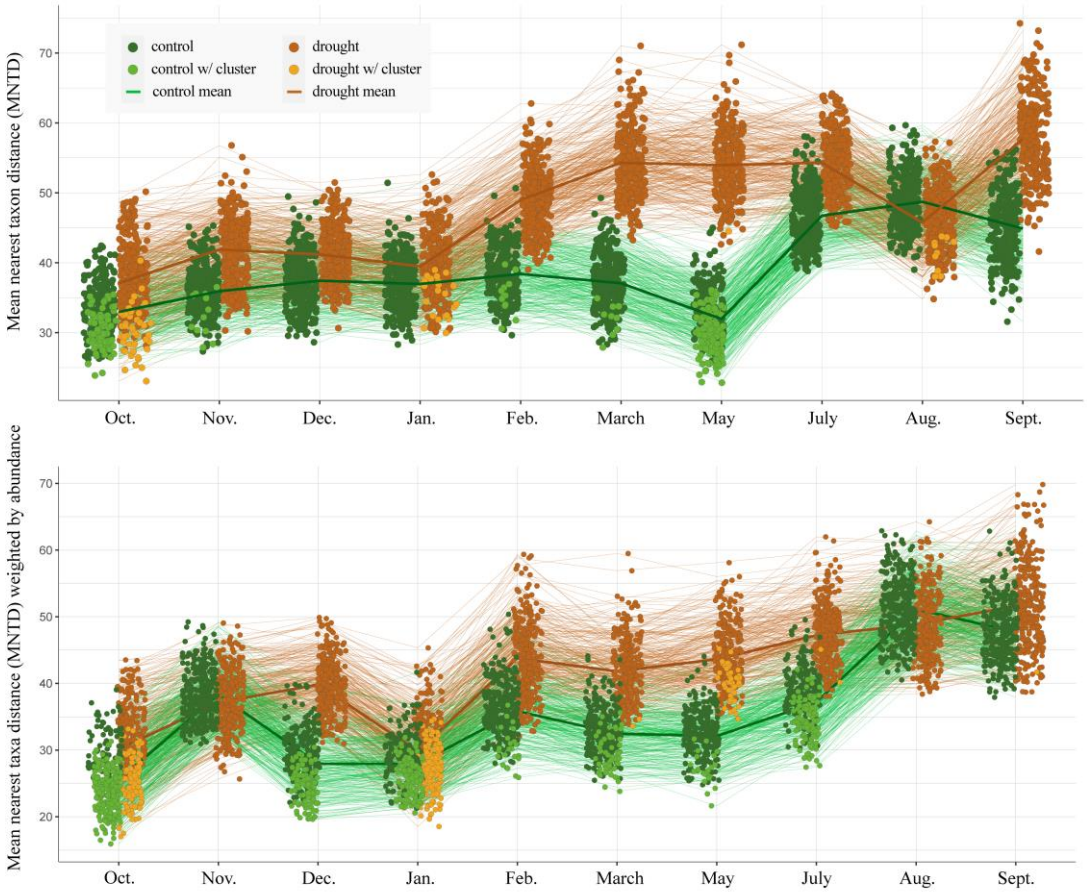
Figure 5. Results from the analyses of mean pairwise distance, with and without considering species abundance. Upper graph: The mean pairwise distance of the communities, not accounting for abundance, over the course of the 10-month experiment. Lower graph: Mean pairwise distance of the communities, weighted with species abundance. Each line and point represent a simulated phylogenetic tree, with the thicker line representing the average mean pairwise distance value for each treatment. The darker colored points represent simulations where the mean pairwise distance was within the expected range given the structure of the phylogenetic tree. Lighter colored points (respectively, light green and light orange), represent simulations where the community exhibited less mean pairwise distance than expected given the structure of the phylogenetic tree, indicating phylogenetic clustering.



Regarding the MNTD metric, the drought community consistently showed higher MNTD values throughout most of the 10-month period (Figure 6). From October to February, the control community shows a consistently lower MNTD than the drought community, with a few simulated trees presenting phylogenetic clustering in both communities in different months (Figure 6). From March to May, the drought group consistently showed a higher MNTD compared to the control community, with the greatest divergence noted in May, when the control presented phylogenetic clustering in most simulated trees (Figure 6). From July to September, the MNTD values of control and drought communities were similar (Figure 6). When we weighed species abundance, the MNTD values were similar, except for December, March and July where drought community presented slightly higher values, no longer presenting the clustering seen in control (Figure 6). MNTD values weighted or not with species

abundance of both the control and drought communities tended to increase over time (Figure 6).

Figure 6. Results from the analyses of mean nearest taxa distance, with and without considering species abundance. Upper graph: The mean nearest taxa distance of the communities, not accounting for abundance, over the course of the 10-month experiment. Lower graph: Mean nearest taxa distance of the communities, weighted with species abundance. Each line and point represent a simulated phylogenetic tree, with the thicker line representing the average mean nearest taxa distance value for each treatment. The darker colored points represent simulations where the mean nearest taxa distance was within the expected range given the structure of the phylogenetic tree. Lighter colored points (respectively, light green and light orange), represent simulations where the community exhibited less mean nearest taxa distance than expected given the structure of the phylogenetic tree, indicating phylogenetic clustering.



With these results, we do not confirm our second prediction that induced drought would select for species with close phylogenetic relationship. In some months the drought community

did present closer phylogenetic distance, but it was not consistent enough throughout the experiment. Considering the mean phylogenetic distances, the control community gradually changed over time presenting phylogenetic clustering only in the first and last months. The drought community presented peaks of high and low mean distances, and the phylogenetic clustering was less frequent. On the other hand, when considering the distance between nearest taxa, the drought community consistently presented higher distances. In addition, both communities presented species loss over time, and it was reflected on the increase of phylogenetic distance between nearest taxa. The inconsistency in the mean distance of drought community suggests that these conditions affect the phylogenetic structure of the community, however, the lack of a distinctive pattern indicates that the response may not be uniform or predictable. Comparing the mean distances with and without abundance weighted, it shows different trends for both control and drought communities, indicating that drought conditions affected the phylogenetic structure at higher levels.

3.5 Discussion

Our study provides evidence that drought conditions cause changes in the phylogenetic structure of Amazonian ant communities, confirming that it can lead to a decrease in phylogenetic richness, primarily driven by a loss of species. This pattern corroborates the observations of previous studies (López-Rubio et al., 2022; Almeida et al., 2023; Delsinne et al., 2013; Li et al., 2019), but different from López-Rubio et al. (2022) which found that phylogenetic diversity increased. Impacts on phylogenetic richness can have ecological implications, because it can buffer against ecological disturbances by increasing the chances of having functionally diverse and complementary species that can maintain ecosystem processes even when some species decline or disappear (Bononi et al., 2018; Hoffmann and Sgrò, 2011;

Carroll and Fox, 2008; Yachi and Loreau, 1999) Its decrease can also result in the loss of unique traits that belong to single branches (Faith et al., 2004), including those not identified yet and non-measurable through current approaches (Tan et al., 2015).

The genera *Blepharidatta* and *Ochetomyrmex*, for instance, which consistently displayed a decrease in abundance throughout the experiment, might be expected to respond uniformly, decreasing in abundance under natural drought conditions. In contrast, species of *Cephalotes* genus and, to some extent, *Rasopone*, demonstrated an overall increase in abundance during the simulated drought conditions, a response that might be expected to occur in natural drought events. However, the overall pattern across all species, genera, and subfamilies was very variable. Therefore, we suggest that the principle of niche conservatism might not be a reliable predictor of responses to drought events for the Amazonian ants. Although we observed a general decrease in species richness and abundance, there was also a wide array of responses, including the increase in abundance of some species, indicating a high degree of variability in drought tolerance traits within this community.

The mean phylogenetic distance of the control community exhibited gradual temporal variation, while drought community displayed peaks, fluctuating between low and high values in consecutive months. These sudden shifts are likely attributable to stochastic sampling of a small fraction of the phylogenetic tree, a consequence of low species richness (Cadotte and Davies, 2016). These observations evidence the impact of species loss on destabilizing the phylogenetic dimension of biodiversity, leading to rapid alternations in the phylogenetic structure of a community over a short temporal scale. These unpredictable changes in phylogenetic structure could lead to decreased ecosystem resilience in the face of further disturbances or stressors. If drought-induced species loss continues, this unpredictability might intensify, potentially resulting in increasingly extreme fluctuations that could further destabilize the community structure and function.

We found that drought conditions decreased the phylogenetic richness through species loss but did not select for species with close phylogenetic relationship, contrary to our expectations. In fact, given the high occurrence of phylogenetic clustering in control community, it surprisingly suggests that it might be the natural state of this community. Based on several studies, it has been hypothesized that phylogenetic overdispersion pattern could be a general pattern for communities at small scales (see discussion in Hendry, 2017). However, in this study, both communities presented phylogenetic clustering in more than one month of the experiment. The drought community presented a phylogenetically clustered pattern due to the predominance of Myrmicinae species while other subfamilies are underrepresented - mostly one species each (Figure 3). Clustering in control community is also a result of this predominance, but other genera were represented by more taxa (Figure 3). This pattern would be traditionally interpreted as a result of abiotic filtering (Webb et al., 2002), where the drought conditions prevented species from surviving while Myrmicinae species have good tolerance to this new environment (Webb et al., 2002). However, these interpretations are only valid when there is niche conservatism, and it should not be assumed once it is highly context dependent (Hendry, 2017). Clustering can also be a result of the difficulty for phylogenetically close species (in this case, the Myrmicinae species) to exclude each other, due to similarity in competitive abilities (Mayfield and Levine, 2010). Local diversification of a particular clade can also favor phylogenetic clustering, because of their high internal relatedness (Pontarp et al., 2012). This is compatible with the group in question, because Myrmicinae is the most diverse and speciose ant subfamily, comprising about half of all ant species described (Ward et al., 2015) and sampled in this experiment. Myrmicinae evolution took place initially in Neotropics (southern Mexico to southern Brazil), where several of its clades diversified (Ward et al., 2015), which can explain why this clade is well-adapted and predominate this environment. This suggests that, in a natural ant community in the Neotropics, Myrmicinae dominates the

landscape and, consequently, the most likely phylogenetic pattern for this group at small scales is the clustering. This highlights the importance of considering the group and biogeographic region in elaborating general phylogenetic patterns at any scale.

We found the general trend of the species loss to increase the phylogenetic distance of nearest taxa (or changes at recent branches), with the highest difference in March and May, months where there were the highest species losses. It is important to highlight that the observed increase in phylogenetic distance, considering the natural clustering tendency, does not signify an increase in phylogenetic diversity. Contrarily, it illustrates the emergence of phylogenetic gaps among species, a direct consequence of species loss and a decrease in genera representation. Thus, this increase in phylogenetic distance should not be extrapolated as a potential increase of functional complementarity. Instead, it reflects the decrease of diversity in species representation and the heightened vulnerability of genera.

Although we found changes in the responses of several genera, it is important to notice that this drought experiment was implemented in a plot located embedded on a preserved natural park in Amazon. Even applying buffer and sampling at the plot center, the community likely receive individuals from the natural surroundings, that act as a species source. In a real situation of intensive drought caused by climate change, the most likely scenario is a much larger area affected by these new conditions, and the influence of the surroundings acting as a species source would decrease.

3.6 Conclusion

Intense drought conditions, much like those experienced by Amazon forests in 2005, resulted in structural changes in the Amazonian ant community. These changes corroborate with the observed in previous studies and models on the impact of climate-change driven drought on

Amazonian forests leading to the loss of biodiversity. We observed that induced drought led to a decrease in the phylogenetic richness of ant community, increase the phylogenetic gaps between species and destabilized the phylogenetic structure of Amazonian ant community. This demonstrates that drought conditions influence both the physiological functions of trees and the biodiversity of key forest inhabitants, such as ants, and how environmental disturbances can lead to changes in forest structure, function, and permeate through various facets of biodiversity, including on phylogenetic diversity. Lastly, this finding also adds to the growing body of evidence highlighting the importance of long-term, manipulative field experiments in understanding how tropical forest communities respond to climate change.

3.7 References

ALMEIDA, R. P. et al. Induced drought strongly affects richness and composition of ground-dwelling ants in the eastern Amazon. **Oecologia**, v. 201, n. 2, 11 ago. 2020.

ANDERSEN, A. Using Ants as bioindicators: Multiscale Issues in Ant Community Ecology. **Conservation Ecology**, v. 1, n. 1, 1997.

ARNAN, X. et al. Increased anthropogenic disturbance and aridity reduce phylogenetic and functional diversity of ant communities in Caatinga dry forest. **Science of The Total Environment**, v. 631-632, p. 429–438, 1 ago. 2018.

BONONI, L. et al. Higher phylogenetic diversity prevents loss of functional diversity caused by successive drying and rewetting cycles. **Antonie van Leeuwenhoek**, v. 111, n. 7, p. 1033–1045, 29 dez. 2017.

BRANDO, P. M. et al. Droughts, Wildfires, and Forest Carbon Cycling: A Pantropical Synthesis. **Annual Review of Earth and Planetary Sciences**, v. 47, n. 1, p. 555–581, 30 maio 2019.

BROECKER, W. S. Climatic Change: Are We on the Brink of a Pronounced Global Warming? **Science**, v. 189, n. 4201, p. 460–463, 8 ago. 1975.

CADOTTE, M. W.; DINNAGE, R.; TILMAN, D. Phylogenetic diversity promotes ecosystem stability. **Ecology**, v. 93, n. sp8, p. S223–S233, ago. 2012.

CADOTTE, M. W.; T. JONATHAN DAVIES. **Phylogenies in Ecology**. [s.l.] Princeton University Press, 2016.

CARROLL, S. P.; FOX, C. W. **Conservation biology : evolution in action**. Oxford ; New York: Oxford University Press, 2008.

COOK, K. H.; VIZY, E. K. Effects of Twenty-First-Century Climate Change on the Amazon Rain Forest. **Journal of Climate**, v. 21, n. 3, p. 542–560, fev. 2008.

COSTA, A. C. L. et al. Stand dynamics modulate water cycling and mortality risk in droughted tropical forest. **Global Change Biology**, v. 24, n. 1, p. 249–258, set. 2017.

DA COSTA, A. C. L. et al. Effect of 7 yr of experimental drought on vegetation dynamics and biomass storage of an eastern Amazonian rainforest. **New Phytologist**, v. 187, n. 3, p. 579–591, 9 jun. 2010.

DEL TORO, I. et al. The little things that run the world revisited: a review of ant-mediated ecosystem services and disservices (Hymenoptera: Formicidae). **Myrmecol. News**, v. 17, n. 10, p. 133–146, 2012.

DELSINNE, T. et al. Ant diversity along a wide rainfall gradient in the Paraguayan dry Chaco. **Journal of Arid Environments**, v. 74, n. 10, p. 1149–1155, out. 2010.

DELSINNE, T.; ARIAS-PENNA, T.; LEPONCE, M. Effect of rainfall exclusion on ant assemblages in montane rainforests of Ecuador. **Basic and Applied Ecology**, v. 14, n. 4, p. 357–365, jun. 2013.

DING, Y. et al. Disturbance regime changes the trait distribution, phylogenetic structure and community assembly of tropical rain forests. **Oikos**, v. 121, n. 8, p. 1263–1270, 21 dez. 2011.

DINNAGE, R. Disturbance Alters the Phylogenetic Composition and Structure of Plant Communities in an Old Field System. **PLoS ONE**, v. 4, n. 9, p. e7071, 18 set. 2009.

DORN, R. I. Ants as a powerful biotic agent of olivine and plagioclase dissolution. **Geology**, v. 42, n. 9, p. 771–774, set. 2014.

EL-BAROUGY, R. et al. Richness, phylogenetic diversity, and abundance all have positive effects on invader performance in an arid ecosystem. **Ecosphere**, v. 11, n. 2, fev. 2020.

FAITH, D. P. Conservation evaluation and phylogenetic diversity. **Biological Conservation**, v. 61, n. 1, p. 1–10, 1992.

FAITH, D. P.; REID, C. A. M.; HUNTER, J. Integrating Phylogenetic Diversity, Complementarity, and Endemism for Conservation Assessment. **Conservation Biology**, v. 18, n. 1, p. 255–261, fev. 2004.

FLORES, B. M.; STAAL, A. Feedback in tropical forests of the Anthropocene. **Global Change Biology**, v. 28, n. 17, p. 5041–5061, 30 jun. 2022.

GRAB, H. et al. Agriculturally dominated landscapes reduce bee phylogenetic diversity and pollination services. **Science**, v. 363, n. 6424, p. 282–284, 17 jan. 2019.

HELMUS, M. R. et al. Communities contain closely related species during ecosystem disturbance. **Ecology Letters**, v. 13, n. 2, p. 162–174, 1 fev. 2010.

HENDRY, A. P. **Eco-evolutionary dynamics**. Princeton: Princeton University Press, 2017.

HOFFMANN, A. A.; SGRÒ, C. M. Climate change and evolutionary adaptation. **Nature**, v. 470, n. 7335, p. 479–485, 23 fev. 2011.

HOOPER, D. U. et al. A Global Synthesis Reveals Biodiversity Loss as a Major Driver of Ecosystem Change. **Nature**, v. 486, n. 7401, p. 105–108, 2 maio 2012.

IPCC. **Managing the risks of extreme events and disasters to advance climate change adaption : special report of the Intergovernmental Panel on Climate Change.** Cambridge: Cambridge University Press, 2012.

LACH, L.; PARR, C. L.; ABBOTT, K. L. **Ant ecology**. Oxford: Oxford University Press, 2010.

LADE, S. J. et al. Human impacts on planetary boundaries amplified by Earth system interactions. **Nature Sustainability**, v. 3, n. 2, p. 119–128, 16 dez. 2019.

LI, D.; MILLER, J. E. D.; HARRISON, S. Climate drives loss of phylogenetic diversity in a grassland community. **Proceedings of the National Academy of Sciences**, v. 116, n. 40, p. 19989–19994, 16 set. 2019.

LÓPEZ-RUBIO, R. et al. Rainy years counteract negative effects of drought on taxonomic, functional and phylogenetic diversity: Resilience in annual plant communities. **Journal of Ecology**, v. 110, n. 10, p. 2308–2320, 6 jul. 2022.

MAYFIELD, M. M.; LEVINE, J. M. Opposing effects of competitive exclusion on the phylogenetic structure of communities. **Ecology Letters**, v. 13, n. 9, p. 1085–1093, 23 jun. 2010.

MIAZAKI, A. S.; GASTAUER, M.; MEIRA-NETO, J. A. A. Environmental severity promotes phylogenetic clustering in campo rupestre vegetation. **Acta Botanica Brasilica**, v. 29, n. 4, p. 561–566, dez. 2015.

MIKAEL PONTARP; RIPA, J.; LUNDBERG, P. On the origin of phylogenetic structure in competitive metacommunities. **Evolutionary ecology research**, v. 14, n. 3, p. 269–284, 1 jan. 2012.

MILLER, E.; FARINE, D.; TRISOS, C. H. Phylogenetic community structure metrics and null models: a review with new methods and software. **Ecography**, v. 40, p. 461–477, 19 mar. 2016.

MOREAU, C. S.; BELL, C. D. Testing the museum versus cradle tropical biological diversity hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the ants. **Evolution**, v. 67, n. 8, p. 2240–2257, 22 abr. 2013.

PARR, C. L. et al. Suppression of savanna ants alters invertebrate composition and influences key ecosystem processes. **Ecology**, v. 97, n. 6, p. 1611–1617, jun. 2016.

PHILLIPS , O. et al. Drought Sensitivity of the Amazon Rainforest. **Science**, v. 323, n. 5919, p. 1344, 2009.

PIO, D. V. et al. Spatial Predictions of Phylogenetic Diversity in Conservation Decision Making. **Conservation Biology**, v. 25, n. 6, p. 1229–1239, 9 nov. 2011.

QI, W. et al. Elevation, moisture and shade drive the functional and phylogenetic meadow communities' assembly in the northeastern Tibetan Plateau. **Community ecology**, v. 16, n. 1, p. 66–75, 1 jun. 2015.

R CORE TEAM. R: A language and environment for statistical computing [Software]. Version 4.3.1. Vienna: R Foundation for Statistical Computing, 2024.

REVELL, L. J. phytools: an R package for phylogenetic comparative biology (and other things). **Methods in Ecology and Evolution**, v. 3, n. 2, p. 217–223, 15 dez. 2011.

ROCKSTRÖM, J.; BAI, X.; DEVRIES, B. Global sustainability: the challenge ahead. **Global Sustainability**, v. 1, 2018.

ROWLAND, L. et al. After more than a decade of soil moisture deficit, tropical rainforest trees maintain photosynthetic capacity, despite increased leaf respiration. **Global Change Biology**, v. 21, n. 12, p. 4662–4672, 22 set. 2015.

ROWLAND, L. et al. Shock and stabilisation following long-term drought in tropical forest from 15 years of litterfall dynamics. **Journal of Ecology**, v. 106, n. 4, p. 1673–1682, 1 fev. 2018.

SCHULTHEISS, P. et al. The abundance, biomass, and distribution of ants on Earth. **Proceedings of the National Academy of Sciences**, v. 119, n. 40, 19 set. 2022.

STEFFEN, W. et al. Planetary boundaries: Guiding human development on a changing planet. **Science**, v. 347, n. 6223, 15 jan. 2015.

STEFFEN, W. et al. Trajectories of the Earth System in the Anthropocene. **Proceedings of the National Academy of Sciences**, v. 115, n. 33, p. 8252–8259, 6 ago. 2018.

TAN, J. et al. Resident-Invader Phylogenetic Relatedness, Not Resident Phylogenetic Diversity, Controls Community Invasibility. **The American Naturalist**, v. 186, n. 1, p. 59–71, jul. 2015.

URBAN, M. C. et al. Improving the forecast for biodiversity under climate change. **Science**, v. 353, n. 6304, p. aad8466–aad8466, 8 set. 2016.

WARD, P. S. et al. The evolution of myrmicine ants: phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera: Formicidae). **Systematic Entomology**, v. 40, n. 1, p. 61–81, 23 jul. 2014.

WEBB, C. O. et al. Phylogenies and Community Ecology. **Annual Review of Ecology and Systematics**, v. 33, n. 1, p. 475–505, nov. 2002.

XU, L.; CHEN, N.; ZHANG, X. Global drought trends under 1.5 and 2 °C warming. **International Journal of Climatology**, v. 39, n. 4, p. 2375–2385, 25 dez. 2018.

YACHI, S.; LOREAU, M. Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis. **Proceedings of the National Academy of Sciences**, v. 96, n. 4, p. 1463–1468, 16 fev. 1999.

4 CONCLUSÃO GERAL

Nesta tese, meus colaboradores e eu investigamos os efeitos de fatores naturais e antropogênicos nas diversidades taxonômica e filogenética das formigas em dois dos mais importantes centros de biodiversidade do mundo: a Floresta Amazônica e a Mata Atlântica. Através de diferentes abordagens, conseguimos explorar os principais fatores que moldam a composição das comunidades de formigas e como as mudanças climáticas podem alterá-las.

No primeiro capítulo, exploramos a influência de variáveis ambientais na composição das comunidades de formigas da Mata Atlântica. Nossos resultados demonstram que, embora as variáveis climáticas expliquem parte da distribuição espacial da diversidade de formigas, os fatores edáficos também emergem como uma influência dominante, mas a combinação entre eles é o que explica a maior parte dessa distribuição. A precipitação durante o mês mais seco do ano e o teor de argila no solo foram identificados como os fatores com maior importância na distribuição da composição de formigas. Esses achados destacam a importância de considerar não apenas as mudanças climáticas, mas também as alterações no uso do solo, que podem modificar significativamente a estrutura e a diversidade dessas comunidades.

No segundo capítulo, investigamos os efeitos de eventos de seca extrema na diversidade filogenética das comunidades de formigas da Floresta Amazônica. Nós encontramos que a seca reduziu a riqueza filogenética de todas as subfamílias, mas não selecionou espécies filogeneticamente próximas. Observamos uma diminuição na riqueza filogenética da comunidade, mas com uma ampla variedade de respostas, inclusive dentro de um mesmo gênero. A seca também levou à mudanças repentinas na estrutura filogenética, que podem levar à diminuição da resiliência e desestabilizar a função das comunidades. Nossos resultados sugerem que a hipótese de conservadorismo de nicho, que assume que espécies

filogeneticamente próximas respondem de maneira semelhante às pressões ambientais, pode não ser um preditor confiável de respostas a eventos de seca para as formigas amazônicas.

De maneira geral, nossos achados reforçam a ideia de que tanto as mudanças climáticas quanto as atividades de uso da terra representam ameaças significativas à biodiversidade das formigas das florestas tropicais. A diversidade dessas comunidades é moldada pela interação entre fatores climáticos e edáficos, e as respostas das formigas a essas mudanças variam amplamente entre espécies e gêneros. Portanto, políticas de conservação e estratégias de manejo devem levar em consideração essa complexidade e a necessidade de proteger tanto o clima quanto a integridade do solo para preservar a biodiversidade e a resiliência desses ecossistemas.

APÊNDICE A

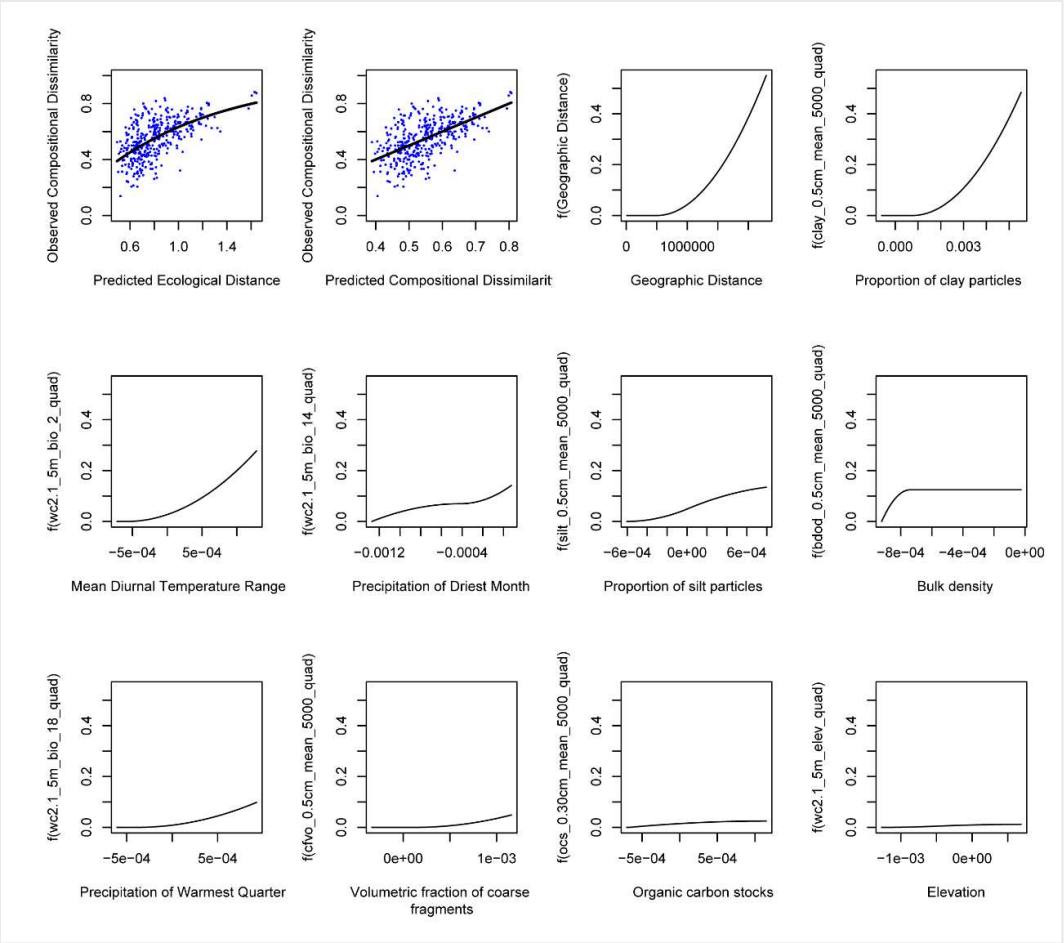
Apêndice referente ao Capítulo 1: "**Soil and climatic effects: How do environmental factors shape the ant community similarity in the Atlantic Forest?**".

Essa seção será incluída como Material Suplementar do artigo.

GDM panels

The GDM panels illustrate the relationships between the environmental variables and the compositional dissimilarity of ant communities within the Atlantic Forest. The x-axis represents the variation of an environmental variable (converted to polynomial), while the y-axis indicates the biological response, according to the model. The peak height of each spline represents the total extent of biological variation across the gradient, highlighting the relative significance of that predictor in driving biological shifts while other variables remain unchanged (i.e., it represents a specific ecological distance) (Mokany 2022). The form of the spline adjustments demonstrates how the pace of biological transformation changes based on its position along the gradient (Mokany 2022). Therefore, each graph depicts how changes in the environmental variables are associated with changes in the diversity of ant communities (Figure 1).

Figure 1. The fitted model (first two panels) and I-splines for all environmental variables selected by the model (remaining panels).



The first panel of Figure 1 illustrates a positive relationship between observed compositional dissimilarity and predicted ecological distance. The scatter plot indicates that as the predicted ecological distance increases, the compositional dissimilarity among communities also increases. However, the relationship is not strictly proportional; the rate of increase in dissimilarity appears to diminish as the ecological distance becomes greater, suggesting a decelerating curve as the ecological gradient extends. The second panel presents a positive linear trend between observed and predicted compositional dissimilarity. The arrangement of data points and line trend suggests that the predictions of the model are consistent with the observed dissimilarities, with the rate of change in observed dissimilarity being directly proportional to the rate of change in predicted dissimilarity. The third panel indicates a relationship between observed compositional dissimilarity and geographic distance. At smaller geographic distances, the compositional dissimilarity remains relatively low, indicating similar

community compositions. However, there is a critical point beyond which the dissimilarity sharply increases, suggesting that beyond this specific spatial threshold, community compositions become significantly more dissimilar as the geographic distance continues to expand.

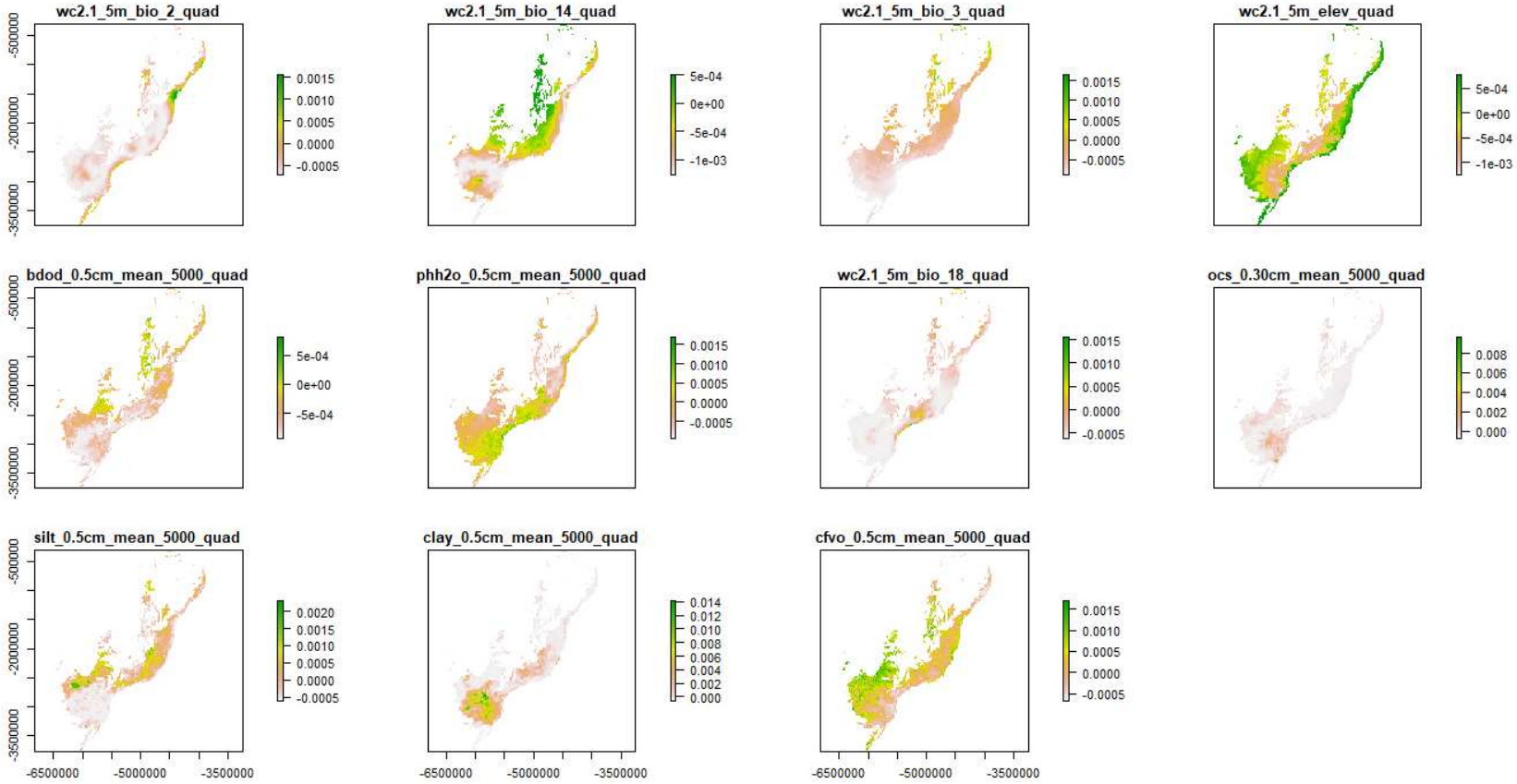
The fourth panel displays the trend between observed compositional dissimilarity and the proportion of clay particles in the soil. The rate of increase in dissimilarity appears to be proportional as the proportion of clay in the soil increases. The fifth panel demonstrates the relationship between observed compositional dissimilarity and mean diurnal temperature range. The curve suggests that compositional dissimilarity increases gradually as the temperature range increase. The sixth panel show the relationship between observed compositional dissimilarity and precipitation during the driest month. The observed compositional dissimilarity exhibits an increase as the amount of precipitation in the driest month increases. The seventh panel depicts the relationship between the observed compositional dissimilarity and the proportion of silt particles. This relationship is characterized by a gradual increase in dissimilarity as the proportion of silt particles increases. The eighth panel presents the relationship between observed compositional dissimilarity and soil bulk density. As bulk density initially increases, we observe a sharp increase in compositional dissimilarity. However, beyond this initial rise, the curve flattens, indicating that further increases in bulk density do not correspond to significant changes in dissimilarity. The ninth panel presents a correlation between observed compositional dissimilarity and precipitation of the warmest quarter. It shows a marginal initial rise in dissimilarity, but only with a considerable increase in precipitation. Even though, the increase in dissimilarity remains low, indicating that while precipitation is a contributing factor, its impact on dissimilarity is relatively restrained. The tenth panel presents a correlation between observed compositional dissimilarity and the volumetric fraction of coarse fragments in soil, with a similar pattern to the last one, but with an even greater threshold

of coarse fragment fraction must be surpassed to observe any discernible increase in dissimilarity. Once this threshold is crossed, the escalation in dissimilarity is minimal, indicating that the volumetric fraction of coarse fragments has a very limited influence on community composition. The eleventh and last panel illustrates the lack of a direct correlation of organic carbon stocks and elevation with ant similarity, as depicted by the flat line in the response axis. This implies that changes in carbon stocks and elevation neither increase nor decrease the dissimilarity among ant communities.

Spatial distribution of the climate and edaphic variables selected by the model

Figure 2 (below) presents the spatial distribution of the variables initially selected by the model. The correspondence between the codes used and the variables is:

- bio 2: Mean Diurnal Range (Mean of monthly (max temp - min temp))
- bio 14: Precipitation of Driest Month
- bio 3: Isothermality (BIO2/BIO7) (×100)
- elev: Elevation
- bdod: Bulk density of the fine earth fraction
- phh2o: soil pH
- bio 18: Precipitation of Warmest Quarter
- ocs: Organic carbon stocks
- silt: Proportion of silt particles (≥ 0.002 mm and ≤ 0.05/0.063 mm) in the fine earth fraction
- clay: Proportion of clay particles (< 0.002 mm) in the fine earth fraction
- cfvo: Volumetric fraction of coarse fragments (> 2 mm)


wc2.1_5m_bio_2_quad
wc2.1_5m_bio_14_quad
wc2.1_5m_bio_3_quad
wc2.1_5m_elev_quad
bdod_0.5cm_mean_5000_quad
phh2o_0.5cm_mean_5000_quad
wc2.1_5m_bio_18_quad
ocs_0.30cm_mean_5000_quad
silt_0.5cm_mean_5000_quad
clay_0.5cm_mean_5000_quad
cfvo_0.5cm_mean_5000_quad
-500000
-2000000
-3500000
-6500000
-5000000
-3500000

Script Report

O conteúdo a seguir é um relatório com o *script* feito no R e utilizado para processar os dados e gerar os resultados presente nesse estudo. Esse conteúdo não será submetido junto ao artigo, é apenas informativo caso os membros da banca queiram acessá-lo. Esse código foi originalmente desenvolvido pelo Prof. Dr. Mario Moura (mariormoura@gmail.com) e adaptado pela autora desta tese para este estudo, sendo ela a responsável por qualquer incoerência em decorrência desta adaptação. Os próximos passos deste estudo incluirão uma verificação cuidadosa deste código pelos coautores.

```
## Installing packages into 'C:/Users/isagv/AppData/Local/R/win-library/4.
3'
## (as 'lib' is unspecified)

## Warning: packages 'BuenColors', 'ENMTML', 'ggsflabel', 'MSDM' are not a
vailable for this version of R
##
## Versions of these packages for your version of R might be available els
ewhere,
## see the ideas at
## https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installin
g-packages

lapply(needed_packages, require, character.only = TRUE)

## Carregando pacotes exigidos: ade4

## Carregando pacotes exigidos: agricolae

## Warning: package 'agricolae' was built under R version 4.3.3
```

```
## Carregando pacotes exigidos: ape

##

## Attaching package: 'ape'

## The following object is masked from 'package:agricolae':

##

##     consensus

## Carregando pacotes exigidos: BAT

## Warning: package 'BAT' was built under R version 4.3.3

## Error: package or namespace load failed for 'BAT' in loadNamespace(i, c
(lib.loc, .libPaths()), versionCheck = vI[[i]]):
##  there is no package called 'recipes'

## Carregando pacotes exigidos: beepr

## Carregando pacotes exigidos: betapart

## Carregando pacotes exigidos: BuenColors

## Warning in library(package, lib.loc = lib.loc, character.only = TRUE,
## logical.return = TRUE, : there is no package called 'BuenColors'

## Carregando pacotes exigidos: cluster

## Carregando pacotes exigidos: CommEcol

## Carregando pacotes exigidos: vegan

## Carregando pacotes exigidos: permute

## Carregando pacotes exigidos: lattice
```

```
## This is vegan 2.6-4

## Carregando pacotes exigidos: picante

## Carregando pacotes exigidos: nlme

## Carregando pacotes exigidos: adespatial

## Warning: package 'adespatial' was built under R version 4.3.3

## Registered S3 methods overwritten by 'adegraphics':
##   method         from
##   biplot.dudi    ade4
##   kplot.foucart  ade4
##   kplot.mcoa     ade4
##   kplot.mfa      ade4
##   kplot.pta      ade4
##   kplot.sepan    ade4
##   kplot.statis   ade4
##   scatter.coa    ade4
##   scatter.dudi   ade4
##   scatter.nipals ade4
##   scatter.pco    ade4
##   score.acm      ade4
##   score.mix      ade4
##   score.pca      ade4
##   screeplot.dudi ade4

## Registered S3 method overwritten by 'spdep':
##   method   from
##   plot.mst ape
```

```
## Registered S3 methods overwritten by 'adespatial':
##   method             from
##   plot.multispati    adegraphics
##   print.multispati   ade4
##   summary.multispati ade4

##
## Attaching package: 'adespatial'

## The following object is masked from 'package:ade4':
##
##     multispati

## Carregando pacotes exigidos: data.table

## Carregando pacotes exigidos: dismo

## Carregando pacotes exigidos: raster

## Warning: package 'raster' was built under R version 4.3.3

## Carregando pacotes exigidos: sp

## Warning: package 'sp' was built under R version 4.3.3

##
## Attaching package: 'raster'

## The following object is masked from 'package:nlme':
##
##     getData

## Carregando pacotes exigidos: doParallel

## Carregando pacotes exigidos: foreach
```

```
## Carregando pacotes exigidos: iterators

## Carregando pacotes exigidos: parallel

## Carregando pacotes exigidos: dplyr

##

## Attaching package: 'dplyr'

## The following objects are masked from 'package:raster':

##

##     intersect, select, union

## The following objects are masked from 'package:data.table':

##

##     between, first, last

## The following object is masked from 'package:nlme':

##

##     collapse

## The following object is masked from 'package:ape':

##

##     where

## The following objects are masked from 'package:stats':

##

##     filter, lag

## The following objects are masked from 'package:base':

##

##     intersect, setdiff, setequal, union
```

```
## Carregando pacotes exigidos: FD

## Carregando pacotes exigidos: geometry

##
## Attaching package: 'FD'

## The following object is masked from 'package:dismo':
##
##     maxent

## Carregando pacotes exigidos: ENMTML

## Warning in library(package, lib.loc = lib.loc, character.only = TRUE,
## logical.return = TRUE, : there is no package called 'ENMTML'

## Carregando pacotes exigidos: flexclust

## Carregando pacotes exigidos: grid

## Carregando pacotes exigidos: modeltools

## Carregando pacotes exigidos: stats4

## Carregando pacotes exigidos: gdm

## Carregando pacotes exigidos: geiger## Carregando pacotes exigidos: phyt
ools

## Carregando pacotes exigidos: maps

##
## Attaching package: 'maps'
```

```
## The following object is masked from 'package:cluster':
##
##     votes.repub
##
## Attaching package: 'phytools'
## The following object is masked from 'package:vegan':
##
##     scores
##
## Attaching package: 'geiger'
## The following object is masked from 'package:raster':
##
##     hdr
## Carregando pacotes exigidos: ggcorrplot
## Warning: package 'ggcorrplot' was built under R version 4.3.3
## Carregando pacotes exigidos: ggplot2
## Carregando pacotes exigidos: ggnewscale
## Warning: package 'ggnewscale' was built under R version 4.3.3
## Carregando pacotes exigidos: ggrepel
## Warning: package 'ggrepel' was built under R version 4.3.3
## Carregando pacotes exigidos: ggsflabel
```

```
## Warning in library(package, lib.loc = lib.loc, character.only = TRUE,
## logical.return = TRUE, : there is no package called 'ggsflabel'

## Carregando pacotes exigidos: ggspatial

## Warning: package 'ggspatial' was built under R version 4.3.3

## Carregando pacotes exigidos: ggtree

## ggtree v3.11.1.001 For help: https://yulab-smu.top/treedata-book/
##
## If you use the ggtree package suite in published research, please cite
## the appropriate paper(s):
##
## Guangchuang Yu, David Smith, Huachen Zhu, Yi Guan, Tommy Tsan-Yuk Lam.
## ggtree: an R package for visualization and annotation of phylogenetic
## trees with their covariates and other associated data. Methods in
## Ecology and Evolution. 2017, 8(1):28-36. doi:10.1111/2041-210X.12628
##
## LG Wang, TTY Lam, S Xu, Z Dai, L Zhou, T Feng, P Guo, CW Dunn, BR
## Jones, T Bradley, H Zhu, Y Guan, Y Jiang, G Yu. treeio: an R package
## for phylogenetic tree input and output with richly annotated and
## associated data. Molecular Biology and Evolution. 2020, 37(2):599-603.
## doi: 10.1093/molbev/msz240
##
## Shuangbin Xu, Lin Li, Xiao Luo, Meijun Chen, Wenli Tang, Li Zhan, Zehan
## Dai, Tommy T. Lam, Yi Guan, Guangchuang Yu. Ggtree: A serialized data
## object for visualization of a phylogenetic tree and annotation data.
## iMeta 2022, 1(4):e56. doi:10.1002/imt2.56
```

```
##
## Attaching package: 'ggtree'

## The following objects are masked from 'package:raster':
##
##     flip, rotate

## The following object is masked from 'package:nlme':
##
##     collapse

## The following object is masked from 'package:ape':
##
##     rotate

## Carregando pacotes exigidos: leafsync

## Carregando pacotes exigidos: mapview

## Warning: package 'mapview' was built under R version 4.3.3

## Carregando pacotes exigidos: metacom

## Carregando pacotes exigidos: MSDM

## Warning in library(package, lib.loc = lib.loc, character.only = TRUE,
## logical.return = TRUE, : there is no package called 'MSDM'## Carregando
pacotes exigidos: ncf

##
## Attaching package: 'ncf'
```

```
## The following object is masked from 'package:vegan':
##
##     mantel.correlog
## The following object is masked from 'package:ape':
##
##     mantel.test
## Carregando pacotes exigidos: PerformanceAnalytics
## Carregando pacotes exigidos: xts
## Warning: package 'xts' was built under R version 4.3.3
## Carregando pacotes exigidos: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
##
## ############################ Warning from 'xts' package ##################
########
## #
#
## # The dplyr lag() function breaks how base R's lag() function is suppos
ed to  #
## # work, which breaks lag(my_xts). Calls to lag(my_xts) that you type or
#
```

```
## # source() into this session won't work correctly.
#
## #
#
## # Use stats::lag() to make sure you're not using dplyr::lag(), or you c
an add #
## # conflictRules('dplyr', exclude = 'lag') to your .Rprofile to stop
#
## # dplyr from breaking base R's lag() function.
### # Code in packages is not affected. It's protected by R's namespace me
chanism #
## # Set `options(xts.warn_dplyr_breaks_lag = FALSE)` to suppress this war
ning.  #
## #
#
## ##############################################################################
########

##
## Attaching package: 'xts'

## The following objects are masked from 'package:dplyr':
##
##     first, last

## The following objects are masked from 'package:data.table':
##
##     first, last
```

```
##
## Attaching package: 'PerformanceAnalytics'
## The following objects are masked from 'package:agricolae':
##
##     kurtosis, skewness
## The following object is masked from 'package:graphics':
##
##     legend
## Carregando pacotes exigidos: pgirmess
## Warning: package 'pgirmess' was built under R version 4.3.3
## Registered S3 method overwritten by 'pgirmess':
##   method        from
##   plot.correlog ncf
##
## Attaching package: 'pgirmess'
## The following object is masked from 'package:ncf':
##
##     correlog
## Carregando pacotes exigidos: phangorn
##
## Attaching package: 'phangorn'
```

```
## The following object is masked from 'package:pgirmess':
##
##     CI

## The following objects are masked from 'package:vegan':
##
##     diversity, treedist

## Carregando pacotes exigidos: plyr

## ------------------------------------------------------------------------
-------

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first,
then dplyr:
## library(plyr); library(dplyr)

## ------------------------------------------------------------------------
-------

##
## Attaching package: 'plyr'

## The following object is masked from 'package:maps':
##
##     ozone

## The following object is masked from 'package:modeltools':
##
##     empty
```

```
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

## Carregando pacotes exigidos: prettyGraphs

## Carregando pacotes exigidos: psych

## Warning: package 'psych' was built under R version 4.3.3

##
## Attaching package: 'psych'

## The following object is masked from 'package:pgirmess':
##
##     shannon

## The following object is masked from 'package:ncf':
##
##     cor2

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha

## The following object is masked from 'package:phytools':
##
##     rescale

## The following object is masked from 'package:raster':
##
##     distance
```

```
## Carregando pacotes exigidos: RColorBrewer

## Carregando pacotes exigidos: reshape2

##

## Attaching package: 'reshape2'

## The following objects are masked from 'package:data.table':

##

##     dcast, melt

## Carregando pacotes exigidos: Rfast

## Warning: package 'Rfast' was built under R version 4.3.3

## Carregando pacotes exigidos: Rcpp

## Carregando pacotes exigidos: RcppZiggurat

## Carregando pacotes exigidos: RcppParallel

## Warning: package 'RcppParallel' was built under R version 4.3.3

##

## Attaching package: 'RcppParallel'

## The following object is masked from 'package:Rcpp':

##

##     LdFlags

##

## Rfast: 2.1.0

## Attaching package: 'Rfast'
```

```
## The following object is masked from 'package:psych':
##
##     skew

## The following object is masked from 'package:dplyr':
##
##     nth

## The following object is masked from 'package:data.table':
##
##     transpose

## The following object is masked from 'package:ape':
##
##     yule

## Carregando pacotes exigidos: installr

##
## Welcome to installr version 0.23.4
##
## More information is available on the installr project website:
## https://github.com/talgalili/installr/
##
## Contact: <tal.galili@gmail.com>
## Suggestions and bug-reports can be submitted at: https://github.com/tal
galili/installr/issues
##
##          To suppress this message use:
##          suppressPackageStartupMessages(library(installr))
```

```
## Carregando pacotes exigidos: rgdal

## Please note that rgdal will be retired during October 2023,
## plan transition to sf/stars/terra functions using GDAL and PROJ
## at your earliest convenience.
## See https://r-spatial.org/r/2023/05/15/evolution4.html and https://gith
ub.com/r-spatial/evolution
## rgdal: version: 1.6-7, (SVN revision 1203)
## Geospatial Data Abstraction Library extensions to R successfully loaded
## Loaded GDAL runtime: GDAL 3.6.2, released 2023/01/02
## Path to GDAL shared files: C:/Users/isagv/AppData/Local/R/win-library/4
.3/rgdal/gdal
##  GDAL does not use iconv for recoding strings.
## GDAL binary built with GEOS: TRUE
## Loaded PROJ runtime: Rel. 9.2.0, March 1st, 2023, [PJ_VERSION: 920]
## Path to PROJ shared files: C:/Users/isagv/AppData/Local/R/win-library/4
.3/rgdal/proj
## PROJ CDN enabled: FALSE
## Linking to sp version:2.0-0
## To mute warnings of possible GDAL/OSR exportToProj4() degradation,
## use options("rgdal_show_exportToProj4_warnings"="none") before loading
sp or rgdal.

## Carregando pacotes exigidos: letsR

## Carregando pacotes exigidos: terra

## Warning: package 'terra' was built under R version 4.3.3

## terra 1.7.71
```

```
##
## Attaching package: 'terra'

## The following object is masked from 'package:rgdal':
##
##     project

## The following object is masked from 'package:installr':
##
##     is.empty

## The following objects are masked from 'package:psych':
##
##     describe, distance, rescale

## The following object is masked from 'package:zoo':
##
##     time<-

## The following objects are masked from 'package:ggtree':
##
##     flip, inset, rotate

## The following object is masked from 'package:phytools':
##
##     rescale

## The following object is masked from 'package:grid':
##
##     depth
```

```
## The following object is masked from 'package:data.table':
##
##     shift

## The following objects are masked from 'package:ape':
##
##     rotate, trans, zoom

## Carregando pacotes exigidos: rgeos

## rgeos version: 0.6-4, (SVN revision 699)
##  GEOS runtime version: 3.11.2-CAPI-1.17.2
##  Please note that rgeos will be retired during October 2023,
## plan transition to sf or terra functions using GEOS at your earliest co
nvenience.
## See https://r-spatial.org/r/2023/05/15/evolution4.html for details.
##  GEOS using OverlayNG
##  Linking to sp version: 2.0-0
##  Polygon checking: TRUE##
## Attaching package: 'rgeos'

## The following object is masked from 'package:dplyr':
##
##     symdiff

## Carregando pacotes exigidos: rnaturalearth

## Warning: package 'rnaturalearth' was built under R version 4.3.3

## Carregando pacotes exigidos: RStoolbox

## Warning: package 'RStoolbox' was built under R version 4.3.3
```

```
## Error: package or namespace load failed for 'RStoolbox' in loadNamespac
e(i, c(lib.loc, .libPaths()), versionCheck = vI[[i]]):
##  there is no package called 'recipes'
## Carregando pacotes exigidos: sf
## Warning: package 'sf' was built under R version 4.3.3
## Linking to GEOS 3.11.2, GDAL 3.8.2, PROJ 9.3.1; sf_use_s2() is TRUE
## Carregando pacotes exigidos: SpatialPack
## Carregando pacotes exigidos: fastmatrix
## Warning: package 'fastmatrix' was built under R version 4.3.2
##
## Attaching package: 'fastmatrix'
## The following object is masked from 'package:psych':
##
##     minkowski
## The following object is masked from 'package:phangorn':
##
##     hadamard
## The following objects are masked from 'package:PerformanceAnalytics':
##
##     kurtosis, skewness
## The following object is masked from 'package:data.table':
##
##     frank
```

```
## The following object is masked from 'package:nlme':
##
##     corAR1

## The following objects are masked from 'package:agricolae':
##
##     kurtosis, skewness

## Carregando pacotes exigidos: spThin

## Carregando pacotes exigidos: spam

## Warning: package 'spam' was built under R version 4.3.3

## Spam version 2.10-0 (2023-10-23) is loaded.
## Type 'help( Spam)' or 'demo( spam)' for a short introduction
## and overview of this package.
## Help for individual functions is also obtained by adding the
## suffix '.spam' to the function name, e.g. 'help( chol.spam)'.

##
## Attaching package: 'spam'

## The following objects are masked from 'package:Rfast':
##
##     permutation, rmvnorm, rmvt

## The following object is masked from 'package:stats4':
##
##     mle
```

```
## The following objects are masked from 'package:base':
##
##     backsolve, forwardsolve
## Carregando pacotes exigidos: fields
## Warning: package 'fields' was built under R version 4.3.3
## Carregando pacotes exigidos: viridisLite
##
## Try help(fields) to get started.
##
## Attaching package: 'fields'
## The following object is masked from 'package:terra':
##
##     describe
## The following object is masked from 'package:psych':
##
##     describe
## The following object is masked from 'package:xts':
##
##     addLegend
## Carregando pacotes exigidos: knitr
## Warning: package 'knitr' was built under R version 4.3.3
##
## Attaching package: 'knitr'
```

```
## The following object is masked from 'package:terra':
##
##     spin
## Carregando pacotes exigidos: stars
## Warning: package 'stars' was built under R version 4.3.3
## Carregando pacotes exigidos: abind
## Carregando pacotes exigidos: stringr
## Carregando pacotes exigidos: svMisc
##
## Attaching package: 'svMisc'
## The following object is masked from 'package:ape':
##
##     def
## The following object is masked from 'package:utils':
##
##     ?
## Carregando pacotes exigidos: tools
## Carregando pacotes exigidos: usdm
## Warning: package 'usdm' was built under R version 4.3.3
##
## Attaching package: 'usdm'
```

```
## The following object is masked from 'package:ncf':
##
##     lisa

## The following object is masked from 'package:nlme':
##
##     Variogram

## Carregando pacotes exigidos: viridis

## Warning: package 'viridis' was built under R version 4.3.3
## Attaching package: 'viridis'## The following object is masked from 'pac
kage:maps':
##
##     unemp

## Carregando pacotes exigidos: weights

## Carregando pacotes exigidos: Hmisc

## Warning: package 'Hmisc' was built under R version 4.3.3

##
## Attaching package: 'Hmisc'

## The following object is masked from 'package:fields':
##
##     describe

## The following object is masked from 'package:fastmatrix':
##
##     bezier
```

```
## The following object is masked from 'package:rgeos':
##
##     translate

## The following objects are masked from 'package:terra':
##
##     describe, mask, units, zoom

## The following object is masked from 'package:psych':
##
##     describe

## The following objects are masked from 'package:plyr':
##
##     is.discrete, summarize

## The following objects are masked from 'package:dplyr':
##
##     src, summarize

## The following objects are masked from 'package:raster':
##
##     mask, zoom

## The following object is masked from 'package:ape':
##
##     zoom

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
## Warning in check_dep_version(): ABI version mismatch:
## lme4 was built with Matrix ABI version 1
## Current Matrix ABI version is 0
## Please re-install lme4 from source or restore original 'Matrix' package

## Carregando pacotes exigidos: tidyverse

## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✔ forcats   1.0.0     ✔ readr     2.1.4
## ✔ lubridate 1.9.2     ✔ tibble    3.2.1
## ✔ purrr     1.0.2     ✔ tidyr     1.3.0
## ── Conflicts ────────────────────────────────────────── tidyverse_confl
icts() ──
## ✖ psych::%+%()         masks ggplot2::%+%()
## ✖ purrr::accumulate()  masks foreach::accumulate()
## ✖ psych::alpha()       masks ggplot2::alpha()
## ✖ plyr::arrange()      masks dplyr::arrange()
## ✖ dplyr::between()     masks data.table::between()
## ✖ ggtree::collapse()   masks dplyr::collapse(), nlme::collapse()
## ✖ purrr::compact()     masks plyr::compact()
## ✖ plyr::count()        masks dplyr::count()
## ✖ plyr::desc()         masks dplyr::desc()
## ✖ tidyr::expand()      masks ggtree::expand()
## ✖ tidyr::extract()     masks terra::extract(), raster::extract()
## ✖ plyr::failwith()     masks dplyr::failwith()
```

```
## ✖ dplyr::filter()      masks stats::filter()
## ✖ xts::first()         masks dplyr::first(), data.table::first()
## ✖ tidyr::gather()      masks ncf::gather()
## ✖ lubridate::hour()    masks data.table::hour()
## ✖ plyr::id()           masks dplyr::id()
## ✖ purrr::is_integer()  masks Rfast::is_integer()
## ✖ lubridate::isoweek() masks data.table::isoweek()
## ✖ dplyr::lag()         masks stats::lag()
## ✖ xts::last()          masks dplyr::last(), data.table::last()
## ✖ purrr::map()         masks maps::map()
## ✖ lubridate::mday()    masks data.table::mday()
## ✖ lubridate::minute()  masks data.table::minute()
## ✖ lubridate::month()   masks data.table::month()
## ✖ plyr::mutate()       masks dplyr::mutate()
## ✖ Rfast::nth()         masks dplyr::nth()
## ✖ lubridate::quarter() masks data.table::quarter()
## ✖ plyr::rename()       masks dplyr::rename()
## ✖ lubridate::second()  masks data.table::second()
## ✖ dplyr::select()      masks raster::select()
## ✖ Hmisc::src()         masks dplyr::src()
## ✖ plyr::summarise()    masks dplyr::summarise()
## ✖ Hmisc::summarize()   masks plyr::summarize(), dplyr::summarize()
## ✖ rgeos::symdiff()     masks dplyr::symdiff()
## ✖ purrr::transpose()   masks Rfast::transpose(), data.table::transpose
```

```
()
## ✖ lubridate::wday()    masks data.table::wday()
## ✖ lubridate::week()    masks data.table::week()
## ✖ purrr::when()        masks foreach::when()
## ✖ dplyr::where()       masks ape::where()
## ✖ lubridate::yday()    masks data.table::yday()
## ✖ lubridate::year()    masks data.table::year()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force
all conflicts to become errors
## Carregando pacotes exigidos: phylosignal
##
##
## Attaching package: 'phylosignal'
##
##
## The following object is masked from 'package:lattice':
##
##     dotplot
##
##
## Carregando pacotes exigidos: adephylo
##
## Carregando pacotes exigidos: phylobase
##
##
## Attaching package: 'phylobase'
```

```
##
##
## The following object is masked from 'package:ggtree':
##
##     MRCA
##
##
## The following object is masked from 'package:phytools':
##
##     readNexus
##
##
## The following object is masked from 'package:ape':
##
##     edges
##
##
## Carregando pacotes exigidos: phylotools
##
## Carregando pacotes exigidos: hrbrthemes
##
## NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use
these themes.
##
##       Please use hrbrthemes::import_roboto_condensed() to install Robot
o Condensed and
##
```

```
##       if Arial Narrow is not on your system, please see https://bit.ly/
arialnarrow
##
## Carregando pacotes exigidos: devtools
##
## Carregando pacotes exigidos: usethis
##
##
## Attaching package: 'devtools'
##
##
## The following object is masked from 'package:permute':
##
##     check
##
##
## Carregando pacotes exigidos: evobiR
##
##
## Attaching package: 'evobiR'
##
##
## The following object is masked from 'package:phangorn':
##
##     AICc
##
```

```
##
## Carregando pacotes exigidos: pez
## Warning in .recacheSubclasses(def@className, def, env): subclasse "ndiM
atrix"
## da classe "replValueSp" não definida; definição não atualizada
library(rmarkdown)
## Warning: package 'rmarkdown' was built under R version 4.3.3
#load("Capitulo_modelagem.RData")




################################################################################
#
#
# DADOS AMBIENTAIS - CLIMÁTICOS E DE SOLO
#
################################################################################
#

# Load data on environmental variables:
path_current_predictors<-"climate/" # Specify the path to bioclim layers f
or LGM CCSM4
bioclim<-raster::stack(list.files(path=path_current_predictors, pattern='.
tif', full.names=T))
bioclim$wc2.1_5m_bio_1
```

```
## class      : RasterLayer
## dimensions : 2160, 4320, 9331200  (nrow, ncol, ncell)
## resolution : 0.08333333, 0.08333333  (x, y)
## extent     : -180, 180, -90, 90  (xmin, xmax, ymin, ymax)
## crs        : +proj=longlat +datum=WGS84 +no_defs
## source     : wc2.1_5m_bio_1.tif
## names      : wc2.1_5m_bio_1
## values     : -54.73946, 31.05112  (min, max)

print(crs(bioclim))## Coordinate Reference System:
## Deprecated Proj.4 representation: +proj=longlat +datum=WGS84 +no_defs
## WKT2 2019 representation:
## GEOGCRS["unknown",
##     DATUM["World Geodetic System 1984",
##         ELLIPSOID["WGS 84",6378137,298.257223563,
##             LENGTHUNIT["metre",1]],
##         ID["EPSG",6326]],
##     PRIMEM["Greenwich",0,
##         ANGLEUNIT["degree",0.0174532925199433],
##         ID["EPSG",8901]],
##     CS[ellipsoidal,2],
##         AXIS["longitude",east,
##             ORDER[1],
##             ANGLEUNIT["degree",0.0174532925199433,
##                 ID["EPSG",9122]]],
##         AXIS["latitude",north,
##             ORDER[2],
```

```
##             ANGLEUNIT["degree",0.0174532925199433,
##                 ID["EPSG",9122]]]]

path_current_predictors_soil<-"soil/" # Specify the path to bioclim layers
for LGM CCSM4
soil<-raster::stack(list.files(path=path_current_predictors_soil, pattern=
'.tif', full.names=T))
soil$bdod_0.5cm_mean_5000

## class      : RasterLayer
## dimensions : 2902, 7962, 23105724  (nrow, ncol, ncell)
## resolution : 5000, 5000  (x, y)
## extent     : -19949750, 19860250, -6149000, 8361000  (xmin, xmax, ymin,
ymax)
## crs        : +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no
_defs
## source     : bdod_0-5cm_mean_5000.tif
## names      : bdod_0.5cm_mean_5000

print(crs(soil))

## Coordinate Reference System:
## Deprecated Proj.4 representation:
##  +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs
## WKT2 2019 representation:
## PROJCRS["unknown",
##     BASEGEOGCRS["unknown",
##         DATUM["World Geodetic System 1984",
##             ELLIPSOID["WGS 84",6378137,298.257223563,
##                 LENGTHUNIT["metre",1]],
```

```
##             ID["EPSG",6326]],
##         PRIMEM["Greenwich",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8901]]],
##     CONVERSION["unknown",
##         METHOD["Interrupted Goode Homolosine"],
##         PARAMETER["Longitude of natural origin",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8802]],
##         PARAMETER["False easting",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8806]],
##         PARAMETER["False northing",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8807]]],
##     CS[Cartesian,2],
##         AXIS["(E)",east,
##             ORDER[1],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]],
##         AXIS["(N)",north,
##             ORDER[2],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]]]

bioclim_projected <- projectRaster(bioclim, soil)

print(res(bioclim_projected))
```

```
## [1] 5000 5000

crs(bioclim_projected)

## Coordinate Reference System:
## Deprecated Proj.4 representation:
##  +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs
## WKT2 2019 representation:
## PROJCRS["unknown",
##     BASEGEOGCRS["unknown",
##         DATUM["World Geodetic System 1984",
##             ELLIPSOID["WGS 84",6378137,298.257223563,
##                 LENGTHUNIT["metre",1]],
##             ID["EPSG",6326]],
##         PRIMEM["Greenwich",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8901]]],
##     CONVERSION["unknown",
##         METHOD["Interrupted Goode Homolosine"],
##         PARAMETER["Longitude of natural origin",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8802]],
##         PARAMETER["False easting",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8806]],
##         PARAMETER["False northing",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8807]]],
```

```
##     CS[Cartesian,2],
##         AXIS["(E)",east,
##             ORDER[1],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]],
##         AXIS["(N)",north,
##             ORDER[2],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]]]

bioclim_projected2 <- aggregate(bioclim_projected, fact=2, fun=mean)
soil_projected2 <- aggregate(soil, fact=2, fun=mean)
print(res(bioclim_projected2))

## [1] 10000 10000

crs(bioclim_projected2)

## Coordinate Reference System:
## Deprecated Proj.4 representation:
##  +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs
## WKT2 2019 representation:
## PROJCRS["unknown",
##     BASEGEOGCRS["unknown",
##         DATUM["World Geodetic System 1984",
##             ELLIPSOID["WGS 84",6378137,298.257223563,
##                 LENGTHUNIT["metre",1]],
##             ID["EPSG",6326]],
##         PRIMEM["Greenwich",0,
##             ANGLEUNIT["degree",0.0174532925199433],
```

```
##             ID["EPSG",8901]]],
##     CONVERSION["unknown",
##         METHOD["Interrupted Goode Homolosine"],
##         PARAMETER["Longitude of natural origin",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8802]],
##         PARAMETER["False easting",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8806]],
##         PARAMETER["False northing",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8807]]],
##     CS[Cartesian,2],
##         AXIS["(E)",east,
##             ORDER[1],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]],
##         AXIS["(N)",north,
##             ORDER[2],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]]]
```

```
print(res(soil_projected2))
```

```
## [1] 10000 10000
```

```
crs(soil_projected2)
```

```
## Coordinate Reference System:
## Deprecated Proj.4 representation:
```

```
##  +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs
## WKT2 2019 representation:
## PROJCRS["unknown",
##     BASEGEOGCRS["unknown",
##         DATUM["World Geodetic System 1984",
##             ELLIPSOID["WGS 84",6378137,298.257223563,
##                 LENGTHUNIT["metre",1]],
##             ID["EPSG",6326]],
##         PRIMEM["Greenwich",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8901]]],
##     CONVERSION["unknown",
##         METHOD["Interrupted Goode Homolosine"],
##         PARAMETER["Longitude of natural origin",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8802]],
##         PARAMETER["False easting",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8806]],
##         PARAMETER["False northing",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8807]]],
##     CS[Cartesian,2],
##         AXIS["(E)",east,
##             ORDER[1],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]],
```

