

# **1. Análise Estatística de Propriedades Físicas do Solo em Diferentes Níveis de Impacto em decorrência de barragem em Brumadinho**

---

Autor: Diogo Paes da Costa

Data: 27/06/2025

## **2. Introdução**

---

O rompimento da barragem de rejeitos em Brumadinho (2019) provocou uma das maiores tragédias ambientais do Brasil, liberando milhões de metros cúbicos de lama que impactaram drasticamente o solo na região. Alterações em suas propriedades físicas – como densidade, porosidade e textura – podem comprometer funções ecológicas essenciais, incluindo infiltração de água, armazenamento de nutrientes e sustentação da vegetação. Embora estudos anteriores tenham documentado efeitos gerais de desastres similares, ainda faltam análises estatísticas robustas que quantifiquem como diferentes níveis de impacto modificam essas características. Este estudo busca preencher essa lacuna, investigando as propriedades físicas do solo em áreas diretamente afetadas (AF) e não afetadas (NA) pelo desastre, utilizando métodos estatísticos avançados para identificar padrões e relações causais.

Partindo do pressuposto de que o impacto da barragem alterou significativamente a estrutura do solo, três hipóteses foram testadas. Primeiro, que as áreas AF exibem diferenças mensuráveis em propriedades como densidade e porosidade em comparação com as NA. Segundo, que o nível de impacto explica uma parcela relevante da variabilidade multivariada dessas propriedades. Por fim, que é possível discriminar estatisticamente entre áreas AF e NA com base em um conjunto de variáveis físicas. Essas hipóteses foram avaliadas por meio de uma abordagem metodológica abrangente, combinando análises univariadas e multivariadas.

Para isso, amostras de solo foram coletadas em locais com diferentes níveis de exposição ao desastre, abrangendo variáveis como densidade do solo, porosidade total, distribuição granulométrica (argila, silte, areia) e condutividade hidráulica. Os dados foram analisados no ambiente R, utilizando modelos mistos (MLM) para comparações univariadas, análises de redundância (RDA) e PERMANOVA para avaliar padrões multivariados, e análise discriminante linear (LDA) para classificação dos níveis de impacto. Técnicas complementares,

como PCA e correlações, ajudaram a explorar relações entre variáveis e identificar os fatores mais influentes.

Os resultados confirmaram diferenças significativas entre áreas AF e NA. Nas análises univariadas, densidade do solo, porosidade total e microporosidade apresentaram variações marcantes ( $p < 0,05$ ), indicando maior compactação e redução de vazios nos solos impactados. A RDA explicou 39,8% da variância total, com clara separação entre os grupos ( $p = 0,001$ ), enquanto a PERMANOVA reforçou essa distinção ( $R^2 = 13,2\%$ ;  $p = 1e-04$ ). A LDA demonstrou alta acurácia (83,3%) na discriminação dos grupos, destacando microporosidade e areia grossa como variáveis-chave. Correlações revelaram que áreas AF tendem a ter menos areia grossa ( $r = -0,613$ ) e mais silte ( $r = 0,467$ ), sugerindo alterações na textura devido à deposição de rejeitos finos.

Em conclusão, este estudo evidencia que o desastre em Brumadinho modificou significativamente as propriedades físicas do solo, com implicações para sua recuperação. O aumento da densidade e a redução da porosidade em áreas AF indicam desafios para a revegetação e a restauração de funções hidrológicas. As técnicas multivariadas aplicadas não apenas validaram essas diferenças, mas também oferecem ferramentas úteis para monitoramento futuro. Os resultados reforçam a necessidade de estratégias de manejo adaptadas a solos degradados por atividades minerárias, destacando a importância de abordagens baseadas em evidências para a recuperação ambiental pós-desastre.

### 3. Material e Métodos

---

#### 3.1. Análises estatísticas

As análises estatísticas foram conduzidas no ambiente de programação R (versão 4.4.0) e no ambiente de desenvolvimento integrado RStudio (versão 2023.06.1 Build 524 © 2009-2023 Posit Software, PBC). Diversos pacotes foram empregados para a manipulação, análise e visualização dos dados, conforme detalhado a seguir.