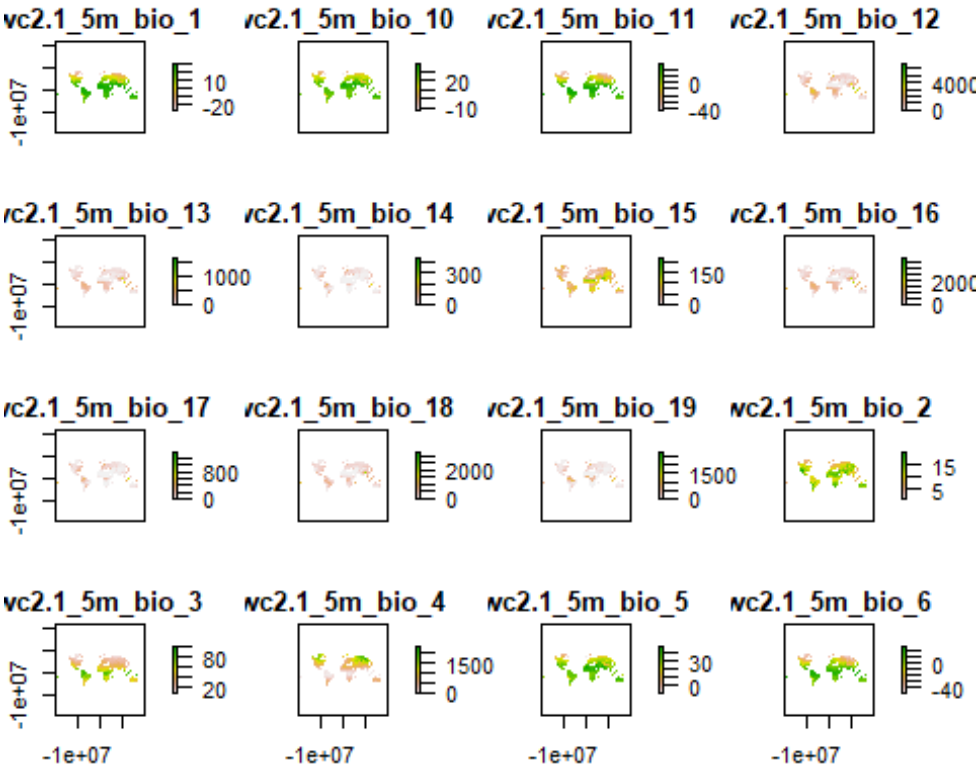
```
##         AXIS["(N)",north,
##             ORDER[2],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]]]
```

```
#soil_projected$bdod_0.5cm_mean_5000
bioclim_soil<-raster::stack(bioclim_projected2, soil_projected2)
print(res(bioclim_soil))
```

```
## [1] 10000 10000
```

```
crs(bioclim_soil)
```

```
plot(bioclim_soil)
```




```
##########################################################################
#
#
```

```r
#DADOS DAS COMUNIDADES
#
#############################################################################
#

#carregar o dataset com stringasfactors as TRUE pra que ele leia as palavr
as como fatores (vai facilitar pra frente)
ants_raw<-read.csv("ATLANTIC_ANTS_dataset.txt", head=T, stringsAsFactors =
T, na.strings=c("","NA"), fileEncoding="UTF-8",sep="\t",dec=".")



#Filtrando os dados:

ants_raw$Method<-as.character(ants_raw$Method)
ants<-filter(ants_raw, Method =="Pitfall")
ants<-filter(ants, Country =="BRAZIL")
ants<-filter(ants, Source.Data =="published manuscript" | Source.Data =="u
npublished data")
ants<-filter(ants, Disturbance =="Undisturbed")

unique_types <- unique(ants$Habitat.Type)
unique_types

print(levels(ants_raw$Habitat.Type))

##   [1] "Agriculture"
##   [2] "Agriculture -crotalaria-"
##   [3] "Agriculture -grass-"
```

```
##   [4] "Agriculture -sugar cane-"
##   [5] "Agroecossystem, coffee plantation, banana trees and atlantic for
est fragments"
##   [6] "Agroforestry"
##   [7] "Agrosystem"
##   [8] "Agrosystem -citrus-"
##   [9] "Agrosystem -cocoa-"
##  [10] "Agrosystem -coconut-"
##  [11] "Agrosystem -coffee-"
##  [12] "Agrosystem -corn-"
##  [13] "Agrosystem -organe-"
##  [14] "Agrosystem -pasture-"
##  [15] "Agrosystem -rubber plant-"
##  [16] "Agrosystem -rubber tree-"
##  [17] "Agrosystem -soybean-"
##  [18] "Agrosystem -sugar cane-"
##  [19] "Agrosystem -tobacco-"
##  [20] "Agrosystem, Pasture, and Atlantic Forest"
##  [21] "Altitudinal Grassland"
##  [22] "Araucaria Reflorestation"
##  [23] "Arboreal Caatinga"
##  [24] "Arboreal Restinga"
##  [25] "Atlantic-Cerrado"
##  [26] "Atlantic Forest"
##  [27] "Atlantic Forest and Caatinga"
##  [28] "Atlantic Forest and Restinga Forest"
##  [29] "Atlantic Forest and Savanna"
```

```
##  [30] "Bamboo Forest"
##  [31] "Bovine Pasture"
##  [32] "Brejo de Altitude"
##  [33] "Bromeliads"
##  [34] "Brush-cleared"
##  [35] "Caatinga"
##  [36] "Campinarana"
##  [37] "Carrasco and Forest"
##  [38] "Cerrado"
##  [39] "Cerrado-Atlantic Forest ecotone"
##  [40] "Cerrado-Campos Rupestres ecotone"
##  [41] "Cerrado and Atlantic Forest"
##  [42] "Cerrado and Forest"
##  [43] "Cerrado and Pasture"
##  [44] "Cerrado and Rupestrian grassland"
##  [45] "Cerrado edge"
##  [46] "Cerrado, Pasture, and Soy cropfield"
##  [47] "Chaco"
##  [48] "Coastal plain dunes"
##  [49] "Dense Ombrophylous Atlantic Forest"
##  [50] "Dense Ombrophylous Forest"
##  [51] "Dense Ombrophylous Forest, Eucalyptus Plantation, and Pasture"
##  [52] "Dense Ombrophylous Montane Atlantic Forest"
##  [53] "Dense Ombrophylous Submontane Atlantic Forest"
##  [54] "Dense Ombrophylous Submontane Forest"
##  [55] "Desert sbrub"
##  [56] "Dry Forest"
```

```
##  [57] "Ecotone Forest and Lake"
##  [58] "Ecotone Low Forest and Lake"
##  [59] "Eucalyptus Plantation"
##  [60] "Eucalyptus Plantation and Acacia Plantation"
##  [61] "Farm field"
##  [62] "Forest"
##  [63] "Forest-like cerradao"
##  [64] "Forest and Cerrado"
##  [65] "Forest and Pasture"
##  [66] "Forest edge"
##  [67] "Forest remnants"
##  [68] "Forestry"
##  [69] "Galery Forest"
##  [70] "Grassland"
##  [71] "Hillside"
##  [72] "Humid Sub-Tropical Forest"
##  [73] "Hygrophillous Forest"
##  [74] "Initial succession"
##  [75] "Iron ore caves and adjacent surface habitats"
##  [76] "Lawn"
##  [77] "Low Tropical Forest"
##  [78] "Lowland Atlantic Forest"
##  [79] "Lowland Atlantic Forest/Restinga"
##  [80] "Lowland, Submontane, and Montane Dense Ombrophilous Forest"
##  [81] "Mangrove"
##  [82] "Mata de Carrasco"
##  [83] "Mata de Cipo"
```

```
##  [84] "Mata Paludosa"
##  [85] "Mixed ombrophilous, semi-deciduous forest, and grasslands with m
ixed ombrophilous forest patches"
##  [86] "Mixed Ombrophylous Forest"
##  [87] "Mixed Ombrophylous Forest with Araucaria"
##  [88] "Mixed Ombrophylous Montane Forest"
##  [89] "Montane Forest"
##  [90] "Oceanic archipelago"
##  [91] "Open and Low Forest"
##  [92] "Open grasslands with Araucaria"
##  [93] "Open vegetation"
##  [94] "Open vegetation and Forest Patch"
##  [95] "Palm swamped area"
##  [96] "Pantanal"
##  [97] "Pasture"
##  [98] "Pasture and Soy cropfield"
##  [99] "Pine Plantation"
## [100] "Plantation"
## [101] "Pre-forest succession and Restinga"
## [102] "Preserved Forest"
## [103] "Primary Forest"
## [104] "Primary Forest, Secondary Forest and Cabruca Agrosystem"
## [105] "Rehabilitation Area"
## [106] "Restinga"
## [107] "Restinga edge"
## [108] "Restinga Forest"
## [109] "Restinga Forest / Coastal Atlantic Forest / Atlantic Forest alon
```

```
g a hillside "
## [110] "Restoration Area"
## [111] "Restoration Monoculture"
## [112] "Riparian forest"
## [113] "Riparian forest and brush-cleared"
## [114] "Riparian forest and Cerradao"
## [115] "Road edge"
## [116] "Rocky Forest"
## [117] "Rupestrian grassland"
## [118] "Rural area"
## [119] "Sand dunes"
## [120] "Scattered tree in pasture"
## [121] "Scrub Forest"
## [122] "Seasonal Semi-deciduous Atlantic Forest"
## [123] "Seasonal Semi-deciduous Atlantic Forest and Caatinga"
## [124] "Seasonal Semi-deciduous Montane Atlantic Forest"
## [125] "Seasonal Semi-deciduous Secondary Atlantic Forest"
## [126] "Seasonal Semi-deciduous Submontane Atlantic Forest"
## [127] "Seasonal Semi-deciduous Submontane Secondary Atlantic Forest"
## [128] "Secondary Atlantic Forest"
## [129] "Secondary Forest"
## [130] "Secondary Forest and Urban Area"
## [131] "Secondary Forest edge"
## [132] "Secondary Forest tree - Ecotone Forest-Lake"
## [133] "Secondary Riparian Vegetation"
## [134] "Small-scale forest and fields, approaching the savannah"
## [135] "Successional Atlantic Forest gradient"
```

```
## [136] "Under rock"
## [137] "Urban Area"
## [138] "Urban Area/Park"
## [139] "Urban Forest"
## [140] "Urban Park"
## [141] "Urban tree"
## [142] "Vineyards"

ants <- dplyr::filter(ants, Habitat.Type %in%
                        c("Forest","Seasonal Semi-deciduous Atlantic Fores
t", "Secondary Forest","Mixed Ombrophylous Forest", "Mixed Ombrophylous Mo
ntane Forest","Dense Ombrophylous Forest","Montane Forest","Dense Ombrophy
lous Submontane Atlantic Forest","Secondary Atlantic Forest", "Seasonal Se
mi-deciduous Submontane Atlantic Forest","Primary Forest", "Atlantic Fores
t","Atlantic Forest and Restinga Forest","Dense Ombrophylous Atlantic Fore
st","Dense Ombrophylous Montane Atlantic Forest","Dense Ombrophylous Submo
ntane Forest","Humid Sub-Tropical Forest","Low Tropical Forest","Lowland A
tlantic Forest","Lowland, Submontane, and Montane Dense Ombrophilous Fores
t","Montane Forest","Preserved Forest","Seasonal Semi-deciduous Atlantic F
orest", "Seasonal Semi-deciduous Montane Atlantic Forest","Seasonal Semi-d
eciduous Secondary Atlantic Forest", "Seasonal Semi-deciduous Submontane A
tlantic Forest", "Seasonal Semi-deciduous Submontane Secondary Atlantic Fo
rest","Secondary Atlantic Forest", "Secondary Forest"))
#criando uma coluna com ID unicos para cada comunidade
Data<-dplyr::select(ants, "Genus", "Longitude.x", "Latitude.y", "ID.codLoc
", "Start.year", "Subfamily", "Habitat.Type", "Disturbance", "Disturbance.
Category.1", "Source.Data", "Municipality", "State", "Pitfall.Duration.h")
```

```
Data$LatLongHT <- paste(Data$Latitude.y, Data$Longitude.x, Data$Habitat.Ty
pe)


Data<-dplyr::select(Data, "LatLongHT", "Genus", "Longitude.x", "Latitude.y
", "ID.codLoc", "Start.year", "Subfamily", "Habitat.Type", "Disturbance",
"Disturbance.Category.1", "Source.Data", "Municipality", "State", "Pitfall
.Duration.h")


Data$LatLongHT <- as.character(Data$LatLongHT)
Data$LatLongHT <- as.factor(Data$LatLongHT)
nlevels(Data$LatLongHT) #74 localidades diferentes, baseados no lat long e
tipo de habitat

## [1] 74

Data<-as.data.table(Data)


# funcaoo para criar uma coluna com ID unicos
Data$LatLongHT <- as.character(Data$LatLongHT)
Data$Municipality <- as.character(Data$Municipality)
Data$Source.Citation <- as.character(Data$Source.Citation)


Data[, ("Site") := ""]
regions <- list()
year <- list ()



for(i in 1:nrow(Data)){
```

```
  newRegion <- TRUE
  region <- Data[i,LatLongHT]

  if(!is.null(regions[[region]]))
    newRegion <- FALSE

  if(newRegion){
    regions[[region]] <- length(regions)+1
    year[[region]] <- Data[i, Start.year]}

  Data[i,Site := paste("site", regions[[region]], sep="")]
  Data[i, Start.year :=year[[region]]]
}



Data$Site<- as.character(Data$Site)
Data$Site<- as.factor(Data$Site)
nlevels(Data$Site)

## [1] 74

Data$Genus <- as.factor(Data$Genus)
Data$Genus <- as.character(Data$Genus)
Data$Genus <- as.factor(Data$Genus)
nlevels(Data$Genus)

## [1] 76
```

```
mata_atlantica<-rgdal::readOGR(dsn="C:/Users/isagv/Documents/Drive/Projeto Modelagem da diversidade filogenética/Shapefiles", layer='union_AF')

## Warning: OGR support is provided by the sf and terra packages among others## Warning: OGR support is provided by the sf and terra packages among others

## Warning: OGR support is provided by the sf and terra packages among others

## Warning: OGR support is provided by the sf and terra packages among others

## Warning: OGR support is provided by the sf and terra packages among others

## Warning: OGR support is provided by the sf and terra packages among others

## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\isagv\Documents\Drive\Projeto Modelagem da diversidade filogenética\Shapefiles", layer: "union_AF"
## with 3 features
## It has 9 fields

Coords<-Data[, c(3,4)]
plot(mata_atlantica)
points(Coords, col="red", pch=16)
```

```
#Alguns pontos estão fora da mata atlântica


#criar um objeto que esteja no sistema wgs84
wgs84<-"+proj=longlat +datum=WGS84 +no_defs"
#cria um conjunto de pontos no espa?o, informando a logitude (x) e latitud
e (y) e informando sistema de proje??o wgs84 com o objeto criado anteriorm
ente.
#? como se fosse um shapefile de pontos georeferenciados. Mas n?o usamos t
oda a base de dados de ocorr?ncia, usamos s? as coordenadas (filtramos as
colunas lat e long)
pts<-sp::SpatialPoints(Data[, c("Longitude.x","Latitude.y")], proj4string
= sp::CRS(wgs84))
#checar quais ocorr?ncias est?o dentro da ?rea de interesse
WithinStudyArea <- sp::over(x=pts, y=mata_atlantica)[, 1]
```

```
#compilar ocorrencias e ?rea de interesse
Data<-cbind(Data, WithinStudyArea)

Data<- Data[!(is.na(Data$WithinStudyArea)),] #filtrando pra conter somente
as observa??es que ca?ram dentro da ?rea de interesse
SiteCoords<-Data[, c(3,4)] # separate a data.frame with coordinates

plot(mata_atlantica, main="mata_atlantica")
points(SiteCoords, col="red", pch=16)
```


```
Data<-CoordinateCleaner::cc_dupl(Data, lon="LatLongHT", lat="Latitude.y",
species="Genus", value="clean")
```

```
## Testing duplicates
## Removed 1634 records.



Data$Site<- as.character(Data$Site)
Data$Site<- as.factor(Data$Site)
nlevels(Data$Site)
## [1] 64
nlevels(Data$Genus)
## [1] 76
#64 sites unicos; 76 generos diferentes


# Convert shapefiles to sf object
Neotropic_shp<-rgdal::readOGR(dsn="C:/Users/isagv/Documents/Drive/Projeto
Modelagem da diversidade filogenética/Shapefiles", layer='wwf_realm') # ch
ange directory as needed
## Warning: OGR support is provided by the sf and terra packages among oth
ers
## Warning: OGR support is provided by the sf and terra packages among oth
ers
## Warning: OGR support is provided by the sf and terra packages among oth
ers
```

```
## Warning: OGR support is provided by the sf and terra packages among oth
ers

## Warning: OGR support is provided by the sf and terra packages among oth
ers

## Warning: OGR support is provided by the sf and terra packages among oth
ers

## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\isagv\Documents\Drive\Projeto Modelagem da diversidad
e filogenética\Shapefiles", layer: "wwf_realm"
## with 6 features
## It has 1 fields

Neotropic_shp<-Neotropic_shp[Neotropic_shp@data$wwf_realm=="Neotropic",]
Neotropic_sf<-sf::st_as_sf(Neotropic_shp) #transformou o shape pro formato
sf pra ficar mais facil
mata_atlantica_sf<-sf::st_as_sf(mata_atlantica)




#unificando os sites que estão num raio de 10km


library(geosphere)

##
## Attaching package: 'geosphere'
##
## The following object is masked from 'package:fastmatrix':
```

```
##
##     geomean

library(dplyr)
library(sf)


# Certificando-se de que as coordenadas são numéricas
Data$Longitude.x <- as.numeric(Data$Longitude.x)
Data$Latitude.y <- as.numeric(Data$Latitude.y)
# Selecionando sites únicos
unique_sites <- distinct(Data, Site, Longitude.x, Latitude.y)
# Convertendo para sf object
coordinates <- st_as_sf(unique_sites, coords = c("Longitude.x", "Latitude.
y"), crs = 4326)
# Calculando a matriz de distância entre todos os pares de sites em metros
dist_matrix <- st_distance(coordinates)
# Convertendo a matriz de distância para km
dist_matrix_km <- dist_matrix / 1000
# Aplicando o clustering hierárquico
clustering <- hclust(as.dist(dist_matrix_km))
# Cortando a árvore de clustering a uma altura de 10 km para obter as regi
ões
clusters <- cutree(clustering, h = 10)
# Atribuindo o ID da região de volta aos sites únicos
unique_sites$regiao <- paste0("regiao", clusters)
# Atribuindo o ID da região de volta ao Data original
```

```
setkey(Data, Site)  # Configurando a chave para a junção
setkey(unique_sites, Site)  # Configurando a chave para a junção
Data <- unique_sites[Data]  # Realizando a junção

num_regioes <- length(unique(Data$regiao))
print(num_regioes) #28 regiões
```

```
## [1] 28
```

```
# STEP 1: LOAD INVENTORY DATA AND CONVERT TABLES FROM LONG TO WIDE FORMATS
################################################################################

#selecionado somente os dados de site vs esp?cies, pra montar a matriz PAM
InventoryData<-dplyr::select(Data, "regiao", "Genus")
InventoryData$Presence<-1
str(InventoryData$regiao)
```

```
##  chr [1:847] "regiao1" "regiao1" "regiao1" "regiao1" "regiao1" "regiao1
" ...
```

```
pam_wide<-reshape2::dcast(InventoryData, regiao ~ Genus, value.var="Presen
ce") # reshape the df, remodelar colunas
```

```
## Aggregation function missing: defaulting to length
```

```
# Get a vector of species name in the same order as in 'PAM_Long' and reor
der columns of pam_wide:
Spp_order<-as.character(unique(InventoryData$Genus)) # specie s names list
ed in the PAM_Long$Sp
```

```r
Spp_order<-c("regiao", Spp_order) # vector with the order of column names
pam_wide<-as.data.frame(pam_wide) # certify it is a data.frame object
pam_wide<-pam_wide[, Spp_order] # reorder columns
pam_wide[is.na(pam_wide)]<-0 # replace NA by 0

# Order sites in the original order of data compilation:
regiao_order<-as.character(unique(InventoryData$regiao))
rownames(pam_wide)<-pam_wide$regiao # it is necessary to name rows
pam_wide[1:10, 1:5] # visualise order of rows
```

```
##            regiao Camponotus Crematogaster Sericomyrmex Atta
## regiao1   regiao1          1             1            1    1
## regiao10 regiao10          1             1            1    1
## regiao11 regiao11          1             1            1    1
## regiao12 regiao12          2             2            0    0
## regiao13 regiao13          4             1            0    3
## regiao14 regiao14          1             0            0    1
## regiao15 regiao15          1             0            0    1
## regiao16 regiao16          2             0            0    1
## regiao17 regiao17          1             0            0    1
## regiao18 regiao18          6             3            0    4
```

```r
pam_wide<-pam_wide[regiao_order, ] # reorder rows
pam_wide[1:10, 1:5] # visualise order of rows again, vai estar na mesma or
dem que em InventoryData
```

```
##            regiao Camponotus Crematogaster Sericomyrmex Atta
## regiao1   regiao1          1             1            1    1
## regiao5   regiao5          6             1            0    1
## regiao6   regiao6          1             1            0    1
## regiao7   regiao7          0             2            0    0
## regiao8   regiao8          0             1            0    0
## regiao2   regiao2          5             5            0    1
## regiao9   regiao9          1             1            1    1
## regiao10 regiao10          1             1            1    1
## regiao11 regiao11          1             1            1    1
## regiao12 regiao12          2             2            0    0
```

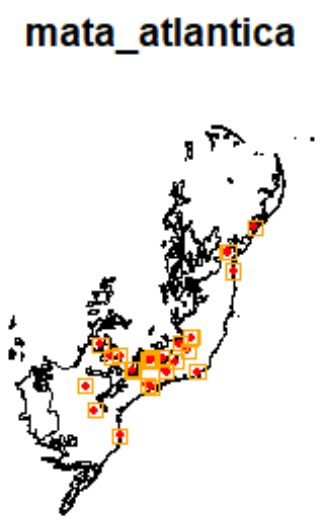
```
pam_wide2 <- data.matrix(pam_wide, rownames.force = NA)
pam_wide3 <- as.data.frame(pam_wide2)
pam_wide4 <- pam_wide3 %>% dplyr::select(!(regiao))
```

```
#############################################################################
#
#
#DADOS AMBIENTAIS DAS REGIÕES
#
#############################################################################
#
```

```
regiaoData<-as.data.table(dplyr::select(Data, "regiao", "Longitude.x", "La
titude.y"))
regiaoData<-CoordinateCleaner::cc_dupl(regiaoData, lon="Longitude.x", lat=
"Latitude.y", species="regiao", value="clean")

## Testing duplicates
## Removed 785 records.

#unificar as regioes em uma mesma coordenada cada
centroids <- aggregate(cbind(Longitude.x, Latitude.y) ~ regiao, regiaoData
, mean, na.rm = TRUE)
# Exibir os resultados
print(centroids)

##      regiao Longitude.x Latitude.y
## 1   regiao1   -37.19481  -10.91156
## 2  regiao10   -49.30010  -21.24170
## 3  regiao11   -50.17360  -20.28250
## 4  regiao12   -48.50090  -27.52613
## 5  regiao13   -46.13708  -21.44319
## 6  regiao14   -45.87657  -21.50417
## 7  regiao15   -45.98824  -21.47028
## 8  regiao16   -45.92847  -21.57958
## 9  regiao17   -45.84056  -21.36278
## 10 regiao18   -43.46404  -20.22177
## 11 regiao19   -44.72590  -22.36962
## 12  regiao2   -39.46714  -12.88206
```

```
## 13 regiao20   -44.95642  -21.37441
## 14 regiao21   -51.45295  -23.56186
## 15 regiao22   -43.86667  -21.66667
## 16 regiao23   -45.89520  -23.64831
## 17 regiao24   -39.02347  -14.41965
## 18 regiao25   -50.63176  -25.50729
## 19 regiao26   -42.86000  -20.75750
## 20 regiao27   -42.55306  -19.68972
## 21 regiao28   -42.62370  -19.76590
## 22  regiao3   -39.44843  -12.96114
## 23  regiao4   -42.03843  -22.42960
## 24  regiao5   -44.62784  -22.42647
## 25  regiao6   -47.48969  -22.35026
## 26  regiao7   -47.42054  -22.30126
## 27  regiao8   -46.19620  -23.49582
## 28  regiao9   -48.55040  -21.13120

crs(centroids)

## [1] NA

Data2 <- tibble::rownames_to_column(pam_wide4, var = "regiao")

Data3 <- left_join(centroids, Data, by= "regiao")

regiaoData <- Data3
```

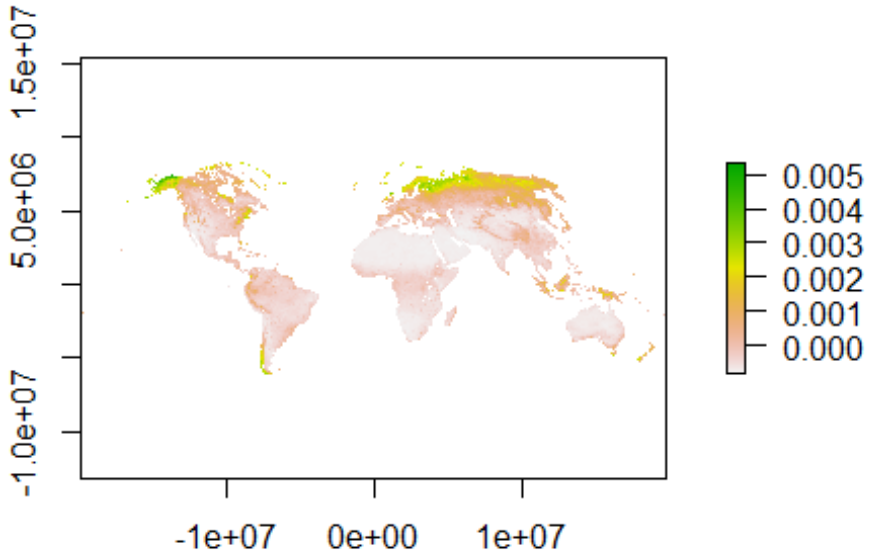
```
#visualizar os sites e regioes
SiteCoords<-Data[, c(2,3)]
regiaoCoords<-centroids[, c(2,3)]
plot(mata_atlantica, main="mata_atlantica")
points(SiteCoords, col="red", pch=16, cex=0.5)
points(regiaoCoords, col="orange", pch=0.5)
```



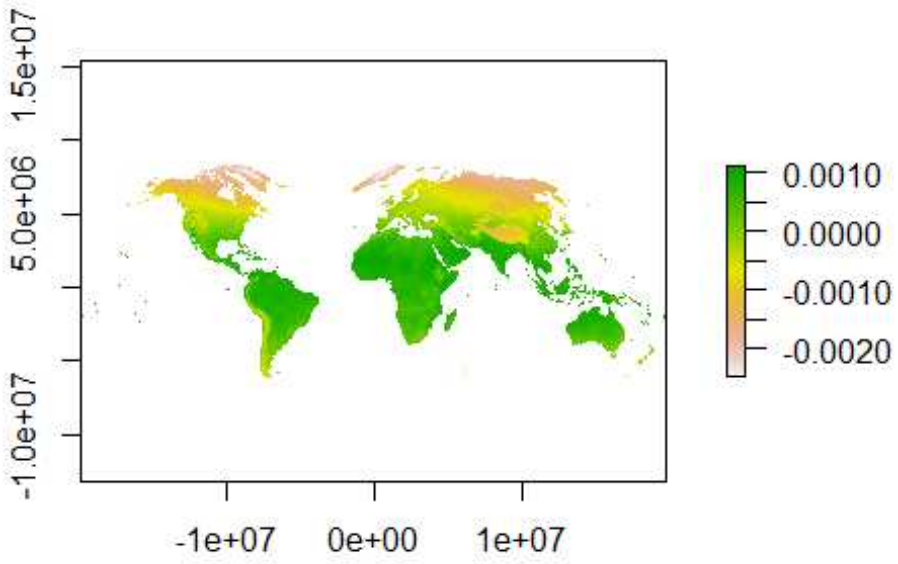

```
#convertendo as variáveis do raster para polinòmios ortogonais
#tem valores NA nas variáveis (provavelmente nas camadas de solo) e esse c
omando
#abaixo dá problema se tiver NA. Como não quero excluir nada, ele vai conv
erter
#em polinômio só os pixels que não forem NA. Depois de extrair as variávei
s
#para cada site, eu removo os sites com NA (o modelo nao gera com NA)
```

```
# Create two empty RasterStacks to store the transformed layers
bioclim_soil_poly_linear <- raster::stack()
bioclim_soil_poly_quad <- raster::stack()

# Loop over each layer in the original RasterStack
for (i in 1:nlayers(bioclim_soil)) {
  # Extract the current layer
  layer <- raster::subset(bioclim_soil, i)

  # Convert the RasterLayer to a vector
  layer_vector <- raster::getValues(layer)

  # Identify the non-NA values
  non_na_values <- !is.na(layer_vector)

  # Calculate the orthogonal polynomial transformation only on non-NA valu
es
  poly_vector <- matrix(NA, nrow=length(layer_vector), ncol=2)
  poly_vector[non_na_values, ] <- stats::poly(layer_vector[non_na_values],
degree=2, raw=FALSE)

  # Convert the vectors back into RasterLayers
  linear_layer <- raster::setValues(layer, poly_vector[,1])
  quad_layer <- raster::setValues(layer, poly_vector[,2])

  # Add the transformed layers to the new RasterStacks
```

```
  bioclim_soil_poly_linear <- raster::addLayer(bioclim_soil_poly_linear, l
inear_layer)
  bioclim_soil_poly_quad <- raster::addLayer(bioclim_soil_poly_quad, quad_
layer)

  # Give specific names to the transformed layers
  names(bioclim_soil_poly_linear)[i] <- paste0(names(bioclim_soil)[i], "_l
inear")
  names(bioclim_soil_poly_quad)[i] <- paste0(names(bioclim_soil)[i], "_qua
d")
}



plot(bioclim_soil_poly_linear[[31]])
```



```
# Combine the two RasterStacks into one
bioclim_soil_poly <- raster::stack(bioclim_soil_poly_linear, bioclim_soil_
poly_quad)
plot(bioclim_soil_poly[[1]])
```




```
# Extrair a primeira camada do RasterStack
first_layer_linear <- raster::subset(bioclim_soil_poly, 1)
# Converter a RasterLayer em um vetor
first_layer_linear_vector <- raster::getValues(first_layer_linear)
# Filtrar os valores NA
first_layer_linear_vector_non_na <- first_layer_linear_vector[!is.na(first
_layer_linear_vector)]
# Imprimir os primeiros 10 valores não-NA
print(first_layer_linear_vector_non_na[1:10])
```

```
##  [1] -0.002088301 -0.002033115 -0.002056927 -0.002066786 -0.002048731
##  [6] -0.002035555 -0.002015101 -0.002003478 -0.002018941 -0.002028389

crs(bioclim_soil_poly)

## Coordinate Reference System:
## Deprecated Proj.4 representation:
##  +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs
## WKT2 2019 representation:
## PROJCRS["unknown",
##     BASEGEOGCRS["unknown",
##         DATUM["World Geodetic System 1984",
##             ELLIPSOID["WGS 84",6378137,298.257223563,
##                 LENGTHUNIT["metre",1]],
##             ID["EPSG",6326]],
##         PRIMEM["Greenwich",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8901]]],
##     CONVERSION["unknown",
##         METHOD["Interrupted Goode Homolosine"],
##         PARAMETER["Longitude of natural origin",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8802]],
##         PARAMETER["False easting",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8806]],
##         PARAMETER["False northing",0,
##             LENGTHUNIT["metre",1],
```

```
##             ID["EPSG",8807]]],
##     CS[Cartesian,2],
##         AXIS["(E)",east,
##             ORDER[1],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]],
##         AXIS["(N)",north,
##             ORDER[2],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]]]
```

```
crs(mata_atlantica)
```

```
## Coordinate Reference System:
## Deprecated Proj.4 representation: +proj=longlat +datum=WGS84 +no_defs
## WKT2 2019 representation:
## GEOGCRS["WGS 84",
##     DATUM["World Geodetic System 1984",
##         ELLIPSOID["WGS 84",6378137,298.257223563,
##             LENGTHUNIT["metre",1]]],
##     PRIMEM["Greenwich",0,
##         ANGLEUNIT["degree",0.0174532925199433]],
##     CS[ellipsoidal,2],
##         AXIS["latitude",north,
##             ORDER[1],
##             ANGLEUNIT["degree",0.0174532925199433]],
##         AXIS["longitude",east,
##             ORDER[2],
```

```
##             ANGLEUNIT["degree",0.0174532925199433]],
##     ID["EPSG",4326]]

library(rgdal)
proj_polygon <- "+proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +n
o_defs"
mata_atlantica_projected <- spTransform(mata_atlantica, CRS(proj_polygon))
crs(bioclim_soil_poly)

## Coordinate Reference System:
## Deprecated Proj.4 representation:
##  +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs
## WKT2 2019 representation:
## PROJCRS["unknown",
##     BASEGEOGCRS["unknown",
##         DATUM["World Geodetic System 1984",
##             ELLIPSOID["WGS 84",6378137,298.257223563,
##                 LENGTHUNIT["metre",1]],
##             ID["EPSG",6326]],
##         PRIMEM["Greenwich",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8901]]],
##     CONVERSION["unknown",
##         METHOD["Interrupted Goode Homolosine"],
##         PARAMETER["Longitude of natural origin",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8802]],
##         PARAMETER["False easting",0,
```

```
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8806]],
##         PARAMETER["False northing",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8807]]],
##     CS[Cartesian,2],
##         AXIS["(E)",east,
##             ORDER[1],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]],
##         AXIS["(N)",north,
##             ORDER[2],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]]]
```

```
crs(mata_atlantica_projected)
```

```
## Coordinate Reference System:
## Deprecated Proj.4 representation:
##  +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs
## WKT2 2019 representation:
## PROJCRS["unknown",
##     BASEGEOGCRS["unknown",
##         DATUM["World Geodetic System 1984",
##             ELLIPSOID["WGS 84",6378137,298.257223563,
##                 LENGTHUNIT["metre",1]],
##             ID["EPSG",6326]],
##         PRIMEM["Greenwich",0,
```

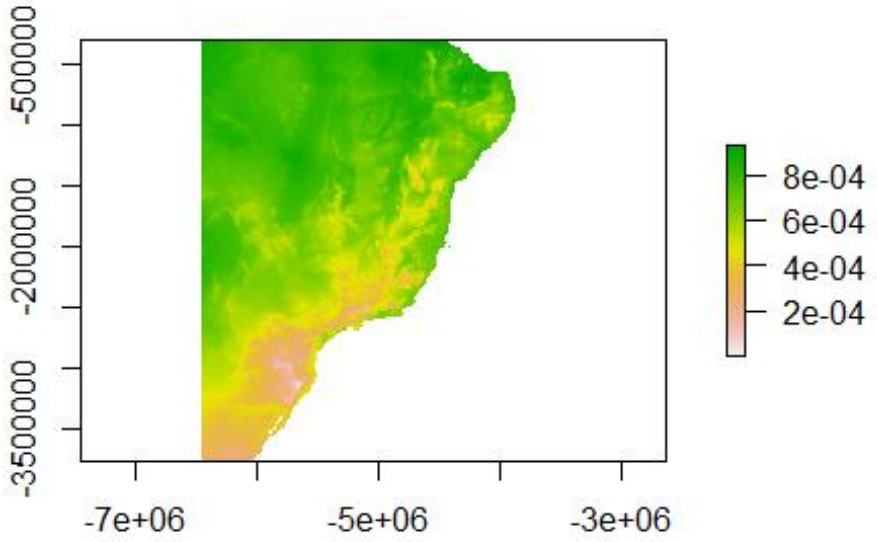
```
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8901]]],
##     CONVERSION["unknown",
##         METHOD["Interrupted Goode Homolosine"],
##         PARAMETER["Longitude of natural origin",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8802]],
##         PARAMETER["False easting",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8806]],
##         PARAMETER["False northing",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8807]]],
##     CS[Cartesian,2],
##         AXIS["(E)",east,
##             ORDER[1],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]],
##         AXIS["(N)",north,
##             ORDER[2],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]]]
```

```
crs (regiaoCoords)
```

```
## [1] NA
```

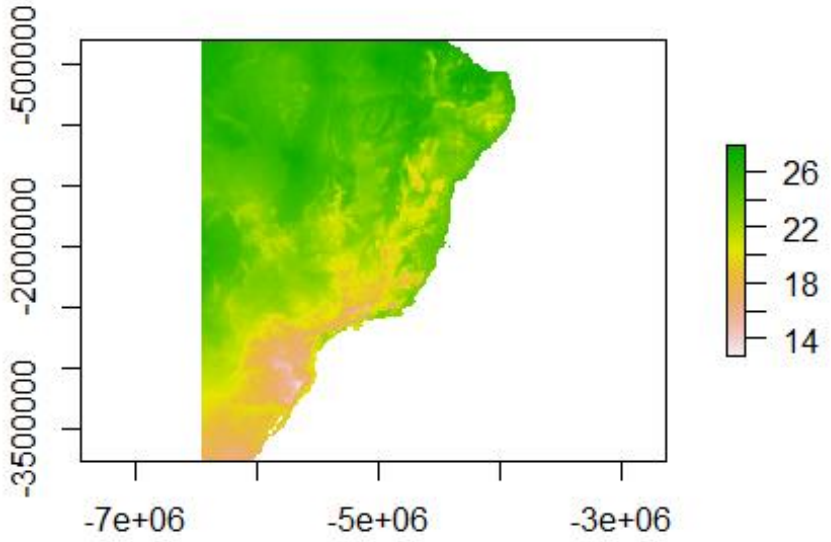
```
str(regiaoCoords)
```

```
## 'data.frame':    28 obs. of  2 variables:
##  $ Longitude.x: num  -37.2 -49.3 -50.2 -48.5 -46.1 ...
##  $ Latitude.y : num  -10.9 -21.2 -20.3 -27.5 -21.4 ...

bioclim_soil_poly<-raster::crop(bioclim_soil_poly, mata_atlantica_projecte
d)
bioclim_soil_poly$wc2.1_5m_bio_1_quad

## class      : RasterLayer
## band       : 32  (of  62  bands)
## dimensions : 345, 285, 98325  (nrow, ncol, ncell)
## resolution : 10000, 10000  (x, y)
## extent     : -6459750, -3609750, -3759000, -309000  (xmin, xmax, ymin,
ymax)
## crs        : +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no
_defs
## source     : r_tmp_2024-10-06_222357.269995_11096_54303.grd
## names      : wc2.1_5m_bio_1_quad
## values     : -0.0008664947, 0.0009063875  (min, max)

plot(bioclim_soil_poly[[1]])
```

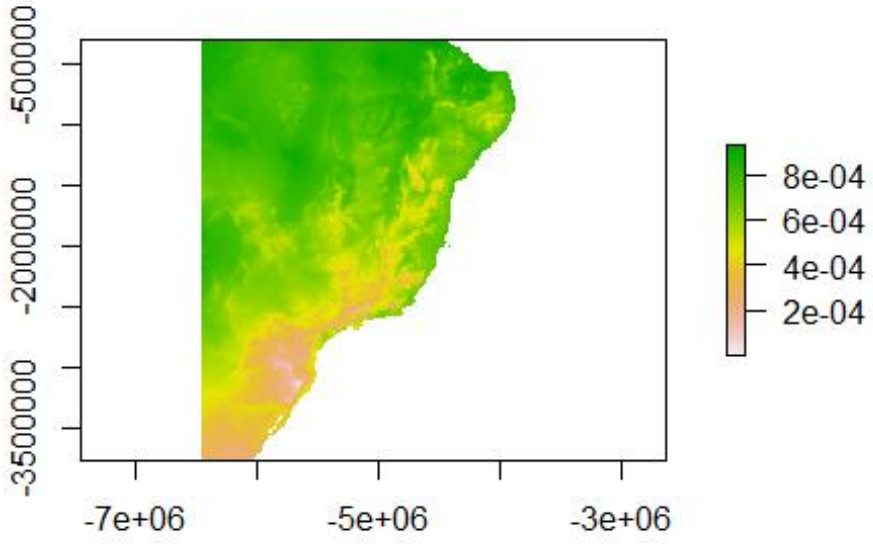



```
bioclim_soil_crop<-raster::crop(bioclim_soil, mata_atlantica_projected)


#plotando a mesma variável nas versões: original, polinomio 1 ordem e 2 or
dem
bioclim_soil_poly@data@names

##  [1] "wc2.1_5m_bio_1_linear"           "wc2.1_5m_bio_10_linear"
##  [3] "wc2.1_5m_bio_11_linear"          "wc2.1_5m_bio_12_linear"
##  [5] "wc2.1_5m_bio_13_linear"          "wc2.1_5m_bio_14_linear"
##  [7] "wc2.1_5m_bio_15_linear"          "wc2.1_5m_bio_16_linear"
##  [9] "wc2.1_5m_bio_17_linear"          "wc2.1_5m_bio_18_linear"
## [11] "wc2.1_5m_bio_19_linear"          "wc2.1_5m_bio_2_linear"
## [13] "wc2.1_5m_bio_3_linear"           "wc2.1_5m_bio_4_linear"
## [15] "wc2.1_5m_bio_5_linear"           "wc2.1_5m_bio_6_linear"
## [17] "wc2.1_5m_bio_7_linear"           "wc2.1_5m_bio_8_linear"
## [19] "wc2.1_5m_bio_9_linear"           "wc2.1_5m_elev_linear"
```

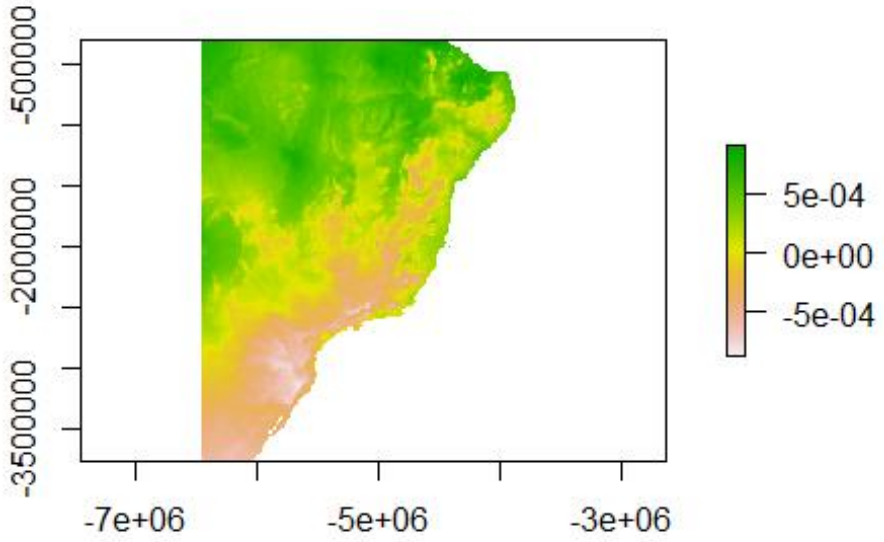
```
## [21] "bdod_0.5cm_mean_5000_linear"     "cec_0.5cm_mean_5000_linear"
## [23] "cfvo_0.5cm_mean_5000_linear"     "clay_0.5cm_mean_5000_linear"
## [25] "nitrogen_0.5cm_mean_5000_linear" "ocd_0.5cm_mean_5000_linear"
## [27] "ocs_0.30cm_mean_5000_linear"     "phh2o_0.5cm_mean_5000_linear"
## [29] "sand_0.5cm_mean_5000_linear"     "silt_0.5cm_mean_5000_linear"
## [31] "soc_0.5cm_mean_5000_linear"      "wc2.1_5m_bio_1_quad"
## [33] "wc2.1_5m_bio_10_quad"            "wc2.1_5m_bio_11_quad"
## [35] "wc2.1_5m_bio_12_quad"            "wc2.1_5m_bio_13_quad"
## [37] "wc2.1_5m_bio_14_quad"            "wc2.1_5m_bio_15_quad"
## [39] "wc2.1_5m_bio_16_quad"            "wc2.1_5m_bio_17_quad"
## [41] "wc2.1_5m_bio_18_quad"            "wc2.1_5m_bio_19_quad"
## [43] "wc2.1_5m_bio_2_quad"             "wc2.1_5m_bio_3_quad"
## [45] "wc2.1_5m_bio_4_quad"             "wc2.1_5m_bio_5_quad"
## [47] "wc2.1_5m_bio_6_quad"             "wc2.1_5m_bio_7_quad"
## [49] "wc2.1_5m_bio_8_quad"             "wc2.1_5m_bio_9_quad"
## [51] "wc2.1_5m_elev_quad"              "bdod_0.5cm_mean_5000_quad"
## [53] "cec_0.5cm_mean_5000_quad"        "cfvo_0.5cm_mean_5000_quad"
## [55] "clay_0.5cm_mean_5000_quad"       "nitrogen_0.5cm_mean_5000_quad"
## [57] "ocd_0.5cm_mean_5000_quad"        "ocs_0.30cm_mean_5000_quad"
## [59] "phh2o_0.5cm_mean_5000_quad"      "sand_0.5cm_mean_5000_quad"
## [61] "silt_0.5cm_mean_5000_quad"       "soc_0.5cm_mean_5000_quad"

plot(bioclim_soil_crop[[1]])
```




```
plot(bioclim_soil_poly[[1]])
```




```
plot(bioclim_soil_poly[[32]])
```




```
coordinates(regiaoCoords) <- ~Longitude.x+Latitude.y

proj4string(regiaoCoords) <- CRS("+proj=longlat +datum=WGS84 +no_defs")

regiaoCoords_transformed <- spTransform(regiaoCoords, CRS(proj4string(bioc

lim_soil_poly)))

crs(regiaoCoords_transformed)

## Coordinate Reference System:

## Deprecated Proj.4 representation:

##  +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs

## WKT2 2019 representation:

## PROJCRS["unknown",

##     BASEGEOGCRS["unknown",

##         DATUM["World Geodetic System 1984",

##             ELLIPSOID["WGS 84",6378137,298.257223563,

##                 LENGTHUNIT["metre",1]],

##             ID["EPSG",6326]],
```

```
##         PRIMEM["Greenwich",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8901]]],
##     CONVERSION["unknown",
##         METHOD["Interrupted Goode Homolosine"],
##         PARAMETER["Longitude of natural origin",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8802]],
##         PARAMETER["False easting",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8806]],
##         PARAMETER["False northing",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8807]]],
##     CS[Cartesian,2],
##         AXIS["(E)",east,
##             ORDER[1],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]],
##         AXIS["(N)",north,
##             ORDER[2],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]]]

# Create a spatial buffer around each site and get the overall extent of t
he study area:
```

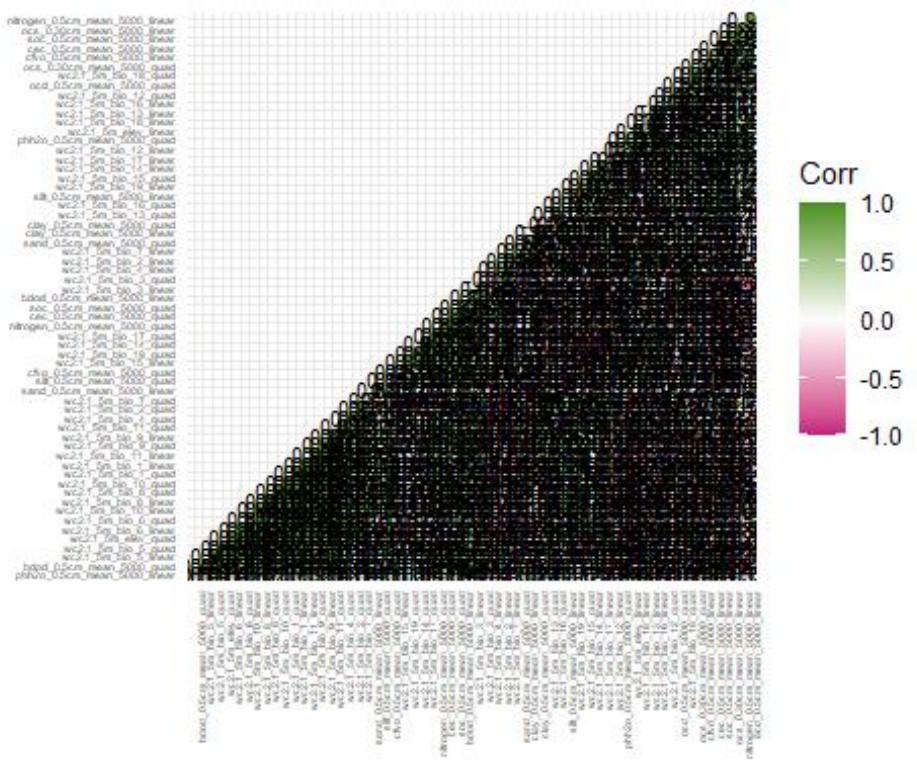
```
#sp::coordinates(regiaoCoords)<- ~ Longitude.x + Latitude.y # convert to s
patial points

# Extract each raster layer in parallel:
cl <- parallel::makePSOCKcluster(parallel::detectCores()-2) # number of co
res in computer
doParallel::registerDoParallel(cl) # set the number of cores to run next s
teps in parallel
getDoParWorkers()

## [1] 2

EnvData<-foreach(i = 1:length(names(bioclim_soil_poly)),
                 .combine = 'cbind',
                 .packages = c("raster", "sp", "rgdal")) %dopar% {

                   # Extract environmental data for each raster layer:
                   extract(bioclim_soil_poly[[i]], regiaoCoords_transforme
d,
                           buffer=6800, # buffer around each site (in mete
rs)
                           fun=mean, # function to compute using raster va
lues
                           na.rm=TRUE, # remove NA values
                           df=TRUE)[,2]
                 }
```

```
# Convert EnvData to data.frame and rename columns:
EnvData<-as.data.frame(EnvData)
variables_names<-gsub("", "", names(bioclim_soil_poly))
names(EnvData)<-variables_names




# Converter o dataframe centroids em um objeto sf
centroids_sf <- st_as_sf(centroids, coords = c("Longitude.x", "Latitude.y"
), crs = 4326)
# Definir a string PROJ.4 para o CRS desejado
crs_proj4 <- "+proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_d
efs"
# Transformar as coordenadas
centroids_transformed <- st_transform(centroids_sf, crs = crs_proj4)
# Converter o objeto sf de volta para data.frame
centroids_df <- as.data.frame(centroids_transformed)
# Extrair as coordenadas transformadas para novas colunas no dataframe
centroids_df$Longitude.x <- st_coordinates(centroids_transformed)[, 1]
centroids_df$Latitude.y <- st_coordinates(centroids_transformed)[, 2]




regiaoCoords2<-centroids_df
# Bind the SiteID information to the EnvData:
EnvData2<-cbind(regiaoCoords2[,c("regiao", "Longitude.x", "Latitude.y")],
```

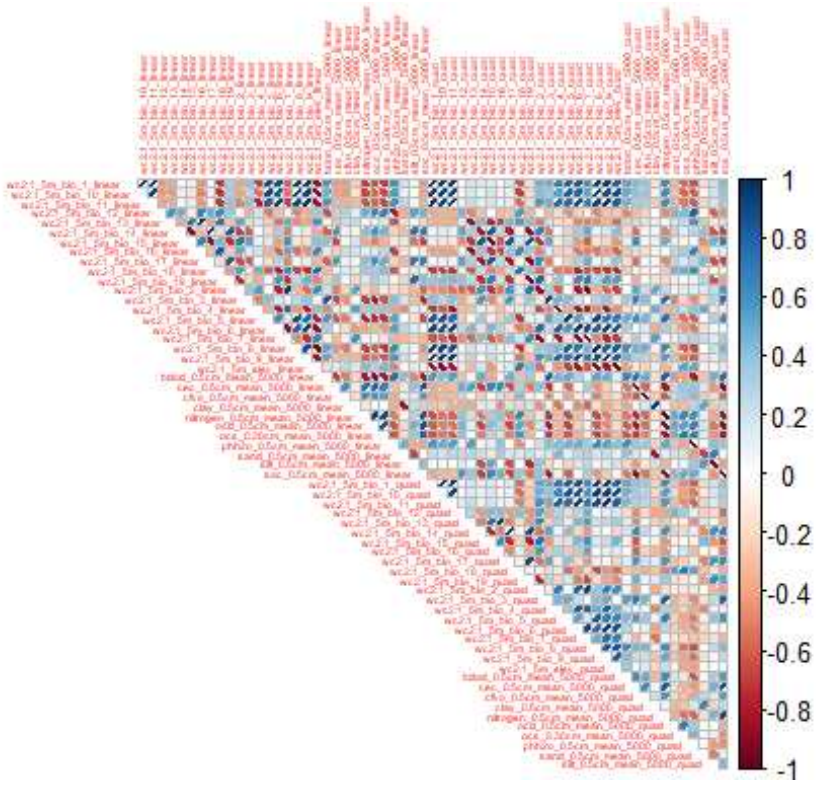
```
EnvData)


EnvData2 <- na.omit(EnvData2)


# Remember what each bioclimatic layer represents:
# BIO1 = Annual Mean Temperature
# BIO2 = Mean temperate Diurnal Range (Mean of monthly (max temp - min tem
p))
# BIO3 = Isothermality (BIO2/BIO7) (×100)
# BIO4 = Temperature Seasonality (standard deviation ×100)
# BIO5 = Max Temperature of Warmest Month
# BIO6 = Min Temperature of Coldest Month
# BIO7 = Temperature Annual Range (BIO5-BIO6)
# BIO8 = Mean Temperature of Wettest Quarter
# BIO9 = Mean Temperature of Driest Quarter
# BIO10 = Mean Temperature of Warmest Quarter
# BIO11 = Mean Temperature of Coldest Quarter
# BIO12 = Annual Precipitation
# BIO13 = Precipitation of Wettest Month
# BIO14 = Precipitation of Driest Month
# BIO15 = Precipitation Seasonality (Coefficient of Variation)
# BIO16 = Precipitation of Wettest Quarter
# BIO17 = Precipitation of Driest Quarter
# BIO18 = Precipitation of Warmest Quarter
# BIO19 = Precipitation of Coldest Quarter
# Elev = Elevation above sea level


```

```
# Save to disk:
fwrite(EnvData2, "EnvData2.csv")
#save(bioclim_poly, file="bioclim_poly.RData")



#STEP 3: PERFORM THE VARIABLE SELECTION TO REDUCE MULTICOLLINEARITY AMONG
PREDICTORS
################################################################################
#



# Check the pairwise Pearson correlation coefficient for each predictor:
pearson_r<-as.data.frame(cor(EnvData, use="complete.obs")) # Pearson corre
lation
range(as.dist(pearson_r)) # range of values for pairwise correlations

## [1] -0.9985582  0.9946916

#[1] -0.9988317  0.9986923


# Plot pairwise correlations:
ggcorrplot::ggcorrplot(pearson_r,
                       method = "square", # symbol shape (circle or square
)
                       type = "lower",
                       show.diag = FALSE,
                       colors = c("#c51b7d", "white", "#4d9221"),
                       outline.color = "gray",
```

```
                        hc.order = TRUE, # hierarchical clustering of varia
bles
                        hc.method = "average", # hierarchical clustering me
thod
                        lab = TRUE, # show correlation coefficients on the
plot
                        lab_col = "black",
                        lab_size = 3,
                        # p.mat = NULL, # matrix of p-values (user provided
)
                        # sig.level = 0.05,
                        tl.cex = 4, # axes labels
                        tl.col = "white", # axes colours
                        tl.srt = 90) # x-axis text orientation
```

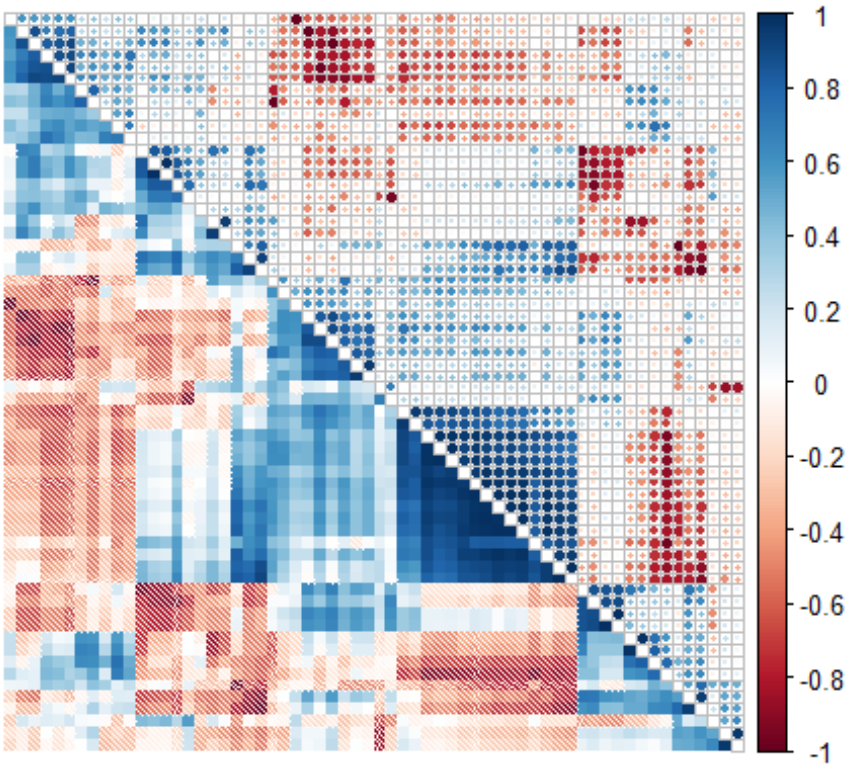


```
                        #digits = 2)


# The package 'corrplot' also does a nice work:
corrplot::corrplot(corr=as.matrix(pearson_r),
                   method="ellipse", # symbol shape (circle, square, ellip
se, pie, number)
                   type="upper", # triangular part to show
                   diag=FALSE,
                   tl.cex = 0.3, # omit diagonal (r = 1)
)
```




```
corrplot::corrplot.mixed(corr=as.matrix(pearson_r),
                         lower="shade",
                         upper="circle",
                         order="hclust",
```

```
                         tl.pos = "n" #esse comando remove axis labels
)
```




```
# See more at: https://cran.r-project.org/web/packages/corrplot/vignettes/
corrplot-intro.html


# Pairwise correlations measure collinearity.
# Use variance inflation factor (VIF) to measure multicollinearity in the
EnvData predictors.
VIF_values <- usdm::vif(EnvData)
VIF_values[order(VIF_values$VIF, decreasing = TRUE), ] # print in decreasi
ng order

##                          Variables VIF
## 1            wc2.1_5m_bio_1_linear Inf
## 2           wc2.1_5m_bio_10_linear Inf
```

```
## 3           wc2.1_5m_bio_11_linear Inf
## 4           wc2.1_5m_bio_12_linear Inf
## 5           wc2.1_5m_bio_13_linear Inf
## 6           wc2.1_5m_bio_14_linear Inf
## 7           wc2.1_5m_bio_15_linear Inf
## 8           wc2.1_5m_bio_16_linear Inf
## 9           wc2.1_5m_bio_17_linear Inf
## 10          wc2.1_5m_bio_18_linear Inf
## 11          wc2.1_5m_bio_19_linear Inf
## 12           wc2.1_5m_bio_2_linear Inf
## 13           wc2.1_5m_bio_3_linear Inf
## 14           wc2.1_5m_bio_4_linear Inf
## 15           wc2.1_5m_bio_5_linear Inf
## 16           wc2.1_5m_bio_6_linear Inf
## 17           wc2.1_5m_bio_7_linear Inf
## 18           wc2.1_5m_bio_8_linear Inf
## 19           wc2.1_5m_bio_9_linear Inf
## 20            wc2.1_5m_elev_linear Inf
## 21     bdod_0.5cm_mean_5000_linear Inf
## 22      cec_0.5cm_mean_5000_linear Inf
## 23     cfvo_0.5cm_mean_5000_linear Inf
## 24     clay_0.5cm_mean_5000_linear Inf
## 25 nitrogen_0.5cm_mean_5000_linear Inf
## 26      ocd_0.5cm_mean_5000_linear Inf
## 27     ocs_0.30cm_mean_5000_linear Inf
## 28    phh2o_0.5cm_mean_5000_linear Inf
## 29     sand_0.5cm_mean_5000_linear Inf
```

```
## 30     silt_0.5cm_mean_5000_linear Inf
## 31      soc_0.5cm_mean_5000_linear Inf
## 32             wc2.1_5m_bio_1_quad Inf
## 33            wc2.1_5m_bio_10_quad Inf
## 34            wc2.1_5m_bio_11_quad Inf
## 35            wc2.1_5m_bio_12_quad Inf
## 36            wc2.1_5m_bio_13_quad Inf
## 37            wc2.1_5m_bio_14_quad Inf
## 38            wc2.1_5m_bio_15_quad Inf
## 39            wc2.1_5m_bio_16_quad Inf
## 40            wc2.1_5m_bio_17_quad Inf
## 41            wc2.1_5m_bio_18_quad Inf
## 42            wc2.1_5m_bio_19_quad Inf
## 43             wc2.1_5m_bio_2_quad Inf
## 44             wc2.1_5m_bio_3_quad Inf
## 45             wc2.1_5m_bio_4_quad Inf
## 46             wc2.1_5m_bio_5_quad Inf
## 47             wc2.1_5m_bio_6_quad Inf
## 48             wc2.1_5m_bio_7_quad Inf
## 49             wc2.1_5m_bio_8_quad Inf
## 50             wc2.1_5m_bio_9_quad Inf
## 51              wc2.1_5m_elev_quad Inf
## 52       bdod_0.5cm_mean_5000_quad Inf
## 53        cec_0.5cm_mean_5000_quad Inf
## 54       cfvo_0.5cm_mean_5000_quad Inf
## 55       clay_0.5cm_mean_5000_quad Inf
## 56   nitrogen_0.5cm_mean_5000_quad Inf
```

```
## 57        ocd_0.5cm_mean_5000_quad Inf
## 58       ocs_0.30cm_mean_5000_quad Inf
## 59      phh2o_0.5cm_mean_5000_quad Inf
## 60       sand_0.5cm_mean_5000_quad Inf
## 61       silt_0.5cm_mean_5000_quad Inf
## 62        soc_0.5cm_mean_5000_quad Inf

#esses valores tem que estar abaixo de 10 para indicarem que nao estao cor
relacionados. Alguns mais
#conservadores consideram abaixo de 5 (é o que vamos usar). Ent?o vamos re
tirando essas vari?veis com valores mais altos e
#roda de novo (pq ele roda usando os dados inteiros, ent?o sempre tem que
rodar de novo)
# VIF measures the proportion by which the variance of a regression coeffi
cient is
# inflated in the presence of other explanatory variables.


# Remove variables iteratively to reduce multicollinearity:


SelectedVars<-list()
SelectedVars[[1]]<-EnvData
VIF_list<-list()
for(i in 1:(ncol(EnvData)-1)){

  # Compute the VIF values:
  VIF_values <- usdm::vif(SelectedVars[[i]])
```

```
  # Reorder predictors in decreasing order of VIF:
  VIF_values<-VIF_values[order(VIF_values$VIF, decreasing = TRUE), ]
  VIF_list[[i]]<-VIF_values

  # Identify variables that should be keep:
  keep_columns<-VIF_values[2:nrow(VIF_values),]$Variables

  # Remove the variable holding the highest VIF:
  SelectedVars[[(i+1)]]<-EnvData[, keep_columns]
}

# Extract the first row (maximum VIF) of each VIF list:
VarSelection<-lapply(VIF_list, function(x) # for each list, apply the func
tion below
  max(x[,2], na.rm=TRUE))

# Identify variable selection does not have predictors with VIF > 5
which(VarSelection <= 5)[1]

## [1] 52

#quais dessas tentativas em que as vari?veis voltaram todas com valores ab
aixo de 5?
EnvData_Filtered <- SelectedVars[[which(VarSelection <= 5)[1]]]
usdm::vif(EnvData_Filtered) #12
```

```
##                      Variables      VIF
## 1         wc2.1_5m_bio_2_quad 4.212702
## 2        wc2.1_5m_bio_14_quad 4.819535
## 3         wc2.1_5m_bio_3_quad 3.212171
## 4          wc2.1_5m_elev_quad 3.507526
## 5   bdod_0.5cm_mean_5000_quad 3.332924
## 6  phh2o_0.5cm_mean_5000_quad 3.362402
## 7        wc2.1_5m_bio_18_quad 3.414823
## 8   ocs_0.30cm_mean_5000_quad 2.783329
## 9   silt_0.5cm_mean_5000_quad 3.097532
## 10  clay_0.5cm_mean_5000_quad 2.860400
## 11  cfvo_0.5cm_mean_5000_quad 1.971071
```

```
# Recompute Pearson correlation coefficient for the selected variables:
pearson_r<-as.data.frame(cor(EnvData_Filtered, use="complete.obs")) # Pear
son correlation
range(as.dist(pearson_r)) # range of values for pairwise correlations
```

```
## [1] -0.6094399  0.7106748
```

```
#[1] -0.7063586  0.7441418


# Export to disk:
#fwrite(EnvData_Filtered, "EnvDataFiltered.csv")
names(EnvData_Filtered)
```

```
##  [1] "wc2.1_5m_bio_2_quad"        "wc2.1_5m_bio_14_quad"
##  [3] "wc2.1_5m_bio_3_quad"        "wc2.1_5m_elev_quad"
##  [5] "bdod_0.5cm_mean_5000_quad"  "phh2o_0.5cm_mean_5000_quad"
##  [7] "wc2.1_5m_bio_18_quad"       "ocs_0.30cm_mean_5000_quad"
```

```
##  [9] "silt_0.5cm_mean_5000_quad"  "clay_0.5cm_mean_5000_quad"
## [11] "cfvo_0.5cm_mean_5000_quad"

# STEP 5: MODEL VARIATION IN SPECIES ASSEMBLAGE USING GDM AND PROJECT TEMP
ORAL CHANGES IN BETA-DIVERSITY
################################################################################
################################################
# STEP 5: MODEL VARIATION IN SPECIES ASSEMBLAGE USING GDM AND PROJECT TEMP
ORAL CHANGES IN BETA-DIVERSITY


# GDM is a constrained modelling approach at the assemblage-level.


# It is necessary to add info on SiteID and geographical coordinates in th
e 'EnvData_Filtered' object:
# Because 'EnvDataFiltered' were obtained in the same order as 'EnvData',
just bind:


EnvData_Filtered2 <- na.omit(EnvData_Filtered)
EnvData_Filtered_regiaoID<-cbind(EnvData2[,1:3], EnvData_Filtered2)


summary(EnvData2)

##     regiao           Longitude.x         Latitude.y       wc2.1_5m_bio_
1_linear
##  Length:28          Min.   :-5807040   Min.   :-3064195   Min.   :0.000
1004
##  Class :character   1st Qu.:-5385467   1st Qu.:-2491757   1st Qu.:0.000
3665
```

```
##  Mode  :character   Median :-5213821   Median :-2388553   Median :0.000
4477
##                     Mean   :-5122621   Mean   :-2307026   Mean   :0.000
4638
##                     3rd Qu.:-4885957   3rd Qu.:-2256147   3rd Qu.:0.000
5839
##                     Max.   :-4186405   Max.   :-1214670   Max.   :0.000
7796
##  wc2.1_5m_bio_10_linear wc2.1_5m_bio_11_linear wc2.1_5m_bio_12_linear
##  Min.   :-3.772e-04     Min.   :0.0003232      Min.   :8.993e-05
##  1st Qu.: 4.115e-05     1st Qu.:0.0005247      1st Qu.:4.978e-04
##  Median : 1.192e-04     Median :0.0005695      Median :7.507e-04
##  Mean   : 1.358e-04     Mean   :0.0005950      Mean   :7.172e-04
##  3rd Qu.: 2.633e-04     3rd Qu.:0.0006902      3rd Qu.:8.177e-04
##  Max.   : 5.019e-04     Max.   :0.0008875      Max.   :1.574e-03
##  wc2.1_5m_bio_13_linear wc2.1_5m_bio_14_linear wc2.1_5m_bio_15_linear
##  Min.   :-0.0002761     Min.   :-2.787e-04     Min.   :-1.115e-03
##  1st Qu.: 0.0005445     1st Qu.:-8.493e-05     1st Qu.:-3.139e-04
##  Median : 0.0007547     Median : 6.074e-05     Median : 1.867e-04
##  Mean   : 0.0007293     Mean   : 3.499e-04     Mean   :-8.565e-05
##  3rd Qu.: 0.0009963     3rd Qu.: 5.717e-04     3rd Qu.: 2.356e-04
##  Max.   : 0.0013093     Max.   : 2.411e-03     Max.   : 4.082e-04
##  wc2.1_5m_bio_16_linear wc2.1_5m_bio_17_linear wc2.1_5m_bio_18_linear
##  Min.   :-0.0002339     Min.   :-3.030e-04     Min.   :0.0001185
##  1st Qu.: 0.0006187     1st Qu.:-6.403e-05     1st Qu.:0.0010442
##  Median : 0.0007775     Median : 2.501e-05     Median :0.0015124
##  Mean   : 0.0007681     Mean   : 3.387e-04     Mean   :0.0013997
```

```
##  3rd Qu.: 0.0009918     3rd Qu.: 5.341e-04     3rd Qu.:0.0016473
##  Max.   : 0.0014570     Max.   : 2.277e-03     Max.   :0.0027574
##  wc2.1_5m_bio_19_linear wc2.1_5m_bio_2_linear wc2.1_5m_bio_3_linear
##  Min.   :-4.023e-04     Min.   :-1.445e-03    Min.   :0.0001547
##  1st Qu.:-2.476e-04     1st Qu.:-5.815e-04    1st Qu.:0.0006910
##  Median :-1.761e-04     Median : 7.936e-05    Median :0.0007852
##  Mean   : 3.108e-05     Mean   :-1.406e-04    Mean   :0.0007273
##  3rd Qu.: 2.831e-04     3rd Qu.: 2.926e-04    3rd Qu.:0.0008030
##  Max.   : 1.440e-03     Max.   : 4.687e-04    Max.   :0.0009865
##  wc2.1_5m_bio_4_linear wc2.1_5m_bio_5_linear wc2.1_5m_bio_6_linear
##  Min.   :-0.0009670    Min.   :-4.210e-04    Min.   :0.0002544
##  1st Qu.:-0.0008206    1st Qu.:-7.739e-05    1st Qu.:0.0004954
##  Median :-0.0007713    Median :-2.537e-05    Median :0.0005467
##  Mean   :-0.0007947    Mean   :-2.368e-05    Mean   :0.0005948
##  3rd Qu.:-0.0007626    3rd Qu.: 5.815e-05    3rd Qu.:0.0006602
##  Max.   :-0.0006495    Max.   : 2.512e-04    Max.   :0.0010290
##  wc2.1_5m_bio_7_linear wc2.1_5m_bio_8_linear wc2.1_5m_bio_9_linear
##  Min.   :-0.0013064    Min.   :-3.759e-05    Min.   :0.0001164
##  1st Qu.:-0.0008905    1st Qu.: 2.926e-04    1st Qu.:0.0003368
##  Median :-0.0007792    Median : 3.852e-04    Median :0.0003657
##  Mean   :-0.0008357    Mean   : 3.810e-04    Mean   :0.0004023
##  3rd Qu.:-0.0006891    3rd Qu.: 4.994e-04    3rd Qu.:0.0004909
##  Max.   :-0.0006492    Max.   : 5.996e-04    Max.   :0.0007910
##  wc2.1_5m_elev_linear bdod_0.5cm_mean_5000_linear cec_0.5cm_mean_5000_l
inear
##  Min.   :-6.084e-04   Min.   :-1.039e-03          Min.   :-1.100e-03
##  1st Qu.:-3.120e-04   1st Qu.:-2.343e-04          1st Qu.:-8.191e-04
```

```
##  Median : 4.464e-05   Median : 9.695e-06          Median :-7.126e-04
##  Mean   : 6.961e-06   Mean   :-2.546e-05          Mean   :-6.572e-04
##  3rd Qu.: 1.848e-04   3rd Qu.: 2.212e-04          3rd Qu.:-4.755e-04
##  Max.   : 1.005e-03   Max.   : 5.681e-04          Max.   :-8.888e-05
##  cfvo_0.5cm_mean_5000_linear clay_0.5cm_mean_5000_linear
##  Min.   :-0.0012099          Min.   :6.953e-05
##  1st Qu.:-0.0008412          1st Qu.:7.853e-04
##  Median :-0.0006978          Median :1.494e-03
##  Mean   :-0.0007074          Mean   :1.498e-03
##  3rd Qu.:-0.0005453          3rd Qu.:2.252e-03
##  Max.   :-0.0003068          Max.   :3.002e-03
##  nitrogen_0.5cm_mean_5000_linear ocd_0.5cm_mean_5000_linear
##  Min.   :-0.0005444              Min.   :-5.236e-04
##  1st Qu.:-0.0004154              1st Qu.:-9.416e-05
##  Median :-0.0002929              Median : 7.544e-05
##  Mean   :-0.0002373              Mean   : 1.609e-04
##  3rd Qu.:-0.0001187              3rd Qu.: 3.430e-04
##  Max.   : 0.0004597              Max.   : 1.052e-03
##  ocs_0.30cm_mean_5000_linear phh2o_0.5cm_mean_5000_linear
##  Min.   :-5.965e-04          Min.   :-0.0012158
##  1st Qu.:-4.209e-05          1st Qu.:-0.0010224
##  Median : 7.069e-05          Median :-0.0009564
##  Mean   : 2.694e-04          Mean   :-0.0008922
##  3rd Qu.: 5.382e-04          3rd Qu.:-0.0007364
##  Max.   : 1.692e-03          Max.   :-0.0004048
##  sand_0.5cm_mean_5000_linear silt_0.5cm_mean_5000_linear
##  Min.   :-0.0015293          Min.   :-0.0013018
```

```
##  1st Qu.:-0.0006608          1st Qu.:-0.0010473
##  Median :-0.0003851          Median :-0.0008621
##  Mean   :-0.0002618          Mean   :-0.0007370
##  3rd Qu.: 0.0002695          3rd Qu.:-0.0005442
##  Max.   : 0.0008953          Max.   : 0.0001420
##  soc_0.5cm_mean_5000_linear wc2.1_5m_bio_1_quad  wc2.1_5m_bio_10_quad
##  Min.   :-0.0004869         Min.   :-7.797e-04   Min.   :-7.179e-04
##  1st Qu.:-0.0003838         1st Qu.:-4.275e-04   1st Qu.:-5.585e-04
##  Median :-0.0003346         Median :-2.844e-04   Median :-4.959e-04
##  Mean   :-0.0002172         Mean   :-2.269e-04   Mean   :-4.510e-04
##  3rd Qu.:-0.0000398         3rd Qu.:-7.697e-06   3rd Qu.:-3.524e-04
##  Max.   : 0.0001741         Max.   : 4.711e-04   Max.   :-3.933e-05
##  wc2.1_5m_bio_11_quad wc2.1_5m_bio_12_quad wc2.1_5m_bio_13_quad
##  Min.   :-6.023e-04   Min.   :-0.0008542   Min.   :-0.0006169
##  1st Qu.:-2.344e-04   1st Qu.:-0.0008498   1st Qu.:-0.0006115
##  Median :-1.376e-04   Median :-0.0008368   Median :-0.0005964
##  Mean   :-5.842e-05   Mean   :-0.0007815   Mean   :-0.0005393
##  3rd Qu.: 1.517e-04   3rd Qu.:-0.0007987   3rd Qu.:-0.0005477
##  Max.   : 7.109e-04   Max.   :-0.0002979   Max.   :-0.0000301
##  wc2.1_5m_bio_14_quad wc2.1_5m_bio_15_quad wc2.1_5m_bio_16_quad
##  Min.   :-1.271e-03   Min.   :-0.0007630   Min.   :-0.0006436
##  1st Qu.:-9.107e-04   1st Qu.:-0.0007615   1st Qu.:-0.0006375
##  Median :-4.202e-04   Median :-0.0007543   Median :-0.0006267
##  Mean   :-5.342e-04   Mean   :-0.0003730   Mean   :-0.0005724
##  3rd Qu.:-2.216e-04   3rd Qu.:-0.0004214   3rd Qu.:-0.0005599
##  Max.   : 7.505e-05   Max.   : 0.0011268   Max.   :-0.0001133
##  wc2.1_5m_bio_17_quad wc2.1_5m_bio_18_quad wc2.1_5m_bio_19_quad
```

```
##  Min.   :-0.0011929   Min.   :-0.0006059   Min.   :-1.038e-03
##  1st Qu.:-0.0008794   1st Qu.:-0.0005535   1st Qu.:-5.984e-04
##  Median :-0.0003768   Median :-0.0003889   Median :-9.581e-05
##  Mean   :-0.0005192   Mean   :-0.0003135   Mean   :-2.443e-04
##  3rd Qu.:-0.0002591   3rd Qu.:-0.0003416   3rd Qu.:-6.570e-08
##  Max.   : 0.0000947   Max.   : 0.0009258   Max.   : 2.230e-04
##  wc2.1_5m_bio_2_quad  wc2.1_5m_bio_3_quad  wc2.1_5m_bio_4_quad
##  Min.   :-0.0007214   Min.   :-0.0008905   Min.   :3.576e-06
##  1st Qu.:-0.0006896   1st Qu.:-0.0005811   1st Qu.:2.349e-04
##  Median :-0.0005825   Median :-0.0004544   Median :2.539e-04
##  Mean   :-0.0003506   Mean   :-0.0004977   Mean   :3.116e-04
##  3rd Qu.:-0.0004266   3rd Qu.:-0.0004280   3rd Qu.:3.646e-04
##  Max.   : 0.0012825   Max.   :-0.0001105   Max.   :7.256e-04
##  wc2.1_5m_bio_5_quad  wc2.1_5m_bio_6_quad  wc2.1_5m_bio_7_quad
##  Min.   :-0.0006711   Min.   :-6.939e-04   Min.   :-9.638e-05
##  1st Qu.:-0.0005782   1st Qu.:-3.100e-04   1st Qu.:-2.930e-05
##  Median :-0.0005494   Median :-2.080e-04   Median : 1.338e-04
##  Mean   :-0.0005288   Mean   :-6.392e-05   Mean   : 2.847e-04
##  3rd Qu.:-0.0004926   3rd Qu.: 4.320e-05   3rd Qu.: 3.589e-04
##  Max.   :-0.0003210   Max.   : 1.096e-03   Max.   : 1.427e-03
##  wc2.1_5m_bio_8_quad  wc2.1_5m_bio_9_quad  wc2.1_5m_elev_quad
##  Min.   :-4.492e-04   Min.   :-0.0008897   Min.   :-1.269e-03
##  1st Qu.:-1.467e-04   1st Qu.:-0.0005692   1st Qu.:-6.876e-04
##  Median :-3.795e-05   Median :-0.0005170   Median :-5.023e-04
##  Mean   :-3.013e-05   Mean   :-0.0004194   Mean   :-3.592e-04
##  3rd Qu.: 1.106e-04   3rd Qu.:-0.0002618   3rd Qu.: 8.141e-05
##  Max.   : 2.539e-04   Max.   : 0.0005341   Max.   : 6.871e-04
```

```
##  bdod_0.5cm_mean_5000_quad cec_0.5cm_mean_5000_quad cfvo_0.5cm_mean_500
0_quad
##  Min.   :-9.231e-04        Min.   :-5.607e-04       Min.   :-3.347e-04
##  1st Qu.:-8.610e-04        1st Qu.:-4.087e-05       1st Qu.:-9.152e-05
##  Median :-7.406e-04        Median : 4.048e-04       Median : 1.229e-04
##  Mean   :-6.597e-04        Mean   : 3.578e-04       Mean   : 1.901e-04
##  3rd Qu.:-5.492e-04        3rd Qu.: 6.379e-04       3rd Qu.: 3.870e-04
##  Max.   :-2.488e-05        Max.   : 1.343e-03       Max.   : 1.157e-03
##  clay_0.5cm_mean_5000_quad nitrogen_0.5cm_mean_5000_quad
##  Min.   :-0.0005933        Min.   :-0.0010142
##  1st Qu.:-0.0003053        1st Qu.:-0.0005958
##  Median : 0.0007399        Median :-0.0003381
##  Mean   : 0.0012710        Mean   :-0.0003674
##  3rd Qu.: 0.0027039        3rd Qu.:-0.0001240
##  Max.   : 0.0055241        Max.   : 0.0001352
##  ocd_0.5cm_mean_5000_quad ocs_0.30cm_mean_5000_quad phh2o_0.5cm_mean_50
00_quad
##  Min.   :-9.057e-04       Min.   :-0.0007021        Min.   :-0.0006803
##  1st Qu.:-8.813e-04       1st Qu.:-0.0006767        1st Qu.:-0.0001753
##  Median :-8.428e-04       Median :-0.0006261        Median : 0.0003166
##  Mean   :-7.056e-04       Mean   :-0.0004343        Mean   : 0.0002094
##  3rd Qu.:-6.163e-04       3rd Qu.:-0.0005064        3rd Qu.: 0.0004852
##  Max.   : 3.285e-05       Max.   : 0.0011482        Max.   : 0.0010453
##  sand_0.5cm_mean_5000_quad silt_0.5cm_mean_5000_quad soc_0.5cm_mean_500
0_quad
##  Min.   :-0.0006244        Min.   :-6.043e-04        Min.   :-8.273e-04
##  1st Qu.:-0.0005396        1st Qu.:-3.602e-04        1st Qu.:-5.790e-04
```

```
##  Median :-0.0003946        Median : 8.816e-06        Median :-9.848e-05
##  Mean   :-0.0002312        Mean   :-2.453e-05        Mean   :-2.712e-04
##  3rd Qu.:-0.0001158        3rd Qu.: 3.001e-04        3rd Qu.:-3.812e-06
##  Max.   : 0.0014851        Max.   : 7.943e-04        Max.   : 2.080e-04

EnvData2$wc2.1_5m_bio_14_quad

##  [1] -9.529959e-04 -1.695247e-04 -1.805601e-04 -1.107412e-03 -3.329895e
-04
##  [6] -4.327002e-04 -4.077657e-04 -4.873488e-04 -3.738628e-04 -1.039218e
-04
## [11] -5.360498e-04 -9.513491e-04 -2.221022e-04 -1.164067e-03 -2.199703e
-04
## [16] -1.271495e-03 -7.517242e-04 -1.263964e-03 -2.728861e-04  7.505380e
-05
## [21]  7.260466e-05 -1.099882e-03 -6.050539e-04 -5.203377e-04 -2.972433e
-04
## [26] -3.141282e-04 -8.970959e-04 -1.695078e-04

EnvData_Filtered2$wc2.1_5m_bio_14_quad

##  [1] -9.529959e-04 -1.695247e-04 -1.805601e-04 -1.107412e-03 -3.329895e
-04
##  [6] -4.327002e-04 -4.077657e-04 -4.873488e-04 -3.738628e-04 -1.039218e
-04
## [11] -5.360498e-04 -9.513491e-04 -2.221022e-04 -1.164067e-03 -2.199703e
-04
## [16] -1.271495e-03 -7.517242e-04 -1.263964e-03 -2.728861e-04  7.505380e
-05
## [21]  7.260466e-05 -1.099882e-03 -6.050539e-04 -5.203377e-04 -2.972433e
```

```
-04
## [26] -3.141282e-04 -8.970959e-04 -1.695078e-04

EnvData_Filtered_regiaoID$wc2.1_5m_bio_14_quad

##  [1] -9.529959e-04 -1.695247e-04 -1.805601e-04 -1.107412e-03 -3.329895e
-04
##  [6] -4.327002e-04 -4.077657e-04 -4.873488e-04 -3.738628e-04 -1.039218e
-04
## [11] -5.360498e-04 -9.513491e-04 -2.221022e-04 -1.164067e-03 -2.199703e
-04
## [16] -1.271495e-03 -7.517242e-04 -1.263964e-03 -2.728861e-04  7.505380e
-05
## [21]  7.260466e-05 -1.099882e-03 -6.050539e-04 -5.203377e-04 -2.972433e
-04
## [26] -3.141282e-04 -8.970959e-04 -1.695078e-04

PAM_Wide <- tibble::rownames_to_column(pam_wide4, var = "regiao")
remaining_ids <- EnvData_Filtered_regiaoID$regiao
PAM_Wide3 <- PAM_Wide[PAM_Wide$regiao %in% remaining_ids, ]
row.names(PAM_Wide3) <- PAM_Wide3$regiao
pam<-as.data.frame(PAM_Wide3[, 1:ncol(PAM_Wide3)])
row_names <- rownames(pam)
pam2 <- as.data.frame(lapply(pam, function(x) as.integer(x > 0)))
rownames(pam2) <- row_names
pam3 <- pam2[,-1]



DissMatrices<-betapart::beta.pair(x = pam3, index.family= "sorensen")
```

```
# It is recommended to take the square-root of diss. matrix without Euclid
ean property:
DissMatrix_sqrt<-sqrt(as.matrix(DissMatrices$beta.sor))
DissMatrix_regiaoID<-data.frame(regiao = EnvData_Filtered_regiaoID$regiao,
DissMatrix_sqrt)
DissMatrix_Coords <- data.frame(EnvData_Filtered_regiaoID[,1:3], DissMatri
x_sqrt)

fwrite(DissMatrix_Coords, "DissMatrix_Coords.txt")
DissMatrix_regiaoID3 <- DissMatrix_regiaoID
DissMatrix_regiaoID3$row_names <- rownames(DissMatrix_regiaoID3)

DissMatrix_regiaoID4 <- DissMatrix_regiaoID3[, c("row_names", setdiff(name
s(DissMatrix_regiaoID3), "row_names"))]

fwrite(DissMatrix_regiaoID4, "DissMatrix_regiaoID4.txt")

#EnvData_Filtered_regiaoID, são as variáveis ambientais filtradas, formato
de tabela, com coords
#bioclim_soil_poly, são todas as variáveis ambientais, formato raster

# It is possible to use the GDM_Model to predict dissimilarity distance be
tween sites based on environmental data.
# Load data on environmental variables for the current scenario, crop, and
mask:
names(bioclim_soil_poly)
```

```
##  [1] "wc2.1_5m_bio_1_linear"           "wc2.1_5m_bio_10_linear"
##  [3] "wc2.1_5m_bio_11_linear"          "wc2.1_5m_bio_12_linear"
##  [5] "wc2.1_5m_bio_13_linear"          "wc2.1_5m_bio_14_linear"
##  [7] "wc2.1_5m_bio_15_linear"          "wc2.1_5m_bio_16_linear"
##  [9] "wc2.1_5m_bio_17_linear"          "wc2.1_5m_bio_18_linear"
## [11] "wc2.1_5m_bio_19_linear"          "wc2.1_5m_bio_2_linear"
## [13] "wc2.1_5m_bio_3_linear"           "wc2.1_5m_bio_4_linear"
## [15] "wc2.1_5m_bio_5_linear"           "wc2.1_5m_bio_6_linear"
## [17] "wc2.1_5m_bio_7_linear"           "wc2.1_5m_bio_8_linear"
## [19] "wc2.1_5m_bio_9_linear"           "wc2.1_5m_elev_linear"
## [21] "bdod_0.5cm_mean_5000_linear"     "cec_0.5cm_mean_5000_linear"
## [23] "cfvo_0.5cm_mean_5000_linear"     "clay_0.5cm_mean_5000_linear"
## [25] "nitrogen_0.5cm_mean_5000_linear" "ocd_0.5cm_mean_5000_linear"
## [27] "ocs_0.30cm_mean_5000_linear"     "phh2o_0.5cm_mean_5000_linear"
## [29] "sand_0.5cm_mean_5000_linear"     "silt_0.5cm_mean_5000_linear"
## [31] "soc_0.5cm_mean_5000_linear"      "wc2.1_5m_bio_1_quad"
## [33] "wc2.1_5m_bio_10_quad"            "wc2.1_5m_bio_11_quad"
## [35] "wc2.1_5m_bio_12_quad"            "wc2.1_5m_bio_13_quad"
## [37] "wc2.1_5m_bio_14_quad"            "wc2.1_5m_bio_15_quad"
## [39] "wc2.1_5m_bio_16_quad"            "wc2.1_5m_bio_17_quad"
## [41] "wc2.1_5m_bio_18_quad"            "wc2.1_5m_bio_19_quad"
## [43] "wc2.1_5m_bio_2_quad"             "wc2.1_5m_bio_3_quad"
## [45] "wc2.1_5m_bio_4_quad"             "wc2.1_5m_bio_5_quad"


## [47] "wc2.1_5m_bio_6_quad"             "wc2.1_5m_bio_7_quad"
## [49] "wc2.1_5m_bio_8_quad"             "wc2.1_5m_bio_9_quad"
```