**Preparação dos Dados e Análise Exploratória:** A etapa inicial de preparação dos dados envolveu a importação de arquivos no formato Excel, utilizando o pacote `readxl` (Wickham & Bryan, 2024). As variáveis categóricas, como "LOCAL", "IMPACTO" e "PSEUDO", foram convertidas para o tipo fator para garantir a correta interpretação em análises subsequentes. Realizou-se uma inspeção preliminar da estrutura dos dados (`str()`, `summary()`) e

a verificação de valores ausentes. Para a análise exploratória, foram geradas estatísticas descritivas das propriedades físicas do solo por nível de impacto (Áreas Afetadas - AF e Áreas Não Afetadas - NA). Visualizações iniciais incluíram boxplots para cada variável física, utilizando ggplot2 (Wickham, 2016) e tidyverse (Wickham et al., 2019), e matrizes de correlação entre as variáveis físicas, visualizadas com corrplot (Wei & Simko, 2021) e GGally (Schloerke et al., 2021) para explorar as relações por grupo de impacto.

**Verificação de Pressupostos para Análises Univariadas:** Previamente às análises univariadas, os pressupostos estatísticos de normalidade e homogeneidade de variâncias foram rigorosamente verificados. A normalidade dos dados foi avaliada por meio do teste de Shapiro-Wilk, aplicado separadamente para cada variável dentro de cada grupo de impacto. Adicionalmente, a normalidade dos resíduos de modelos lineares simples foi testada, e gráficos de probabilidade normal (QQ-plots) foram gerados para inspeção visual com qqplotr (Almeida et al., 2024) para visualização. A homogeneidade de variâncias foi verificada utilizando o teste de Levene, implementado através do pacote car (Fox & Weisberg, 2019). Em casos de violação dos pressupostos, foram testadas e aplicadas transformações de dados (e.g., logit para percentuais, log1p, raiz quadrada, inversa, quadrado) para otimizar o atendimento às suposições dos modelos paramétricos.

**Análise Univariada com Modelos Mistos:** Para a análise univariada do efeito do impacto em cada propriedade física do solo, foram empregados Modelos Lineares Mistos (MLM). A remoção de outliers foi realizada de forma controlada, identificando e excluindo os valores mais extremos por variável e grupo. Os MLMs foram ajustados utilizando o pacote lme4 (Bates et al., 2015), com o nível de impacto como fator fixo e o local de amostragem como efeito aleatório. A significância dos efeitos foi avaliada por meio de análises de variância (ANOVA) com o pacote lmerTest (Kuznetsova et al., 2017), que fornece p-valores para os efeitos fixos. Médias ajustadas e comparações múltiplas foram obtidas com o pacote emmeans (Lenth, 2023). Os gráficos de boxplots e barras combinados foram gerados com ggplot2 e patchwork (Pedersen, 2023), indicando a significância estatística.

**Análise de Redundância (RDA):** A Análise de Redundância (RDA) foi utilizada para investigar a relação entre o conjunto de propriedades físicas do solo (variáveis resposta) e os fatores "IMPACTO" e "LOCAL" (variáveis explicativas). As variáveis resposta foram padronizadas antes da análise. A colinearidade entre as variáveis explicativas foi avaliada por meio de matrizes de correlação e Fatores de Inflação da Variância (VIF). A RDA foi executada

com o pacote *vegan* (Oksanen et al., 2022). Testes de significância global, por eixo e por termo foram realizados (*anova.cca*). A partição da variância (*varpart*) foi empregada para quantificar a contribuição única e compartilhada dos fatores "IMPACTO" e "LOCAL" na explicação da variabilidade das propriedades do solo. Os resultados foram visualizados em biplots, utilizando *ggplot2* e *ggvegan* (Oksanen, 2022).

**Análise de Permutação Multivariada (PERMANOVA):** Para avaliar diferenças significativas na composição multivariada das propriedades físicas do solo entre as áreas afetadas e não afetadas, foi realizada uma Análise de Variância Permutacional Multivariada (PERMANOVA). Uma matriz de distância euclidiana foi calculada a partir dos dados das propriedades físicas padronizadas (*vegdist*). A PERMANOVA foi executada utilizando a função *adonis2* do pacote *vegan*, com permutações restritas dentro de cada local de amostragem (*strata*) para controlar a dependência espacial. A homogeneidade de dispersões multivariadas, um pressuposto da PERMANOVA, foi verificada com o teste *betadisper* (*vegan*). A visualização da ordenação multivariada foi feita através de um Escalonamento Multidimensional Não-Métrico (NMDS), calculado com *metaMDS* (*vegan*) e plotado com *ggplot2*.