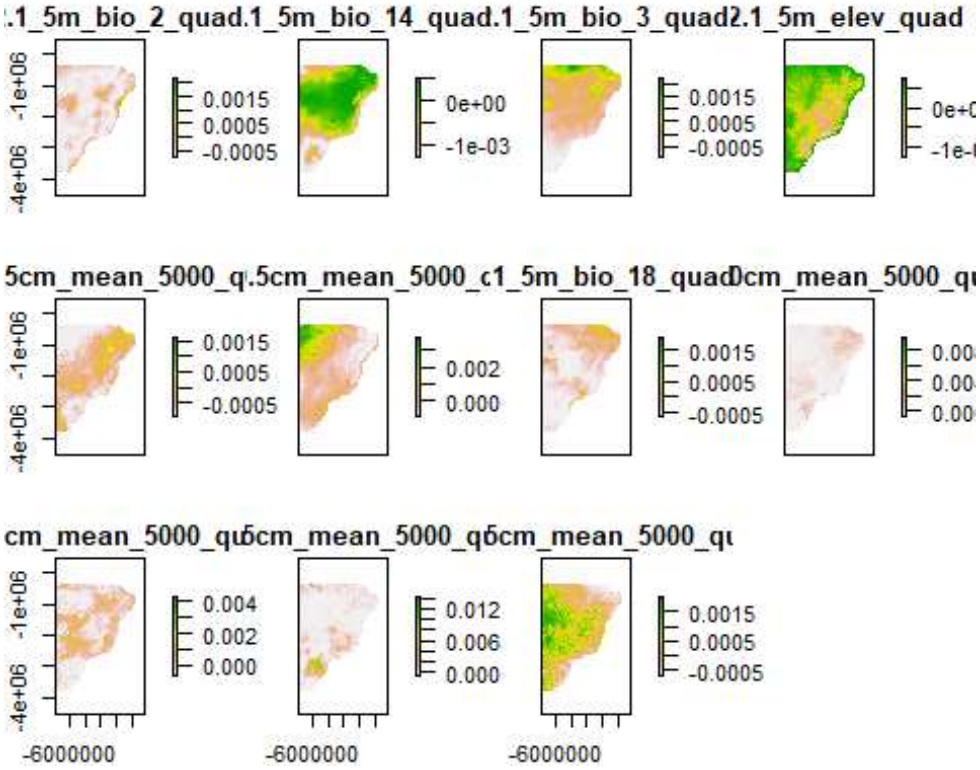
```
## [51] "wc2.1_5m_elev_quad"              "bdod_0.5cm_mean_5000_quad"
## [53] "cec_0.5cm_mean_5000_quad"        "cfvo_0.5cm_mean_5000_quad"
## [55] "clay_0.5cm_mean_5000_quad"       "nitrogen_0.5cm_mean_5000_quad"
## [57] "ocd_0.5cm_mean_5000_quad"        "ocs_0.30cm_mean_5000_quad"
## [59] "phh2o_0.5cm_mean_5000_quad"      "sand_0.5cm_mean_5000_quad"
## [61] "silt_0.5cm_mean_5000_quad"       "soc_0.5cm_mean_5000_quad"
```

```
bioclim_current<-raster::crop(bioclim_soil_poly, mata_atlantica_projected)
# the output is a raster brick

bioclim_current<-raster::stack(bioclim_current) # convert raster brick to
formal class raster stack
bioclim_names<-gsub("", "", names(bioclim_current)) # rename raster layers
as in 'EnvData'
names(bioclim_current)<-bioclim_names
names(bioclim_current)
```

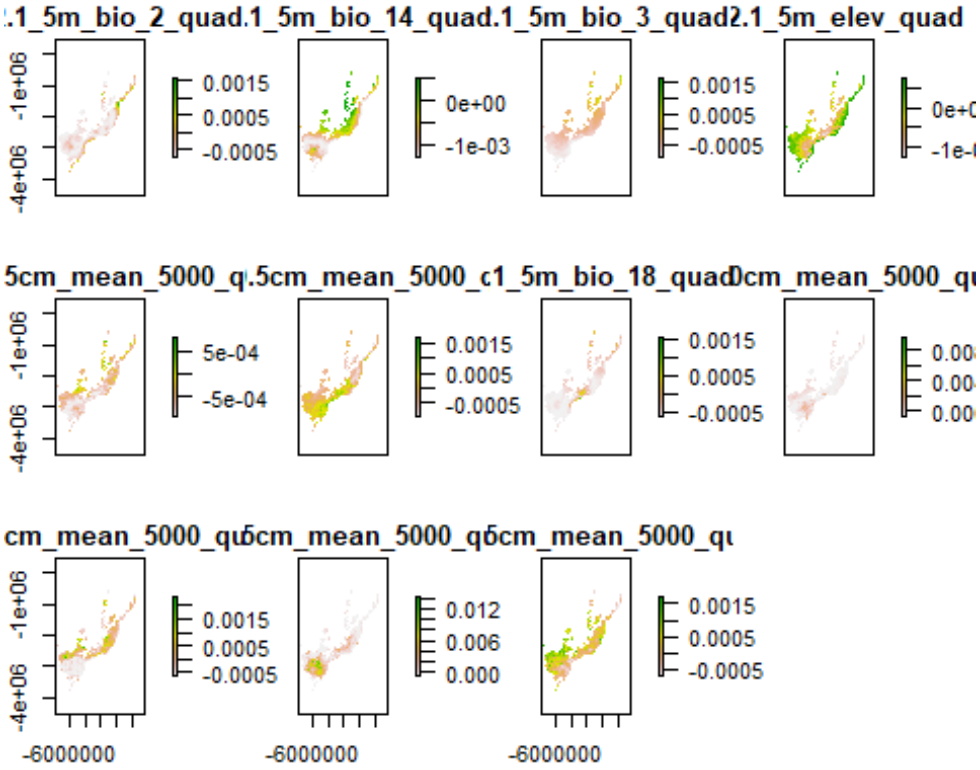
```
##  [1] "wc2.1_5m_bio_1_linear"           "wc2.1_5m_bio_10_linear"
##  [3] "wc2.1_5m_bio_11_linear"          "wc2.1_5m_bio_12_linear"
##  [5] "wc2.1_5m_bio_13_linear"          "wc2.1_5m_bio_14_linear"
##  [7] "wc2.1_5m_bio_15_linear"          "wc2.1_5m_bio_16_linear"
##  [9] "wc2.1_5m_bio_17_linear"          "wc2.1_5m_bio_18_linear"
## [11] "wc2.1_5m_bio_19_linear"          "wc2.1_5m_bio_2_linear"

## [13] "wc2.1_5m_bio_3_linear"           "wc2.1_5m_bio_4_linear"
## [15] "wc2.1_5m_bio_5_linear"           "wc2.1_5m_bio_6_linear"
## [17] "wc2.1_5m_bio_7_linear"           "wc2.1_5m_bio_8_linear"
```

```
## [19] "wc2.1_5m_bio_9_linear"           "wc2.1_5m_elev_linear"
## [21] "bdod_0.5cm_mean_5000_linear"     "cec_0.5cm_mean_5000_linear"
## [23] "cfvo_0.5cm_mean_5000_linear"     "clay_0.5cm_mean_5000_linear"
## [25] "nitrogen_0.5cm_mean_5000_linear" "ocd_0.5cm_mean_5000_linear"
## [27] "ocs_0.30cm_mean_5000_linear"     "phh2o_0.5cm_mean_5000_linear"
## [29] "sand_0.5cm_mean_5000_linear"     "silt_0.5cm_mean_5000_linear"
## [31] "soc_0.5cm_mean_5000_linear"      "wc2.1_5m_bio_1_quad"
## [33] "wc2.1_5m_bio_10_quad"            "wc2.1_5m_bio_11_quad"
## [35] "wc2.1_5m_bio_12_quad"            "wc2.1_5m_bio_13_quad"
## [37] "wc2.1_5m_bio_14_quad"            "wc2.1_5m_bio_15_quad"
## [39] "wc2.1_5m_bio_16_quad"            "wc2.1_5m_bio_17_quad"
## [41] "wc2.1_5m_bio_18_quad"            "wc2.1_5m_bio_19_quad"
## [43] "wc2.1_5m_bio_2_quad"             "wc2.1_5m_bio_3_quad"
## [45] "wc2.1_5m_bio_4_quad"             "wc2.1_5m_bio_5_quad"
## [47] "wc2.1_5m_bio_6_quad"             "wc2.1_5m_bio_7_quad"
## [49] "wc2.1_5m_bio_8_quad"             "wc2.1_5m_bio_9_quad"
## [51] "wc2.1_5m_elev_quad"              "bdod_0.5cm_mean_5000_quad"
## [53] "cec_0.5cm_mean_5000_quad"        "cfvo_0.5cm_mean_5000_quad"
## [55] "clay_0.5cm_mean_5000_quad"       "nitrogen_0.5cm_mean_5000_quad"
## [57] "ocd_0.5cm_mean_5000_quad"        "ocs_0.30cm_mean_5000_quad"
## [59] "phh2o_0.5cm_mean_5000_quad"      "sand_0.5cm_mean_5000_quad"
## [61] "silt_0.5cm_mean_5000_quad"       "soc_0.5cm_mean_5000_quad"

# It is possible to get predictor data from a raster stack, but PAM_Wide n
eed to have coordinates:
#PAM_Wide4_SiteID_coords<-cbind(OrthoPolyData[,1:3], PAM_Wide4)


EnvData_Selected <- EnvData_Filtered_regiaoID %>%
```

```
  select(regiao, 2:3)


# Realizar a união
PAM_Wide_coords <- left_join(EnvData_Selected, PAM_Wide3, by = "regiao")



#agora preciso filtrar bioclim_current para conter somente as variáveis
#que tem VIF abaixo de 5, ou seja, as que estão em EnvData_Filtered


# Get the names of the variables in the EnvData_Filtered object
selected_var_names <- colnames(EnvData_Filtered)


# Filter the layers of the bioclim_soil_poly raster stack
bioclim_current_filtered <- bioclim_current[[selected_var_names]]


# Print out the names of the filtered raster stack to verify
print(names(bioclim_current_filtered))

##  [1] "wc2.1_5m_bio_2_quad"        "wc2.1_5m_bio_14_quad"
##  [3] "wc2.1_5m_bio_3_quad"        "wc2.1_5m_elev_quad"
##  [5] "bdod_0.5cm_mean_5000_quad"  "phh2o_0.5cm_mean_5000_quad"
##  [7] "wc2.1_5m_bio_18_quad"       "ocs_0.30cm_mean_5000_quad"
##  [9] "silt_0.5cm_mean_5000_quad"  "clay_0.5cm_mean_5000_quad"
## [11] "cfvo_0.5cm_mean_5000_quad"

plot(bioclim_current_filtered)
```




```
bioclim_current_filtered_mask <- raster::mask (bioclim_current_filtered, m
ata_atlantica_projected)
bioclim_current_filtered <- raster::stack (bioclim_current_filtered_mask)
plot(bioclim_current_filtered)
```

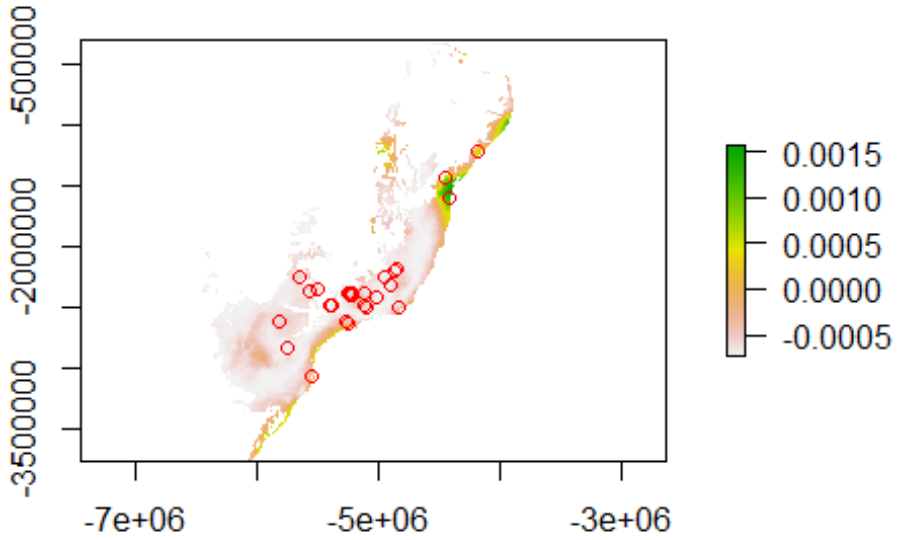



```
extent(bioclim_current_filtered)

## class       : Extent

## xmin        : -6459750

## xmax        : -3609750

## ymin        : -3759000

## ymax        : -309000

crs(regiaoCoords_transformed)



## Coordinate Reference System:

## Deprecated Proj.4 representation:

##  +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs

## WKT2 2019 representation:

## PROJCRS["unknown",
```

```
##     BASEGEOGCRS["unknown",
##         DATUM["World Geodetic System 1984",
##             ELLIPSOID["WGS 84",6378137,298.257223563,
##                 LENGTHUNIT["metre",1]],
##             ID["EPSG",6326]],
##         PRIMEM["Greenwich",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8901]]],
##     CONVERSION["unknown",
##         METHOD["Interrupted Goode Homolosine"],
##         PARAMETER["Longitude of natural origin",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8802]],
##         PARAMETER["False easting",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8806]],
##         PARAMETER["False northing",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8807]]],
##     CS[Cartesian,2],
##         AXIS["(E)",east,
##             ORDER[1],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]],
##         AXIS["(N)",north,
##             ORDER[2],
```

```
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]]]

crs(bioclim_current_filtered)

## Coordinate Reference System:
## Deprecated Proj.4 representation:
##  +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs
## WKT2 2019 representation:
## PROJCRS["unknown",
##     BASEGEOGCRS["unknown",
##         DATUM["World Geodetic System 1984",
##             ELLIPSOID["WGS 84",6378137,298.257223563,
##                 LENGTHUNIT["metre",1]],
##             ID["EPSG",6326]],
##         PRIMEM["Greenwich",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8901]]],
##     CONVERSION["unknown",
##         METHOD["Interrupted Goode Homolosine"],
##         PARAMETER["Longitude of natural origin",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8802]],
##         PARAMETER["False easting",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8806]],
##         PARAMETER["False northing",0,
##             LENGTHUNIT["metre",1],
```

```
##             ID["EPSG",8807]]],
##     CS[Cartesian,2],
##         AXIS["(E)",east,
##             ORDER[1],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]],
##         AXIS["(N)",north,
##             ORDER[2],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]]]

extent(regiaoCoords_transformed)

## class      : Extent
## xmin       : -5807040
## xmax       : -4186405
## ymin       : -3064195
## ymax       : -1214670

extent(bioclim_current_filtered)

## class      : Extent
## xmin       : -6459750
## xmax       : -3609750
## ymin       : -3759000
## ymax       : -309000

plot(bioclim_current_filtered[[1]])
points(regiaoCoords_transformed, col = "red")
```




```
print(res(bioclim_current_filtered))

## [1] 10000 10000

#GERANDO O MODELO GERAL (CLIMA + SOLO)


gdmData_r<-gdm::formatsitepair(bioData = PAM_Wide_coords,
                               bioFormat = 1,
                               dist="bray",
                               abundance=FALSE,
                               siteColumn= "regiao",
                               XColumn = "Longitude.x",
                               YColumn = "Latitude.y",
                               predData = bioclim_current_filtered, # raste
r stack
                               weightType = "equal",
```

```
                                custWeights = NULL,
                                sampleSites = 1,
                                )
```

```
## Warning in gdm::formatsitepair(bioData = PAM_Wide_coords, bioFormat = 1
, : When using rasters for prediction data, sites are assigned to the
##               cells in which they are located and then aggreagted as ne
cessary (e.g.,
##               if more than one site falls in the same raster cell - com
mon for rasters
##               with large cells).
```

```
# The gdmData object named 'gdmData_r' cannot have NA values.
# Before computing the GDM, double check if there are NA values:
sum(is.na(gdmData_r)) #0
```

```
## [1] 297
```

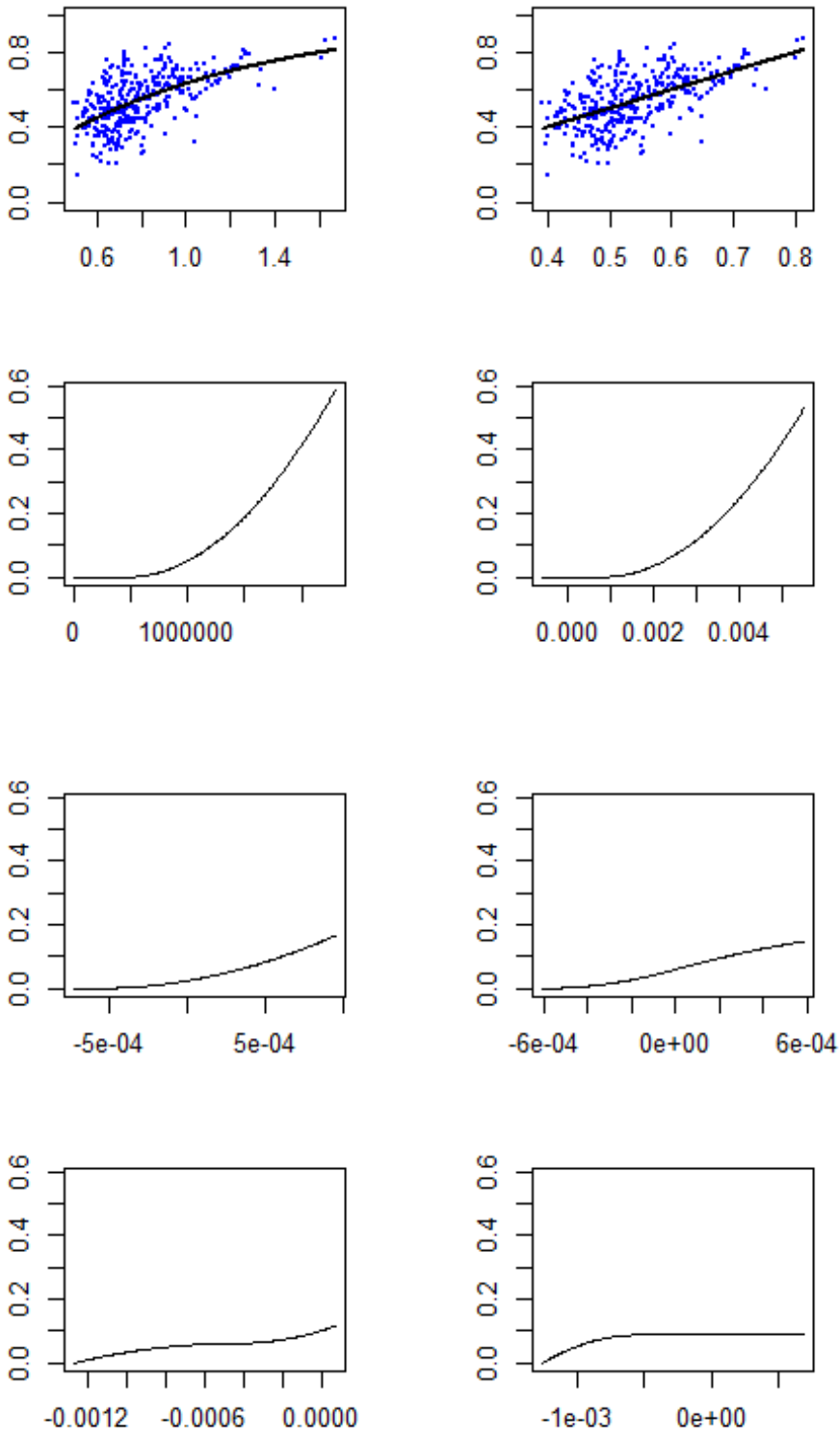
```
gdmData_r <- na.omit(gdmData_r)
```

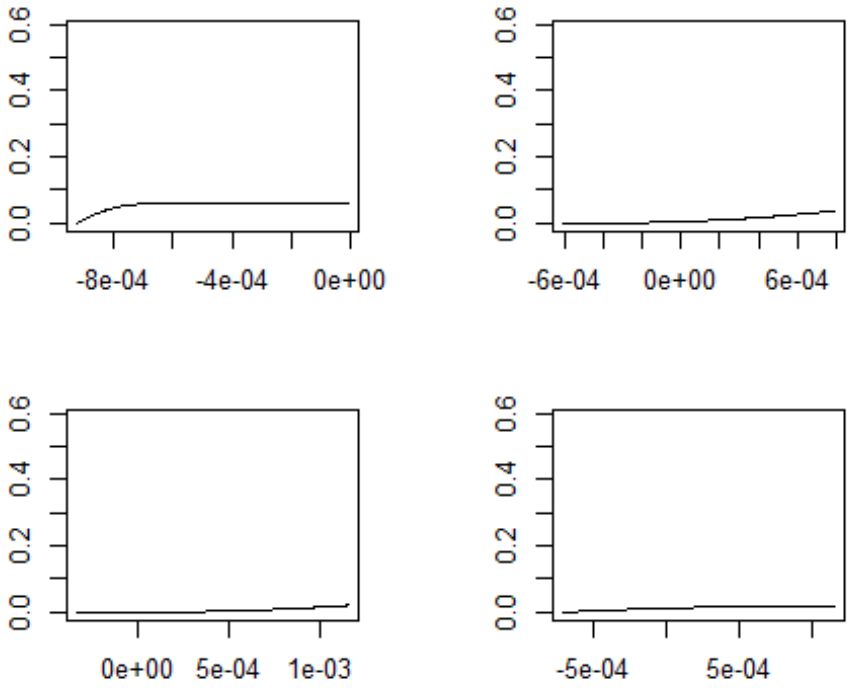
```
# Compute the GDM using predictors as a raster stack (assemblages falling
within the same pixel are pooled):
gdm_Model_r<-gdm::gdm(data = gdmData_r, geo=TRUE)
gdm_Model_r$explained
```

```
## [1] 32.05404
```

```
#[1] 32.3636
```

```
plot(gdm_Model_r)
```


0.0
0.4
0.8
0.6
1.0
1.4
0.4
0.5
0.6
0.7
0.8
0.0
0.2
0.4
0.6
0
1000000
0.000
0.002
0.004
-5e-04
5e-04
-6e-04
0e+00
6e-04
-0.0012
-0.0006
0.0000
-1e-03
0e+00




```
######


#GERANDO O MODELO SÓ COM CLIMA


bioclim_current_filtered_clima <- subset(bioclim_current_filtered, c("wc2.
1_5m_bio_2_quad", "wc2.1_5m_bio_14_quad", "wc2.1_5m_bio_3_quad", "wc2.1_5m
_bio_18_quad"))



gdmData_r_clima<-gdm::formatsitepair(bioData = PAM_Wide_coords,
                                     bioFormat = 1,
                                     dist="bray",
                                     abundance=FALSE,
                                     siteColumn= "regiao",
                                     XColumn = "Longitude.x",
```

```
                                YColumn = "Latitude.y",
                                predData = bioclim_current_filtered_clima,
# raster stack
                                weightType = "equal",
                                custWeights = NULL,
                                sampleSites = 1,
)
```

```
## Warning in gdm::formatsitepair(bioData = PAM_Wide_coords, bioFormat = 1
, : When using rasters for prediction data, sites are assigned to the
##               cells in which they are located and then aggreagted as ne
cessary (e.g.,
##               if more than one site falls in the same raster cell - com
mon for rasters
##               with large cells).
```

```
# The gdmData object named 'gdmData_r' cannot have NA values.
# Before computing the GDM, double check if there are NA values:
sum(is.na(gdmData_r_clima)) #0
```

```
## [1] 108
```

```
gdmData_r_clima <- na.omit(gdmData_r_clima)


# Compute the GDM using predictors as a raster stack (assemblages falling
within the same pixel are pooled):
gdm_Model_r_clima<-gdm::gdm(data = gdmData_r_clima, geo=TRUE)
gdm_Model_r_clima$explained
```

```
## [1] 18.08701
```

```
#[1] 18.08701



##################

#GERANDO O MODELO SÓ SOLO


bioclim_current_filtered_solo <- subset(bioclim_current_filtered, c("bdod_
0.5cm_mean_5000_quad", "phh2o_0.5cm_mean_5000_quad", "ocs_0.30cm_mean_5000
_quad", "silt_0.5cm_mean_5000_quad", "clay_0.5cm_mean_5000_quad", "cfvo_0.
5cm_mean_5000_quad"))



gdmData_r_solo<-gdm::formatsitepair(bioData = PAM_Wide_coords,
                                    bioFormat = 1,
                                    dist="bray",
                                    abundance=FALSE,
                                    siteColumn= "regiao",
                                    XColumn = "Longitude.x",
                                    YColumn = "Latitude.y",
                                    predData = bioclim_current_filtered_solo, #
raster stack
                                    weightType = "equal",
                                    custWeights = NULL,
                                    sampleSites = 1,
)
```

```
## Warning in gdm::formatsitepair(bioData = PAM_Wide_coords, bioFormat = 1
, : When using rasters for prediction data, sites are assigned to the
##               cells in which they are located and then aggreagted as ne
cessary (e.g.,
##               if more than one site falls in the same raster cell - com
mon for rasters
##               with large cells).

# The gdmData object named 'gdmData_r' cannot have NA values.
# Before computing the GDM, double check if there are NA values:
sum(is.na(gdmData_r_solo)) #0

## [1] 162

gdmData_r_solo <- na.omit(gdmData_r_solo)


# Compute the GDM using predictors as a raster stack (assemblages falling
within the same pixel are pooled):
gdm_Model_r_solo<-gdm::gdm(data = gdmData_r_solo, geo=TRUE)
gdm_Model_r_solo$explained

## [1] 29.59122

#[1] 29.59122


######



#GERANDO O MODELO SÓ GEOGRÁFICO
```

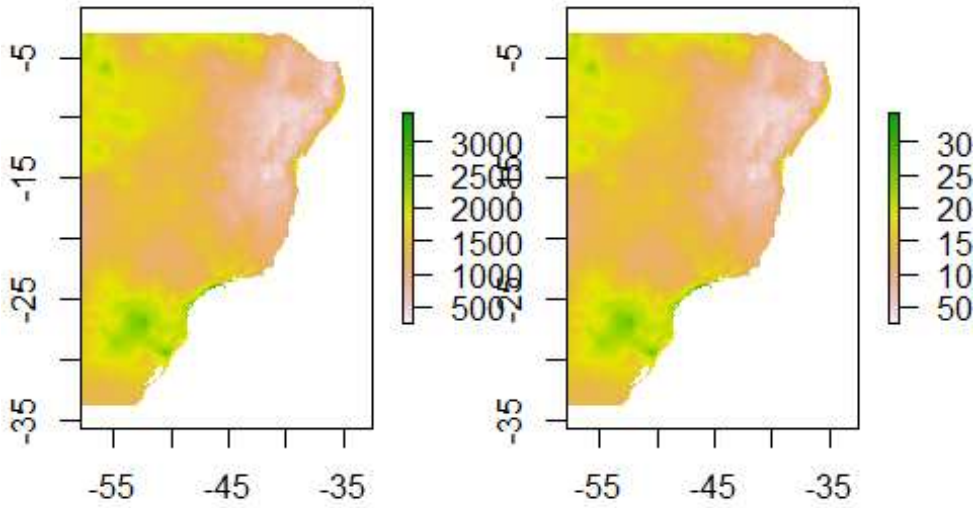
```
bioclim_current_filtered_geo <- subset(bioclim_current_filtered, c("wc2.1_
5m_elev_quad"))



gdmData_r_geo<-gdm::formatsitepair(bioData = PAM_Wide_coords,
                                   bioFormat = 1,
                                   dist="bray",
                                   abundance=FALSE,
                                   siteColumn= "regiao",
                                   XColumn = "Longitude.x",
                                   YColumn = "Latitude.y",
                                   predData = bioclim_current_filtered_ge
o, # raster stack
                                   weightType = "equal",
                                   custWeights = NULL,
                                   sampleSites = 1,
)



## Warning in gdm::formatsitepair(bioData = PAM_Wide_coords, bioFormat = 1
, : When using rasters for prediction data, sites are assigned to the
##               cells in which they are located and then aggreagted as ne
cessary (e.g.,
##               if more than one site falls in the same raster cell - com
mon for rasters
##               with large cells).
```

```
# The gdmData object named 'gdmData_r' cannot have NA values.
# Before computing the GDM, double check if there are NA values:
sum(is.na(gdmData_r_geo)) #0

## [1] 27

gdmData_r_geo <- na.omit(gdmData_r_geo)


# Compute the GDM using predictors as a raster stack (assemblages falling
within the same pixel are pooled):
gdm_Model_r_geo<-gdm::gdm(data = gdmData_r_geo, geo=TRUE)
gdm_Model_r_geo$explained

## [1] 14.70137

#[1] 14.70137


# It is possible to reproject expected changes in biological dissimilarity
between different time periods.
# Load data on environmental variables for the future scenario, crop, and
mask:
GlobalProjection<-"climate_future/wc2.1_5arcmin_bioc_CNRM-CM6-1_ssp585_.ti
f"


cl<-parallel::makePSOCKcluster(detectCores()-1, type="SOCK")
doParallel::registerDoParallel(cl)
getDoParWorkers()

## [1] 3
```

```
bioclim_future<-foreach(i = 1:19,# change maximum 'i' if necessary
                        .combine = 'c', # bind output of each iteration as
lists
                        .packages = c("raster", "sp", "rgdal")) %dopar% {

                          # Load one raster band from the global projectio
n i:
                          SingleBand<-raster::raster(GlobalProjection, ban
d=i)

                          # Crop the raster bands:
                          SingleBand<-raster::crop(SingleBand, mata_atlant
ica)

                          }
parallel::stopCluster(cl)
plot(mata_atlantica)

# Order layers of the raster stack in the same order as in 'bioclim_curren
t' object:
bioclim_order<-as.integer(gsub("wc2.1_5m_bio_", "", names(bioclim)[1:19]))
bioclim_names <- sprintf("bio_%g", bioclim_order)
bioclim_future2 <- bioclim_future[bioclim_order]
names(bioclim_future2) <- bioclim_names
names(bioclim_future2)

##  [1] "bio_1"  "bio_10" "bio_11" "bio_12" "bio_13" "bio_14" "bio_15" "bi
o_16"
```

```
##  [9] "bio_17" "bio_18" "bio_19" "bio_2"  "bio_3"  "bio_4"  "bio_5"  "bi
o_6"
## [17] "bio_7"  "bio_8"  "bio_9"

names(bioclim)

##  [1] "wc2.1_5m_bio_1"  "wc2.1_5m_bio_10" "wc2.1_5m_bio_11" "wc2.1_5m_bi
o_12"
##  [5] "wc2.1_5m_bio_13" "wc2.1_5m_bio_14" "wc2.1_5m_bio_15" "wc2.1_5m_bi
o_16"
##  [9] "wc2.1_5m_bio_17" "wc2.1_5m_bio_18" "wc2.1_5m_bio_19" "wc2.1_5m_bi
o_2"
## [13] "wc2.1_5m_bio_3"  "wc2.1_5m_bio_4"  "wc2.1_5m_bio_5"  "wc2.1_5m_bi
o_6"
## [17] "wc2.1_5m_bio_7"  "wc2.1_5m_bio_8"  "wc2.1_5m_bio_9"  "wc2.1_5m_el
ev"

# Confirm the new order of layers in the raster stack:
par(mfrow = c(1, 2))
```

```
plot(bioclim_future[[12]]) # annual precipitation - bio12

plot(bioclim_future2[[4]]) # annual precipitation - bio12
```



```r
par(mfrow = c(1, 1))
#é a mesma camada. ok




names(bioclim)
```

```
##  [1] "wc2.1_5m_bio_1"  "wc2.1_5m_bio_10" "wc2.1_5m_bio_11" "wc2.1_5m_bi
o_12"
##  [5] "wc2.1_5m_bio_13" "wc2.1_5m_bio_14" "wc2.1_5m_bio_15" "wc2.1_5m_bi
o_16"
##  [9] "wc2.1_5m_bio_17" "wc2.1_5m_bio_18" "wc2.1_5m_bio_19" "wc2.1_5m_bi
o_2"
## [13] "wc2.1_5m_bio_3"  "wc2.1_5m_bio_4"  "wc2.1_5m_bio_5"  "wc2.1_5m_bi
o_6"
## [17] "wc2.1_5m_bio_7"  "wc2.1_5m_bio_8"  "wc2.1_5m_bio_9"  "wc2.1_5m_el
ev"
```

```r
# Replace the 'bioclim_future' object and proceed with the mask procedure:
bioclim_future<-bioclim_future2
bioclim_future<-do.call(raster::stack, unname(bioclim_future))
names(bioclim_future)
```

```
##  [1] "bio01.1"  "bio01.2"  "bio01.3"  "bio01.4"  "bio01.5"  "bio01.6"
##  [7] "bio01.7"  "bio01.8"  "bio01.9"  "bio01.10" "bio01.11" "bio01.12"
## [13] "bio01.13" "bio01.14" "bio01.15" "bio01.16" "bio01.17" "bio01.18"
## [19] "bio01.19"
```

```r
bioclim_future <- projectRaster(bioclim_future, bioclim_current_filtered)
topo_layer<-raster::raster("climate/wc2.1_5m_elev.tif")
```

```
topo_layer <- projectRaster(topo_layer, bioclim_current_filtered)
topo_layer<-raster::crop(topo_layer, bioclim_current_filtered)
bioclim_future<-raster::stack(bioclim_future, topo_layer)
plot(bioclim_future[[1]])
bioclim_names2<-gsub("", "", names(bioclim))
names(bioclim_future)<-bioclim_names2 # rename layers of 'bioclim_future' object
names(bioclim)
```

```
##  [1] "wc2.1_5m_bio_1"  "wc2.1_5m_bio_10" "wc2.1_5m_bio_11" "wc2.1_5m_bio_12"
##  [5] "wc2.1_5m_bio_13" "wc2.1_5m_bio_14" "wc2.1_5m_bio_15" "wc2.1_5m_bio_16"
##  [9] "wc2.1_5m_bio_17" "wc2.1_5m_bio_18" "wc2.1_5m_bio_19" "wc2.1_5m_bio_2"
## [13] "wc2.1_5m_bio_3"  "wc2.1_5m_bio_4"  "wc2.1_5m_bio_5"  "wc2.1_5m_bio_6"
## [17] "wc2.1_5m_bio_7"  "wc2.1_5m_bio_8"  "wc2.1_5m_bio_9"  "wc2.1_5m_elev"
```

```
names(bioclim_future)
```

```
##  [1] "wc2.1_5m_bio_1"  "wc2.1_5m_bio_10" "wc2.1_5m_bio_11" "wc2.1_5m_bio_12"
##  [5] "wc2.1_5m_bio_13" "wc2.1_5m_bio_14" "wc2.1_5m_bio_15" "wc2.1_5m_bio_16"
```

```
##  [9] "wc2.1_5m_bio_17" "wc2.1_5m_bio_18" "wc2.1_5m_bio_19" "wc2.1_5m_bio_2"
```

```
## [13] "wc2.1_5m_bio_3"  "wc2.1_5m_bio_4"  "wc2.1_5m_bio_5"  "wc2.1_5m_bi
o_6"
## [17] "wc2.1_5m_bio_7"  "wc2.1_5m_bio_8"  "wc2.1_5m_bio_9"  "wc2.1_5m_el
ev"

# Load data on soil variables:
path_current_predictors_soil<-"soil/" # Specify the path to bioclim layers
for LGM CCSM4
soil<-raster::stack(list.files(path=path_current_predictors_soil, pattern=
'.tif', full.names=T))
soil$bdod_0.5cm_mean_5000

## class      : RasterLayer
## dimensions : 2902, 7962, 23105724  (nrow, ncol, ncell)
## resolution : 5000, 5000  (x, y)
## extent     : -19949750, 19860250, -6149000, 8361000  (xmin, xmax, ymin,
ymax)
## crs        : +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no
_defs
## source     : bdod_0-5cm_mean_5000.tif
## names      : bdod_0.5cm_mean_5000

soil_projected <- projectRaster(soil, bioclim_future) #projetando pra fica
r com configurações igual dos worldclim


bioclim_soil_future<-raster::stack(bioclim_future, soil_projected)
bioclim_soil_future<-raster::crop(bioclim_soil_future, bioclim_current)
bioclim_soil_future2<-raster::stack(bioclim_soil_future) # convert from ra
ster brick to stack
```

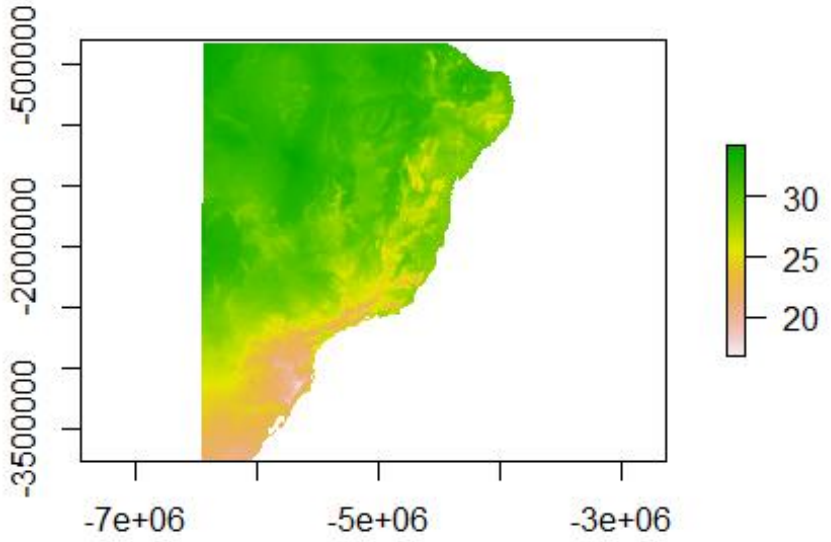
```
bioclim_soil_order<-as.integer(gsub("wc2.1_5m_bio_", "", names(bioclim_soi
l)[1:19]))
bioclim_soil_names2<-gsub("", "", names(bioclim_soil))
names(bioclim_soil_future2)<-bioclim_soil_names2 # rename layers of 'biocl
im_future' object
names(bioclim_soil)

##  [1] "wc2.1_5m_bio_1"           "wc2.1_5m_bio_10"
##  [3] "wc2.1_5m_bio_11"          "wc2.1_5m_bio_12"
##  [5] "wc2.1_5m_bio_13"          "wc2.1_5m_bio_14"
##  [7] "wc2.1_5m_bio_15"          "wc2.1_5m_bio_16"
##  [9] "wc2.1_5m_bio_17"          "wc2.1_5m_bio_18"
## [11] "wc2.1_5m_bio_19"          "wc2.1_5m_bio_2"
## [13] "wc2.1_5m_bio_3"           "wc2.1_5m_bio_4"
## [15] "wc2.1_5m_bio_5"           "wc2.1_5m_bio_6"
## [17] "wc2.1_5m_bio_7"           "wc2.1_5m_bio_8"
## [19] "wc2.1_5m_bio_9"           "wc2.1_5m_elev"
## [21] "bdod_0.5cm_mean_5000"     "cec_0.5cm_mean_5000"
## [23] "cfvo_0.5cm_mean_5000"     "clay_0.5cm_mean_5000"
## [25] "nitrogen_0.5cm_mean_5000" "ocd_0.5cm_mean_5000"
## [27] "ocs_0.30cm_mean_5000"     "phh2o_0.5cm_mean_5000"


## [29] "sand_0.5cm_mean_5000"     "silt_0.5cm_mean_5000"
## [31] "soc_0.5cm_mean_5000"

names(bioclim_soil_future2)

##  [1] "wc2.1_5m_bio_1"           "wc2.1_5m_bio_10"
##  [3] "wc2.1_5m_bio_11"          "wc2.1_5m_bio_12"
```

```
##  [5] "wc2.1_5m_bio_13"          "wc2.1_5m_bio_14"
##  [7] "wc2.1_5m_bio_15"          "wc2.1_5m_bio_16"
##  [9] "wc2.1_5m_bio_17"          "wc2.1_5m_bio_18"
## [11] "wc2.1_5m_bio_19"          "wc2.1_5m_bio_2"
## [13] "wc2.1_5m_bio_3"           "wc2.1_5m_bio_4"
## [15] "wc2.1_5m_bio_5"           "wc2.1_5m_bio_6"
## [17] "wc2.1_5m_bio_7"           "wc2.1_5m_bio_8"
## [19] "wc2.1_5m_bio_9"           "wc2.1_5m_elev"
## [21] "bdod_0.5cm_mean_5000"     "cec_0.5cm_mean_5000"
## [23] "cfvo_0.5cm_mean_5000"     "clay_0.5cm_mean_5000"
## [25] "nitrogen_0.5cm_mean_5000" "ocd_0.5cm_mean_5000"
## [27] "ocs_0.30cm_mean_5000"     "phh2o_0.5cm_mean_5000"
## [29] "sand_0.5cm_mean_5000"     "silt_0.5cm_mean_5000"
## [31] "soc_0.5cm_mean_5000"
```

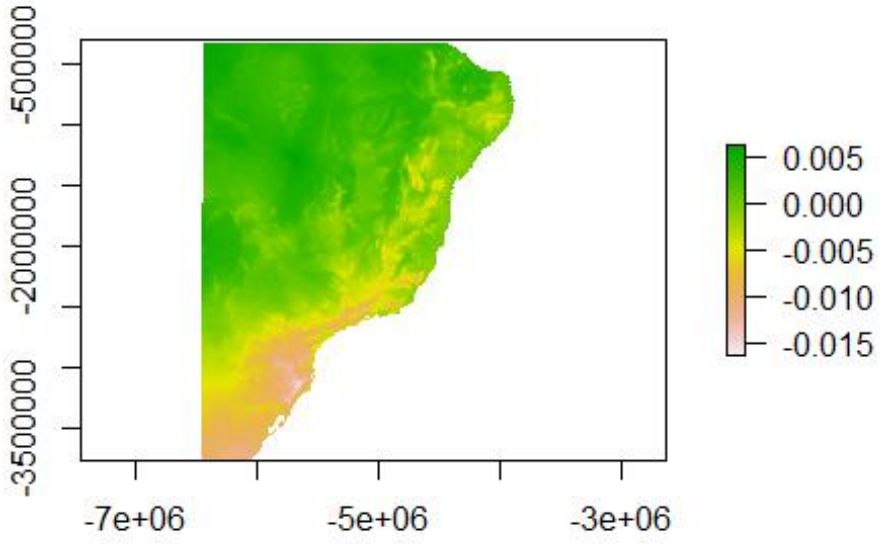
```
####só que aqui são os dados originais, não são polinomios!
#vou usar o mesmo código que usei pra gerar os polinômios com os dados do
presente



# Create two empty RasterStacks to store the transformed layers
bioclim_future_poly_linear <- raster::stack()
bioclim_future_poly_quad <- raster::stack()

# Loop over each layer in the original RasterStack
for (i in 1:nlayers(bioclim_soil_future2)) {
  # Extract the current layer
```

```
  layer <- raster::subset(bioclim_soil_future2, i)

  # Convert the RasterLayer to a vector
  layer_vector <- raster::getValues(layer)

  # Identify the non-NA values
  non_na_values <- !is.na(layer_vector)

  # Calculate the orthogonal polynomial transformation only on non-NA valu
es
  poly_vector <- matrix(NA, nrow=length(layer_vector), ncol=2)
  poly_vector[non_na_values, ] <- stats::poly(layer_vector[non_na_values],
degree=2, raw=FALSE)

  # Convert the vectors back into RasterLayers
  linear_layer <- raster::setValues(layer, poly_vector[,1])
  quad_layer <- raster::setValues(layer, poly_vector[,2])

  # Add the transformed layers to the new RasterStacks
  bioclim_future_poly_linear <- raster::addLayer(bioclim_future_poly_linea
r, linear_layer)
  bioclim_future_poly_quad <- raster::addLayer(bioclim_future_poly_quad, q
uad_layer)

  # Give specific names to the transformed layers
  names(bioclim_future_poly_linear)[i] <- paste0(names(bioclim_soil_future
2)[i], "_linear")
```

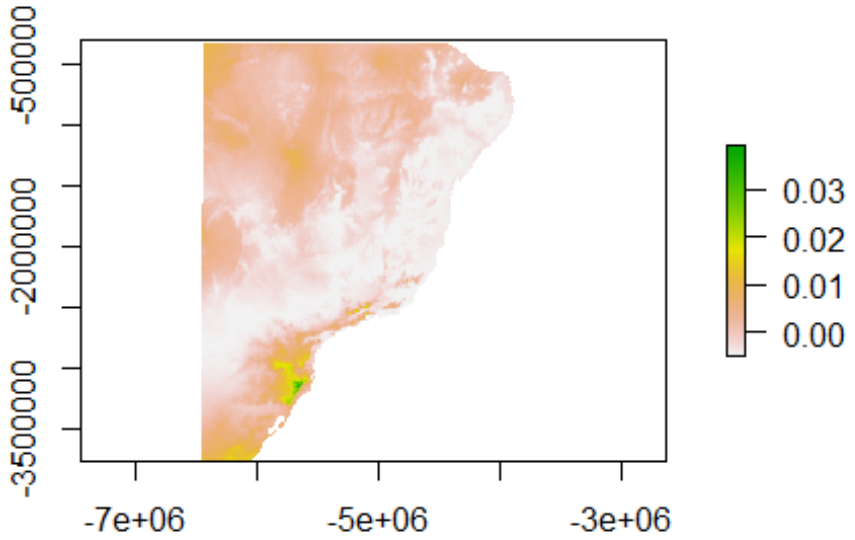
```
  names(bioclim_future_poly_quad)[i] <- paste0(names(bioclim_soil_future2)
[i], "_quad")
}

# Combine the two RasterStacks into one
bioclim_future_poly <- raster::stack(bioclim_future_poly_linear, bioclim_f
uture_poly_quad)

##str(bioclim_future_poly)

# Extrair a primeira camada do RasterStack
first_layer_linear <- raster::subset(bioclim_future_poly, 1)
# Converter a RasterLayer em um vetor
first_layer_linear_vector <- raster::getValues(first_layer_linear)
# Filtrar os valores NA
first_layer_linear_vector_non_na <- first_layer_linear_vector[!is.na(first
_layer_linear_vector)]
# Imprimir os primeiros 10 valores não-NA

print(first_layer_linear_vector_non_na[1:10])

##  [1] 0.006112310 0.006171936 0.006178458 0.006189985 0.006003824 0.0059
23643
##  [7] 0.005884882 0.005958239 0.006026196 0.005972479

#plotando a mesma variável nas versões: original, polinomio 1 ordem e 2 or
dem
plot(bioclim_future[[1]])
```




```
plot(bioclim_future_poly[[1]])
```




```
plot(bioclim_future_poly[[32]])
```




```
#agora vou filtrar o bioclim_future_poly


# Get the names of the variables in the EnvData_Filtered object
selected_var_names <- colnames(EnvData_Filtered)


# Filter the layers of the bioclim_soil_poly raster stack
bioclim_future_poly_filtered <- bioclim_future_poly[[selected_var_names]]


# Print out the names of the filtered raster stack to verify
print(names(bioclim_future_poly_filtered))

##  [1] "wc2.1_5m_bio_2_quad"        "wc2.1_5m_bio_14_quad"
##  [3] "wc2.1_5m_bio_3_quad"        "wc2.1_5m_elev_quad"
##  [5] "bdod_0.5cm_mean_5000_quad"  "phh2o_0.5cm_mean_5000_quad"
##  [7] "wc2.1_5m_bio_18_quad"       "ocs_0.30cm_mean_5000_quad"
```

```
##  [9] "silt_0.5cm_mean_5000_quad"  "clay_0.5cm_mean_5000_quad"
## [11] "cfvo_0.5cm_mean_5000_quad"

#cortar os objetos pra ficar só mata_atlantica.


crs(mata_atlantica_projected)



## Coordinate Reference System:
## Deprecated Proj.4 representation:
##  +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs
## WKT2 2019 representation:
## PROJCRS["unknown",
##     BASEGEOGCRS["unknown",
##         DATUM["World Geodetic System 1984",
##             ELLIPSOID["WGS 84",6378137,298.257223563,
##                 LENGTHUNIT["metre",1]],
##             ID["EPSG",6326]],
##         PRIMEM["Greenwich",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8901]]],
##     CONVERSION["unknown",
##         METHOD["Interrupted Goode Homolosine"],
##         PARAMETER["Longitude of natural origin",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8802]],
##         PARAMETER["False easting",0,
##             LENGTHUNIT["metre",1],
```

```
##             ID["EPSG",8806]],
##         PARAMETER["False northing",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8807]]],
##     CS[Cartesian,2],
##         AXIS["(E)",east,
##             ORDER[1],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]],
##         AXIS["(N)",north,
##             ORDER[2],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]]]
```

```
crs(bioclim_future_poly_filtered)
```

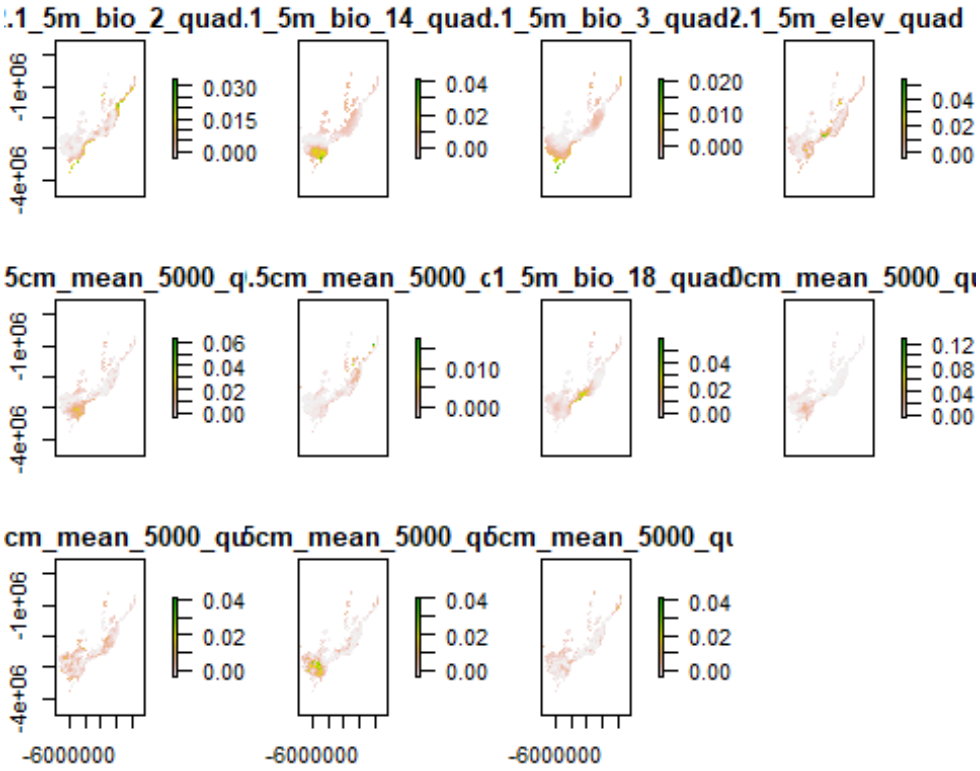
```
## Coordinate Reference System:
## Deprecated Proj.4 representation:
##  +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs
## WKT2 2019 representation:
## PROJCRS["unknown",
##     BASEGEOGCRS["unknown",
##         DATUM["World Geodetic System 1984",
##             ELLIPSOID["WGS 84",6378137,298.257223563,
##                 LENGTHUNIT["metre",1]],
##             ID["EPSG",6326]],
##         PRIMEM["Greenwich",0,
##             ANGLEUNIT["degree",0.0174532925199433],
```

```
##             ID["EPSG",8901]]],
##     CONVERSION["unknown",
##         METHOD["Interrupted Goode Homolosine"],
##         PARAMETER["Longitude of natural origin",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8802]],
##         PARAMETER["False easting",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8806]],
##         PARAMETER["False northing",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8807]]],
##     CS[Cartesian,2],
##         AXIS["(E)",east,
##             ORDER[1],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]],
##         AXIS["(N)",north,
##             ORDER[2],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]]]
```

```
extent(bioclim_current_filtered)
```

```
## class      : Extent
## xmin       : -6459750
## xmax       : -3609750
```

```
## ymin       : -3759000

## ymax       : -309000

extent(bioclim_future_poly_filtered)

## class      : Extent

## xmin       : -6459750

## xmax       : -3609750

## ymin       : -3759000

## ymax       : -309000

bioclim_future_poly_filtered_mask <- raster::mask(bioclim_future_poly_filt

ered, mata_atlantica_projected)

bioclim_future_poly_filtered <- raster::stack (bioclim_future_poly_filtere

d_mask)

plot(bioclim_future_poly_filtered) #12 layers
```



```r
print(res(bioclim_future_poly_filtered))
```

```
## [1] 10000 10000
```

```r
print(res(bioclim_current_filtered))
```

```
## [1] 10000 10000
```

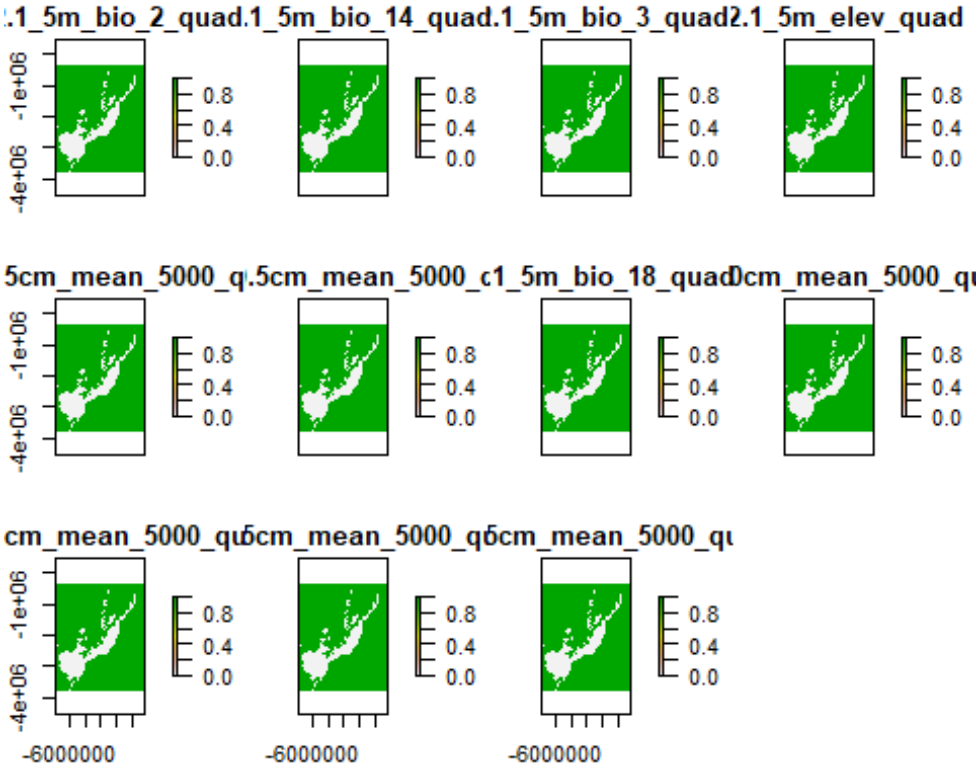
```r
#calculando a qntde de NA nos objetos
na_count_current <- calc(bioclim_current_filtered, fun = function(x) sum(i
s.na(x)))
na_count_future <- calc(bioclim_future_poly_filtered, fun = function(x) su
m(is.na(x)))
sum_na_current <- cellStats(na_count_current, stat = "sum")


sum_na_future <- cellStats(na_count_future, stat = "sum")


print(sum_na_current) #[1] 920558
```

```
## [1] 920570
```

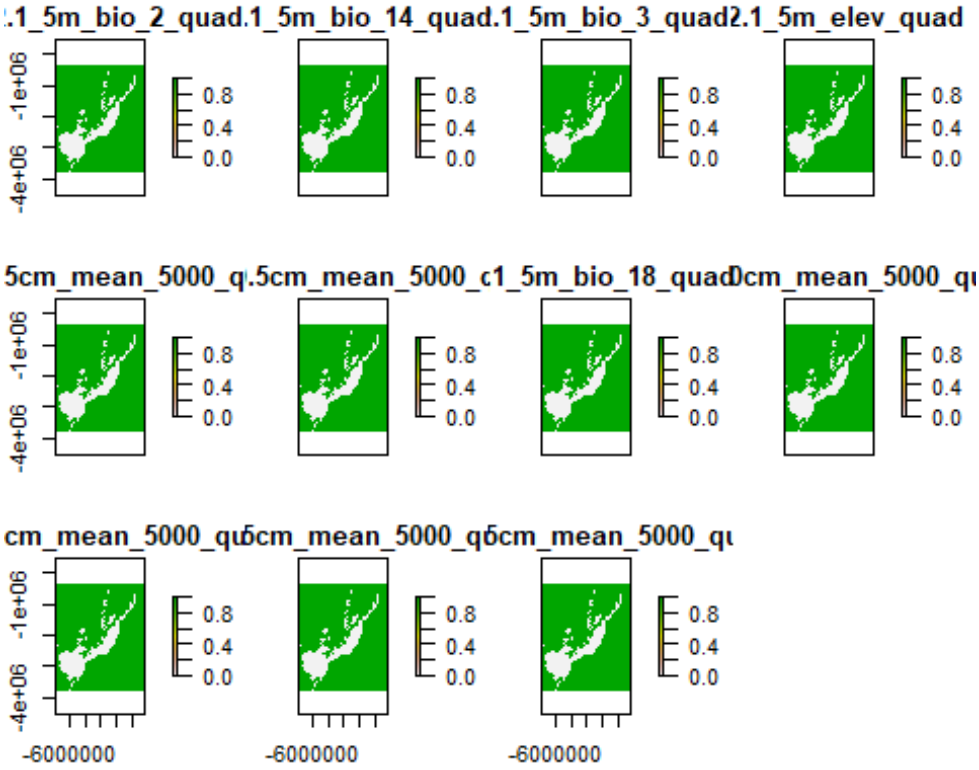
```r
print(sum_na_future) #[1] 921067 a quantidade de pixels com NA é diferente
```

```
## [1] 921076
```

```r
# Criar uma nova camada que é TRUE onde bioclim_soil_poly_filtered_crop é
NA e FALSE onde não é
na_layer_crop <- is.na(bioclim_current_filtered)
plot(na_layer_crop)
```




```
na_layer_crop <- is.na(bioclim_future_poly_filtered)



plot(na_layer_crop)
```




```
# Para o objeto 'bioclim_soil_poly_filtered_crop'
for(i in 1:nlayers(bioclim_current_filtered)){
  print(paste("Layer", i, "of bioclim_current_filtered has", ncell(bioclim
_current_filtered[[i]]), "pixels"))
}
```

```
## [1] "Layer 1 of bioclim_current_filtered has 98325 pixels"
## [1] "Layer 2 of bioclim_current_filtered has 98325 pixels"
## [1] "Layer 3 of bioclim_current_filtered has 98325 pixels"
## [1] "Layer 4 of bioclim_current_filtered has 98325 pixels"
## [1] "Layer 5 of bioclim_current_filtered has 98325 pixels"
## [1] "Layer 6 of bioclim_current_filtered has 98325 pixels"
## [1] "Layer 7 of bioclim_current_filtered has 98325 pixels"
## [1] "Layer 8 of bioclim_current_filtered has 98325 pixels"
## [1] "Layer 9 of bioclim_current_filtered has 98325 pixels"
```

```
## [1] "Layer 10 of bioclim_current_filtered has 98325 pixels"
## [1] "Layer 11 of bioclim_current_filtered has 98325 pixels"
```

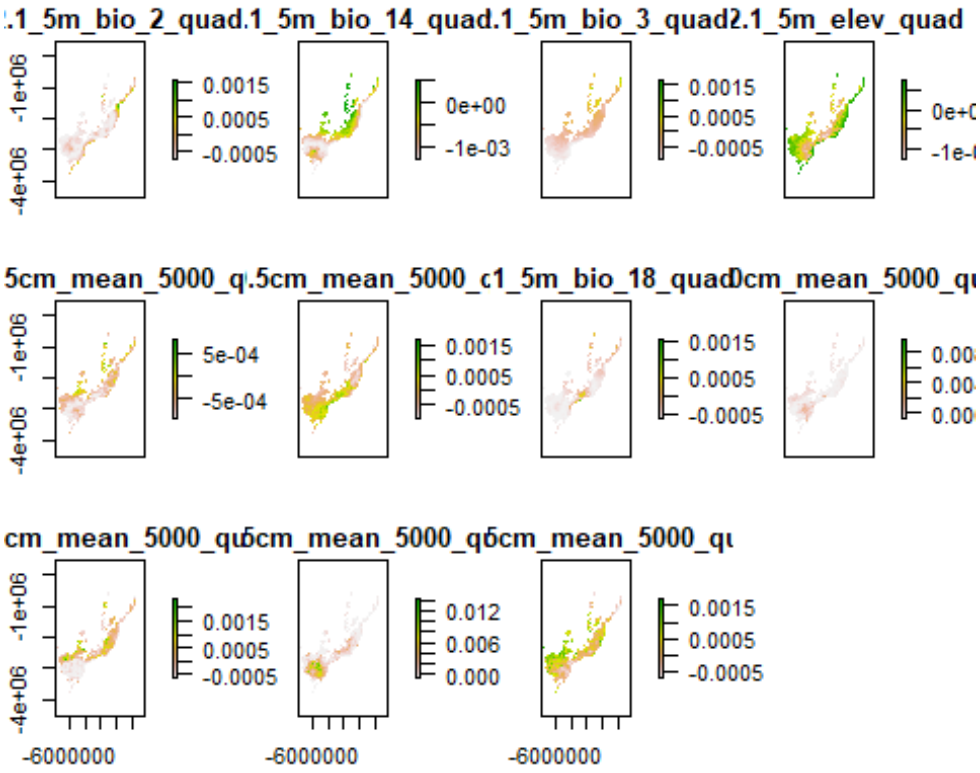
```
# Para o objeto 'bioclim_future_poly_filtered'
for(i in 1:nlayers(bioclim_future_poly_filtered)){
  print(paste("Layer", i, "of bioclim_future_poly_filtered has", ncell(bio
clim_future_poly_filtered[[i]]), "pixels"))
}
```

```
## [1] "Layer 1 of bioclim_future_poly_filtered has 98325 pixels"
## [1] "Layer 2 of bioclim_future_poly_filtered has 98325 pixels"
## [1] "Layer 3 of bioclim_future_poly_filtered has 98325 pixels"
## [1] "Layer 4 of bioclim_future_poly_filtered has 98325 pixels"
## [1] "Layer 5 of bioclim_future_poly_filtered has 98325 pixels"
## [1] "Layer 6 of bioclim_future_poly_filtered has 98325 pixels"
## [1] "Layer 7 of bioclim_future_poly_filtered has 98325 pixels"
## [1] "Layer 8 of bioclim_future_poly_filtered has 98325 pixels"
## [1] "Layer 9 of bioclim_future_poly_filtered has 98325 pixels"
## [1] "Layer 10 of bioclim_future_poly_filtered has 98325 pixels"
## [1] "Layer 11 of bioclim_future_poly_filtered has 98325 pixels"
```

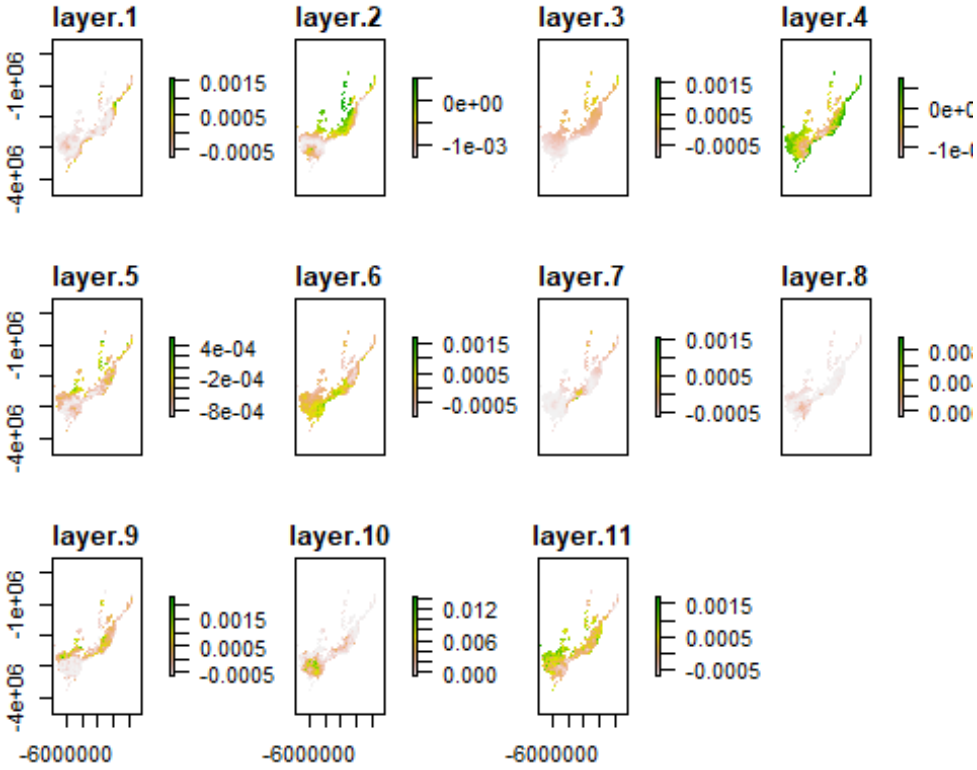
```
mask_na <- function(x, y) {
  na_mask <- is.na(x) | is.na(y)
  x[na_mask] <- NA
  return(x)
}
```

```
bioclim_current_filtered_clean <- overlay(bioclim_current_filtered, biocli
m_future_poly_filtered, fun = mask_na)
```

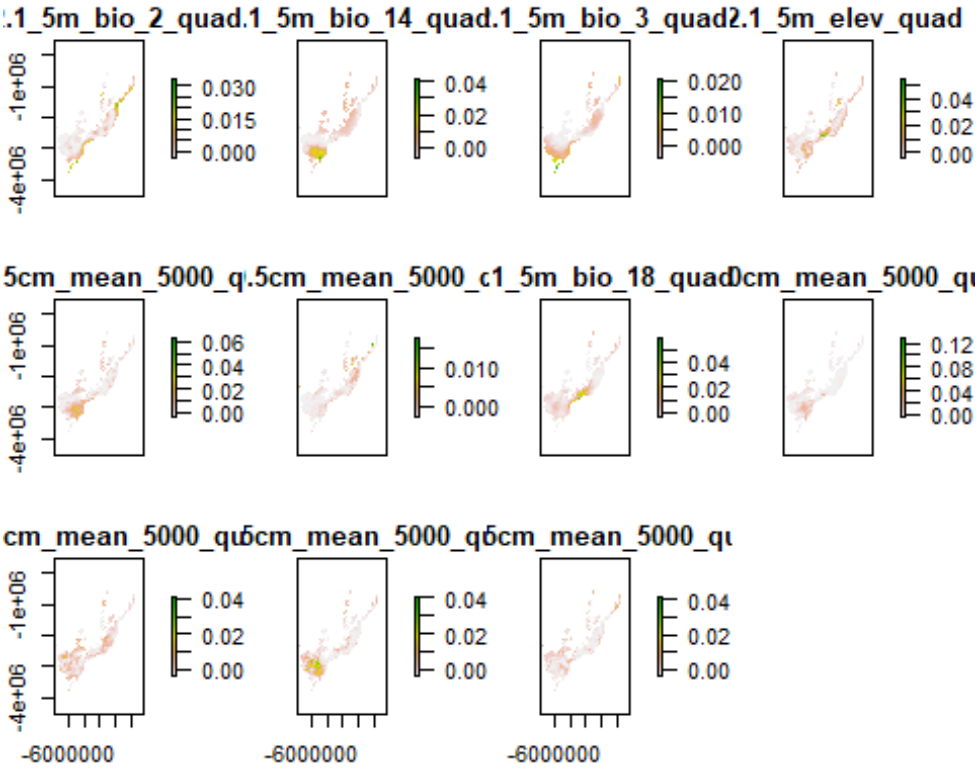
```
bioclim_future_poly_filtered_clean <- overlay(bioclim_future_poly_filtered
, bioclim_current_filtered, fun = mask_na)
na_count_current <- sum(is.na(values(bioclim_current_filtered_clean)))
na_count_future <- sum(is.na(values(bioclim_future_poly_filtered_clean)))
print(na_count_current)

## [1] 921076

print(na_count_future)

## [1] 921076

plot(bioclim_current_filtered)
```



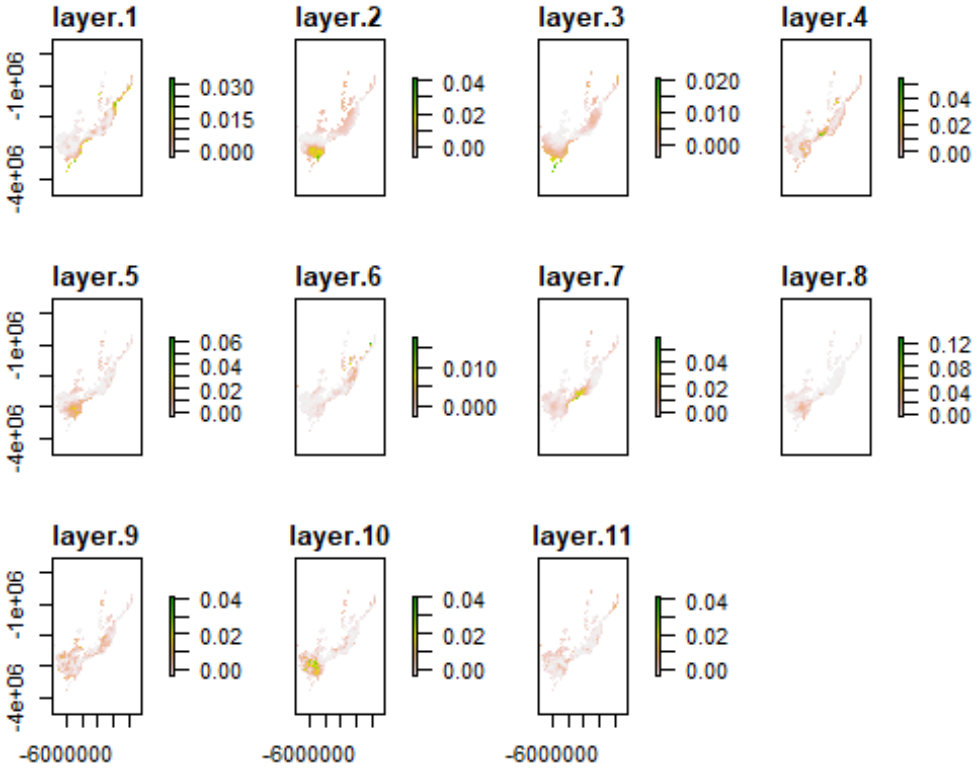

```
plot(bioclim_current_filtered_clean)
```




```
plot(bioclim_future_poly_filtered)
```




```
plot(bioclim_future_poly_filtered_clean)
```




```
crs(bioclim_current_filtered)

## Coordinate Reference System:
## Deprecated Proj.4 representation:
##  +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs
## WKT2 2019 representation:
## PROJCRS["unknown",
##     BASEGEOGCRS["unknown",
##         DATUM["World Geodetic System 1984",
##             ELLIPSOID["WGS 84",6378137,298.257223563,
##                 LENGTHUNIT["metre",1]],
##             ID["EPSG",6326]],
##         PRIMEM["Greenwich",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8901]]],
##     CONVERSION["unknown",
```

```
##         METHOD["Interrupted Goode Homolosine"],
##         PARAMETER["Longitude of natural origin",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8802]],
##         PARAMETER["False easting",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8806]],
##         PARAMETER["False northing",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8807]]],
##     CS[Cartesian,2],
##         AXIS["(E)",east,
##             ORDER[1],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]],
##         AXIS["(N)",north,
##             ORDER[2],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]]]
```

```
crs(bioclim_current_filtered_clean)
```

```
## Coordinate Reference System:
## Deprecated Proj.4 representation:
##  +proj=igh +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs
## WKT2 2019 representation:
## PROJCRS["unknown",
##     BASEGEOGCRS["unknown",
```

```
##         DATUM["World Geodetic System 1984",
##             ELLIPSOID["WGS 84",6378137,298.257223563,
##                 LENGTHUNIT["metre",1]],
##             ID["EPSG",6326]],
##         PRIMEM["Greenwich",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8901]]],
##     CONVERSION["unknown",
##         METHOD["Interrupted Goode Homolosine"],
##         PARAMETER["Longitude of natural origin",0,
##             ANGLEUNIT["degree",0.0174532925199433],
##             ID["EPSG",8802]],
##         PARAMETER["False easting",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8806]],
##         PARAMETER["False northing",0,
##             LENGTHUNIT["metre",1],
##             ID["EPSG",8807]]],
##     CS[Cartesian,2],
##         AXIS["(E)",east,
##             ORDER[1],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]],
##         AXIS["(N)",north,
##             ORDER[2],
##             LENGTHUNIT["metre",1,
##                 ID["EPSG",9001]]]]
```

```
print(res(bioclim_current_filtered))

## [1] 10000 10000

print(res(bioclim_current_filtered_clean))

## [1] 10000 10000

extent(bioclim_current_filtered)

## class      : Extent
## xmin       : -6459750
## xmax       : -3609750
## ymin       : -3759000
## ymax       : -309000

extent(bioclim_current_filtered_clean)

## class      : Extent
## xmin       : -6459750
## xmax       : -3609750
## ymin       : -3759000
## ymax       : -309000

#gerando o modelo de novo, com o novo raster


gdmData_r<-gdm::formatsitepair(bioData = PAM_Wide_coords,
                               bioFormat = 1,
                               dist="bray",
                               abundance=FALSE,
                               siteColumn= "regiao",
                               XColumn = "Longitude.x",
```

```
                                YColumn = "Latitude.y",
                                predData = bioclim_current_filtered_clean,
# raster stack
                                weightType = "equal",
                                custWeights = NULL,
                                sampleSites = 1,
)
```

```
## Warning in gdm::formatsitepair(bioData = PAM_Wide_coords, bioFormat = 1
, : When using rasters for prediction data, sites are assigned to the
##               cells in which they are located and then aggreagted as ne
cessary (e.g.,
##               if more than one site falls in the same raster cell - com
mon for rasters
##               with large cells).
```

```
# The gdmData object named 'gdmData_r' cannot have NA values.
# Before computing the GDM, double check if there are NA values:
sum(is.na(gdmData_r)) #297
```

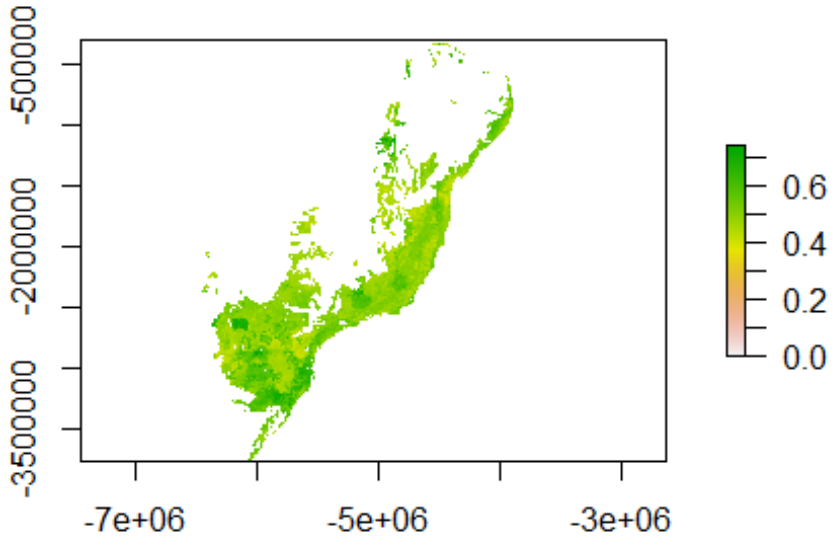
```
## [1] 297
```

```
gdmData_r <- na.omit(gdmData_r)
```

```
# Compute the GDM using predictors as a raster stack (assemblages falling
within the same pixel are pooled):
gdm_Model_r<-gdm::gdm(data = gdmData_r, geo=TRUE)
gdm_Model_r$explained
```

```
## [1] 32.05404
```

```
#[1] 32.05404


# Predict dissimilarity distance between sites based on environmental data
.
GDM_Prediction <- gdm:::predict.gdm(object = gdm_Model_r, # gdm object (th
e output of gdm::gdm function)
                                    data = bioclim_current_filtered_clean,
# raster stack
                                    time = TRUE,  # project using raster st
ack
                                    predRasts = bioclim_future_poly_filtere
d_clean # raster stack with future/past conditions
                                    )


#save.image(file = "Rdata_predict.gdm.RData")



# Visualise expected changes in biological dissimilarity between current a
nd future scenarios:
plot(GDM_Prediction)
points(SiteCoords, col="red", pch=16, cex=0.5)
points(regiaoCoords, col="orange", pch=0.5)
```




```
summary(GDM_Prediction)

##                layer.1
## Min.    -7.241961e-04
## 1st Qu.  4.694229e-01
## Median   5.012670e-01
## 3rd Qu.  5.393601e-01
## Max.     7.418983e-01
## NA's     8.367500e+04

# OBS.: gdm package does not check for differences in coordinate reference
system. Use the same crs for all spatial data.


# Export to disk:
#raster::writeRaster(GDM_Prediction, filename="GDM_Output_TemporalBeta.tif
", format="GTiff", overwrite=TRUE)
```

```
#raster::writeRaster(raster::brick(bioclim_soil_poly_filtered_crop_clean), filename="bioclim_soil_poly_filtered_crop_clean.tif", format="GTiff", overwrite=TRUE)
#raster::writeRaster(raster::brick(bioclim_soil_future_poly_filtered_clean), filename="bioclim_soil_future_poly_filtered_clean.tif", format="GTiff", overwrite=TRUE)
#raster::writeRaster(bioclim_soil_future_poly_filtered_clean, "AF_bioclim_soil_future2.tif", format = "GTiff", overwrite = TRUE)

# OBS.: we will lose layer names in writing raster files individually or as a raster brick.



# Export as RData:
save(bioclim_current_filtered_clean, file="bioclim_current_filtered_clean.RData")
save(bioclim_future_poly_filtered_clean, file="bioclim_future_poly_filtered_clean.RData")

#####



# STEP 6: MODEL VARIATION IN SPECIES ASSEMBLAGE USING GDM AND MAP SPATIAL PATTERNS OF BETA-DIVERSITY
################################################################################
##############################################
```

```
# STEP 6: MODEL VARIATION IN SPECIES ASSEMBLAGE USING GDM AND MAP SPATIAL
PATTERNS OF BETA-DIVERSITY



# Create the gdmData object to run generalised dissimilarity modelling (do
es not work with raster stack):
gdmData_Diss<-gdm::formatsitepair(bioData = PAM_Wide_coords , # table with
beta-diversity data #antes era matriz PAM
                                  bioFormat = 1, # 1=PAM_wide, 2=PAM_long,
3=Dissimilarity matrix, 4=site-pair table #formato da tabela
                                  dist="bray", #aqui estava como NULL, mas
aí ele não rodava, coloquei bray
                                  abundance=FALSE, # presence-absence data
or abundance-data
                                  siteColumn= "regiao", # name of the colu
mn holding sites ID
                                  XColumn = "Longitude.x", # name of the c
olumn holding longitude values
                                  YColumn = "Latitude.y", # same for latit
ude
                                  predData = EnvData_Filtered_regiaoID, #
table with EnvData
                                  weightType = "equal", # weighting for si
tes, "richness" or "custom"
                                  custWeights = NULL, # SiteID + customize
d weights (e.g., sampling effort)
                                  sampleSites = 1, # number between 0-1, f
```

```
raction of sites to be used (in case of memory limitations)
)



# Build the GDM using the gdmData object:
gdm_Model_Diss<-gdm::gdm(data = gdmData_Diss, geo=TRUE) #construir o model
o
str(gdm_Model_Diss) # summary list

## List of 16
##  $ dataname    : symbol gdmData_Diss
##  $ geo         : logi TRUE
##  $ sample      : int 378
##  $ gdmdeviance : num 19.9
##  $ nulldeviance: num 30.5
##  $ explained   : num 34.6
##  $ intercept   : num 0.474
##  $ predictors  : chr [1:12] "Geographic" "wc2.1_5m_bio_2_quad" "wc2.1_5
m_bio_14_quad" "wc2.1_5m_bio_3_quad" ...
##  $ coefficients: num [1:36] 0 0 0.557 0 0 ...
##  $ knots       : num [1:36] 8.90e+03 4.96e+05 2.29e+06 -7.21e-04 -5.82e
-04 ...
##  $ sumCoeff    : NULL
##  $ splines     : num [1:12] 3 3 3 3 3 3 3 3 3 3 ...
##  $ creationdate: chr "Sun Oct  6 22:29:42 2024"
##  $ observed    : num [1:378] 0.6 0.667 0.684 0.625 0.769 ...
##  $ predicted   : num [1:378] 0.659 0.658 0.685 0.615 0.597 ...
```

```
##  $ ecological  : num [1:378] 1.076 1.074 1.155 0.955 0.91 ...
##  - attr(*, "class")= chr [1:2] "gdm" "list"

gdm_Model_Diss$explained # % explained

## [1] 34.63167

#[1] 34.63167


# Transforming from environmental space (raw predictors, 'EnvData') to bio
logical space.
# Since the column 'SiteID' is not a predictor, it is necessary to remove
it from the EnvData:
names(gdmData_Diss) # the order of variables matters #compara??o par a par
de todos os locais

##  [1] "distance"                      "weights"
##  [3] "s1.xCoord"                     "s1.yCoord"
##  [5] "s2.xCoord"                     "s2.yCoord"
##  [7] "s1.wc2.1_5m_bio_2_quad"        "s1.wc2.1_5m_bio_14_quad"
##  [9] "s1.wc2.1_5m_bio_3_quad"        "s1.wc2.1_5m_elev_quad"
## [11] "s1.bdod_0.5cm_mean_5000_quad"  "s1.phh2o_0.5cm_mean_5000_quad"
## [13] "s1.wc2.1_5m_bio_18_quad"       "s1.ocs_0.30cm_mean_5000_quad"
## [15] "s1.silt_0.5cm_mean_5000_quad"  "s1.clay_0.5cm_mean_5000_quad"
## [17] "s1.cfvo_0.5cm_mean_5000_quad"  "s2.wc2.1_5m_bio_2_quad"
## [19] "s2.wc2.1_5m_bio_14_quad"       "s2.wc2.1_5m_bio_3_quad"
## [21] "s2.wc2.1_5m_elev_quad"         "s2.bdod_0.5cm_mean_5000_quad"
## [23] "s2.phh2o_0.5cm_mean_5000_quad" "s2.wc2.1_5m_bio_18_quad"
## [25] "s2.ocs_0.30cm_mean_5000_quad"  "s2.silt_0.5cm_mean_5000_quad"
## [27] "s2.clay_0.5cm_mean_5000_quad"  "s2.cfvo_0.5cm_mean_5000_quad"
```

```
names(EnvData_Filtered_regiaoID) # need to have the same variable order as
'gdmData_Diss'

##  [1] "regiao"                    "Longitude.x"
##  [3] "Latitude.y"                "wc2.1_5m_bio_2_quad"
##  [5] "wc2.1_5m_bio_14_quad"      "wc2.1_5m_bio_3_quad"
##  [7] "wc2.1_5m_elev_quad"        "bdod_0.5cm_mean_5000_quad"
##  [9] "phh2o_0.5cm_mean_5000_quad" "wc2.1_5m_bio_18_quad"
## [11] "ocs_0.30cm_mean_5000_quad"  "silt_0.5cm_mean_5000_quad"
## [13] "clay_0.5cm_mean_5000_quad"  "cfvo_0.5cm_mean_5000_quad"

EnvData_Adjusted <- subset(EnvData_Filtered_regiaoID, select = -c(regiao))


# Proceed with the transformation (the output is of same class as the obje
ct informed in the data argument):
EnvData_Transformed <- gdm::gdm.transform(model = gdm_Model_Diss, data = E
nvData_Adjusted)


# The transformed predictors can be used to inform the partial contributio
n of each predictor in producing the GDM.
# Partial contribution is given by the y-axis range (0-1) for each partial
response curve (arquivo pdf). todos os preditores est?o nessa mesma escala
.


#Essa maplitude que eles exibem ? utilizada pra representar a import?ncia
de cada um deles. A dist?ncia geogr?fica ? o mais importante deles, pois ?
o mais alto
```

```
summary(EnvData_Transformed)# as vari?ves que tem valores zero ? pq n?o ex
plicam nada no modelo

##   Longitude.x       Latitude.y     wc2.1_5m_bio_2_quad wc2.1_5m_bio_14_
quad
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.000e+00   Min.   :0.00000
##  1st Qu.:0.1024   1st Qu.:0.1391   1st Qu.:0.000e+00   1st Qu.:0.04648
##  Median :0.1441   Median :0.1641   Median :1.500e-07   Median :0.06966
##  Mean   :0.1663   Mean   :0.1839   Mean   :2.844e-02   Mean   :0.06607
##  3rd Qu.:0.2237   3rd Qu.:0.1963   3rd Qu.:1.991e-03   3rd Qu.:0.08153
##  Max.   :0.3937   Max.   :0.4493   Max.   :2.803e-01   Max.   :0.14352
##  wc2.1_5m_bio_3_quad wc2.1_5m_elev_quad bdod_0.5cm_mean_5000_quad
##  Min.   :0           Min.   :0.000000   Min.   :0.00000
##  1st Qu.:0           1st Qu.:0.002675   1st Qu.:0.07021
##  Median :0           Median :0.004650   Median :0.12141
##  Mean   :0           Mean   :0.005673   Mean   :0.09524
##  3rd Qu.:0           3rd Qu.:0.009990   3rd Qu.:0.12437
##  Max.   :0           Max.   :0.011866   Max.   :0.12437
##  phh2o_0.5cm_mean_5000_quad wc2.1_5m_bio_18_quad ocs_0.30cm_mean_5000_q
uad
##  Min.   :0                  Min.   :0.000e+00    Min.   :0.0000000
##  1st Qu.:0                  1st Qu.:0.000e+00    1st Qu.:0.0001144
##  Median :0                  Median :3.940e-06    Median :0.0010225


##  Mean   :0                  Mean   :7.335e-03    Mean   :0.0045753
##  3rd Qu.:0                  3rd Qu.:1.305e-04    3rd Qu.:0.0041203
##  Max.   :0                  Max.   :9.934e-02    Max.   :0.0249063
```

```
##  silt_0.5cm_mean_5000_quad clay_0.5cm_mean_5000_quad cfvo_0.5cm_mean_50
00_quad
##  Min.   :0.000000          Min.   :0.0000000         Min.   :0.000e+00
##  1st Qu.:0.008076          1st Qu.:0.0000000         1st Qu.:0.000e+00
##  Median :0.050622          Median :0.0000042         Median :5.680e-06
##  Mean   :0.052220          Mean   :0.0557173         Mean   :5.765e-03
##  3rd Qu.:0.092421          3rd Qu.:0.0826512         3rd Qu.:3.564e-03
##  Max.   :0.134881          Max.   :0.4904129         Max.   :4.946e-02

# Higher range for a given predictor indicates greater influence on the pr
edicted dissimilarity:
PredMinMax<-Rfast::colMinsMaxs(EnvData_Transformed) # extract min and max
values #n?o d? o nome das colunas
PredRange<-PredMinMax[2,] - PredMinMax[1,] # get the range
names(PredRange)<-names(EnvData_Adjusted) # rename o nome das colunas
PredRange # print

##                Longitude.x                 Latitude.y
##                 0.39367382                 0.44927419
##        wc2.1_5m_bio_2_quad       wc2.1_5m_bio_14_quad
##                 0.28029702                 0.14351822
##        wc2.1_5m_bio_3_quad         wc2.1_5m_elev_quad
##                 0.00000000                 0.01186588
##  bdod_0.5cm_mean_5000_quad phh2o_0.5cm_mean_5000_quad
##                 0.12436994                 0.00000000
##       wc2.1_5m_bio_18_quad  ocs_0.30cm_mean_5000_quad
##                 0.09934095                 0.02490628
##  silt_0.5cm_mean_5000_quad  clay_0.5cm_mean_5000_quad
```

```
##                 0.13488088                 0.49041290
##  cfvo_0.5cm_mean_5000_quad
##                 0.04945587
```

```
# The partial contribution of each predictor (PredRange) equals the sum of
their coefficients: cada preditor tem 3 medidas (pq usamos 3 splines), que
s?o os coeficientes. deu 48 coeficientes no final por isso.
```

```
gdm_Model_Diss$predictors # visualise predictor names
```

```
##  [1] "Geographic"                "wc2.1_5m_bio_2_quad"
##  [3] "wc2.1_5m_bio_14_quad"      "wc2.1_5m_bio_3_quad"
##  [5] "wc2.1_5m_elev_quad"        "bdod_0.5cm_mean_5000_quad"
##  [7] "phh2o_0.5cm_mean_5000_quad" "wc2.1_5m_bio_18_quad"
##  [9] "ocs_0.30cm_mean_5000_quad"  "silt_0.5cm_mean_5000_quad"
## [11] "clay_0.5cm_mean_5000_quad"  "cfvo_0.5cm_mean_5000_quad"
```

```
gdm_Model_Diss$coefficients[(length(gdm_Model_Diss$coefficients)-2):
                              length(gdm_Model_Diss$coefficients)] #os 3 c
oefs for last predictor in EnvData
```

```
## [1] 0.00000000 0.00000000 0.04945587
```

```
PredRange[length(PredRange)]
```

```
## cfvo_0.5cm_mean_5000_quad
##                 0.04945587
```

```
# Quantify model significance and variable importance/significance in gdm
using matrix permutation:
```

```
#se j? fez a analise de modelo antes (no inicio do script, com ana?ise de
multicolenearidade) e ja selecionou só as variaveis selecionadas, usar pre
dSelect = FALSE
#predSelect = FALSE vai verificar a importancia do modelo todo. When predS
elect = FALSE results will be returned only for a model fit with all predi
ctors
gdm_VariableImportance<-gdm::gdm.varImp(spTable = gdmData_Diss,
                                        geo = TRUE,
                                        predSelect = FALSE,
                                        nPerm = 1000,
                                        parallel = FALSE,
                                        cores = 7,
                                        sampleSites = 1,
                                        sampleSitePairs = 1,
                                        outFile = NULL)

## Fitting initial model with all 12 predictors...
## Sum of I-spline coefficients for predictor wc2.1_5m_bio_3_quad = 0
## Sum of I-spline coefficients for predictor phh2o_0.5cm_mean_5000_quad =
0
## Removing wc2.1_5m_bio_3_quad and proceeding with permutation testing...
## Removing phh2o_0.5cm_mean_5000_quad and proceeding with permutation tes
ting...
## Creating 1000 permuted site-pair tables...
## Starting model assessment...
## Percent deviance explained by the full model =  34.632
## Fitting GDMs to the permuted site-pair tables...
```

```
## Assessing importance of geographic distance...
## Assessing importance of wc2.1_5m_bio_2_quad...
## Assessing importance of wc2.1_5m_bio_14_quad...
## Assessing importance of wc2.1_5m_elev_quad...
## Assessing importance of bdod_0.5cm_mean_5000_quad...
## Assessing importance of wc2.1_5m_bio_18_quad...
## Assessing importance of ocs_0.30cm_mean_5000_quad...
## Assessing importance of silt_0.5cm_mean_5000_quad...
## Assessing importance of clay_0.5cm_mean_5000_quad...
## Assessing importance of cfvo_0.5cm_mean_5000_quad...
## Backwards elimination not selected by user (predSelect=F). Ceasing asse
ssment.
## Percent deviance explained by final model = 34.632
## Final set of predictors returned:
## Geographic
## wc2.1_5m_bio_2_quad
## wc2.1_5m_bio_14_quad
## wc2.1_5m_elev_quad
## bdod_0.5cm_mean_5000_quad
## wc2.1_5m_bio_18_quad
## ocs_0.30cm_mean_5000_quad
## silt_0.5cm_mean_5000_quad
## clay_0.5cm_mean_5000_quad
## cfvo_0.5cm_mean_5000_quad

#na primeira vez não rodou, mas botei pra rodar desde o gdmData_Diss até a
qui ao mesmo tempo e foi.
save(gdm_VariableImportance, file="gdm_VariableImportance_1000nperm.RData"
```

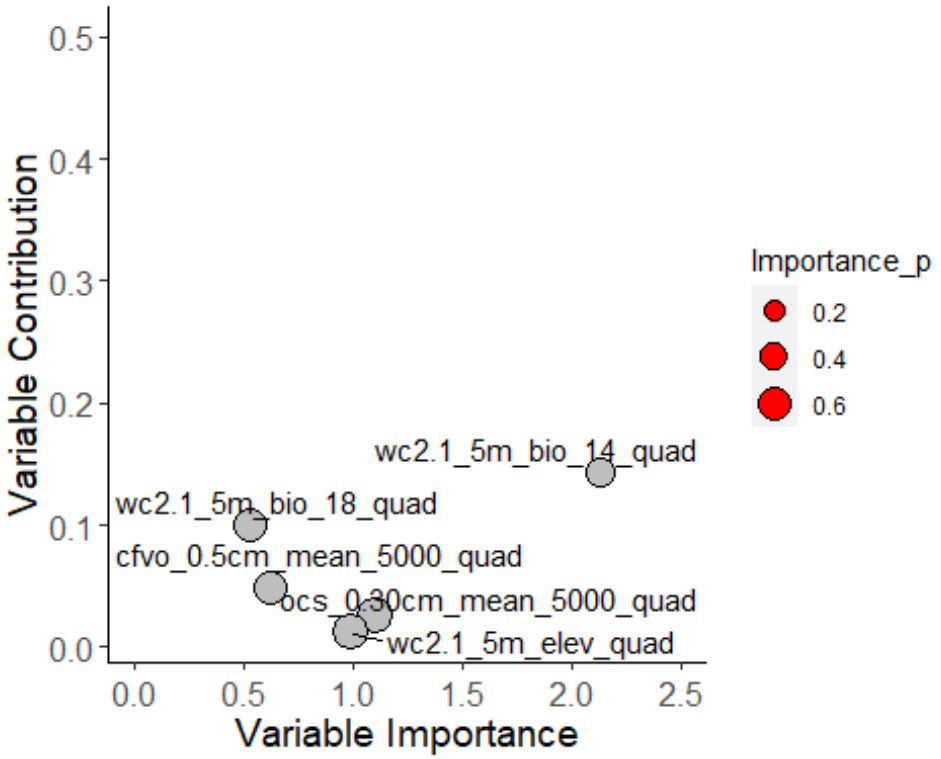
```
)
str(gdm_VariableImportance)

## List of 4
##  $ Model assessment    :'data.frame':    4 obs. of  1 variable:
##   ..$ All predictors: num [1:4] 19.937 34.632 0.001 1000
##  $ Predictor Importance:'data.frame':    10 obs. of  1 variable:
##   ..$ All predictors: num [1:10] 8.926 7.244 2.135 0.987 3.789 ...
##  $ Predictor p-values  :'data.frame':    10 obs. of  1 variable:
##   ..$ All predictors: num [1:10] 0.004 0.252 0.477 0.729 0.397 0.626 0.
716 0.396 0.087 0.601
##  $ Model Convergence   :'data.frame':    10 obs. of  1 variable:
##   ..$ All predictors: num [1:10] 1000 1000 1000 1000 1000 1000 1000 100
0 1000 1000

# Full model deviance, % deviance explained, p-value, nPerm applied:
gdm_VariableImportance[[1]] # print three first columns of the first list,
All predictors significa o modelo com 19 variaveis; -2 ? o com 18 etc

##                            All predictors
## Model deviance                     19.937
## Percent deviance explained         34.632
## Model p-value                       0.001
## Fitted permutations              1000.000

#save.image("Capitulo_modelagem.RData")


# Variable importance = percent change in deviance explained between a mod
el fit with and without that variable:
```

```
gdm_VariableImportance[[2]] # variable importance, import?ncia de cada pre
ditor, o que afeta mais a varia??o da biota de uma regi?o

##                            All predictors
## Geographic                          8.926
## wc2.1_5m_bio_2_quad                 7.244
## wc2.1_5m_bio_14_quad                2.135
## wc2.1_5m_elev_quad                  0.987
## bdod_0.5cm_mean_5000_quad           3.789
## wc2.1_5m_bio_18_quad                0.525
## ocs_0.30cm_mean_5000_quad           1.102
## silt_0.5cm_mean_5000_quad           3.298
## clay_0.5cm_mean_5000_quad          25.354
## cfvo_0.5cm_mean_5000_quad           0.618

gdm_VariableImportance[[3]] # variable significance extracted from permuta
tions used

##                            All predictors
## Geographic                          0.004
## wc2.1_5m_bio_2_quad                 0.252
## wc2.1_5m_bio_14_quad                0.477
## wc2.1_5m_elev_quad                  0.729
## bdod_0.5cm_mean_5000_quad           0.397
## wc2.1_5m_bio_18_quad                0.626
## ocs_0.30cm_mean_5000_quad           0.716
## silt_0.5cm_mean_5000_quad           0.396
## clay_0.5cm_mean_5000_quad           0.087
## cfvo_0.5cm_mean_5000_quad           0.601
```

```
gdm_VariableImportance[[4]] # number of permutations used to calculate the
statistics

##                            All predictors
## Geographic                           1000
## wc2.1_5m_bio_2_quad                  1000
## wc2.1_5m_bio_14_quad                 1000
## wc2.1_5m_elev_quad                   1000
## bdod_0.5cm_mean_5000_quad            1000
## wc2.1_5m_bio_18_quad                 1000
## ocs_0.30cm_mean_5000_quad            1000
## silt_0.5cm_mean_5000_quad            1000
## clay_0.5cm_mean_5000_quad            1000
## cfvo_0.5cm_mean_5000_quad            1000

# Compare predictor contribution and predictor importance:


PredContribImport<-merge(data.frame(Predictor=row.names(gdm_VariableImport
ance[[2]]),
                                    Importance=gdm_VariableImportance[[2]]
[,1],
                                    Importance_p=gdm_VariableImportance[[3
]][,1]),  # first list for illustrative purposes

                         data.frame(Predictor=row.names(as.data.frame(Pred
Range)), # predictor contribution
                                    Contrib=PredRange),
```

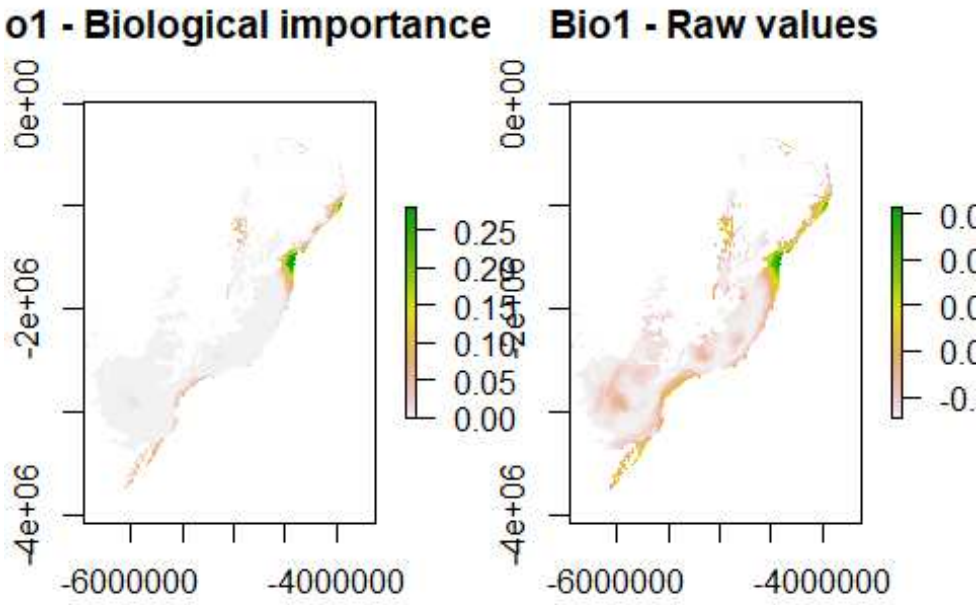
```
                          by="Predictor", all.x=TRUE)


# Visualize
ggplot(data=PredContribImport, aes(x=Importance, y=Contrib)) +
  geom_point(shape=21, aes(size=Importance_p), color="black", fill="gray")
+ # add points
  geom_point(data=PredContribImport[PredContribImport$Importance_p<=0.1,],
# filter only p-value <= 0.10
             shape=21, fill="red", color="black", aes(size=Importance_p))
+
  ggrepel::geom_text_repel(label=PredContribImport$Predictor, # add predic
tor names near the respective symbol
                           nudge_x=0.1, nudge_y=0.01) + # displace label b
y nudge values
  xlab("Variable Importance") + ylab("Variable Contribution") +
  theme(panel.grid.minor = element_blank(), # remove minor gridlines
        panel.grid.major = element_blank(), # remove major gridlines
        panel.background = element_blank(), # white background
        axis.line = element_line(colour="black"), # axis lines aesthetitcs
        axis.text.y = element_text(hjust=1, vjust=0.5, angle=0, size=12),
        axis.text.x = element_text(hjust=0.5, vjust=0.5, angle=0, size=12)
,
        axis.title = element_text(hjust=0.5, vjust=0.5, angle=0, size=14)
  ) +
  scale_x_continuous(limits=c(0,2.5))

## Warning: Removed 5 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).

## Warning: Removed 5 rows containing missing values (`geom_text_repel()`)

.
```




```
# Transform geographical and environmental predictor to their respective b
iological importance:
EnvRaster_Transformed <- gdm::gdm.transform(model = gdm_Model_Diss, data =
bioclim_current_filtered)

## Warning in gdm::gdm.transform(model = gdm_Model_Diss, data =
## bioclim_current_filtered): Extracted data from rasters contained NAs. T
hese
## were automatically removed from the data object to be transformed.
```

```
par(mfrow = c(1, 2))
plot(EnvRaster_Transformed[[3]], main="Bio1 - Biological importance")
plot(bioclim_current_filtered[[1]], main="Bio1 - Raw values")
```



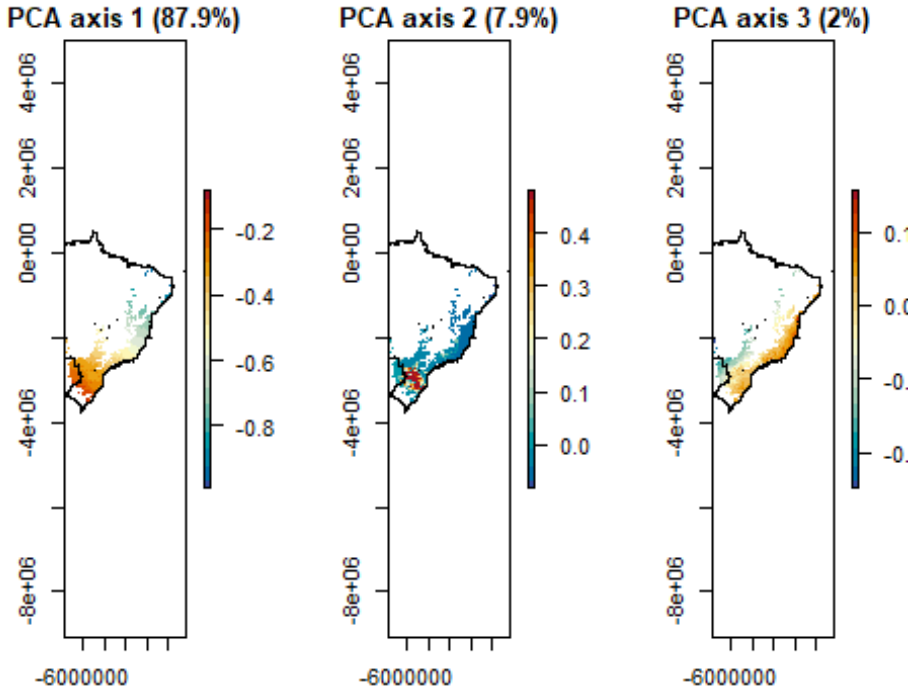

```
par(mfrow = c(1, 1))


# Convert the layers of 'EnvRaster_Transformed' into a dataframe for subse
quent analysis:


EnvRaster_NotNA <- na.omit(raster::getValues(EnvRaster_Transformed)) # omi
t NA's from raster
dim(EnvRaster_NotNA) # 4072 raster cells hold numeric values, and all rema
ining cells (21596) are filled with NA's

## [1] 14577    11
```

```r
# Perform a Principal Component Analysis to get the loadings for each laye
rs of 'EnvRaster_Transformed':
#posi??o na dimens?o em que cada preditor est?. cada pixel pode ser locali
zado nesse espa?o 3d, a dista de um pro outro, ? a medida de dissimilarida
de estimada.
#isso que vai ser usado pra analisar a diversidade beta nesse espa?o multi
dimensional
#como os 22 eixos podem nao ser independentes, vamos usar o emsmo racionci
iono pra analise funcionao: passar um pca antes de fazer dist euclidiana
#mas as metricas aqui sao as caracteristcas das localidades.
#vamos aplicar pca pra garantir que os eixos sao independentes, ortogonais
#depois vamos sumarizar, pra saber o quanto de explica??o cada eixo do pca
tem

#usando os eixos, posso prever
EnvData_PCA <- prcomp(EnvRaster_NotNA, center=FALSE, scale=FALSE)
# Reescaling transformed predictors will remove their partial contribution
(do not do it).

# Percentage of variation in 'EnvRaster_Transformed' captured by the first
three PCA axes:
eigs <- EnvData_PCA$sdev^2
axis1 <- round(eigs[1] / sum(eigs), 3)*100 # eigenvalue 1
axis2 <- round(eigs[2] / sum(eigs), 3)*100 # eigenvalue 2
axis3 <- round(eigs[3] / sum(eigs), 3)*100 # eigenvalue 3
axis1 + axis2 + axis3 # cumulative proportion of variation exaplined by th
e three fist axes
```

```
## [1] 97.8

# Apply the PCA loadings to the 'EnvRaster_Transformed' raster stack:
#os dados transformados s? que resumidos nos eixos pca, transcrevemos os v
alores de pca para dados espaciais. passamos os rasters no pca. agora pode
mos visualizar esses eixos pca num plot espacial
EnvRaster_PCA <- raster::predict(EnvRaster_Transformed, EnvData_PCA, index
=1:3)



brasil_shapefile <- readOGR(dsn = "C:/Users/isagv/Documents/Drive/Projeto
Modelagem da diversidade filogenética/Análises9/clima", layer = "brasil")

## Warning: OGR support is provided by the sf and terra packages among oth
ers

## Warning: OGR support is provided by the sf and terra packages among oth
ers

## Warning: OGR support is provided by the sf and terra packages among oth
ers

## Warning: OGR support is provided by the sf and terra packages among oth
ers

## Warning: OGR support is provided by the sf and terra packages among oth
ers

## Warning: OGR support is provided by the sf and terra packages among oth
ers
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\isagv\Documents\Drive\Projeto Modelagem da diversidad
e filogenética\Análises9\clima", layer: "brasil"
## with 1 features
## It has 5 fields

# Certifique-se de que o sistema de coordenadas do shapefile seja o mesmo
dos rasters
brasil_shapefile <- spTransform(brasil_shapefile, crs(EnvRaster_Transforme
d))




# Definir a extensão do plot com base no shapefile do Brasil
ext <- extent(brasil_shapefile)

# Ajustar margens e inicializar gráficos corretamente
par(mfrow = c(1, 3)) # três eixos
par(mar = c(5, 5, 2, 2))

# Definir limites manuais
xlim <- c(ext@xmin, ext@xmax)
ylim <- c(ext@ymin, ext@ymax)

# PCA axis 1
plot(EnvRaster_PCA[[1]], xlim = xlim, ylim = ylim, col = grDevices::hcl.co
lors(n = 20, palette = "RdYlBu", rev = TRUE), main = paste("PCA axis 1 (",
```