**Análise Discriminante Linear (LDA):** A Análise Discriminante Linear (LDA) foi aplicada para construir um modelo que maximizasse a separação entre as áreas afetadas e não afetadas com base nas propriedades físicas do solo. Previamente à LDA, a colinearidade entre as variáveis preditoras foi rigorosamente avaliada. Em casos de colinearidade severa, uma Análise de Componentes Principais (PCA) foi utilizada como etapa de pré-processamento para transformar as variáveis originais em um conjunto menor de componentes ortogonais, garantindo a estabilidade do modelo LDA. O modelo LDA foi ajustado utilizando o pacote *MASS* (Venables & Ripley, 2002). A acurácia do modelo foi avaliada por meio de validação cruzada (10-fold cross-validation) com o pacote *caret* (Kuhn, 2023), e a importância das variáveis na discriminação foi determinada pelos coeficientes da função discriminante. Gráficos de scores e histogramas foram gerados com *ggplot2* para visualizar a separação dos grupos.

**Análise de Correlação entre Propriedades Físicas e Nível de Impacto:** A análise de correlação entre as propriedades físicas do solo e o nível de impacto foi realizada utilizando a correlação ponto-bisserial, onde a variável "IMPACTO" foi codificada numericamente (0 para não afetado, 1 para afetado). Além disso, foram construídas matrizes de correlação separadas

para as áreas afetadas e não afetadas, permitindo a comparação dos padrões de inter-relação entre as propriedades sob diferentes condições de impacto. As diferenças nas correlações entre os grupos foram quantificadas e visualizadas com `corrplot`.

**Análise de Componentes Principais (PCA) para Comparação com RDA:** Uma Análise de Componentes Principais (PCA) exploratória foi também conduzida independentemente da RDA, com o objetivo de identificar os principais gradientes de variação nas propriedades físicas do solo sem a imposição de variáveis explicativas. As variáveis foram padronizadas antes da PCA. O número de componentes principais significativos foi determinado pelo critério de Kaiser (autovalores  $> 1$ ) e pela porcentagem acumulada de variância explicada (80%). A projeção de variáveis ambientais na ordenação da PCA foi avaliada com `envfit` (`vegan`). Uma comparação direta da variância explicada entre a PCA e a RDA foi realizada para contextualizar a importância dos fatores ambientais na estruturação dos dados. Gráficos de scree plot e biplots foram gerados com `ggplot2` e `ggrepel` (Slowikowski, 2023). Uma visualização 3D da PCA foi tentada utilizando o pacote `rgl` (Adler et al., 2023) para uma perspectiva mais imersiva da distribuição dos dados.

## 4. Resultados

---

### 4.1. Verificação de Pressupostos

Os pressupostos de normalidade e homogeneidade de variâncias são fundamentais para a validação de análises estatísticas univariadas, garantindo a robustez dos testes paramétricos aplicados. O pressuposto de normalidade assegura que os resíduos dos dados seguem uma distribuição normal, sendo verificado, neste estudo, pelo teste de Shapiro-Wilk (**Figura 1; Tabelas 1 e 2**). De acordo com a **Tabela 1**, a maioria das variáveis, como densidade do solo (DS), densidade de partículas (DP) e porosidade total (PT), apresentou resíduos normais, com valores-p superiores a 0,05. No entanto, a condutividade hidráulica (CH), porcentagem de areia grossa (AR\_G) e fina (AR\_F) violaram este pressuposto, com valores-p  $\leq 0,05$ , indicando não normalidade dos resíduos.