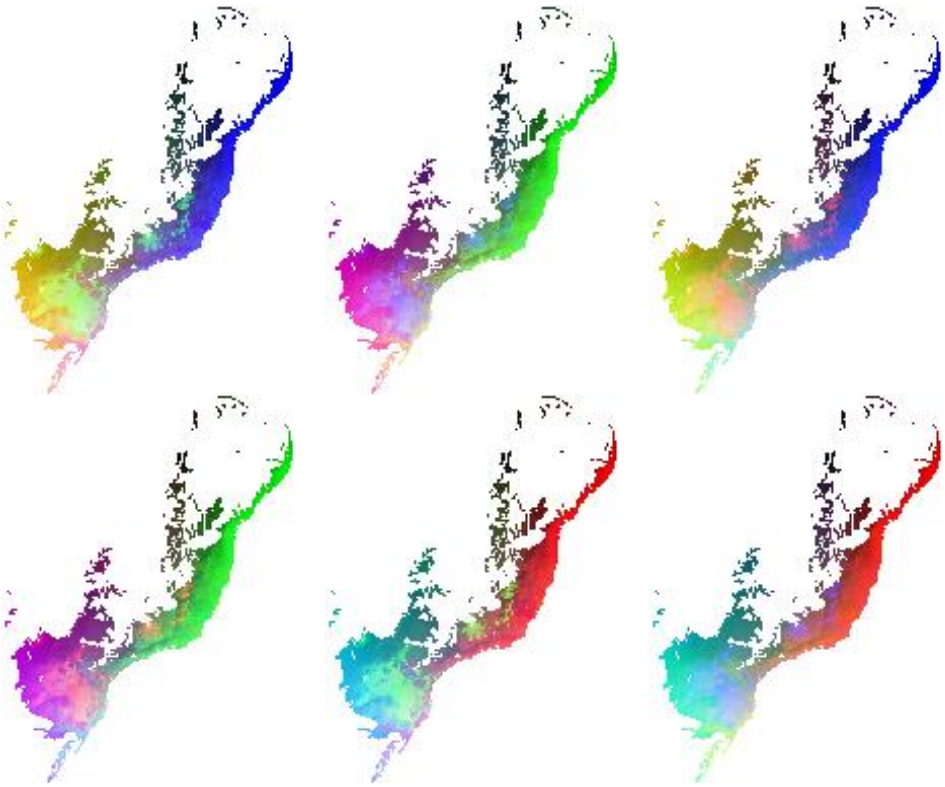
```
axis1, "%)", sep = ""))
plot(brasil_shapefile, add = TRUE, border = 'black')


# PCA axis 2
plot(EnvRaster_PCA[[2]], xlim = xlim, ylim = ylim, col = grDevices::hcl.co
lors(n = 20, palette = "RdYlBu", rev = TRUE), main = paste("PCA axis 2 (",
axis2, "%)", sep = ""))
plot(brasil_shapefile, add = TRUE, border = 'black')


# PCA axis 3
plot(EnvRaster_PCA[[3]], xlim = xlim, ylim = ylim, col = grDevices::hcl.co
lors(n = 20, palette = "RdYlBu", rev = TRUE), main = paste("PCA axis 3 (",
axis3, "%)", sep = ""))
plot(brasil_shapefile, add = TRUE, border = 'black')
```



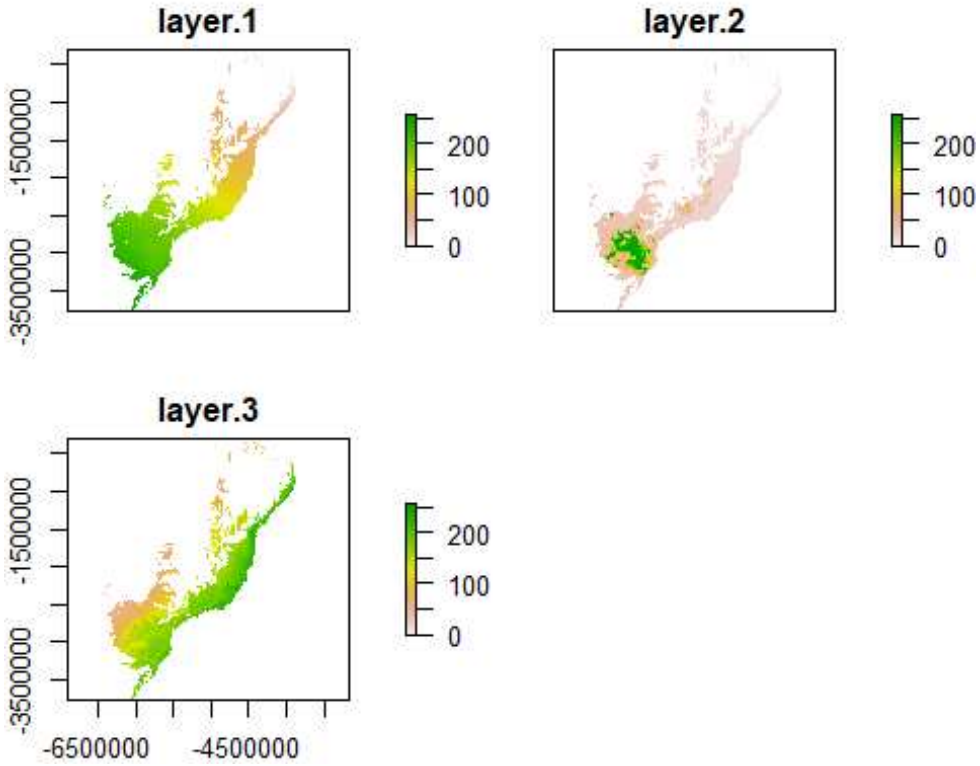
```
par(mfrow = c(1, 1))
```

```
# Rescale rasters do vary from 0 to 1 and multiple by 255 (extent of each
dimension of the RGB colour space):
#ajustando os valores pra encaixar no padr?o rgb. cada camada vai ser lido
como banda. as cores parecidas vao ser comundiades similares. a tonalidade
? proporcinal a dissimilridade. tons diferentes rpesentam comunidades dife
rentes

EnvRaster_PCA[[1]] <- (EnvRaster_PCA[[1]]-EnvRaster_PCA[[1]]@data@min)/
  (EnvRaster_PCA[[1]]@data@max-EnvRaster_PCA[[1]]@data@min)*255

EnvRaster_PCA[[2]] <- (EnvRaster_PCA[[2]]-EnvRaster_PCA[[2]]@data@min)/
  (EnvRaster_PCA[[2]]@data@max-EnvRaster_PCA[[2]]@data@min)*255

EnvRaster_PCA[[3]] <- (EnvRaster_PCA[[3]]-EnvRaster_PCA[[3]]@data@min)/
  (EnvRaster_PCA[[3]]@data@max-EnvRaster_PCA[[3]]@data@min)*255

# Plot RGB raster.  Use stretch option to increase contrast:
#esses todos s?o a mesma coisa, mas mudando as tonaldiade pra ver qual con
trasta melhor
par(mfrow = c(2, 3))
```

```
raster::plotRGB(EnvRaster_PCA, r=1, g=2, b=3, stretch='hist')

raster::plotRGB(EnvRaster_PCA, r=1, g=3, b=2, stretch='hist')

raster::plotRGB(EnvRaster_PCA, r=2, g=1, b=3, stretch='hist')

raster::plotRGB(EnvRaster_PCA, r=2, g=3, b=1, stretch='hist')

raster::plotRGB(EnvRaster_PCA, r=3, g=2, b=1, stretch='hist')

raster::plotRGB(EnvRaster_PCA, r=3, g=1, b=2, stretch='hist')
```



```
par(mfrow = c(1, 1))


# Convert the list of raster layers into a RGB raster:

EnvRaster_RGBStack<-raster::stack(EnvRaster_PCA)


plot(EnvRaster_PCA)
```




```
# Export to disk:
save(EnvRaster_PCA, file =  "EnvRaster_PCA.RData")
# Salvar cada camada do raster PCA como arquivos TIFF
writeRaster(EnvRaster_PCA[[1]], filename = "PCA_axis1.tif", format = "GTif
f", overwrite = TRUE)
writeRaster(EnvRaster_PCA[[2]], filename = "PCA_axis2.tif", format = "GTif
f", overwrite = TRUE)
writeRaster(EnvRaster_PCA[[3]], filename = "PCA_axis3.tif", format = "GTif
f", overwrite = TRUE)

# Salvar o stack inteiro como um arquivo multi-banda TIFF
writeRaster(EnvRaster_PCA, filename = "PCA_axes.tif", format = "GTiff", ov
erwrite = TRUE)
```

```
#####



# STEP 7: USING GDM TO PROJECT SPATIAL CHANGES IN BETA-DIVERSITY ACROSS TI
ME #mapear as mudanças
################################################################################
################################################
# STEP 7: USING GDM TO PROJECT SPATIAL CHANGES IN BETA-DIVERSITY ACROSS TI
ME

#os arquivos com dados atuais e futuros:
#bioclim_future_poly_filtered_clean
#bioclim_current_filtered_clean



# Transform geographical and environmental predictor to their respective b
iological importance and remove NA raster values:
CurrEnv_Transformed <- gdm.transform(model = gdm_Model_Diss, data = biocli
m_current_filtered_clean)

## Warning in gdm.transform(model = gdm_Model_Diss, data =
## bioclim_current_filtered_clean): Extracted data from rasters contained
NAs.
## These were automatically removed from the data object to be transformed
.
```

```
FutEnv_Transformed <- gdm.transform(model = gdm_Model_Diss, data = bioclim
_future_poly_filtered_clean) #os nomes dos preditores tem que ser igual, s
en?o n?o roda

## Warning in gdm.transform(model = gdm_Model_Diss, data =
## bioclim_future_poly_filtered_clean): Extracted data from rasters contai
ned NAs.
## These were automatically removed from the data object to be transformed
.

CurrEnv_NotNA <- na.omit(raster::getValues(CurrEnv_Transformed))
FutEnv_NotNA <- na.omit(raster::getValues(FutEnv_Transformed))


# PCA of environmental conditions get the loadings for each layers of 'Cur
rEnv_Transformed':
EnvData_PCA <- prcomp(CurrEnv_NotNA, center=FALSE, scale=FALSE)


# Apply the PCA loadings to the 'CurrEnv_Transformed' raster stack:
CurrEnv_PCA <- raster::predict(CurrEnv_Transformed, EnvData_PCA, index=1:n
col(EnvData_PCA$rotation))
FutEnv_PCA <- raster::predict(FutEnv_Transformed, EnvData_PCA, index=1:nco
l(EnvData_PCA$rotation))
# Use index = 1:ncol(EnvData_PCA$rotation) to predict site scores for all
PCA axes


# Convert raster projection to equal area:
#window size ? pra detemrinar quantos metros de distancia ate o vizinho
equalareaproj<-"+proj=cea +lon_0=0 +lat_ts=30 +x_0=0 +y_0=0 +datum=WGS84 +
```

```
units=m +no_defs" # cylindrical equal area projection
CurrEnv_cea<-raster::projectRaster(CurrEnv_PCA, crs=equalareaproj, method=
"ngb")
FutEnv_cea<-raster::projectRaster(FutEnv_PCA, crs=equalareaproj, method="n
gb")
window_size<-ceiling(max(raster::res(CurrEnv_cea[[1]])[2])/1000)*1000 # me
ters

# Convert raster brick to a data.frame in which each column is a PCA axis:
CurrEnv_axes<-raster::rasterToPoints(CurrEnv_cea)
FutEnv_axes<-raster::rasterToPoints(FutEnv_cea)

#install.packages("igraph")
#install.packages("adespatial")
#install.packages("gmp")
#install.packages("C:/Users/isagv/Downloads/CommEcol_1.7.1.tar.gz", repos
= NULL, type = "source")

#save.image("Capitulo_modelagem.RData")
# Compute the average distance in the PCA space between the focal cell and
its surroundings.
# Select the focal cell and the respective (immediate) adjacent cells:
#a coordenada ta em metros, a partir do meridiano (x) e linha do equadror
(y)
FocalCell <- CommEcol::select.window(xf=CurrEnv_axes[1000,1], yf=CurrEnv_a
xes[1000,2], radius=window_size, xydata=CurrEnv_axes)
```

```
# Focal cell is the number 1 in the output of CommEcol::select.window
# 7 | 8 | 9
# 5 | 1 | 6
# 2 | 3 | 4

# Compute the Euclidean distance between all selected cells:
FocalCell_dist<-as.matrix(dist(FocalCell[,3:ncol(FocalCell)], method="eucl
idean"))
FocalCell_dist

##             1           2           3           4           5
6
## 1 0.000000000 0.005260243 0.001481849 0.008687524 0.006536051 0.0035069
08
## 2 0.005260243 0.000000000 0.003836163 0.003468460 0.001321452 0.0017554
34
## 3 0.001481849 0.003836163 0.000000000 0.007228252 0.005084860 0.0020849
85
## 4 0.008687524 0.003468460 0.007228252 0.000000000 0.002156808 0.0051907
94
## 5 0.006536051 0.001321452 0.005084860 0.002156808 0.000000000 0.0030348
18
## 6 0.003506908 0.001755434 0.002084985 0.005190794 0.003034818 0.0000000
00
## 7 0.000301312 0.005142144 0.001316803 0.008544986 0.006400319 0.0033869
48
```

```
## 8 0.003833726 0.001531447 0.002367498 0.004860830 0.002722083 0.0005025
36
##             7           8
## 1 0.000301312 0.003833726
## 2 0.005142144 0.001531447
## 3 0.001316803 0.002367498
## 4 0.008544986 0.004860830
## 5 0.006400319 0.002722083
## 6 0.003386948 0.000502536
## 7 0.000000000 0.003684292
## 8 0.003684292 0.000000000
```

```
# Make the matrix diagonal NA (i.e., the distance of the focal cell to its
elf):
diag(FocalCell_dist)<-NA
FocalCell_dist
```

```
##             1           2           3           4           5
6
## 1          NA 0.005260243 0.001481849 0.008687524 0.006536051 0.0035069
08
## 2 0.005260243          NA 0.003836163 0.003468460 0.001321452 0.0017554
34
## 3 0.001481849 0.003836163          NA 0.007228252 0.005084860 0.0020849
85
## 4 0.008687524 0.003468460 0.007228252          NA 0.002156808 0.0051907
94
```

```
## 5 0.006536051 0.001321452 0.005084860 0.002156808          NA 0.0030348
18
## 6 0.003506908 0.001755434 0.002084985 0.005190794 0.003034818
NA
## 7 0.000301312 0.005142144 0.001316803 0.008544986 0.006400319 0.0033869
48
## 8 0.003833726 0.001531447 0.002367498 0.004860830 0.002722083 0.0005025
36
##             7           8
## 1 0.000301312 0.003833726
## 2 0.005142144 0.001531447
## 3 0.001316803 0.002367498
## 4 0.008544986 0.004860830
## 5 0.006400319 0.002722083
## 6 0.003386948 0.000502536
## 7          NA 0.003684292
## 8 0.003684292          NA
```

```
# Calculate the average distance between the focal cell and all surroundin
g cells:
AvgDist_FocalCell<-mean(FocalCell_dist[1,], na.rm=TRUE)
AvgDist_FocalCell
```

```
## [1] 0.004229659
```

```
# Similarly to the framework explored when using multiple-site beta divers
ity metrics, the number of cells may affect AvgDist_FocalCell.
# Since beta-diversity is expected to increase with an area effect, more c
```

```
ells can increase the AvgDist_FocalCell (the estimated beta-diversity).
# Perform a subsampling approach to control the number of cells included i
n computations.

str(FocalCell)

##  num [1:8, 1:4] -3932546 -3932546 -3923416 -3941676 -3923416 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : NULL
##   ..$ : chr [1:4] "x" "y" "layer.9" "layer.10"

str(CurrEnv_axes)

##  num [1:14477, 1:13] -3941676 -3932546 -3923416 -3914286 -3905156 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : NULL
##   ..$ : chr [1:13] "x" "y" "layer.1" "layer.2" ...

# Prepare R for parallel computation:
cl<-makePSOCKcluster(detectCores()-1, type="SOCK")
registerDoParallel(cl)
getDoParWorkers()

## [1] 3



# Extract spatial pattern of estimated beta-diversity in parallel:
SpatialBetaCurr <- foreach(i = 1:nrow(CurrEnv_axes),
                           .export = 'c',
                           .packages = c("base", "stats", "CommEcol", "dat
```

```
a.table")) %dopar% {


                            # Select the focal cell and its surroundings:
                            FocalCell <- CommEcol::select.window(xf=CurrE
nv_axes[i,1], yf=CurrEnv_axes[i,2], radius=window_size, xydata=CurrEnv_axe
s)
                            FocalCell <- as.data.frame(FocalCell)


                            # Proceed with computations only if focal cel
l has at least four adjacent cells and SampledCell has at least 3 columns:
                            if(nrow(FocalCell) >= 4 && ncol(FocalCell) >=
3) {
                              # Subsample sites within the regional cells
.
                              Subsampled_Cells <- list()


                              for(j in 1:10){ # resampling 10 times


                                # Subsample sites within the regional cel
l:
                                set.seed(j)
                                SampledCell <- FocalCell[sample(x = c(1:n
row(FocalCell)),
                                                                size = 4,
                                                                replace =
FALSE),]
```

```
                                # Compute the Euclidean distance between
all subsampled cells:
                                FocalCell_dist <- as.matrix(dist(SampledC
ell[, 3:ncol(SampledCell)], method = "euclidean"))

                                # Make the matrix diagonal NA (i.e., the
distance of the focal cell to itself):
                                diag(FocalCell_dist) <- NA

                                # Calculate the average distance between
the focal cell and all surrounding cells:
                                Subsampled_Cells[[j]] <- mean(FocalCell_d
ist[1,], na.rm=TRUE)
                            }

                            Subsampled_Cells <- unlist(Subsampled_Cells
) # bind outputs for each iteration
                            Subsampled_Cells <- mean(Subsampled_Cells,
na.rm=T) # extract the average value across iterations

                            # Store in a data.frame:
                            SpatialBeta <- data.frame(x = CurrEnv_axes[
i,1], # longitude
                                                      y = CurrEnv_axes[
i,2], # latitude
                                                      EstBeta = Subsamp
led_Cells # estimated beta
```

```
                                )


                                # Return output:
                                SpatialBeta
                              } else {
                                SpatialBeta <- data.frame(x = CurrEnv_axes[
i,1], # longitude
                                                          y = CurrEnv_axes[
i,2], # latitude
                                                          EstBeta = NA # NA
for estimated beta
                                )
                                SpatialBeta
                              }
                            }



# Same as above, but for the future scenario:
# Extract spatial pattern of estimated beta-diversity in parallel:
SpatialBetaFut <- foreach(i = 1:nrow(FutEnv_axes),
                          .export = 'c',
                          .packages = c("base", "stats", "CommEcol", "data
.table")) %dopar% {

                            # Select the focal cell and its surroundings:
                            FocalCell <- CommEcol::select.window(xf=FutEnv
_axes[i,1], yf=FutEnv_axes[i,2], radius=window_size, xydata=FutEnv_axes)
```

```
                            FocalCell <- as.data.frame(FocalCell)


                            # Proceed with computations only if focal cell
has at least four adjacent cells and SampledCell has at least 3 columns:
                            if(nrow(FocalCell) >= 4 && ncol(FocalCell) >=
3) {
                              # Subsample sites within the regional cells.
                              Subsampled_Cells <- list()

                              for(j in 1:10){ # resampling 10 times

                                # Subsample sites within the regional cell
:
                                set.seed(j)
                                SampledCell <- FocalCell[sample(x = c(1:nr
ow(FocalCell)),
                                                                size = 4,
                                                                replace =
FALSE),]

                                # Compute the Euclidean distance between a
ll subsampled cells:
                                FocalCell_dist <- as.matrix(dist(SampledCe
ll[, 3:ncol(SampledCell)], method = "euclidean"))

                                # Make the matrix diagonal NA (i.e., the d
istance of the focal cell to itself):
```

```
                                diag(FocalCell_dist) <- NA

                                # Calculate the average distance between t
he focal cell and all surrounding cells:
                                Subsampled_Cells[[j]] <- mean(FocalCell_di
st[1,], na.rm=TRUE)
                              }

                              Subsampled_Cells <- unlist(Subsampled_Cells)
# bind outputs for each iteration
                              Subsampled_Cells <- mean(Subsampled_Cells, n
a.rm=T) # extract the average value across iterations

                              # Store in a data.frame:
                              SpatialBeta <- data.frame(x = FutEnv_axes[i,
1], # longitude
                                                        y = FutEnv_axes[i,
2], # latitude
                                                        EstBeta = Subsampl
ed_Cells # estimated beta
                              )

                              # Return output:
                              SpatialBeta
                            } else {
                              SpatialBeta <- data.frame(x = FutEnv_axes[i,
1], # longitude
```

```
                                                    y = FutEnv_axes[i,
2], # latitude
                                                    EstBeta = NA # NA
for estimated beta
                                )
                                SpatialBeta
                              }
                            }



# Unbind lists into a single data.frame:
SpatialBetaCurr<-rbindlist(SpatialBetaCurr)
SpatialBetaFut<-rbindlist(SpatialBetaFut)
# os NA é pq naquele local não tinha dado climatico suficiente pra estimar
o beta
# OBS: 'SpatialBetaCurr' and 'SpatialBetaFut' were derived from raster wit
h the same extent, resolution, and crs.


# Compute spatio-temporal changes in species composition (biotic homogeniz
ation):
#valores positivos ? que aumentou diferene?a, valores negativos reduziu a
diferen?a (homogeneizou)
SpatioTemporalPhyloBeta<-data.frame(x=SpatialBetaCurr$x,
                                    y=SpatialBetaCurr$y,
                                    BetaCurr=SpatialBetaCurr$EstBeta,
                                    BetaFut=SpatialBetaFut$EstBeta,
                                    DeltaBeta=(SpatialBetaFut$EstBeta - Sp
```

```
atialBetaCurr$EstBeta))


# Save the results:
SpatioTemporalPhyloBeta$CellId <- 1:nrow(SpatioTemporalPhyloBeta)
fwrite(SpatioTemporalPhyloBeta, "SpatioTemporalBetaGDM.csv")


#####




#save.image("Capitulo_modelagem.RData")









#-----------------------
```

```
plot(mata_atlantica)
```

```
library(sf)
library(ggplot2)
library(readxl)

# Carregar os dados de dissimilaridade
dissimilarity_matrix <- read_excel("Matrix de distância dissimilaridade.xl
sx")

dissimilarity_matrix <- as.data.frame(dissimilarity_matrix)

# Definir a primeira coluna como os nomes das linhas
rownames(dissimilarity_matrix) <- dissimilarity_matrix[[1]]
```

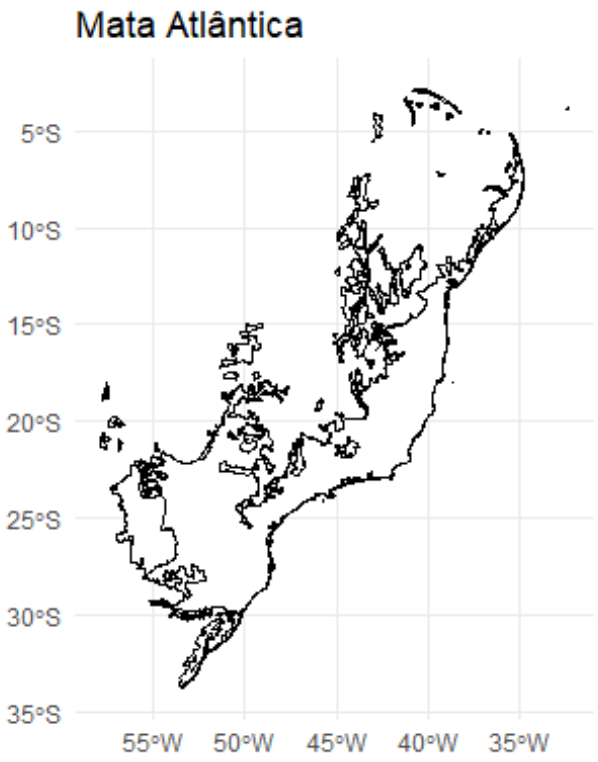
```r
# Remover a primeira coluna do data.frame, já que agora é usada como nomes de linha
dissimilarity_matrix <- dissimilarity_matrix[,-1]

dim(dissimilarity_matrix)

## [1] 28 28

# Supondo que 'dissimilarity_matrix' ainda tem os nomes das linhas corretos
# Primeiro, vamos copiar os nomes das linhas para uma nova coluna
dissimilarity_matrix$regiao <- rownames(dissimilarity_matrix)

# Agora, normalizamos a matriz novamente, mas mantendo a coluna 'regiao'
dissimilarity_matrix_normalized <- as.data.frame(lapply(dissimilarity_matrix, function(x) {
  if(is.numeric(x)) x / max(x, na.rm = TRUE) else x
}))



# Selecionar apenas as colunas de região, longitude e latitude de 'DissMatrix_Coords'
coords_selected <- DissMatrix_Coords %>% select(regiao, Longitude.x, Latitude.y)

# Remover a coluna 'regiao' temporariamente para calcular a média
temp_matrix <- dissimilarity_matrix_normalized[, -which(names(dissimilarity_matrix_normalized) == "regiao")]
```

```r
# Calcular a média de dissimilaridade para cada linha
dissimilarity_matrix_normalized$mean_dissimilarity <- rowMeans(temp_matrix
, na.rm = TRUE)

# Agora combinamos as coordenadas com a matriz normalizada
combined_data <- left_join(coords_selected, dissimilarity_matrix_normalize
d, by = "regiao")

# Garantindo que as coordenadas sejam convertidas corretamente para geomet
ria
coordinates <- st_as_sf(combined_data, coords = c("Longitude.x", "Latitude
.y"), crs = st_crs(mata_atlantica))

# Verificar a estrutura do objeto 'coordinates'
print(str(coordinates))
```

```
## Classes 'sf' and 'data.frame':   28 obs. of  31 variables:
##  $ regiao           : chr  "regiao1" "regiao10" "regiao11" "regiao12"
...
##  $ regiao1          : num  0 0.828 0.873 0.884 0.845 ...
##  $ regiao5          : num  0.927 0.761 0.74 0.584 0.681 ...
##  $ regiao6          : num  0.954 0.712 0.769 0.609 0.562 ...
##  $ regiao7          : num  0.91 0.864 0.876 0.806 0.732 ...

##  $ regiao8          : num  0.866 0.842 0.852 0.742 0.802 ...
##  $ regiao2          : num  0.81 0.77 0.778 0.864 0.798 ...
```

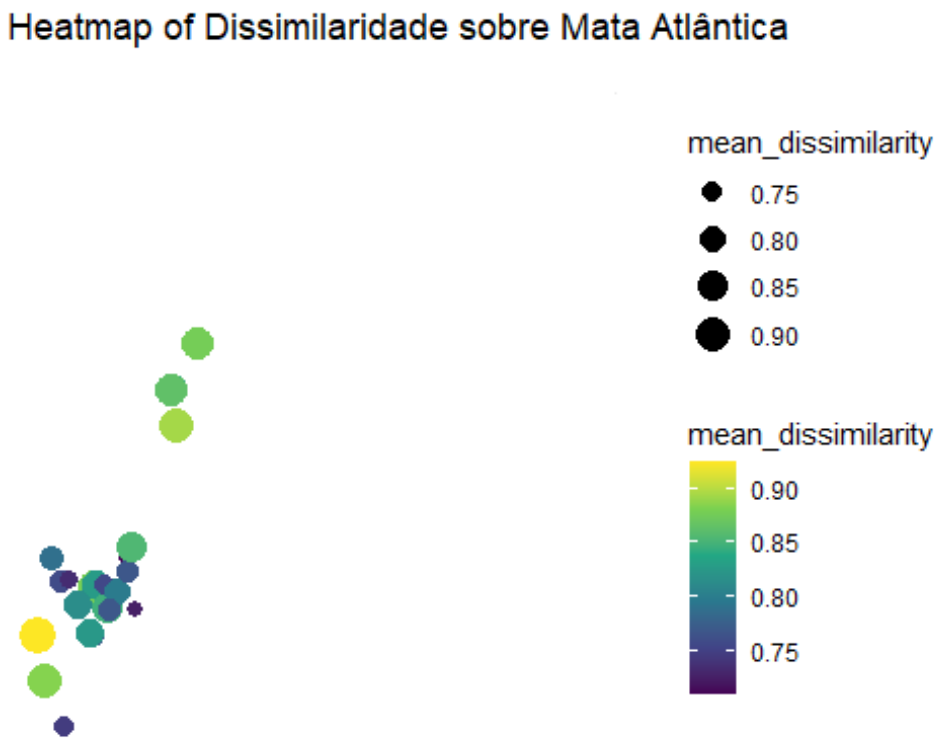
```
##  $ regiao9           : num  0.873 0.438 0.701 0.691 0.692 ...
##  $ regiao10          : num  0.863 0 0.624 0.673 0.708 ...
##  $ regiao11          : num  0.892 0.612 0 0.662 0.758 ...
##  $ regiao12          : num  0.95 0.694 0.696 0 0.608 ...
##  $ regiao13          : num  0.968 0.778 0.85 0.648 0 ...
##  $ regiao14          : num  0.994 0.855 0.835 0.764 0.736 ...
##  $ regiao3           : num  0.867 0.89 0.873 0.917 0.791 ...
##  $ regiao15          : num  0.856 0.989 0.944 0.962 0.877 ...
##  $ regiao16          : num  0.921 0.866 0.913 0.834 0.746 ...
##  $ regiao17          : num  0.931 0.814 0.89 0.862 0.806 ...
##  $ regiao18          : num  0.946 0.541 0.652 0.675 0.63 ...
##  $ regiao19          : num  0.968 0.89 0.872 0.805 0.872 ...
##  $ regiao20          : num  0.982 0.879 0.925 0.886 0.692 ...
##  $ regiao21          : num  0.996 0.955 0.946 0.894 0.854 ...
##  $ regiao22          : num  0.959 0.819 0.854 0.636 0.677 ...
##  $ regiao23          : num  0.987 0.8 0.777 0.591 0.711 ...
##  $ regiao24          : num  0.838 0.86 0.904 0.886 0.849 ...
##  $ regiao25          : num  0.846 0.946 1 0.912 0.867 ...
##  $ regiao26          : num  0.896 0.76 0.764 0.793 0.73 ...
##  $ regiao27          : num  0.879 0.574 0.606 0.627 0.766 ...
##  $ regiao28          : num  0.854 0.766 0.808 0.913 0.913 ...
##  $ regiao4           : num  0.837 0.718 0.755 0.748 0.714 ...
##  $ mean_dissimilarity: num  0.877 0.758 0.788 0.745 0.729 ...
##  $ geometry          :sfc_POINT of length 28; first list element:  'XY'
num  -4186405 -1214670
##  - attr(*, "sf_column")= chr "geometry"
##  - attr(*, "agr")= Factor w/ 3 levels "constant","aggregate",..: NA NA
```

```
NA NA NA NA NA NA NA NA ...
##   ..- attr(*, "names")= chr [1:30] "regiao" "regiao1" "regiao5" "regiao
6" ...
## NULL

mata_atlantica_sf <- st_as_sf(mata_atlantica)


ggplot() +
  geom_sf(data = mata_atlantica_sf, fill = "white", color = "black") +
  theme_minimal() +
  labs(title = "Mata Atlântica")
```




```
ggplot() +
  geom_sf(data = mata_atlantica_sf, fill = "white", color = "black") +
  geom_sf(data = coordinates, aes(size = mean_dissimilarity, color = mean_
dissimilarity)) +
```

```
  scale_color_viridis_c() +
  theme_minimal() +
  labs(title = "Heatmap of Dissimilaridade sobre Mata Atlântica")
```



APÊNDICE B

Apêndice referente ao Capítulo 2: "**Does an induced drought change the phylogenetic structure of Amazonian ant communities?**".

Essa seção não será incluída na submissão do artigo, é apenas para se caso os membros da banca queiram acessá-lo.

Script Report

O conteúdo a seguir é o *script report* em linguagem R utilizado para processar os dados e gerar os resultados presentes nesse estudo.

```
library(ape)
library(dplyr)
library(stringr)
library(tidyverse)
library(phylosignal)
library(adephylo)
library(phylobase)
library(phylotools)
library(phytools)
library(geiger)
library(vegan)
library(picante)


library(geiger)
library(ggtree)
```

```
library(ggplot2)
library(hrbrthemes)
library(abind)
library(devtools)
library(MicEco)
library(evobiR)
library(tidyverse)
library(phangorn)
library(pez)
#library(PDcalc)
#install.packages("remotes")
#library(remotes)
#remotes::install_github("Russel88/MicEco")
#if(!requireNamespace("BiocManager")){
#  install.packages("BiocManager")
#}
#BiocManager::install("phyloseq")



tree.moreau<-read.tree("moreau.txt")
plotTree(tree.moreau)

tree.moreau$tip.label[tree.moreau$tip.label=="Acromyrmex_versicolor"]<-"Ac
romyrmex"
tree.moreau$tip.label[tree.moreau$tip.label=="Anochetus_madagascarensis"]<
-"Anochetus"
tree.moreau$tip.label[tree.moreau$tip.label=="Apterostigma_auriculatum"]<-
```

```
"Apterostigma"
tree.moreau$tip.label[tree.moreau$tip.label=="Azteca_ovaticeps"]<-"Azteca"
tree.moreau$tip.label[tree.moreau$tip.label=="Basiceros_manni"]<-"Basicero
s"
tree.moreau$tip.label[tree.moreau$tip.label=="Brachymyrmex_depilis"]<-"Bra
chymyrmex"
tree.moreau$tip.label[tree.moreau$tip.label=="Camponotus_conithorax"]<-"Ca
mponotus"
tree.moreau$tip.label[tree.moreau$tip.label=="Centromyrmex_feae__CSM"]<-"C
entromyrmex"
tree.moreau$tip.label[tree.moreau$tip.label=="Cephalotes_unimaculatus__CSM
"]<-"Cephalotes"
tree.moreau$tip.label[tree.moreau$tip.label=="Crematogaster_navajoa__CSM"]
<-"Crematogaster"
tree.moreau$tip.label[tree.moreau$tip.label=="Daceton_armigerum"]<-"Daceto
n"
tree.moreau$tip.label[tree.moreau$tip.label=="Dolichoderus_scabridus"]<-"D
olichoderus"
tree.moreau$tip.label[tree.moreau$tip.label=="Eciton_vagans"]<-"Eciton"
tree.moreau$tip.label[tree.moreau$tip.label=="Ectatomma_opaciventre"]<-"Ec
tatomma"
tree.moreau$tip.label[tree.moreau$tip.label=="Gnamptogenys_striatula"]<-"G
namptogenys"
tree.moreau$tip.label[tree.moreau$tip.label=="Hypoponera_inexorata__CSM"]<
-"Hypoponera"
tree.moreau$tip.label[tree.moreau$tip.label=="Labidus_spininodis__CSM"]<-"
Labidus"
```

```
tree.moreau$tip.label[tree.moreau$tip.label=="Megalomyrmex_latreillei__CSM
"]<-"Megalomyrmex"
tree.moreau$tip.label[tree.moreau$tip.label=="Monomorium_ergatogyna"]<-"Mo
nomorium"
tree.moreau$tip.label[tree.moreau$tip.label=="Myrmicocrypta_cf_infuscata"]
<-"Myrmicocrypta"
tree.moreau$tip.label[tree.moreau$tip.label=="Neivamyrmex_nigrescens__CSM"
]<-"Neivamyrmex"
tree.moreau$tip.label[tree.moreau$tip.label=="Nomamyrmex_esenbecki__CSM"]<
-"Nomamyrmex"
tree.moreau$tip.label[tree.moreau$tip.label=="Odontomachus_clarus__CSM"]<-
"Odontomachus"
tree.moreau$tip.label[tree.moreau$tip.label=="Pachycondyla_stigma__CSM"]<-
"Pachycondyla"
tree.moreau$tip.label[tree.moreau$tip.label=="Pheidole_hyatti"]<-"Pheidole
"
tree.moreau$tip.label[tree.moreau$tip.label=="Pseudomyrmex_gracilis"]<-"Ps
eudomyrmex"
tree.moreau$tip.label[tree.moreau$tip.label=="Solenopsis_invicta__CSM"]<-"
Solenopsis"
tree.moreau$tip.label[tree.moreau$tip.label=="Strumigenys_dicomas"]<-"Stru
migenys"
tree.moreau$tip.label[tree.moreau$tip.label=="'Atta_sp.__CSM'"]<-"Atta"
tree.moreau$tip.label[tree.moreau$tip.label=="'Carebara_sp.__CSM'"]<-"Care
bara"
tree.moreau$tip.label[tree.moreau$tip.label=="'Cyphomyrmex_sp.__CSM'"]<-"C
yphomyrmex"
```

```
tree.moreau$tip.label[tree.moreau$tip.label=="'Sericomyrmex_sp.__CSM'"]<-"
Sericomyrmex"
tree.moreau$tip.label[tree.moreau$tip.label=="'Wasmannia_sp.__CSM'"]<-"Was
mannia"

#Os proximos sao os generos que nao tinham originalmente na arvore, entao
substitui pelo genero mais proximo que estava presente

tree.moreau$tip.label[tree.moreau$tip.label=="Acanthognathus_ocellatus"]<-
"Blepharidatta"
tree.moreau$tip.label[tree.moreau$tip.label=="Pogonomyrmex_maricopa__CSM"]
<-"Hylomyrma"
tree.moreau$tip.label[tree.moreau$tip.label=="Cryptopone_gilva__CSM"]<-"Ne
oponera"
tree.moreau$tip.label[tree.moreau$tip.label=="'Diacamma_sp.__CSM'"]<-"Maya
ponera"
tree.moreau$tip.label[tree.moreau$tip.label=="'Paratrechina_sp.__CSM'"]<-"
Nylanderia"
tree.moreau$tip.label[tree.moreau$tip.label=="Tranopelta_subterranea__CSM"
]<-"Ochetomyrmex"
tree.moreau$tip.label[tree.moreau$tip.label=="'Eurhopalothrix_sp.__CSM'"]<
-"Octostruma"

tips <- c("Acromyrmex","Anochetus","Apterostigma", "Azteca","Basiceros", "
Brachymyrmex",
          "Camponotus", "Centromyrmex", "Cephalotes", "Crematogaster", "Da
ceton","Dolichoderus",
```

```
          "Eciton","Ectatomma", "Gnamptogenys", "Hypoponera", "Labidus","M
egalomyrmex",
          "Monomorium", "Myrmicocrypta", "Neivamyrmex", "Nomamyrmex","Odon
tomachus",
          "Pachycondyla", "Pheidole","Pseudomyrmex","Solenopsis","Strumige
nys",
          "Atta","Carebara","Cyphomyrmex","Sericomyrmex","Wasmannia",
          "Blepharidatta","Hylomyrma","Neoponera","Mayaponera","Nylanderia
",
          "Ochetomyrmex","Octostruma")

tree.temp1 <- keep.tip(tree.moreau, tips)

plotTree(tree.temp1)
is.ultrametric(tree.temp1)

## inserindo as outras espC)cies na C!rvore
nodelabels(bg=NULL, frame=NULL)
#no no de sericomyrmex, mas serico pra fora
tree <- bind.tip(tree.temp1,tip.label = "Trachymyrmex", where=62, position
=1)
plotTree(tree)
nodelabels(bg=NULL, frame=NULL)
tiplabels()



#no n? grande que inclui odontomachus
```

```
tree <- bind.tip(tree,tip.label = "Rasopone", where=79, position=1)
plotTree(tree)

# irmaod de basiceros + octostruma
tree <- bind.tip(tree,tip.label = "Rhopalothrix", where=55, position=1)
plotTree(tree)
nodelabels(bg=NULL, frame=NULL)



#irmao de solenopsis e megalo
tree <- bind.tip(tree,tip.label = "Rogeria", where=69, position=1)
plotTree(tree)
nodelabels(bg=NULL, frame=NULL)

tree #44 tips
tree$tip.label
plotTree(tree)
is.phylo(tree)



#as espC)cies que precisam ser incluC-das nos gC*neros:

tips2 <- c("Trachymyrmex_MPEG02","Trachymyrmex_MPEG04","Trachymyrmex_MPEG0
5",
           "Rogeria_MPEG02","Rogeria_MPEG04","Acromyrmex_MPEG02","Anochetu
s_MPEG02",
           "Azteca_MPEG02","Azteca_MPEG03","Azteca_MPEG04","Azteca_MPEG05"
```

```
,
            "Brachymyrmex_MPEG02","Camponotus_prmelatoticus","Camponotus_MP
EG01",
            "Camponotus_MPEG02","Camponotus_MPEG05","Carebara_MPEG02","Care
bara_MPEG03",
            "Cephalotes_MPEG03","Cephalotes_MPEG01","Crematogaster_crinosa"
,
            "Crematogaster_flavosensitiva","Crematogaster_prlongispina","Cr
ematogaster_sotobosque","Crematogaster_MPEG01","Crematogaster_MPEG02","Cre
matogaster_MPEG04",
            "Cyphomyrmex_prminutus","Cyphomyrmex_peltatus","Dolichoderus_bi
spinosus",
            "Dolichoderus_decollatus","Dolichoderus_lutosus","Dolichoderus_
MPEG01",
            "Dolichoderus_MPEG03","Eciton_MPEG02","Eciton_MPEG03","Eciton_r
apax",
            "Ectatomma_lugens","Gnamptogenys_concinna","Gnamptogenys_horni"
,"Gnamptogenys_sulcata","Gnamptogenys_tortuolosa","Hypoponera_MPEG02","Lab
idus_praedator",
            "Labidus_spininodis","Megalomyrmex_MPEG03","Megalomyrmex_MPEG04
",
            "Myrmicocrypta_MPEG02","Myrmicocrypta_MPEG04","Neivamyrmex_MPEG
02",
            "Neivamyrmex_MPEG03","Neivamyrmex_MPEG04","Neivamyrmex_MPEG06",
"Neivamyrmex_pilosus","Neivamyrmex_pseudops","Neoponera_cfcommutata","Neop
onera_MPEG01",
            "Neoponera_verenae","Nomamyrmex_hartigii","Nylanderia_MPEG02","
```

```
Nylanderia_MPEG03","Nylanderia_MPEG04","Nylanderia_MPEG05","Nylanderia_MPE
G08","Ochetomyrmex_semipolitus","Odontomachus_MPEG02","Odontomachus_MPEG03
","Odontomachus_MPEG04",
            "Pheidole_grDiligensMPEG01","Pheidole_grDiligensMPEG02","Pheido
le_grDiligensMPEG03","Pheidole_grDiligensMPEG05","Pheidole_grDiligensMPEG0
6","Pheidole_grFallaxMPEG01","Pheidole_grFallaxMPEG02","Pheidole_grFallaxM
PEG03","Pheidole_grFlavensMPEG01","Pheidole_grTristisMPEG01","Pheidole_grT
ristisMPEG02","Pheidole_biconstricta","Pheidole_bruesi","Pheidole_cfaraneo
ides","Pheidole_dolon","Pheidole_fimbriata","Pheidole_fowleri","Pheidole_f
racticeps","Pheidole_jeannei","Pheidole_MPEG02","Pheidole_MPEG03","Pheidol
e_MPEG04","Pheidole_MPEG08","Pheidole_MPEG09","Pheidole_MPEG10","Pheidole_
MPEG11","Pheidole_MPEG12","Pheidole_MPEG13","Pheidole_MPEG15","Pheidole_sc
olioceps","Pheidole_subarmata","Pheidole_synarmata","Pheidole_vorax","Pseu
domyrmex_MPEG02","Pseudomyrmex_MPEG04","Pseudomyrmex_MPEG06","Sericomyrmex
_MPEG02","Solenopsis_MPEG02","Solenopsis_MPEG03","Solenopsis_MPEG04","Sole
nopsis_MPEG05","Solenopsis_MPEG06","Solenopsis_MPEG07","Strumigenys_elonga
ta","Strumigenys_MPEG01","Strumigenys_MPEG02","Strumigenys_perparva","Stru
migenys_zeteki","Wasmannia_MPEG01","Octostruma_betschi","Acromyrmex_MPEG01
",
            "Anochetus_MPEG01",
            "Apterostigma_MPEG01",
            "Azteca_MPEG01",
            "Basiceros_militaris",
            "Brachymyrmex_MPEG01",
            "Camponotus_femoratus",
            "Centromyrmex_brachycola",
            "Cephalotes_atratus",
```

```
            "Crematogaster_brasiliensis",
            "Daceton_armigerum",
            "Dolichoderus_attelaboides",
            "Eciton_MPEG01",
            "Ectatomma_edentatum",
            "Gnamptogenys_MPEG02",
            "Hypoponera_MPEG01",
            "Labidus_coecus",
            "Megalomyrmex_MPEG01",
            "Monomorium_cffloricola",
            "Myrmicocrypta_MPEG01",
            "Neivamyrmex_MPEG01",
            "Nomamyrmex_esenbeckii",
            "Odontomachus_MPEG01",
            "Pachycondyla_crassinoda",
            "Pheidole_MPEG01",
            "Pseudomyrmex_MPEG01","Solenopsis_MPEG01",
            "Strumigenys_denticulata",
            "Atta_MPEG01","Carebara_MPEG01","Cyphomyrmex_laevigatus",
            "Sericomyrmex_MPEG01","Wasmannia_auropunctata",
            "Blepharidatta_brasiliensis","Hylomyrma_MPEG03",
            "Neoponera_apicalis","Mayaponera_constricta",
            "Nylanderia_MPEG01","Ochetomyrmex_neopolitus",
            "Octostruma_iheringi", "Rasopone_prpergandei",
            "Rhopalothrix_MPEG01","Rogeria_MPEG01","Trachymyrmex_MPEG01")

######
```

```
library(phytools)

shuffle_species_within_genera <- function(genus_phylogeny, species_list, n
um_simulations) {
  # Extract the tips from the input phylogeny
  tips <- genus_phylogeny$tip.label

  # Split the species names into genus and species
  species_split <- strsplit(tips, "_")
  genera <- sapply(tips2, "[", 1)
  species <- sapply(tips, "[", 2)

  # Create a list to store the simulated trees
  trees <- list()

  # Loop through the number of simulations
  for (i in 1:num_simulations) {
    # Shuffle the species within each genus
    shuffled_species <- unlist(lapply(unique(genera), function(x) {
      genus_species <- species[genera == x]
      shuffled_species <- sample(genus_species, length(genus_species))
      paste0(x, "_", shuffled_species)
    }))

    # Create a new tree with the shuffled species
    new_tree <- genus.to.species.tree(genus_phylogeny, shuffled_species)
```

```
    # Add the tree to the list
    trees[[i]] <- new_tree
  }

  return(trees)
}

simulated_trees <- shuffle_species_within_genera(tree, tips2, 300)



plot(simulated_trees[[3]])
simulated_trees[[5]]$tip.label



#########################

subparcelas.centro<-read.csv("subparcelas.centro2.csv",sep=";",header=T,ro
w.names=NULL)
subparcelas.centro<- data.frame(subparcelas.centro[,-1], row.names=subparc
elas.centro[,1])



tree_list <- simulated_trees
community_matrix <- subparcelas.centro

# Loop through each tree in the list
```

```
for (i in 1:length(tree_list)) {
  # Remove "_NA" from species names
  tree_list[[i]]$tip.label <- gsub("_NA$", "", tree_list[[i]]$tip.label)
  # Remove underscore from tip labels
  tree_list[[i]]$tip.label <- gsub("_", "", tree_list[[i]]$tip.label)
}

plot(tree_list[[1]])



tree_list[[1]]$tip.label [1:5]
colnames(subparcelas.centro) [1:5]



subparcelas.centro2<-as.data.frame(t(subparcelas.centro))
subparcelas.centro3<-evobiR::ReorderData(tree_list[[1]], subparcelas.centr
o2, taxa.names="row names")
subparcelas.centro4<-as.data.frame(t(subparcelas.centro3))

tree_list[[1]]$tip.label [1:5]
colnames(subparcelas.centro4) [1:5]

match.phylo.comm(tree_list[[1]], subparcelas.centro4)
name.check(tree_list[[1]], data.names=colnames(subparcelas.centro4))
saveRDS(tree_list, file = "tree_list.rds")
write.csv(subparcelas.centro4, file="subparcelas.centro4.csv")
```

```
##################

## AN??LISES

##################


tree_list <- readRDS(file = "tree_list.rds")
subparcelas.centro4<-read.csv("subparcelas.centro4.csv",sep=",",header=T,row.names=NULL)
subparcelas.centro4<- data.frame(subparcelas.centro4[,-1], row.names=subparcelas.centro4[,1])
rownames(subparcelas.centro4)
summary(subparcelas.centro4)

#### SES PD
ses.pd_list <- list() # create an empty list to store the results

for (i in 1:length(tree_list)) {
  tree <- tree_list[[i]]
  ses_pd <- ses.pd(subparcelas.centro4, tree, null.model="richness", run=500)
  ses.pd_list[[i]] <- ses_pd # append the result to the list
}
```

```r
ses.pd_list[[1]][["pd.obs.p"]] #primeira simulaC'C#o
ses.pd_list[[2]][["pd.obs.p"]] #segunda simulaC'C#o

saveRDS(ses.pd_list, file = "ses_pd_list.rds")



################## SES MPD
ses.mpd_list <- list() # create an empty list to store the results

for (i in 1:length(tree_list)) {
  tree <- tree_list[[i]]
  ses_mpd <- ses.mpd(subparcelas.centro4, cophenetic(tree), null.model = "
taxa.labels", abundance.weighted = FALSE, runs = 500)
  ses.mpd_list[[i]] <- ses_mpd # append the result to the list
}

saveRDS(ses.mpd_list, file = "ses_mpd_list.rds")

################## SES MNTD
ses.mntd_list <- list() # create an empty list to store the results

for (i in 1:length(tree_list)) {
  tree <- tree_list[[i]]
  ses_mntd <- ses.mntd(subparcelas.centro4, cophenetic(tree), null.model =
"taxa.labels", abundance.weighted = FALSE, runs = 500)
  ses.mntd_list[[i]] <- ses_mntd # append the result to the list
```

```
}

saveRDS(ses.mntd_list, file = "ses_mntd_list.rds")

################## SES MPD - abund
ses.mpd_ab_list <- list() # create an empty list to store the results

for (i in 1:length(tree_list)) {
  tree <- tree_list[[i]]
  ses_mpd_ab <- ses.mpd(subparcelas.centro4, cophenetic(tree), null.model
= "taxa.labels", abundance.weighted = TRUE, runs = 500)
  ses.mpd_ab_list[[i]] <- ses_mpd_ab # append the result to the list
}

saveRDS(ses.mpd_ab_list, file = "ses_mpd_ab_list.rds")

################## SES MNTD -  abund
ses.mntd_ab_list <- list() # create an empty list to store the results

for (i in 1:length(tree_list)) {
  tree <- tree_list[[i]]
  ses_mntd_ab <- ses.mntd(subparcelas.centro4, cophenetic(tree), null.mode
l = "taxa.labels", abundance.weighted = TRUE, runs = 500)
  ses.mntd_ab_list[[i]] <- ses_mntd_ab # append the result to the list
}

saveRDS(ses.mntd_ab_list, file = "ses_mntd_ab_list.rds")
```

```
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✔ dplyr     1.1.2     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.2     ✔ tibble    3.2.1
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ✔ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_confl
icts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force
all conflicts to become errors

# ou se preferir carregar apenas pacotes específicos:
library(dplyr)
library(tidyr)
library(ape)

##
## Attaching package: 'ape'
##
## The following object is masked from 'package:dplyr':
##
##     where
```

```
library(dplyr)
library(stringr)
library(tidyverse)
library(phylosignal)
library(adephylo)

## Carregando pacotes exigidos: ade4

library(phylobase)

##
## Attaching package: 'phylobase'
##
## The following object is masked from 'package:ape':
##
##     edges

library(phylotools)
library(phytools)

## Carregando pacotes exigidos: maps
##
## Attaching package: 'maps'
##
## The following object is masked from 'package:purrr':
##


##     map
##
##
```

```
## Attaching package: 'phytools'

##

## The following object is masked from 'package:phylobase':

##

##     readNexus

library(geiger)

library(vegan)

## Carregando pacotes exigidos: permute

## Carregando pacotes exigidos: lattice

##

## Attaching package: 'lattice'

##

## The following object is masked from 'package:phylosignal':

##

##     dotplot

##

## This is vegan 2.6-4

##

## Attaching package: 'vegan'

##

## The following object is masked from 'package:phytools':

##

##     scores

library(picante)

## Carregando pacotes exigidos: nlme

##
```

```
## Attaching package: 'nlme'
##
## The following object is masked from 'package:dplyr':
##
##     collapse

library(geiger)
library(ggtree)

## ggtree v3.11.1.001 For help: https://yulab-smu.top/treedata-book/
##
## If you use the ggtree package suite in published research, please cite
## the appropriate paper(s):
##
## Guangchuang Yu, David Smith, Huachen Zhu, Yi Guan, Tommy Tsan-Yuk Lam.
## ggtree: an R package for visualization and annotation of phylogenetic
## trees with their covariates and other associated data. Methods in
## Ecology and Evolution. 2017, 8(1):28-36. doi:10.1111/2041-210X.12628
##
## G Yu. Data Integration, Manipulation and Visualization of Phylogenetic
## Trees (1st ed.). Chapman and Hall/CRC. 2022. ISBN: 9781032233574
##
## Guangchuang Yu, Tommy Tsan-Yuk Lam, Huachen Zhu, Yi Guan. Two methods
## for mapping and visualizing associated data on phylogeny using ggtree.
## Molecular Biology and Evolution. 2018, 35(12):3041-3043.
## doi:10.1093/molbev/msy194
##
## Attaching package: 'ggtree'
```

```
##
## The following object is masked from 'package:nlme':
##
##     collapse
##
## The following object is masked from 'package:phylobase':
##
##     MRCA
##
## The following object is masked from 'package:ape':
##
##     rotate
##
## The following object is masked from 'package:tidyr':
##
##     expand

library(ggplot2)
library(hrbrthemes)

## NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use
these themes.
##       Please use hrbrthemes::import_roboto_condensed() to install Robot
o Condensed and
##       if Arial Narrow is not on your system, please see https://bit.ly/
arialnarrow

library(abind)
library(devtools)
```

```
## Carregando pacotes exigidos: usethis
##
## Attaching package: 'devtools'
##
## The following object is masked from 'package:permute':
##
##     check
```

```
library(MicEco)
```

```
## Found more than one class "Annotated" in cache; using the first, from n
amespace 'RNeXML'
## Also defined by 'S4Vectors'
## Found more than one class "Annotated" in cache; using the first, from n
amespace 'RNeXML'
## Also defined by 'S4Vectors'
## Found more than one class "Annotated" in cache; using the first, from n
amespace 'RNeXML'
## Also defined by 'S4Vectors'
## Found more than one class "Annotated" in cache; using the first, from n
amespace 'RNeXML'
## Also defined by 'S4Vectors'
## Found more than one class "Annotated" in cache; using the first, from n
amespace 'RNeXML'
## Also defined by 'S4Vectors'
## Found more than one class "Annotated" in cache; using the first, from n
amespace 'RNeXML'
## Also defined by 'S4Vectors'
```

```
library(evobiR)

library(tidyverse)

library(phangorn)

##

## Attaching package: 'phangorn'

##

## The following object is masked from 'package:evobiR':

##

##     AICc

##

## The following objects are masked from 'package:vegan':

##

##     diversity, treedist

library(pez)

## Warning in .recacheSubclasses(def@className, def, env): subclasse "ndiM

atrix"

## da classe "replValueSp" não definida; definição não atualizada

######################



### RESULTADOS E GR??FICOS



######################

tree_list <- readRDS(file = "tree_list.rds")
```

```r
ses.pd_list <- readRDS(file = "ses_pd_list.rds")
ses.mpd_list <- readRDS(file = "ses_mpd_list.rds")
ses.mpd_ab_list <- readRDS(file = "ses_mpd_ab_list.rds")
ses.mntd_list <- readRDS(file = "ses_mntd_list.rds")
ses.mntd_ab_list <- readRDS(file = "ses_mntd_ab_list.rds")
```

```r
#acrescentando coluna com tratamento, pra plotar depois:
ses.pd_list[[1]]
```

```
##                   ntaxa   pd.obs pd.rand.mean pd.rand.sd pd.obs.rank
## ControleOutubro      55 1550.812     1896.612   113.0024           1
## ControleNovembro     54 1602.227     1885.634   121.7559           8
## ControleDezembro     62 1978.590     2029.675   119.7369         160

## ControleJaneiro      62 1941.790     2024.256   121.1063         124
## ControleFevereiro    45 1489.033     1703.394   118.2680          17
## ControleMarco        50 1632.594     1809.526   117.3491          39
## ControleMaio         54 1686.131     1885.120   119.0781          28
## ControleJulho        34 1297.337     1448.311   121.6187          55
## ControleAgosto       36 1342.932     1512.078   117.4356          42
## ControleSetembro     35 1234.632     1468.002   119.7911          17
## SecaOutubro          47 1722.681     1748.907   122.4178         221
## SecaNovembro         51 1796.535     1828.531   124.4204         201
```

```
## SecaDezembro          53 1792.024     1862.288   121.9587          136
## SecaJaneiro           41 1259.022     1616.138   119.2938            1
## SecaFevereiro         36 1236.658     1505.807   120.9576            7
## SecaMarco             33 1317.147     1421.698   127.4026          102
## SecaMaio              23 1002.579     1157.250   112.5682           45
## SecaJulho             26 1137.236     1231.055   118.1068          106
## SecaAgosto            30 1115.322     1347.773   118.7479           16
## SecaSetembro          28 1297.700     1295.011   118.9524          247
##                        pd.obs.z    pd.obs.p runs
## ControleOutubro   -3.06011058 0.001996008  500
## ControleNovembro  -2.32766153 0.015968064  500
## ControleDezembro  -0.42664573 0.319361277  500
## ControleJaneiro   -0.68094130 0.247504990  500
## ControleFevereiro -1.81250653 0.033932136  500
## ControleMarco     -1.50774005 0.077844311  500
## ControleMaio      -1.67107577 0.055888224  500
## ControleJulho     -1.24137383 0.109780439  500
## ControleAgosto    -1.44032712 0.083832335  500
## ControleSetembro  -1.94814422 0.033932136  500
## SecaOutubro       -0.21423710 0.441117764  500
## SecaNovembro      -0.25715873 0.401197605  500
## SecaDezembro      -0.57612603 0.271457086  500
## SecaJaneiro       -2.99358336 0.001996008  500
## SecaFevereiro     -2.22515496 0.013972056  500
## SecaMarco         -0.82064103 0.203592814  500
## SecaMaio          -1.37401638 0.089820359  500
## SecaJulho         -0.79435483 0.211576846  500
```

```
## SecaAgosto         -1.95752239 0.031936128  500
## SecaSetembro        0.02259951 0.493013972  500

# create a vector of treatment labels
treatment_labels <- c(rep("Control", 10), rep("Drought", 10))
# loop through each dataframe in the list and add the treatment column
for (i in 1:length(ses.pd_list)) {
  ses.pd_list[[i]]$treatment <- treatment_labels
}
month_labels <- c("Outubro","Novembro","Dezembro","Janeiro","Fevereiro","M
arco","Maio", "Julho","Agosto","Setembro","Outubro","Novembro","Dezembro",
"Janeiro","Fevereiro","Marco","Maio", "Julho","Agosto","Setembro")
# loop through each dataframe in the list and add the month column
for (i in 1:length(ses.pd_list)) {
  ses.pd_list[[i]]$month <- month_labels
}
ses.pd_list[[1]]

##                   ntaxa   pd.obs pd.rand.mean pd.rand.sd pd.obs.rank
## ControleOutubro      55 1550.812     1896.612   113.0024           1
## ControleNovembro     54 1602.227     1885.634   121.7559           8
## ControleDezembro     62 1978.590     2029.675   119.7369         160
## ControleJaneiro      62 1941.790     2024.256   121.1063         124
## ControleFevereiro    45 1489.033     1703.394   118.2680          17
## ControleMarco        50 1632.594     1809.526   117.3491          39
## ControleMaio         54 1686.131     1885.120   119.0781          28
## ControleJulho        34 1297.337     1448.311   121.6187          55
## ControleAgosto       36 1342.932     1512.078   117.4356          42
```

```
## ControleSetembro     35 1234.632     1468.002   119.7911          17
## SecaOutubro          47 1722.681     1748.907   122.4178         221
## SecaNovembro         51 1796.535     1828.531   124.4204         201
## SecaDezembro         53 1792.024     1862.288   121.9587         136
## SecaJaneiro          41 1259.022     1616.138   119.2938           1
## SecaFevereiro        36 1236.658     1505.807   120.9576           7
## SecaMarco            33 1317.147     1421.698   127.4026         102
## SecaMaio             23 1002.579     1157.250   112.5682          45
## SecaJulho            26 1137.236     1231.055   118.1068         106
## SecaAgosto           30 1115.322     1347.773   118.7479          16
## SecaSetembro         28 1297.700     1295.011   118.9524         247
##                        pd.obs.z    pd.obs.p runs treatment     month
## ControleOutubro   -3.06011058 0.001996008  500   Control   Outubro
## ControleNovembro  -2.32766153 0.015968064  500   Control  Novembro
## ControleDezembro  -0.42664573 0.319361277  500   Control  Dezembro
## ControleJaneiro   -0.68094130 0.247504990  500   Control   Janeiro
## ControleFevereiro -1.81250653 0.033932136  500   Control Fevereiro
## ControleMarco     -1.50774005 0.077844311  500   Control     Marco
## ControleMaio      -1.67107577 0.055888224  500   Control      Maio
## ControleJulho     -1.24137383 0.109780439  500   Control     Julho
## ControleAgosto    -1.44032712 0.083832335  500   Control    Agosto
## ControleSetembro  -1.94814422 0.033932136  500   Control  Setembro
## SecaOutubro       -0.21423710 0.441117764  500   Drought   Outubro
## SecaNovembro      -0.25715873 0.401197605  500   Drought  Novembro
## SecaDezembro      -0.57612603 0.271457086  500   Drought  Dezembro
## SecaJaneiro       -2.99358336 0.001996008  500   Drought   Janeiro
## SecaFevereiro     -2.22515496 0.013972056  500   Drought Fevereiro
```

```
## SecaMarco          -0.82064103 0.203592814  500   Drought     Marco
## SecaMaio           -1.37401638 0.089820359  500   Drought      Maio
## SecaJulho          -0.79435483 0.211576846  500   Drought     Julho
## SecaAgosto         -1.95752239 0.031936128  500   Drought    Agosto
## SecaSetembro        0.02259951 0.493013972  500   Drought  Setembro

ses.mpd_list[[1]]

##                   ntaxa  mpd.obs mpd.rand.mean mpd.rand.sd mpd.obs.rank
## ControleOutubro      55 148.7402      174.0779    7.495038            2
## ControleNovembro     54 156.3168      173.5487    7.729169           14
## ControleDezembro     62 162.1780      173.6573    6.560087           25
## ControleJaneiro      62 166.0482      173.5350    7.053449           72
## ControleFevereiro    45 157.9871      174.0021    8.963887           25
## ControleMarco        50 170.0231      174.2026    8.083694          151
## ControleMaio         54 170.8332      173.9182    7.751706          173
## ControleJulho        34 171.8133      173.5437   10.274882          204
## ControleAgosto       36 157.4501      173.9314   10.593004           35
## ControleSetembro     35 150.0032      174.0315   10.573265           14
## SecaOutubro          47 172.0292      173.6687    8.701570          204
## SecaNovembro         51 159.2645      173.4400    7.757113           17
## SecaDezembro         53 166.6749      173.7322    7.792033           89
## SecaJaneiro          41 145.4934      173.5312    9.536485            4
## SecaFevereiro        36 157.3442      173.1876   10.815150           43
## SecaMarco            33 163.7737      173.0771   11.258243           90
## SecaMaio             23 162.8449      174.2055   13.108832           96
## SecaJulho            26 163.9515      173.1473   12.047162          106
## SecaAgosto           30 147.8371      173.3134   11.738387           14
```

```
## SecaSetembro          28 186.4344      173.1122   12.608134          427
##                     mpd.obs.z   mpd.obs.p runs
## ControleOutubro    -3.3805960 0.003992016  500
## ControleNovembro   -2.2294635 0.027944112  500
## ControleDezembro   -1.7498701 0.049900200  500
## ControleJaneiro    -1.0614352 0.143712575  500
## ControleFevereiro -1.7866197 0.049900200  500
## ControleMarco      -0.5170273 0.301397206  500
## ControleMaio       -0.3979781 0.345309381  500
## ControleJulho      -0.1684043 0.407185629  500
## ControleAgosto     -1.5558661 0.069860279  500
## ControleSetembro   -2.2725563 0.027944112  500
## SecaOutubro        -0.1884160 0.407185629  500
## SecaNovembro       -1.8274217 0.033932136  500
## SecaDezembro       -0.9057007 0.177644711  500
## SecaJaneiro        -2.9400565 0.007984032  500
## SecaFevereiro      -1.4649237 0.085828343  500
## SecaMarco          -0.8263634 0.179640719  500
## SecaMaio           -0.8666376 0.191616766  500
## SecaJulho          -0.7633221 0.211576846  500
## SecaAgosto         -2.1703345 0.027944112  500
## SecaSetembro        1.0566337 0.852295409  500

# create a vector of treatment labels
treatment_labels <- c(rep("Control", 10), rep("Drought", 10))
# loop through each dataframe in the list and add the treatment column
for (i in 1:length(ses.mpd_list)) {
  ses.mpd_list[[i]]$treatment <- treatment_labels
```

```
}
month_labels <- c("Outubro","Novembro","Dezembro","Janeiro","Fevereiro","M
arco","Maio", "Julho","Agosto","Setembro","Outubro","Novembro","Dezembro",
"Janeiro","Fevereiro","Marco","Maio", "Julho","Agosto","Setembro")
# loop through each dataframe in the list and add the month column
for (i in 1:length(ses.mpd_list)) {
  ses.mpd_list[[i]]$month <- month_labels
}
ses.mpd_list[[1]]
```

```
##                   ntaxa  mpd.obs mpd.rand.mean mpd.rand.sd mpd.obs.rank
## ControleOutubro      55 148.7402      174.0779    7.495038            2
## ControleNovembro     54 156.3168      173.5487    7.729169           14
## ControleDezembro     62 162.1780      173.6573    6.560087           25
## ControleJaneiro      62 166.0482      173.5350    7.053449           72
## ControleFevereiro    45 157.9871      174.0021    8.963887           25
## ControleMarco        50 170.0231      174.2026    8.083694          151
## ControleMaio         54 170.8332      173.9182    7.751706          173
## ControleJulho        34 171.8133      173.5437   10.274882          204
## ControleAgosto       36 157.4501      173.9314   10.593004           35
## ControleSetembro     35 150.0032      174.0315   10.573265           14
## SecaOutubro          47 172.0292      173.6687    8.701570          204
## SecaNovembro         51 159.2645      173.4400    7.757113           17
## SecaDezembro         53 166.6749      173.7322    7.792033           89
## SecaJaneiro          41 145.4934      173.5312    9.536485            4
## SecaFevereiro        36 157.3442      173.1876   10.815150           43
## SecaMarco            33 163.7737      173.0771   11.258243           90
## SecaMaio             23 162.8449      174.2055   13.108832           96
```

```
## SecaJulho             26 163.9515      173.1473   12.047162          106
## SecaAgosto            30 147.8371      173.3134   11.738387           14
## SecaSetembro          28 186.4344      173.1122   12.608134          427
##                     mpd.obs.z   mpd.obs.p runs treatment     month
## ControleOutubro    -3.3805960 0.003992016  500   Control   Outubro
## ControleNovembro   -2.2294635 0.027944112  500   Control  Novembro
## ControleDezembro   -1.7498701 0.049900200  500   Control  Dezembro
## ControleJaneiro    -1.0614352 0.143712575  500   Control   Janeiro
## ControleFevereiro -1.7866197 0.049900200  500   Control Fevereiro
## ControleMarco      -0.5170273 0.301397206  500   Control     Marco
## ControleMaio       -0.3979781 0.345309381  500   Control      Maio
## ControleJulho      -0.1684043 0.407185629  500   Control     Julho
## ControleAgosto     -1.5558661 0.069860279  500   Control    Agosto
## ControleSetembro   -2.2725563 0.027944112  500   Control  Setembro
## SecaOutubro        -0.1884160 0.407185629  500   Drought   Outubro
## SecaNovembro       -1.8274217 0.033932136  500   Drought  Novembro
## SecaDezembro       -0.9057007 0.177644711  500   Drought  Dezembro
## SecaJaneiro        -2.9400565 0.007984032  500   Drought   Janeiro
## SecaFevereiro      -1.4649237 0.085828343  500   Drought Fevereiro
## SecaMarco          -0.8263634 0.179640719  500   Drought     Marco
## SecaMaio           -0.8666376 0.191616766  500   Drought      Maio
## SecaJulho          -0.7633221 0.211576846  500   Drought     Julho
## SecaAgosto         -2.1703345 0.027944112  500   Drought    Agosto
## SecaSetembro        1.0566337 0.852295409  500   Drought  Setembro

ses.mpd_ab_list[[1]]
```

```
##                    ntaxa  mpd.obs mpd.rand.mean mpd.rand.sd mpd.obs.rank
## ControleOutubro       55 135.5913      167.7220   10.790471            2
## ControleNovembro      54 138.0286      167.7936   11.187074            5
## ControleDezembro      62 138.2121      168.5203   10.155032            2
## ControleJaneiro       62 143.1300      168.8446    9.115449            2
## ControleFevereiro     45 147.2440      168.1458   10.675892           19
## ControleMarco         50 146.8744      168.1724   10.178609           17
## ControleMaio          54 152.0622      168.0798   10.561221           39
## ControleJulho         34 160.4264      166.9260   11.666486          140
## ControleAgosto        36 147.2514      166.2676   12.168330           35
## ControleSetembro      35 136.6780      165.9151   12.283241            8
## SecaOutubro           47 148.7035      165.8399   12.797326           53
## SecaNovembro          51 153.8098      167.2933   11.693212           75
## SecaDezembro          53 161.4689      167.2751   11.663091          150
## SecaJaneiro           41 124.0891      165.1672   12.602530            1
## SecaFevereiro         36 145.2261      165.8803   13.478171           35
## SecaMarco             33 140.4043      165.7482   13.415485           18
## SecaMaio              23 152.6662      162.3488   16.016778          137
## SecaJulho             26 162.5802      162.3725   15.324270          244
## SecaAgosto            30 152.7206      165.3616   13.240500           90
## SecaSetembro          28 172.4967      164.0830   14.538610          337
##                     mpd.obs.z   mpd.obs.p runs
## ControleOutubro   -2.97769080 0.003992016  500
## ControleNovembro  -2.66066110 0.009980040  500
## ControleDezembro  -2.98455332 0.003992016  500
## ControleJaneiro   -2.82098906 0.003992016  500
## ControleFevereiro -1.95784539 0.037924152  500
```

```
## ControleMarco     -2.09243488 0.033932136  500
## ControleMaio      -1.51664362 0.077844311  500
## ControleJulho     -0.55711953 0.279441118  500
## ControleAgosto    -1.56275708 0.069860279  500
## ControleSetembro  -2.38024577 0.015968064  500
## SecaOutubro       -1.33905959 0.105788423  500
## SecaNovembro      -1.15311161 0.149700599  500
## SecaDezembro      -0.49782290 0.299401198  500
## SecaJaneiro       -3.25951281 0.001996008  500
## SecaFevereiro     -1.53241808 0.069860279  500
## SecaMarco         -1.88915890 0.035928144  500
## SecaMaio          -0.60452642 0.273453094  500
## SecaJulho          0.01355977 0.487025948  500
## SecaAgosto        -0.95472412 0.179640719  500
## SecaSetembro       0.57871423 0.672654691  500
```

```
# create a vector of treatment labels
treatment_labels <- c(rep("Control", 10), rep("Drought", 10))
# loop through each dataframe in the list and add the treatment column
for (i in 1:length(ses.mpd_ab_list)) {
  ses.mpd_ab_list[[i]]$treatment <- treatment_labels
}
month_labels <- c("Outubro","Novembro","Dezembro","Janeiro","Fevereiro","M
arco","Maio", "Julho","Agosto","Setembro","Outubro","Novembro","Dezembro",
"Janeiro","Fevereiro","Marco","Maio", "Julho","Agosto","Setembro")
# loop through each dataframe in the list and add the month column
```

```r
for (i in 1:length(ses.mpd_ab_list)) {
  ses.mpd_ab_list[[i]]$month <- month_labels
}
ses.mpd_ab_list[[1]]
```

```
##                   ntaxa  mpd.obs mpd.rand.mean mpd.rand.sd mpd.obs.rank
## ControleOutubro      55 135.5913      167.7220   10.790471            2
## ControleNovembro     54 138.0286      167.7936   11.187074            5
## ControleDezembro     62 138.2121      168.5203   10.155032            2
## ControleJaneiro      62 143.1300      168.8446    9.115449            2
## ControleFevereiro    45 147.2440      168.1458   10.675892           19
## ControleMarco        50 146.8744      168.1724   10.178609           17
## ControleMaio         54 152.0622      168.0798   10.561221           39
## ControleJulho        34 160.4264      166.9260   11.666486          140
## ControleAgosto       36 147.2514      166.2676   12.168330           35
## ControleSetembro     35 136.6780      165.9151   12.283241            8


## SecaOutubro          47 148.7035      165.8399   12.797326           53
## SecaNovembro         51 153.8098      167.2933   11.693212           75
## SecaDezembro         53 161.4689      167.2751   11.663091          150
## SecaJaneiro          41 124.0891      165.1672   12.602530            1
## SecaFevereiro        36 145.2261      165.8803   13.478171           35
## SecaMarco            33 140.4043      165.7482   13.415485           18
## SecaMaio             23 152.6662      162.3488   16.016778          137
## SecaJulho            26 162.5802      162.3725   15.324270          244
## SecaAgosto           30 152.7206      165.3616   13.240500           90
## SecaSetembro         28 172.4967      164.0830   14.538610          337
```

```
##                      mpd.obs.z   mpd.obs.p runs treatment     month
## ControleOutubro   -2.97769080 0.003992016  500   Control   Outubro
## ControleNovembro  -2.66066110 0.009980040  500   Control  Novembro
## ControleDezembro  -2.98455332 0.003992016  500   Control  Dezembro
## ControleJaneiro   -2.82098906 0.003992016  500   Control   Janeiro
## ControleFevereiro -1.95784539 0.037924152  500   Control Fevereiro
## ControleMarco     -2.09243488 0.033932136  500   Control     Marco
## ControleMaio      -1.51664362 0.077844311  500   Control      Maio
## ControleJulho     -0.55711953 0.279441118  500   Control     Julho
## ControleAgosto    -1.56275708 0.069860279  500   Control    Agosto
## ControleSetembro  -2.38024577 0.015968064  500   Control  Setembro
## SecaOutubro       -1.33905959 0.105788423  500   Drought   Outubro
## SecaNovembro      -1.15311161 0.149700599  500   Drought  Novembro
## SecaDezembro      -0.49782290 0.299401198  500   Drought  Dezembro
## SecaJaneiro       -3.25951281 0.001996008  500   Drought   Janeiro
## SecaFevereiro     -1.53241808 0.069860279  500   Drought Fevereiro
## SecaMarco         -1.88915890 0.035928144  500   Drought     Marco
## SecaMaio          -0.60452642 0.273453094  500   Drought      Maio
## SecaJulho          0.01355977 0.487025948  500   Drought     Julho
## SecaAgosto        -0.95472412 0.179640719  500   Drought    Agosto
## SecaSetembro       0.57871423 0.672654691  500   Drought  Setembro

ses.mntd_list[[1]]

##                   ntaxa mntd.obs mntd.rand.mean mntd.rand.sd mntd.obs.r
ank
## ControleOutubro      55 25.21369       38.28567     4.930350
4
```

```
## ControleNovembro     54 30.84906       38.81500     5.107128
29
## ControleDezembro     62 36.68020       35.55805     4.418387
304
## ControleJaneiro      62 34.76799       35.38670     4.574711
220
## ControleFevereiro    45 38.64479       42.65519     6.071510
133
## ControleMarco        50 32.29773       40.10390     5.229892
50
## ControleMaio         54 28.84469       38.22524     5.140843
15
## ControleJulho        34 40.86686       50.11261     7.947213
67
## ControleAgosto       36 42.95888       48.97626     7.652783
106
## ControleSetembro     35 43.07472       49.81794     7.527518
91
## SecaOutubro          47 38.40851       41.88378     5.514726
130
## SecaNovembro         51 42.52200       39.89666     5.229796
345
## SecaDezembro         53 37.96472       39.08608     4.890585
211
## SecaJaneiro          41 34.80631       45.33745     6.484663
26
## SecaFevereiro        36 43.76009       48.28810     7.230685
```

```
136
## SecaMarco            33 46.31108       51.36345     8.299139
131
## SecaMaio             23 52.91137       61.89314    11.318034
111
## SecaJulho            26 52.25831       58.44974    10.518967
134
## SecaAgosto           30 45.30582       53.73804     8.613674
82
## SecaSetembro         28 51.57348       55.74717     9.217412
162
##                   mntd.obs.z  mntd.obs.p runs
## ControleOutubro   -2.6513295 0.007984032  500
## ControleNovembro  -1.5597700 0.057884232  500
## ControleDezembro   0.2539724 0.606786427  500
## ControleJaneiro   -0.1352448 0.439121756  500
## ControleFevereiro -0.6605284 0.265469062  500
## ControleMarco     -1.4926058 0.099800399  500
## ControleMaio      -1.8247105 0.029940120  500
## ControleJulho     -1.1633951 0.133732535  500
## ControleAgosto    -0.7862998 0.211576846  500
## ControleSetembro  -0.8958096 0.181636727  500
## SecaOutubro       -0.6301792 0.259481038  500
## SecaNovembro       0.5019958 0.688622754  500
## SecaDezembro      -0.2292892 0.421157685  500
## SecaJaneiro       -1.6240067 0.051896208  500
## SecaFevereiro     -0.6262215 0.271457086  500
```

```
## SecaMarco          -0.6087827 0.261477046  500
## SecaMaio           -0.7935807 0.221556886  500
## SecaJulho          -0.5885968 0.267465070  500
## SecaAgosto         -0.9789340 0.163672655  500
## SecaSetembro       -0.4528045 0.323353293  500

# create a vector of treatment labels
treatment_labels <- c(rep("Control", 10), rep("Drought", 10))
# loop through each dataframe in the list and add the treatment column
for (i in 1:length(ses.mntd_list)) {
  ses.mntd_list[[i]]$treatment <- treatment_labels
}
month_labels <- c("Outubro","Novembro","Dezembro","Janeiro","Fevereiro","M
arco","Maio", "Julho","Agosto","Setembro","Outubro","Novembro","Dezembro",
"Janeiro","Fevereiro","Marco","Maio", "Julho","Agosto","Setembro")


# loop through each dataframe in the list and add the month column
for (i in 1:length(ses.mntd_list)) {
  ses.mntd_list[[i]]$month <- month_labels
}
ses.mntd_list[[1]]

##                   ntaxa mntd.obs mntd.rand.mean mntd.rand.sd mntd.obs.r
ank
## ControleOutubro      55 25.21369       38.28567     4.930350
4
## ControleNovembro     54 30.84906       38.81500     5.107128
29
```

```
## ControleDezembro     62 36.68020       35.55805     4.418387
304
## ControleJaneiro      62 34.76799       35.38670     4.574711
220
## ControleFevereiro    45 38.64479       42.65519     6.071510
133
## ControleMarco        50 32.29773       40.10390     5.229892
50
## ControleMaio         54 28.84469       38.22524     5.140843
15
## ControleJulho        34 40.86686       50.11261     7.947213
67
## ControleAgosto       36 42.95888       48.97626     7.652783
106


## ControleSetembro     35 43.07472       49.81794     7.527518
91
## SecaOutubro          47 38.40851       41.88378     5.514726
130
## SecaNovembro         51 42.52200       39.89666     5.229796
345
## SecaDezembro         53 37.96472       39.08608     4.890585
211
## SecaJaneiro          41 34.80631       45.33745     6.484663
26
## SecaFevereiro        36 43.76009       48.28810     7.230685
136
```

```
## SecaMarco             33 46.31108       51.36345     8.299139
131
## SecaMaio              23 52.91137       61.89314    11.318034
111
## SecaJulho             26 52.25831       58.44974    10.518967
134
## SecaAgosto            30 45.30582       53.73804     8.613674
82
## SecaSetembro          28 51.57348       55.74717     9.217412
162
##                   mntd.obs.z  mntd.obs.p runs treatment     month
## ControleOutubro   -2.6513295 0.007984032  500   Control   Outubro
## ControleNovembro  -1.5597700 0.057884232  500   Control  Novembro
## ControleDezembro   0.2539724 0.606786427  500   Control  Dezembro
## ControleJaneiro   -0.1352448 0.439121756  500   Control   Janeiro
## ControleFevereiro -0.6605284 0.265469062  500   Control Fevereiro
## ControleMarco     -1.4926058 0.099800399  500   Control     Marco
## ControleMaio      -1.8247105 0.029940120  500   Control      Maio
## ControleJulho     -1.1633951 0.133732535  500   Control     Julho
## ControleAgosto    -0.7862998 0.211576846  500   Control    Agosto
## ControleSetembro  -0.8958096 0.181636727  500   Control  Setembro
## SecaOutubro       -0.6301792 0.259481038  500   Drought   Outubro
## SecaNovembro       0.5019958 0.688622754  500   Drought  Novembro
## SecaDezembro      -0.2292892 0.421157685  500   Drought  Dezembro
## SecaJaneiro       -1.6240067 0.051896208  500   Drought   Janeiro
## SecaFevereiro     -0.6262215 0.271457086  500   Drought Fevereiro
## SecaMarco         -0.6087827 0.261477046  500   Drought     Marco
```

```
## SecaMaio          -0.7935807 0.221556886  500   Drought      Maio
## SecaJulho         -0.5885968 0.267465070  500   Drought     Julho
## SecaAgosto        -0.9789340 0.163672655  500   Drought    Agosto
## SecaSetembro      -0.4528045 0.323353293  500   Drought  Setembro

ses.mntd_ab_list[[1]]

##                   ntaxa mntd.obs mntd.rand.mean mntd.rand.sd mntd.obs.r
ank
## ControleOutubro      55 20.08896       38.13568     7.253302
2
## ControleNovembro     54 34.53989       38.69068     7.247617
150
## ControleDezembro     62 26.12549       35.22145     6.477265
32
## ControleJaneiro      62 24.48068       35.49784     6.356646
15
## ControleFevereiro    45 35.36287       43.13793     7.604933
71
## ControleMarco        50 27.54649       40.13038     7.065716
12
## ControleMaio         54 28.50956       38.26694     7.048411
32
## ControleJulho        34 31.49668       50.15790     9.707043
12
## ControleAgosto       36 49.20518       49.40516    10.332817
261
## ControleSetembro     35 45.20159       49.05395     9.705019
```

```
177
## SecaOutubro          47 31.67676       41.62714     9.224607
63
## SecaNovembro         51 39.65188       39.77655     8.595148
258
## SecaDezembro         53 37.94792       38.41233     8.230026
248
## SecaJaneiro          41 26.62896       45.05095    10.175945
13
## SecaFevereiro        36 41.73545       49.52066    10.535833
120
## SecaMarco            33 34.92427       51.14670    11.293239
38
## SecaMaio             23 41.74924       62.60717    14.901832
40
## SecaJulho            26 47.47093       59.21592    13.426780
97
## SecaAgosto           30 52.26935       54.57323    11.556669
227
## SecaSetembro         28 46.77256       56.67456    12.590074
112
##                    mntd.obs.z  mntd.obs.p runs
## ControleOutubro   -2.48806879 0.003992016  500
## ControleNovembro  -0.57271095 0.299401198  500
## ControleDezembro  -1.40428887 0.063872255  500
## ControleJaneiro   -1.73317159 0.029940120  500
## ControleFevereiro -1.02236994 0.141716567  500
```

```
## ControleMarco     -1.78097899 0.023952096  500
## ControleMaio      -1.38433830 0.063872255  500
## ControleJulho     -1.92244190 0.023952096  500
## ControleAgosto    -0.01935355 0.520958084  500
## ControleSetembro  -0.39694584 0.353293413  500
## SecaOutubro       -1.07867805 0.125748503  500
## SecaNovembro      -0.01450499 0.514970060  500
## SecaDezembro      -0.05642859 0.495009980  500
## SecaJaneiro       -1.81034663 0.025948104  500
## SecaFevereiro     -0.73892745 0.239520958  500
## SecaMarco         -1.43647303 0.075848303  500
## SecaMaio          -1.39968891 0.079840319  500
## SecaJulho         -0.87474374 0.193612774  500
## SecaAgosto        -0.19935491 0.453093812  500
## SecaSetembro      -0.78649234 0.223552894  500

# create a vector of treatment labels
treatment_labels <- c(rep("Control", 10), rep("Drought", 10))
# loop through each dataframe in the list and add the treatment column
for (i in 1:length(ses.mntd_ab_list)) {
  ses.mntd_ab_list[[i]]$treatment <- treatment_labels
}
month_labels <- c("Outubro","Novembro","Dezembro","Janeiro","Fevereiro","M
arco","Maio", "Julho","Agosto","Setembro","Outubro","Novembro","Dezembro",
"Janeiro","Fevereiro","Marco","Maio", "Julho","Agosto","Setembro")
# loop through each dataframe in the list and add the month column
for (i in 1:length(ses.mntd_ab_list)) {
  ses.mntd_ab_list[[i]]$month <- month_labels
```

```
}
ses.mntd_ab_list[[1]]

##                   ntaxa mntd.obs mntd.rand.mean mntd.rand.sd mntd.obs.r
ank
## ControleOutubro      55 20.08896       38.13568     7.253302
2
## ControleNovembro     54 34.53989       38.69068     7.247617
150
## ControleDezembro     62 26.12549       35.22145     6.477265
32
## ControleJaneiro      62 24.48068       35.49784     6.356646
15
## ControleFevereiro    45 35.36287       43.13793     7.604933
71
## ControleMarco        50 27.54649       40.13038     7.065716
12
## ControleMaio         54 28.50956       38.26694     7.048411
32
## ControleJulho        34 31.49668       50.15790     9.707043
12
## ControleAgosto       36 49.20518       49.40516    10.332817
261
## ControleSetembro     35 45.20159       49.05395     9.705019
177
## SecaOutubro          47 31.67676       41.62714     9.224607
63
## SecaNovembro         51 39.65188       39.77655     8.595148
```

```
258
## SecaDezembro         53 37.94792       38.41233     8.230026
248
## SecaJaneiro          41 26.62896       45.05095    10.175945
13
## SecaFevereiro        36 41.73545       49.52066    10.535833
120
## SecaMarco            33 34.92427       51.14670    11.293239
38
## SecaMaio             23 41.74924       62.60717    14.901832
40
## SecaJulho            26 47.47093       59.21592    13.426780
97
## SecaAgosto           30 52.26935       54.57323    11.556669
227
## SecaSetembro         28 46.77256       56.67456    12.590074
112
##                    mntd.obs.z  mntd.obs.p runs treatment     month
## ControleOutubro   -2.48806879 0.003992016  500   Control   Outubro
## ControleNovembro  -0.57271095 0.299401198  500   Control  Novembro
## ControleDezembro  -1.40428887 0.063872255  500   Control  Dezembro
## ControleJaneiro   -1.73317159 0.029940120  500   Control   Janeiro
## ControleFevereiro -1.02236994 0.141716567  500   Control Fevereiro
## ControleMarco     -1.78097899 0.023952096  500   Control     Marco
## ControleMaio      -1.38433830 0.063872255  500   Control      Maio
## ControleJulho     -1.92244190 0.023952096  500   Control     Julho
## ControleAgosto    -0.01935355 0.520958084  500   Control    Agosto
```

```
## ControleSetembro  -0.39694584 0.353293413  500   Control  Setembro
## SecaOutubro       -1.07867805 0.125748503  500   Drought   Outubro
## SecaNovembro      -0.01450499 0.514970060  500   Drought  Novembro
## SecaDezembro      -0.05642859 0.495009980  500   Drought  Dezembro
## SecaJaneiro       -1.81034663 0.025948104  500   Drought   Janeiro
## SecaFevereiro     -0.73892745 0.239520958  500   Drought Fevereiro
## SecaMarco         -1.43647303 0.075848303  500   Drought     Marco
## SecaMaio          -1.39968891 0.079840319  500   Drought      Maio
## SecaJulho         -0.87474374 0.193612774  500   Drought     Julho
## SecaAgosto        -0.19935491 0.453093812  500   Drought    Agosto
## SecaSetembro      -0.78649234 0.223552894  500   Drought  Setembro

##########################



##### Gr??ficos



##########################

#    PD

# Create empty lists to store the mean values
control_means <- list()
drought_means <- list()

# Iterate over each month
for (month in c("Outubro", "Novembro", "Dezembro", "Janeiro", "Fevereiro",
```

```
"Marco", "Maio", "Julho", "Agosto", "Setembro")) {
  control_values <- numeric()
  drought_values <- numeric()

  # Iterate over each element in ses.mpd_list
  for (i in 1:length(ses.pd_list)) {
    # Extract the corresponding dataframe for the current element
    df <- ses.pd_list[[i]]

    # Filter the dataframe for the current month and treatment
    month_filter <- df$month == month
    control_filter <- df$treatment == "Control"
    drought_filter <- df$treatment == "Drought"

    # Extract the mpd.obs values for the current month and treatment
    control_pd_obs <- df$pd.obs[month_filter & control_filter]
    drought_pd_obs <- df$pd.obs[month_filter & drought_filter]

    # Append the values to the respective lists
    control_values <- c(control_values, control_pd_obs)
    drought_values <- c(drought_values, drought_pd_obs)
  }

  # Calculate the mean for each treatment and store the results in the res
pective lists
  control_means[[month]] <- mean(control_values)
  drought_means[[month]] <- mean(drought_values)
```

```r
}


# Print the mean values for each month in control treatment
cat("Mean values in control treatment:\n")

## Mean values in control treatment:

for (month in names(control_means)) {
  cat(month, ": ", control_means[[month]], "\n")
}



## Outubro :  1712.223

## Novembro :  1714.417

## Dezembro :  2062.077

## Janeiro :  2025.416

## Fevereiro :  1536.854

## Marco :  1751.261

## Maio :  1760.671

## Julho :  1392.558

## Agosto :  1432.104

## Setembro :  1314.483

# Print the mean values for each month in drought treatment
cat("\nMean values in drought treatment:\n")

##

## Mean values in drought treatment:
```

```r
for (month in names(drought_means)) {
  cat(month, ": ", drought_means[[month]], "\n")
}
```

```
## Outubro :  1729.735
## Novembro :  1831.841
## Dezembro :  1897.991
## Janeiro :  1372.201
## Fevereiro :  1334.179
## Marco :  1425.811
## Maio :  1033.27
## Julho :  1177.358
## Agosto :  1166.697
## Setembro :  1365.576
```

```r
# Create a data frame for control means
control_df <- data.frame(
  Month = names(control_means),
  Control_Mean = unlist(control_means)
)



# Create a data frame for drought means
drought_df <- data.frame(
  Month = names(drought_means),
  Drought_Mean = unlist(drought_means)
)
```

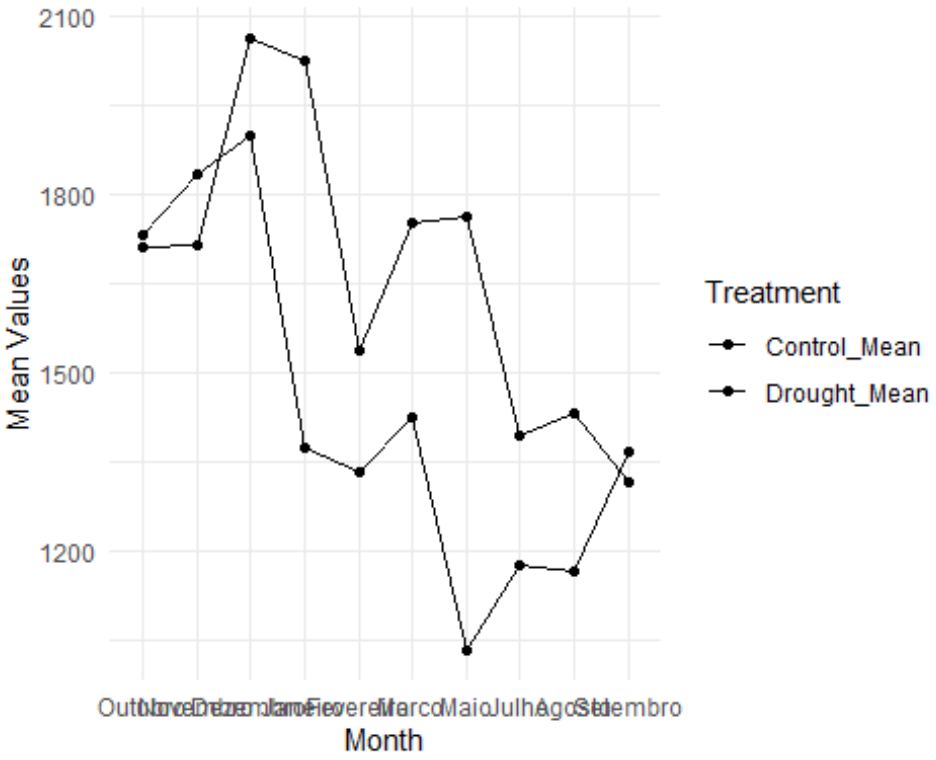
```r
# Merge the control and drought data frames
result_df <- merge(control_df, drought_df, by = "Month")


# Print the resulting dataframe
print(result_df)
```

```
##        Month Control_Mean Drought_Mean
## 1     Agosto     1432.104     1166.697
## 2   Dezembro     2062.077     1897.991
## 3  Fevereiro     1536.854     1334.179
## 4    Janeiro     2025.416     1372.201
## 5      Julho     1392.558     1177.358
## 6       Maio     1760.671     1033.270
## 7      Marco     1751.261     1425.811
## 8   Novembro     1714.417     1831.841
## 9    Outubro     1712.223     1729.735
## 10  Setembro     1314.483     1365.576
```

```r
# Create a data frame for control means
control_df <- data.frame(
  Month = names(control_means),
  Control_Mean = unlist(control_means)
)


# Create a data frame for drought means
drought_df <- data.frame(
  Month = names(drought_means),
  Drought_Mean = unlist(drought_means)
```

```
)

# Merge the control and drought data frames
result_df <- merge(control_df, drought_df, by = "Month")
write.table(result_df, file = "result_pd.csv", sep = ",", row.names = FALS
E)

# Reshape the data into long format
result_long <- result_df %>%
  gather(key = "Treatment", value = "Mean", -Month)

# Print the resulting long-format dataframe
print(result_long)

##        Month    Treatment     Mean
## 1     Agosto Control_Mean 1432.104
## 2   Dezembro Control_Mean 2062.077
## 3  Fevereiro Control_Mean 1536.854
## 4    Janeiro Control_Mean 2025.416
## 5      Julho Control_Mean 1392.558
## 6       Maio Control_Mean 1760.671
## 7      Marco Control_Mean 1751.261
## 8   Novembro Control_Mean 1714.417
## 9    Outubro Control_Mean 1712.223
## 10  Setembro Control_Mean 1314.483
## 11    Agosto Drought_Mean 1166.697
## 12  Dezembro Drought_Mean 1897.991
## 13 Fevereiro Drought_Mean 1334.179
```

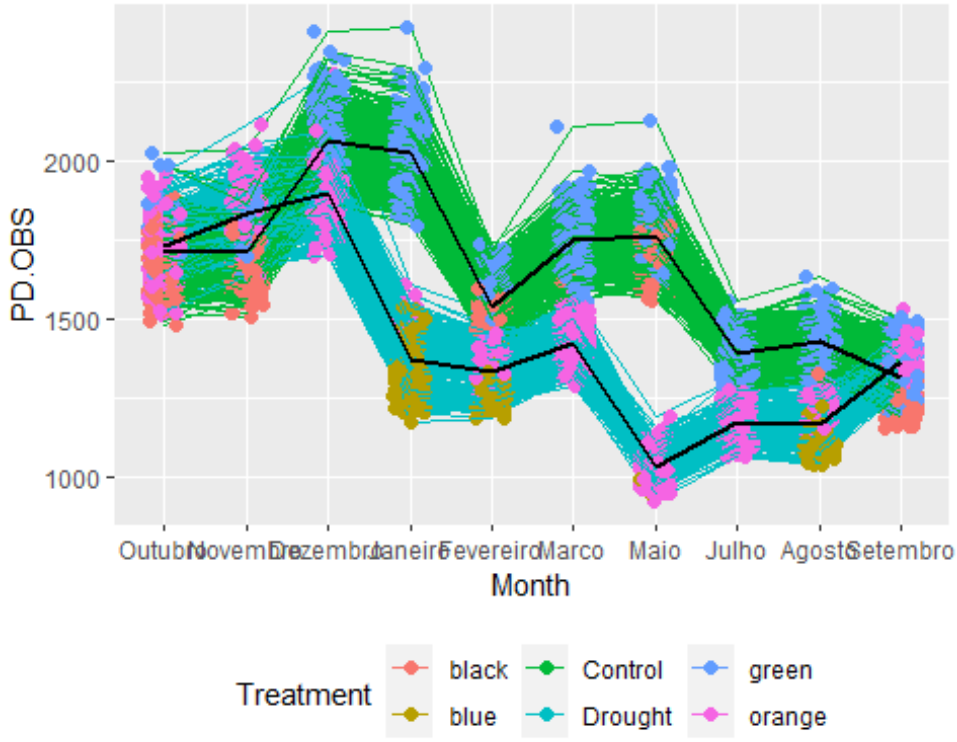
```
## 14   Janeiro Drought_Mean 1372.201
## 15     Julho Drought_Mean 1177.358
## 16      Maio Drought_Mean 1033.270
## 17     Marco Drought_Mean 1425.811
## 18  Novembro Drought_Mean 1831.841
## 19   Outubro Drought_Mean 1729.735
## 20  Setembro Drought_Mean 1365.576

# Define the desired order of months
month_order <- c("Outubro", "Novembro", "Dezembro", "Janeiro", "Fevereiro"
, "Marco", "Maio", "Julho", "Agosto", "Setembro")


# Convert "Month" to a factor with ordered levels
result_long$Month <- factor(result_long$Month, levels = month_order)


# Plotting
ggplot(result_long, aes(x = Month, y = Mean, color = Treatment, group = Tr
eatment)) +
  geom_line() +
  geom_point() +
  labs(x = "Month", y = "Mean Values", color = "Treatment") +
  scale_color_manual(values = c("black", "black")) +
  theme_minimal()
```




```
# Create an empty plot object
p <- ggplot() + labs(x = "Month", y = "PD.OBS", color = "Treatment")


plot_data <- function(df, plot) {
  plot <- plot +
    geom_line(data = df, aes(x = factor(month, levels = c("Outubro","Novembro","Dezembro","Janeiro","Fevereiro","Marco","Maio", "Julho","Agosto","Setembro")), y = pd.obs, color = treatment, group = treatment)) +
    geom_point(data = df, aes(x = factor(month, levels = c("Outubro","Novembro","Dezembro","Janeiro","Fevereiro","Marco","Maio", "Julho","Agosto","Setembro")), y = pd.obs, color = ifelse(treatment == "Control" & pd.obs.p < 0.025,"black",

ifelse(treatment == "Control" & pd.obs.p >= 0.025, "green",
```

```
ifelse(treatment == "Drought" & pd.obs.p < 0.025, "blue", "orange")


)
    )
    ), size = 2, shape = 16, position = position_jitter(width = 0.2, heigh
t = 0)) +
    labs(x = "Month", y = "PD.OBS") +
    theme(legend.position = "bottom")
  return(plot)
}




# Loop through each dataframe in the list and add the data to the plot obj
ect
for (i in 1:length(ses.pd_list)) {
  p <- plot_data(ses.pd_list[[i]], p)
}


# Add the lines representing the means from the separate dataframe
p <- p +
  geom_line(data = result_long, aes(x = Month, y = Mean, color = Treatment
, group = Treatment), size = 1, color="black")

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4
.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning w
as
## generated.

# Display the final plot
print(p)
```




```
############################
#########################

#    MPD

# Create empty lists to store the mean values
control_means <- list()
drought_means <- list()
```

```r
# Iterate over each month
for (month in c("Outubro", "Novembro", "Dezembro", "Janeiro", "Fevereiro",
"Marco", "Maio", "Julho", "Agosto", "Setembro")) {
  control_values <- numeric()
  drought_values <- numeric()



  # Iterate over each element in ses.mpd_list
  for (i in 1:length(ses.mpd_list)) {
    # Extract the corresponding dataframe for the current element
    df <- ses.mpd_list[[i]]

    # Filter the dataframe for the current month and treatment
    month_filter <- df$month == month
    control_filter <- df$treatment == "Control"
    drought_filter <- df$treatment == "Drought"

    # Extract the mpd.obs values for the current month and treatment
    control_mpd_obs <- df$mpd.obs[month_filter & control_filter]
    drought_mpd_obs <- df$mpd.obs[month_filter & drought_filter]

    # Append the values to the respective lists
    control_values <- c(control_values, control_mpd_obs)
    drought_values <- c(drought_values, drought_mpd_obs)
  }

  # Calculate the mean for each treatment and store the results in the res
```

```
pective lists
  control_means[[month]] <- mean(control_values)
  drought_means[[month]] <- mean(drought_values)
}

# Print the mean values for each month in control treatment
cat("Mean values in control treatment:\n")

## Mean values in control treatment:

for (month in names(control_means)) {
  cat(month, ": ", control_means[[month]], "\n")
}

## Outubro :  151.0467
## Novembro :  158.3879
## Dezembro :  164.4789
## Janeiro :  167.7866
## Fevereiro :  161.4048
## Marco :  172.5715
## Maio :  172.0292
## Julho :  174.4787
## Agosto :  159.7077
## Setembro :  154.0294

# Print the mean values for each month in drought treatment
cat("\nMean values in drought treatment:\n")

##
## Mean values in drought treatment:
```

```r
for (month in names(drought_means)) {
  cat(month, ": ", drought_means[[month]], "\n")
}
```

```
## Outubro :  172.8402
## Novembro :  160.9453
## Dezembro :  169.2248
## Janeiro :  149.7855
## Fevereiro :  160.4568
## Marco :  166.091
## Maio :  164.2443
## Julho :  166.3492
## Agosto :  150.9673
## Setembro :  187.6169
```

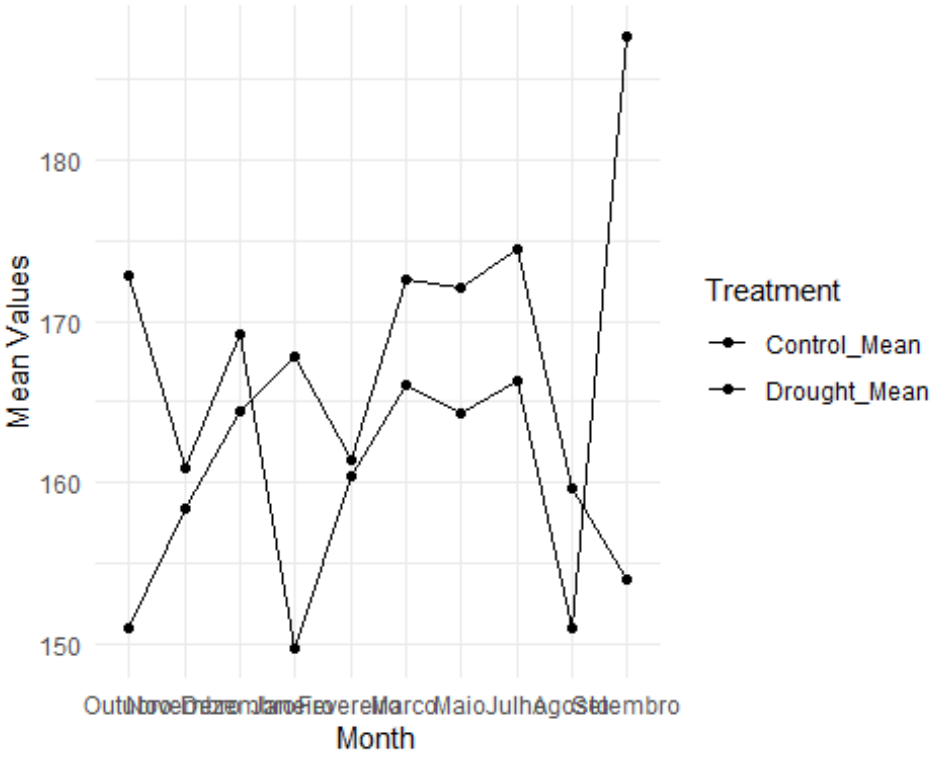
```r
# Create a data frame for control means
control_df <- data.frame(
  Month = names(control_means),
  Control_Mean = unlist(control_means)
)

# Create a data frame for drought means
drought_df <- data.frame(
  Month = names(drought_means),
  Drought_Mean = unlist(drought_means)
)

# Merge the control and drought data frames
```

```r
result_df <- merge(control_df, drought_df, by = "Month")



# Print the resulting dataframe
print(result_df)

##        Month Control_Mean Drought_Mean
## 1     Agosto     159.7077     150.9673
## 2   Dezembro     164.4789     169.2248
## 3  Fevereiro     161.4048     160.4568
## 4    Janeiro     167.7866     149.7855
## 5      Julho     174.4787     166.3492
## 6       Maio     172.0292     164.2443
## 7      Marco     172.5715     166.0910
## 8   Novembro     158.3879     160.9453
## 9    Outubro     151.0467     172.8402
## 10  Setembro     154.0294     187.6169

# Create a data frame for control means
control_df <- data.frame(
  Month = names(control_means),
  Control_Mean = unlist(control_means)
)


# Create a data frame for drought means
drought_df <- data.frame(
  Month = names(drought_means),
  Drought_Mean = unlist(drought_means)
```

```
)

# Merge the control and drought data frames
result_df <- merge(control_df, drought_df, by = "Month")
write.table(result_df, file = "result_mpd.csv", sep = ",", row.names = FAL
SE)

# Reshape the data into long format
result_long <- result_df %>%
  gather(key = "Treatment", value = "Mean", -Month)

# Print the resulting long-format dataframe
print(result_long)
```

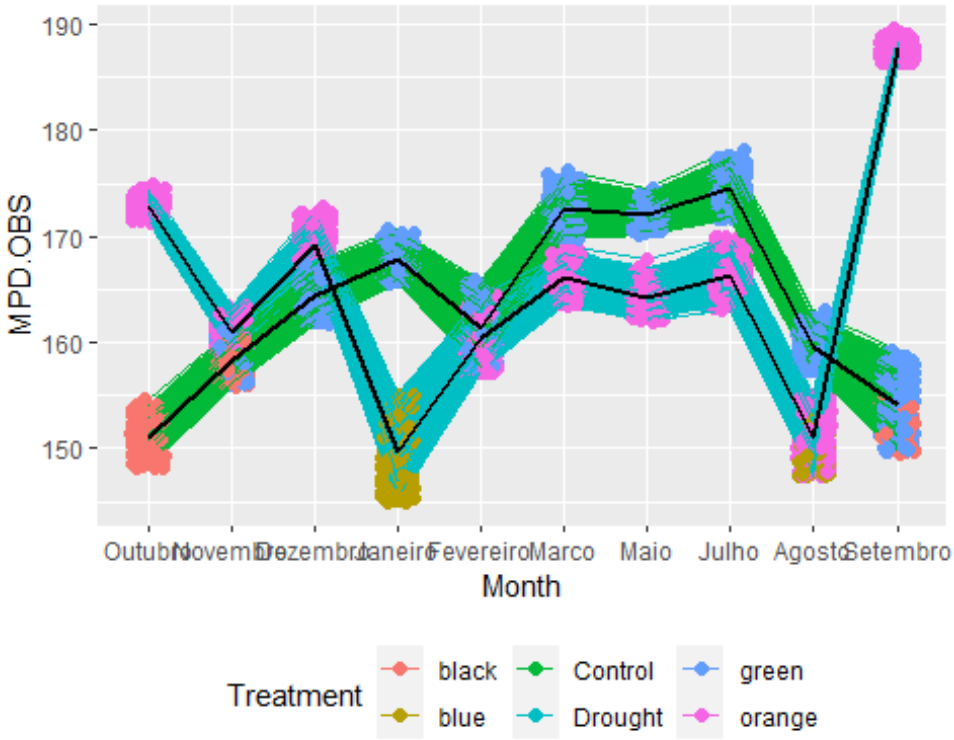
```
##        Month    Treatment     Mean
## 1     Agosto Control_Mean 159.7077
## 2   Dezembro Control_Mean 164.4789
## 3  Fevereiro Control_Mean 161.4048
## 4    Janeiro Control_Mean 167.7866
## 5      Julho Control_Mean 174.4787
## 6       Maio Control_Mean 172.0292
## 7      Marco Control_Mean 172.5715
## 8   Novembro Control_Mean 158.3879
## 9    Outubro Control_Mean 151.0467
## 10  Setembro Control_Mean 154.0294
## 11    Agosto Drought_Mean 150.9673
## 12  Dezembro Drought_Mean 169.2248
## 13 Fevereiro Drought_Mean 160.4568
```

```
## 14   Janeiro Drought_Mean 149.7855
## 15     Julho Drought_Mean 166.3492
## 16      Maio Drought_Mean 164.2443
## 17     Marco Drought_Mean 166.0910
## 18  Novembro Drought_Mean 160.9453
## 19   Outubro Drought_Mean 172.8402
## 20  Setembro Drought_Mean 187.6169

# Define the desired order of months
month_order <- c("Outubro", "Novembro", "Dezembro", "Janeiro", "Fevereiro"
, "Marco", "Maio", "Julho", "Agosto", "Setembro")

# Convert "Month" to a factor with ordered levels
result_long$Month <- factor(result_long$Month, levels = month_order)

# Plotting
ggplot(result_long, aes(x = Month, y = Mean, color = Treatment, group = Tr
eatment)) +
  geom_line() +
  geom_point() +
  labs(x = "Month", y = "Mean Values", color = "Treatment") +
  scale_color_manual(values = c("black", "black")) +
  theme_minimal()
```




```
# Create an empty plot object
p <- ggplot() + labs(x = "Month", y = "MPD.OBS", color = "Treatment")


plot_data <- function(df, plot) {
  plot <- plot +
    geom_line(data = df, aes(x = factor(month, levels = c("Outubro","Novembro","Dezembro","Janeiro","Fevereiro","Marco","Maio", "Julho","Agosto","Setembro")), y = mpd.obs, color = treatment, group = treatment)) +
    geom_point(data = df, aes(x = factor(month, levels = c("Outubro","Novembro","Dezembro","Janeiro","Fevereiro","Marco","Maio", "Julho","Agosto","Setembro")), y = mpd.obs, color = ifelse(treatment == "Control" & mpd.obs.p < 0.025,"black",



ifelse(treatment == "Control" & mpd.obs.p >= 0.025, "green",
```

```
ifelse(treatment == "Drought" & mpd.obs.p < 0.025, "blue", "orange")


)
    )
    ), size = 2, shape = 16, position = position_jitter(width = 0.2, heigh
t = 0)) +
    labs(x = "Month", y = "MPD.OBS") +
    theme(legend.position = "bottom")
  return(plot)
}



# Loop through each dataframe in the list and add the data to the plot obj
ect
for (i in 1:length(ses.mpd_list)) {
  p <- plot_data(ses.mpd_list[[i]], p)
}

# Add the lines representing the means from the separate dataframe
p <- p +
  geom_line(data = result_long, aes(x = Month, y = Mean, color = Treatment
, group = Treatment), size = 1, color="black")

# Display the final plot
print(p)
```




```
############################



#########################

#    MPD.AB

# Create empty lists to store the mean values
control_means <- list()
drought_means <- list()

# Iterate over each month
for (month in c("Outubro", "Novembro", "Dezembro", "Janeiro", "Fevereiro",
"Marco", "Maio", "Julho", "Agosto", "Setembro")) {
  control_values <- numeric()
```

```
  drought_values <- numeric()

  # Iterate over each element in ses.mntd_list
  for (i in 1:length(ses.mpd_ab_list)) {
    # Extract the corresponding dataframe for the current element
    df <- ses.mpd_ab_list[[i]]

    # Filter the dataframe for the current month and treatment
    month_filter <- df$month == month
    control_filter <- df$treatment == "Control"
    drought_filter <- df$treatment == "Drought"

    # Extract the mntd.obs values for the current month and treatment
    control_mpd_obs <- df$mpd.obs[month_filter & control_filter]
    drought_mpd_obs <- df$mpd.obs[month_filter & drought_filter]

    # Append the values to the respective lists
    control_values <- c(control_values, control_mpd_obs)
    drought_values <- c(drought_values, drought_mpd_obs)
  }

  # Calculate the mean for each treatment and store the results in the res
pective lists
  control_means[[month]] <- mean(control_values)
  drought_means[[month]] <- mean(drought_values)
}
```

```r
# Print the mean values for each month in control treatment
cat("Mean values in control treatment:\n")

## Mean values in control treatment:

for (month in names(control_means)) {
  cat(month, ": ", control_means[[month]], "\n")
}

## Outubro :  138.4728

## Novembro :  140.1168

## Dezembro :  142.0264

## Janeiro :  146.2357

## Fevereiro :  151.5151

## Marco :  151.4258

## Maio :  154.2337

## Julho :  163.4079

## Agosto :  149.8624

## Setembro :  141.2214

# Print the mean values for each month in drought treatment
cat("\nMean values in drought treatment:\n")

##

## Mean values in drought treatment:

for (month in names(drought_means)) {
  cat(month, ": ", drought_means[[month]], "\n")
}
```

```
## Outubro :  149.5787
## Novembro :  155.0806
## Dezembro :  163.3339
## Janeiro :  128.3119
## Fevereiro :  147.9651
## Marco :  143.1624
## Maio :  154.5266
## Julho :  164.2301
## Agosto :  154.415
## Setembro :  173.4446
```

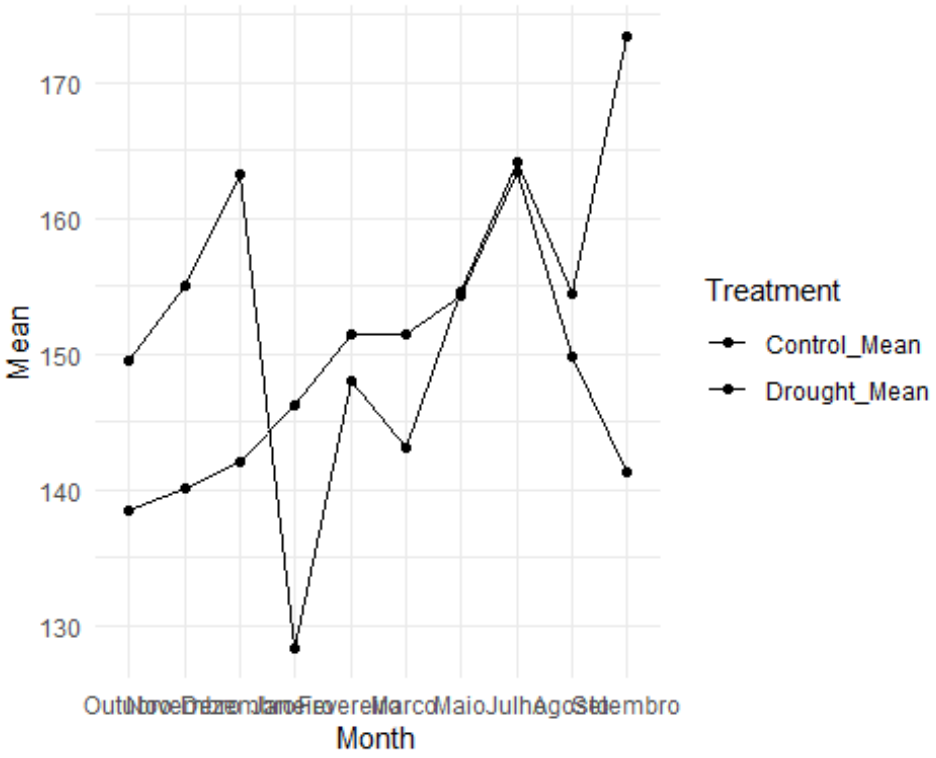
```r
# Create a data frame for control means
control_df <- data.frame(
  Month = names(control_means),
  Control_Mean = unlist(control_means)
)

# Create a data frame for drought means
drought_df <- data.frame(
  Month = names(drought_means),
  Drought_Mean = unlist(drought_means)
)

# Merge the control and drought data frames
result_df <- merge(control_df, drought_df, by = "Month")

# Print the resulting dataframe
print(result_df)
```

```
##        Month Control_Mean Drought_Mean
## 1     Agosto     149.8624     154.4150
## 2   Dezembro     142.0264     163.3339
## 3  Fevereiro     151.5151     147.9651
## 4    Janeiro     146.2357     128.3119
## 5      Julho     163.4079     164.2301
## 6       Maio     154.2337     154.5266
## 7      Marco     151.4258     143.1624
## 8   Novembro     140.1168     155.0806
## 9    Outubro     138.4728     149.5787
## 10  Setembro     141.2214     173.4446

# Create a data frame for control means
control_df <- data.frame(
  Month = names(control_means),
  Control_Mean = unlist(control_means)
)


# Create a data frame for drought means
drought_df <- data.frame(
  Month = names(drought_means),
  Drought_Mean = unlist(drought_means)
)


# Merge the control and drought data frames
result_df <- merge(control_df, drought_df, by = "Month")
write.table(result_df, file = "result_mpd.ab.csv", sep = ",", row.names =
FALSE)
```

```r
# Reshape the data into long format
result_long <- result_df %>%
  gather(key = "Treatment", value = "Mean", -Month)

# Print the resulting long-format dataframe
print(result_long)
```

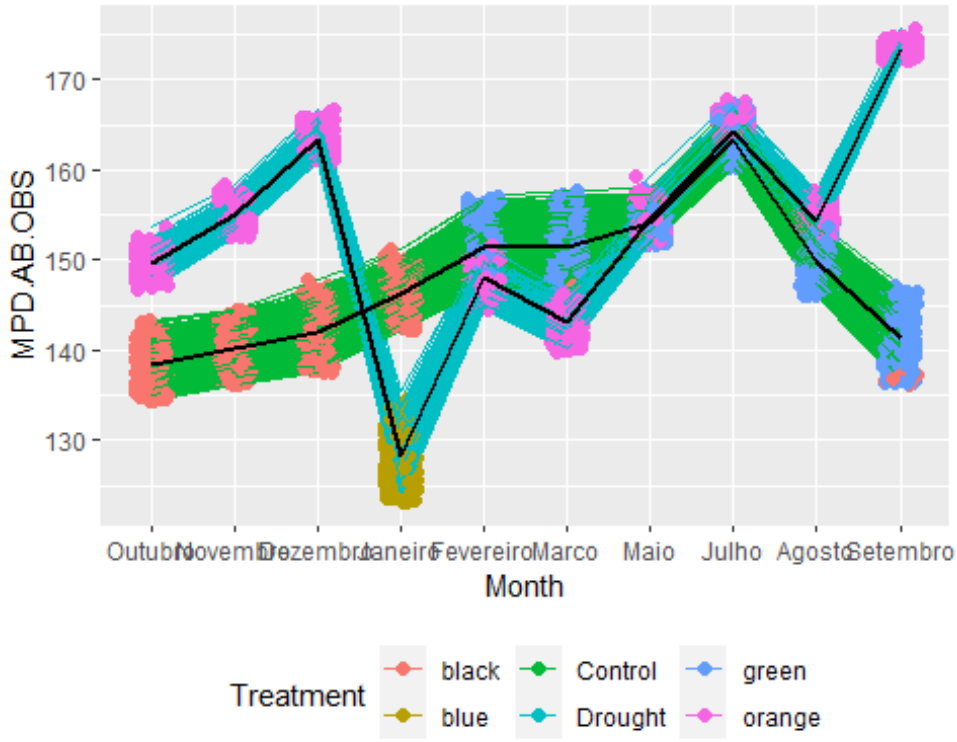
```
##        Month    Treatment     Mean
## 1     Agosto Control_Mean 149.8624
## 2   Dezembro Control_Mean 142.0264
## 3  Fevereiro Control_Mean 151.5151
## 4    Janeiro Control_Mean 146.2357
## 5      Julho Control_Mean 163.4079
## 6       Maio Control_Mean 154.2337
## 7      Marco Control_Mean 151.4258
## 8   Novembro Control_Mean 140.1168
## 9    Outubro Control_Mean 138.4728
## 10  Setembro Control_Mean 141.2214
## 11    Agosto Drought_Mean 154.4150
## 12  Dezembro Drought_Mean 163.3339

## 13 Fevereiro Drought_Mean 147.9651
## 14   Janeiro Drought_Mean 128.3119
## 15     Julho Drought_Mean 164.2301
## 16      Maio Drought_Mean 154.5266
## 17     Marco Drought_Mean 143.1624
```

```
## 18  Novembro Drought_Mean 155.0806
## 19   Outubro Drought_Mean 149.5787
## 20  Setembro Drought_Mean 173.4446

# Define the desired order of months
month_order <- c("Outubro", "Novembro", "Dezembro", "Janeiro", "Fevereiro"
, "Marco", "Maio", "Julho", "Agosto", "Setembro")


# Convert "Month" to a factor with ordered levels
result_long$Month <- factor(result_long$Month, levels = month_order)


# Plotting
ggplot(result_long, aes(x = Month, y = Mean, color = Treatment, group = Tr
eatment)) +
  geom_line() +
  geom_point() +
  scale_color_manual(values = c("black", "black")) +
  theme_minimal()
```




```
# Create an empty plot object
p <- ggplot() + labs(x = "Month", y = "MPD.AB.OBS", color = "Treatment")


plot_data <- function(df, plot) {
  plot <- plot +
    geom_line(data = df, aes(x = factor(month, levels = c("Outubro","Novembro","Dezembro","Janeiro","Fevereiro","Marco","Maio", "Julho","Agosto","Setembro")), y = mpd.obs, color = treatment, group = treatment)) +
    geom_point(data = df, aes(x = factor(month, levels = c("Outubro","Novembro","Dezembro","Janeiro","Fevereiro","Marco","Maio", "Julho","Agosto","Setembro")), y = mpd.obs, color = ifelse(treatment == "Control" & mpd.obs.p < 0.025, "black",



ifelse(treatment == "Control" & mpd.obs.p >= 0.025, "green",
```

```
ifelse(treatment == "Drought" & mpd.obs.p < 0.025, "blue", "orange")


)
    )
    ), size = 2, shape = 16, position = position_jitter(width = 0.2, heigh
t = 0)) +
    labs(x = "Month", y = "MPD.AB.OBS") +
    theme(legend.position = "bottom")
  return(plot)
}

# Loop through each dataframe in the list and add the data to the plot obj
ect
for (i in 1:length(ses.mpd_ab_list)) {
  p <- plot_data(ses.mpd_ab_list[[i]], p)
}

# Add the lines representing the means from the separate dataframe
p <- p +
  geom_line(data = result_long, aes(x = Month, y = Mean, color = Treatment
, group = Treatment), size = 1, color="black")

# Display the final plot
print(p)
```




```
############################


#########################


#    MNTD


# Create empty lists to store the mean values
control_means <- list()
drought_means <- list()


# Iterate over each month
for (month in c("Outubro", "Novembro", "Dezembro", "Janeiro", "Fevereiro",
"Marco", "Maio", "Julho", "Agosto", "Setembro")) {
  control_values <- numeric()
```

```
  drought_values <- numeric()

  # Iterate over each element in ses.mntd_list
  for (i in 1:length(ses.mntd_list)) {
    # Extract the corresponding dataframe for the current element
    df <- ses.mntd_list[[i]]

    # Filter the dataframe for the current month and treatment
    month_filter <- df$month == month
    control_filter <- df$treatment == "Control"
    drought_filter <- df$treatment == "Drought"

    # Extract the mntd.obs values for the current month and treatment
    control_mntd_obs <- df$mntd.obs[month_filter & control_filter]
    drought_mntd_obs <- df$mntd.obs[month_filter & drought_filter]

    # Append the values to the respective lists
    control_values <- c(control_values, control_mntd_obs)
    drought_values <- c(drought_values, drought_mntd_obs)
  }

  # Calculate the mean for each treatment and store the results in the res
pective lists
  control_means[[month]] <- mean(control_values)
  drought_means[[month]] <- mean(drought_values)
}
```

```r
# Print the mean values for each month in control treatment
cat("Mean values in control treatment:\n")

## Mean values in control treatment:

for (month in names(control_means)) {
  cat(month, ": ", control_means[[month]], "\n")
}

## Outubro :  32.60426
## Novembro :  35.66211
## Dezembro :  37.20796
## Janeiro :  36.64151
## Fevereiro :  37.9714
## Marco :  36.61483
## Maio :  31.623
## Julho :  46.05422
## Agosto :  48.2814
## Setembro :  44.56775

# Print the mean values for each month in drought treatment
cat("\nMean values in drought treatment:\n")

##
## Mean values in drought treatment:

for (month in names(drought_means)) {
  cat(month, ": ", drought_means[[month]], "\n")
}
```

```
## Outubro :  36.93275
## Novembro :  41.40491
## Dezembro :  40.8492
## Janeiro :  39.07137
## Fevereiro :  48.34089
## Marco :  53.85924
## Maio :  53.41318
## Julho :  53.94365
## Agosto :  45.41083
## Setembro :  56.89096
```

```r
# Create a data frame for control means
control_df <- data.frame(
  Month = names(control_means),
  Control_Mean = unlist(control_means)
)

# Create a data frame for drought means
drought_df <- data.frame(
  Month = names(drought_means),
  Drought_Mean = unlist(drought_means)
)

# Merge the control and drought data frames
result_df <- merge(control_df, drought_df, by = "Month")
```

```r
# Print the resulting dataframe
print(result_df)
```

```
##        Month Control_Mean Drought_Mean
## 1     Agosto     48.28140     45.41083
## 2   Dezembro     37.20796     40.84920
## 3  Fevereiro     37.97140     48.34089
## 4    Janeiro     36.64151     39.07137
## 5      Julho     46.05422     53.94365
## 6       Maio     31.62300     53.41318
## 7      Marco     36.61483     53.85924
## 8   Novembro     35.66211     41.40491
## 9    Outubro     32.60426     36.93275
## 10  Setembro     44.56775     56.89096
```

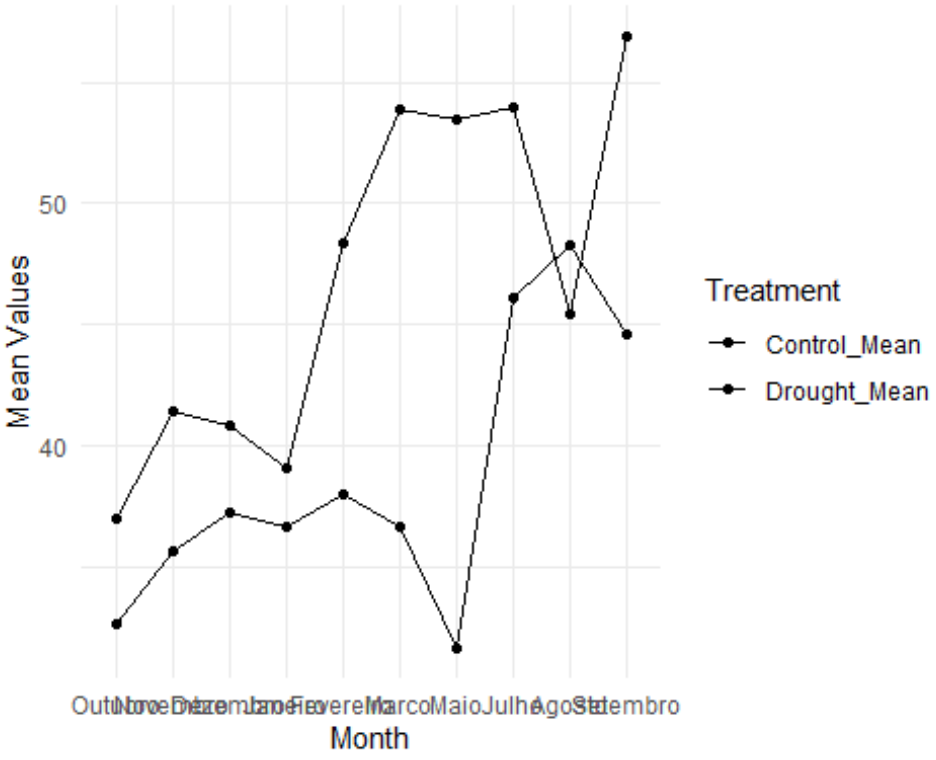
```r
library(tidyverse)

# Create a data frame for control means
control_df <- data.frame(
  Month = names(control_means),
  Control_Mean = unlist(control_means)
)

# Create a data frame for drought means
drought_df <- data.frame(
  Month = names(drought_means),
  Drought_Mean = unlist(drought_means)
```

```
)

# Merge the control and drought data frames
result_df <- merge(control_df, drought_df, by = "Month")
write.table(result_df, file = "result_mntd.csv", sep = ",", row.names = FA
LSE)

# Reshape the data into long format
result_long <- result_df %>%
  gather(key = "Treatment", value = "Mean", -Month)

# Print the resulting long-format dataframe
print(result_long)
```

```
##        Month    Treatment     Mean
## 1     Agosto Control_Mean 48.28140
## 2   Dezembro Control_Mean 37.20796
## 3  Fevereiro Control_Mean 37.97140
## 4    Janeiro Control_Mean 36.64151
## 5      Julho Control_Mean 46.05422
## 6       Maio Control_Mean 31.62300
## 7      Marco Control_Mean 36.61483
## 8   Novembro Control_Mean 35.66211
## 9    Outubro Control_Mean 32.60426
## 10  Setembro Control_Mean 44.56775
## 11    Agosto Drought_Mean 45.41083
```

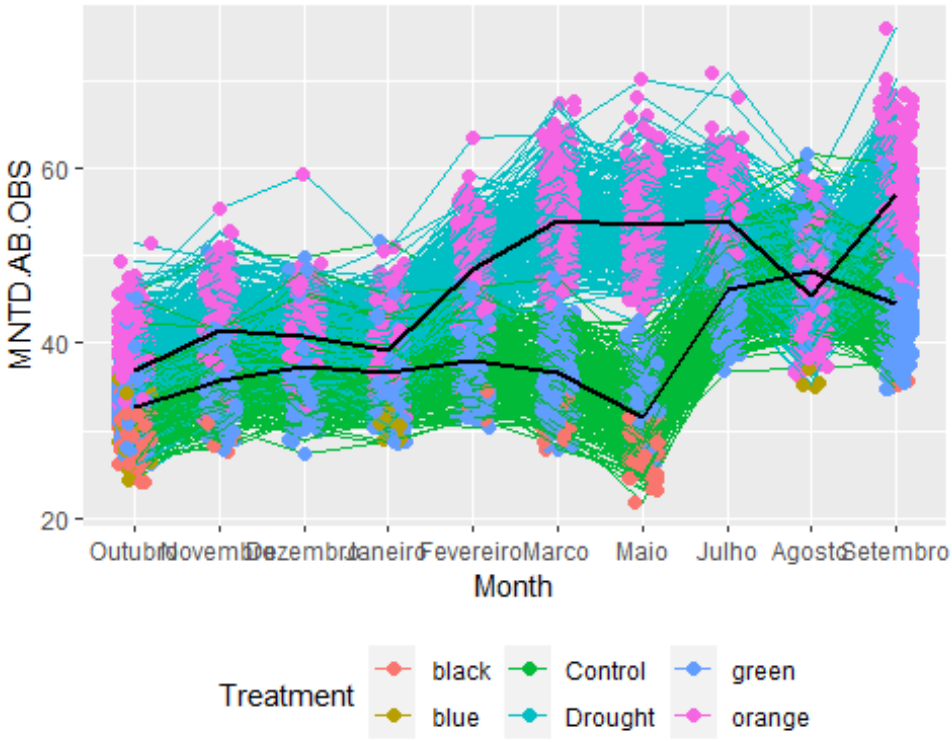
```
## 12  Dezembro Drought_Mean 40.84920
## 13 Fevereiro Drought_Mean 48.34089
## 14   Janeiro Drought_Mean 39.07137
## 15     Julho Drought_Mean 53.94365
## 16      Maio Drought_Mean 53.41318
## 17     Marco Drought_Mean 53.85924
## 18  Novembro Drought_Mean 41.40491
## 19   Outubro Drought_Mean 36.93275
## 20  Setembro Drought_Mean 56.89096
```

```r
# Define the desired order of months
month_order <- c("Outubro", "Novembro", "Dezembro", "Janeiro", "Fevereiro"
, "Marco", "Maio", "Julho", "Agosto", "Setembro")

# Convert "Month" to a factor with ordered levels
result_long$Month <- factor(result_long$Month, levels = month_order)

# Plotting
ggplot(result_long, aes(x = Month, y = Mean, color = Treatment, group = Tr
eatment)) +
  geom_line() +
  geom_point() +
  labs(x = "Month", y = "Mean Values", color = "Treatment") +
  scale_color_manual(values = c("black", "black")) +
  theme_minimal()
```




```
# Create an empty plot object
p <- ggplot() + labs(x = "Month", y = "MNTD.OBS", color = "Treatment")


plot_data <- function(df, plot) {
  plot <- plot +
    geom_line(data = df, aes(x = factor(month, levels = c("Outubro","Novembro","Dezembro","Janeiro","Fevereiro","Marco","Maio", "Julho","Agosto","Setembro")), y = mntd.obs, color = treatment, group = treatment)) +
    geom_point(data = df, aes(x = factor(month, levels = c("Outubro","Novembro","Dezembro","Janeiro","Fevereiro","Marco","Maio", "Julho","Agosto","Setembro")), y = mntd.obs, color = ifelse(treatment == "Control" & mntd.obs.p < 0.025, "black",



ifelse(treatment == "Control" & mntd.obs.p >= 0.025, "green",
```

```
ifelse(treatment == "Drought" & mntd.obs.p < 0.025, "blue", "orange")


)
    )
    ), size = 2, shape = 16, position = position_jitter(width = 0.2, heigh
t = 0)) +
    labs(x = "Month", y = "MNTD.AB.OBS") +
    theme(legend.position = "bottom")
  return(plot)
}

# Loop through each dataframe in the list and add the data to the plot obj
ect
for (i in 1:length(ses.mntd_list)) {
  p <- plot_data(ses.mntd_list[[i]], p)
}

# Add the lines representing the means from the separate dataframe
p <- p +
  geom_line(data = result_long, aes(x = Month, y = Mean, color = Treatment
, group = Treatment), size = 1, color="black")

# Display the final plot
print(p)
```




```
############################



##########################

#    MNTD.AB

# Create empty lists to store the mean values
control_means <- list()
drought_means <- list()



# Iterate over each month
for (month in c("Outubro", "Novembro", "Dezembro", "Janeiro", "Fevereiro",
"Marco", "Maio", "Julho", "Agosto", "Setembro")) {
```

```
  control_values <- numeric()
  drought_values <- numeric()

  # Iterate over each element in ses.mntd_list
  for (i in 1:length(ses.mntd_ab_list)) {
    # Extract the corresponding dataframe for the current element
    df <- ses.mntd_ab_list[[i]]

    # Filter the dataframe for the current month and treatment
    month_filter <- df$month == month
    control_filter <- df$treatment == "Control"
    drought_filter <- df$treatment == "Drought"

    # Extract the mntd.obs values for the current month and treatment
    control_mntd_obs <- df$mntd.obs[month_filter & control_filter]
    drought_mntd_obs <- df$mntd.obs[month_filter & drought_filter]

    # Append the values to the respective lists
    control_values <- c(control_values, control_mntd_obs)
    drought_values <- c(drought_values, drought_mntd_obs)
  }

  # Calculate the mean for each treatment and store the results in the res
pective lists
  control_means[[month]] <- mean(control_values)
  drought_means[[month]] <- mean(drought_values)
}
```

```r
# Print the mean values for each month in control treatment
cat("Mean values in control treatment:\n")

## Mean values in control treatment:

for (month in names(control_means)) {
  cat(month, ": ", control_means[[month]], "\n")
}

## Outubro :  25.57698

## Novembro :  38.61012

## Dezembro :  27.9044

## Janeiro :  27.98786

## Fevereiro :  35.87972

## Marco :  32.39776

## Maio :  32.10551

## Julho :  37.13733

## Agosto :  51.47304

## Setembro :  47.57772

# Print the mean values for each month in drought treatment
cat("\nMean values in drought treatment:\n")

##

## Mean values in drought treatment:
```

```r
for (month in names(drought_means)) {
  cat(month, ": ", drought_means[[month]], "\n")
}
```

```
## Outubro :  30.09473
## Novembro :  36.94409
## Dezembro :  39.82929
## Janeiro :  29.535
## Fevereiro :  43.67623
## Marco :  41.58784
## Maio :  43.50412
## Julho :  47.19179
## Agosto :  49.28111
## Setembro :  51.29785
```

```r
# Create a data frame for control means
control_df <- data.frame(
  Month = names(control_means),
  Control_Mean = unlist(control_means)
)


# Create a data frame for drought means
drought_df <- data.frame(
  Month = names(drought_means),
  Drought_Mean = unlist(drought_means)
)
# Merge the control and drought data frames
result_df <- merge(control_df, drought_df, by = "Month")
```

```r
# Print the resulting dataframe
print(result_df)
```

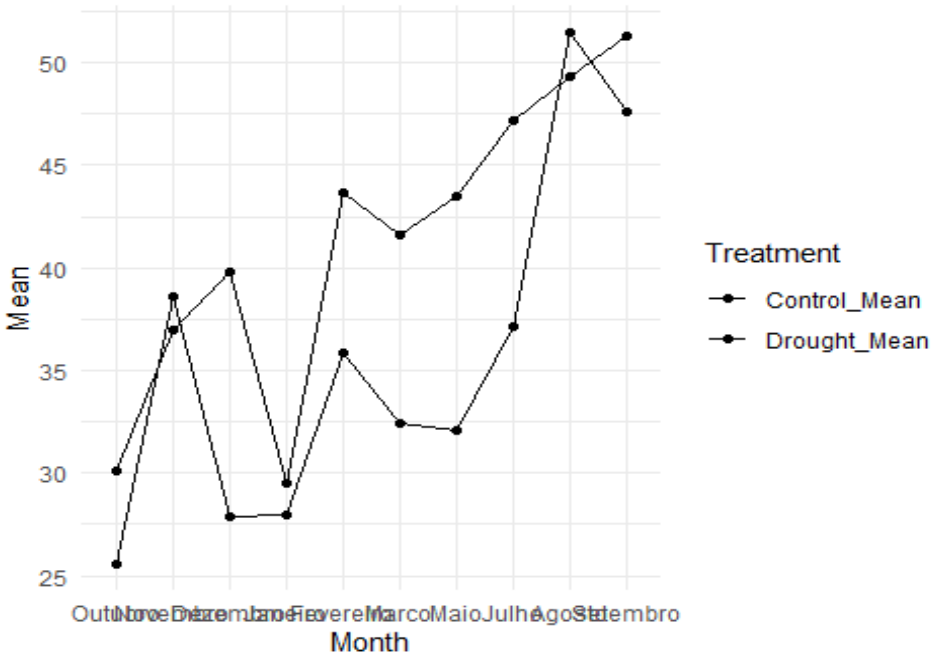
```
##        Month Control_Mean Drought_Mean
## 1     Agosto     51.47304     49.28111
## 2   Dezembro     27.90440     39.82929
## 3  Fevereiro     35.87972     43.67623
## 4    Janeiro     27.98786     29.53500
## 5      Julho     37.13733     47.19179
## 6       Maio     32.10551     43.50412
## 7      Marco     32.39776     41.58784
## 8   Novembro     38.61012     36.94409
## 9    Outubro     25.57698     30.09473
## 10  Setembro     47.57772     51.29785
```

```r
library(tidyverse)
# Create a data frame for control means
control_df <- data.frame(
  Month = names(control_means),
  Control_Mean = unlist(control_means)
)
# Create a data frame for drought means
drought_df <- data.frame(
  Month = names(drought_means),
  Drought_Mean = unlist(drought_means)
)
# Merge the control and drought data frames
```

```r
result_df <- merge(control_df, drought_df, by = "Month")
write.table(result_df, file = "result_mntd.ab.csv", sep = ",", row.names =
FALSE)
result_long <- result_df %>%
  gather(key = "Treatment", value = "Mean", -Month)

# Print the resulting long-format dataframe
print(result_long)
```

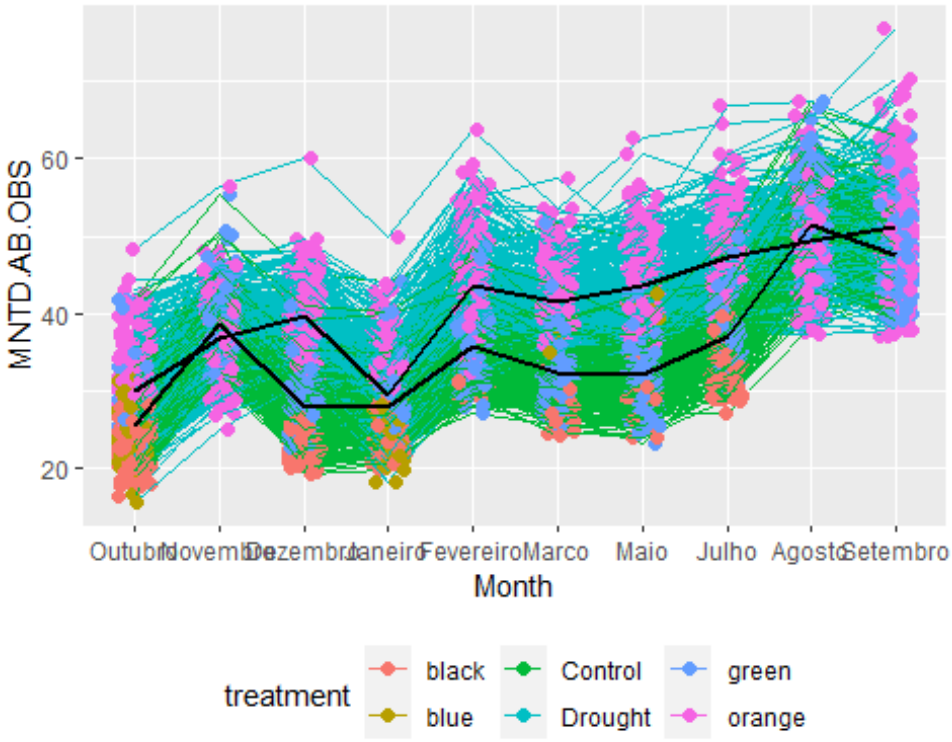
```
##        Month    Treatment     Mean
## 1     Agosto Control_Mean 51.47304
## 2   Dezembro Control_Mean 27.90440
## 3  Fevereiro Control_Mean 35.87972
## 4    Janeiro Control_Mean 27.98786
## 5      Julho Control_Mean 37.13733
## 6       Maio Control_Mean 32.10551
## 7      Marco Control_Mean 32.39776
## 8   Novembro Control_Mean 38.61012
## 9    Outubro Control_Mean 25.57698
## 10  Setembro Control_Mean 47.57772
## 11    Agosto Drought_Mean 49.28111
## 12  Dezembro Drought_Mean 39.82929
## 13 Fevereiro Drought_Mean 43.67623
## 14   Janeiro Drought_Mean 29.53500
## 15     Julho Drought_Mean 47.19179
## 16      Maio Drought_Mean 43.50412
```

```
## 17     Marco Drought_Mean 41.58784
## 18  Novembro Drought_Mean 36.94409
## 19   Outubro Drought_Mean 30.09473
## 20  Setembro Drought_Mean 51.29785

# Define the desired order of months
month_order <- c("Outubro", "Novembro", "Dezembro", "Janeiro", "Fevereiro"
, "Marco", "Maio", "Julho", "Agosto", "Setembro")


# Convert "Month" to a factor with ordered levels
result_long$Month <- factor(result_long$Month, levels = month_order)


# Plotting
ggplot(result_long, aes(x = Month, y = Mean, color = Treatment, group = Tr
eatment)) +
  geom_line() +
  geom_point() +
  scale_color_manual(values = c("black", "black")) +
  theme_minimal()
```




```
# Create an empty plot object
p <- ggplot()

plot_data <- function(df, plot) {
  plot <- plot +
    geom_line(data = df, aes(x = factor(month, levels = c("Outubro","Novembro","Dezembro","Janeiro","Fevereiro","Marco","Maio", "Julho","Agosto","Setembro")), y = mntd.obs, color = treatment, group = treatment)) +
    geom_point(data = df, aes(x = factor(month, levels = c("Outubro","Novembro","Dezembro","Janeiro","Fevereiro","Marco","Maio", "Julho","Agosto","Setembro")), y = mntd.obs, color = ifelse(treatment == "Control" & mntd.obs.p < 0.025, "black",

ifelse(treatment == "Control" & mntd.obs.p >= 0.025, "green",

ifelse(treatment == "Drought" & mntd.obs.p < 0.025, "blue", "orange")
```

```
)
    )
    ), size = 2, shape = 16, position = position_jitter(width = 0.2, heigh
t = 0)) +
    labs(x = "Month", y = "MNTD.AB.OBS") +
    theme(legend.position = "bottom")
  return(plot)
}

# Loop through each dataframe in the list and add the data to the plot obj
ect
for (i in 1:length(ses.mntd_ab_list)) {
  p <- plot_data(ses.mntd_ab_list[[i]], p)
}

# Add the lines representing the means from the separate dataframe
p <- p +
  geom_line(data = result_long, aes(x = Month, y = Mean, color = Treatment
, group = Treatment), size = 1, color="black")

# Display the final plot
print(p)
```




```
############################


#HEATMAP


library(gplots)

## Warning: package 'gplots' was built under R version 4.3.3

## Registered S3 method overwritten by 'gplots':
##   method    from
##   plot.venn eulerr
##
## Attaching package: 'gplots'
##
## The following object is masked from 'package:stats':
```

```
##
##     lowess

tree_list <- readRDS(file = "tree_list.rds")
subparcelas.centro4<-read.csv("subparcelas.centro4.csv",sep=",",header=T,row.names=NULL)
month_order <- c("outubro","novembro","dezembro","janeiro","fevereiro","marco","maio", "julho","agosto","setembro")
subparcelas.controle<-read.csv("subparcelas.controle.csv",sep=";",header=T)
subparcelas.controle$month <- factor(subparcelas.controle$month, levels = month_order)
subparcelas.seca<-read.csv("subparcelas.seca.csv",sep=";",header=T)
subparcelas.seca$month <- factor(subparcelas.seca$month, levels = month_order)



# Calculate the difference (assuming the dataframes are of the same shape and aligned)
subparcelas.controle <- subparcelas.controle[,-(1:3)]
subparcelas.seca <- subparcelas.seca[,-(1:3)]
diff <- subparcelas.seca - subparcelas.controle


tree2<-read.tree("tree1.newick")
tree2$tip.label

##   [1] "Crematogastercrinosa"        "Crematogasterflavosensitiva"
##   [3] "Crematogasterprlongispina"   "Crematogastersotobosque"
```

```
##   [5] "CrematogasterMPEG01"         "CrematogasterMPEG02"
##   [7] "CrematogasterMPEG04"         "Crematogasterbrasiliensis"
##   [9] "CarebaraMPEG02"              "CarebaraMPEG03"
##  [11] "CarebaraMPEG01"              "Strumigenyselongata"
##  [13] "StrumigenysMPEG01"           "StrumigenysMPEG02"
##  [15] "Strumigenysperparva"         "Strumigenyszeteki"
##  [17] "Strumigenysdenticulata"      "Octostrumabetschi"
##  [19] "Octostrumaiheringi"          "Basicerosmilitaris"
##  [21] "RhopalothrixMPEG01"          "CephalotesMPEG03"
##  [23] "CephalotesMPEG01"            "Cephalotesatratus"
##  [25] "PheidolegrDiligensMPEG01"    "PheidolegrDiligensMPEG02"
##  [27] "PheidolegrDiligensMPEG03"    "PheidolegrDiligensMPEG05"
##  [29] "PheidolegrDiligensMPEG06"    "PheidolegrFallaxMPEG01"
##  [31] "PheidolegrFallaxMPEG02"      "PheidolegrFallaxMPEG03"
##  [33] "PheidolegrFlavensMPEG01"     "PheidolegrTristisMPEG01"
##  [35] "PheidolegrTristisMPEG02"     "Pheidolebiconstricta"
##  [37] "Pheidolebruesi"              "Pheidolecfaraneoides"
##  [39] "Pheidoledolon"               "Pheidolefimbriata"
##  [41] "Pheidolefowleri"             "Pheidolefracticeps"
##  [43] "Pheidolejeannei"             "PheidoleMPEG02"
##  [45] "PheidoleMPEG03"              "PheidoleMPEG04"
##  [47] "PheidoleMPEG08"              "PheidoleMPEG09"
##  [49] "PheidoleMPEG10"              "PheidoleMPEG11"
##  [51] "PheidoleMPEG12"              "PheidoleMPEG13"
##  [53] "PheidoleMPEG15"              "Pheidolescolioceps"
##  [55] "Pheidolesubarmata"           "Pheidolesynarmata"
##  [57] "Pheidolevorax"               "PheidoleMPEG01"
```

```
##  [59] "WasmanniaMPEG01"              "Wasmanniaauropunctata"
##  [61] "Blepharidattabrasiliensis"    "ApterostigmaMPEG01"
##  [63] "MyrmicocryptaMPEG02"          "MyrmicocryptaMPEG04"
##  [65] "MyrmicocryptaMPEG01"          "Ochetomyrmexsemipolitus"
##  [67] "Ochetomyrmexneopolitus"       "Cyphomyrmexprminutus"
##  [69] "Cyphomyrmexpeltatus"          "Cyphomyrmexlaevigatus"
##  [71] "AttaMPEG01"                   "AcromyrmexMPEG02"
##  [73] "AcromyrmexMPEG01"             "TrachymyrmexMPEG02"
##  [75] "TrachymyrmexMPEG04"           "TrachymyrmexMPEG05"
##  [77] "TrachymyrmexMPEG01"           "SericomyrmexMPEG02"
##  [79] "SericomyrmexMPEG01"           "Dacetonarmigerum"
##  [81] "Monomoriumcffloricola"        "MegalomyrmexMPEG03"
##  [83] "MegalomyrmexMPEG04"           "MegalomyrmexMPEG01"
##  [85] "SolenopsisMPEG02"             "SolenopsisMPEG03"
##  [87] "SolenopsisMPEG04"             "SolenopsisMPEG05"
##  [89] "SolenopsisMPEG06"             "SolenopsisMPEG07"
##  [91] "SolenopsisMPEG01"             "RogeriaMPEG02"
##  [93] "RogeriaMPEG04"                "RogeriaMPEG01"
##  [95] "HylomyrmaMPEG03"              "Gnamptogenysconcinna"
##  [97] "Gnamptogenyshorni"            "Gnamptogenyssulcata"
##  [99] "Gnamptogenystortuolosa"       "GnamptogenysMPEG02"
## [101] "Ectatommalugens"              "Ectatommaedentatum"
## [103] "Camponotusprmelatoticus"      "CamponotusMPEG01"
## [105] "CamponotusMPEG02"             "CamponotusMPEG05"
## [107] "Camponotusfemoratus"          "NylanderiaMPEG02"
## [109] "NylanderiaMPEG03"             "NylanderiaMPEG04"
## [111] "NylanderiaMPEG05"             "NylanderiaMPEG08"
```

```
## [113] "NylanderiaMPEG01"            "BrachymyrmexMPEG02"
## [115] "BrachymyrmexMPEG01"          "PseudomyrmexMPEG02"
## [117] "PseudomyrmexMPEG04"          "PseudomyrmexMPEG06"
## [119] "PseudomyrmexMPEG01"          "AztecaMPEG02"
## [121] "AztecaMPEG03"                "AztecaMPEG04"
## [123] "AztecaMPEG05"                "AztecaMPEG01"
## [125] "Dolichoderusbispinosus"      "Dolichoderusdecollatus"
## [127] "Dolichoderuslutosus"         "DolichoderusMPEG01"
## [129] "DolichoderusMPEG03"          "Dolichoderusattelaboides"
## [131] "Nomamyrmexhartigii"          "Nomamyrmexesenbeckii"
## [133] "EcitonMPEG02"                "EcitonMPEG03"
## [135] "Ecitonrapax"                 "EcitonMPEG01"
## [137] "Labiduspraedator"            "Labidusspininodis"
## [139] "Labiduscoecus"               "NeivamyrmexMPEG02"
## [141] "NeivamyrmexMPEG03"           "NeivamyrmexMPEG04"
## [143] "NeivamyrmexMPEG06"           "Neivamyrmexpilosus"
## [145] "Neivamyrmexpseudops"         "NeivamyrmexMPEG01"
## [147] "Neoponeracfcommutata"        "NeoponeraMPEG01"
## [149] "Neoponeraverenae"            "Neoponeraapicalis"
## [151] "Pachycondylacrassinoda"      "Rasoponeprpergandei"
## [153] "Mayaponeraconstricta"        "AnochetusMPEG02"
## [155] "AnochetusMPEG01"             "OdontomachusMPEG02"
## [157] "OdontomachusMPEG03"          "OdontomachusMPEG04"
## [159] "OdontomachusMPEG01"          "Centromyrmexbrachycola"
## [161] "HypoponeraMPEG02"            "HypoponeraMPEG01"
```

```
tree2$tip.label [1:5]
```

```
## [1] "Crematogastercrinosa"         "Crematogasterflavosensitiva"
## [3] "Crematogasterprlongispina"    "Crematogastersotobosque"
## [5] "CrematogasterMPEG01"

colnames(diff) [1:5]

## [1] "Basicerosmilitaris"         "RhopalothrixMPEG01"
## [3] "Blepharidattabrasiliensis" "ApterostigmaMPEG01"
## [5] "AttaMPEG01"

diff2<-as.data.frame(t(diff))
matriz_presenca_ausencia <- ifelse(diff2 > 0, 1, ifelse(diff2 < 0, -1, 0))


plot_tree <- ggtree(tree_list[[2]]) +
  geom_tiplab(size=1)



plot_heatmap <- ggtree::gheatmap(plot_tree, matriz_presenca_ausencia, offs
et=8, width=0.4, hjust=0.5, font.size=1.6)


# Here, 'value' should be replaced by the name of the column in diff2 that
is being plotted.
plot_heatmap <- plot_heatmap + scale_fill_gradient2(low = "red", mid = "wh
ite", high = "green", midpoint = 0)

## Scale for fill is already present.
## Adding another scale for fill, which will replace the existing scale.
```