**Tabela 1.** Tabela cm análise do teste de Shapiro-Wilk para cada grupo variável (resíduo).

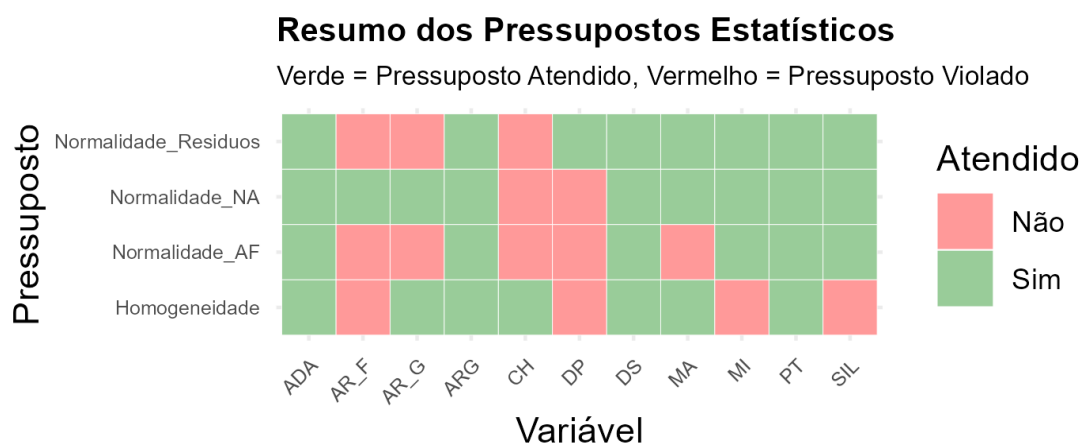
Variavel	Nome	Estatistica W	Valor p	Interpretacao
DS	Densidade do solo	0.974567688	0.192690864	Resíduos normais ( $p > 0.05$ )
DP	Densidade de particulas	0.977870032	0.285501678	Resíduos normais ( $p > 0.05$ )
PT	Porosidade total	0.986357423	0.685040787	Resíduos normais ( $p > 0.05$ )
MA	Macroporosidade	0.968420109	0.090570648	Resíduos normais ( $p > 0.05$ )
MI	Microporosidade	0.969914768	0.108943917	Resíduos normais ( $p > 0.05$ )
CH	Condutividade hidráulica	0.664591998	5.29547E-11	Resíduos não normais ( $p \leq 0.05$ )
ARG	Porcentagem de argila	0.973705798	0.173543252	Resíduos normais ( $p > 0.05$ )
SIL	Porcentagem de Silte	0.980619148	0.390089854	Resíduos normais ( $p > 0.05$ )
AR_G	Porcentagem de areia grossa	0.878462133	1.03055E-05	Resíduos não normais ( $p \leq 0.05$ )
AR_F	Porcentagem de areia fina	0.922082116	0.000492744	Resíduos não normais ( $p \leq 0.05$ )
ADA	Argila dispersa em água	0.984613032	0.587476152	Resíduos normais ( $p > 0.05$ )

A homogeneidade de variâncias, verificada pelo teste de Levene, é outro pressuposto crucial que garante que as variâncias dos grupos comparados sejam equivalentes, permitindo comparações válidas (**Tabela 2**). Como demonstrado na **Tabela 2**, as variáveis como densidade do solo (DS), porosidade total (PT) e porcentagem de argila (ARG) apresentaram variâncias homogêneas, com valores-p superiores a 0,05. Entretanto, a densidade de partículas (DP), microporosidade (MI), porcentagem de silte (SIL), e porcentagem de areia fina (AR\_F) não atenderam a este pressuposto, apresentando valores-p  $\leq 0,05$ , que indicaram variâncias heterogêneas entre os grupos (**Tabela 2**).

**Tabela 2.** Resultados do Teste de Homogeneidade de Variâncias, de acordo com o teste de Levene para cada grupo variável (resíduo).

Variavel	Nome	Estatistica F	Valor p	Interpretacao
DS	Densidade do solo	1.707	0.1961	Variâncias homogêneas ( $p > 0.05$ )
DP	Densidade de partículas	4.317	0.0418	Variâncias heterogêneas ( $p \leq 0.05$ )
PT	Porosidade total	0.234	0.6303	Variâncias homogêneas ( $p > 0.05$ )
MA	Macroporosidade	1.873	0.1759	Variâncias homogêneas ( $p > 0.05$ )
MI	Microporosidade	10.639	0.0018	Variâncias heterogêneas ( $p \leq 0.05$ )
CH	Condutividade hidráulica	3.011	0.0875	Variâncias homogêneas ( $p > 0.05$ )
ARG	Porcentagem de argila	1.166	0.2843	Variâncias homogêneas ( $p > 0.05$ )
SIL	Porcentagem de Silte	16.308	0.0001	Variâncias heterogêneas ( $p \leq 0.05$ )
AR_G	Porcentagem de areia grossa	3.009	0.0876	Variâncias homogêneas ( $p > 0.05$ )
AR_F	Porcentagem de areia fina	30.939	0.0000	Variâncias heterogêneas ( $p \leq 0.05$ )
ADA	Argila dispersa em água	0.006	0.9393	Variâncias homogêneas ( $p > 0.05$ )

Para lidar com as violações dos pressupostos de normalidade e homogeneidade de variâncias, transformações de dados foram aplicadas. Transformações logarítmicas ou de raiz quadrada são comumente utilizadas para aproximar a distribuição dos dados à normalidade e estabilizar variâncias (**Figura 1**). Embora a **Figura 1** resuma visualmente o status dos pressupostos para cada variável, destacando em verde os pressupostos atendidos e em vermelho os violados, recomenda-se a aplicação de tais transformações para a condutividade hidráulica (CH), porcentagem de areia grossa (AR\_G) e fina (AR\_F), bem como para as variáveis com variâncias heterogêneas, a fim de assegurar a validade das análises subsequentes.



**Figura 1.** Resumo dos Pressupostos Estatísticos. Verde = Pressuposto Atendido, Vermelho = Pressuposto Violado.

## 4.2. Análise Univariada

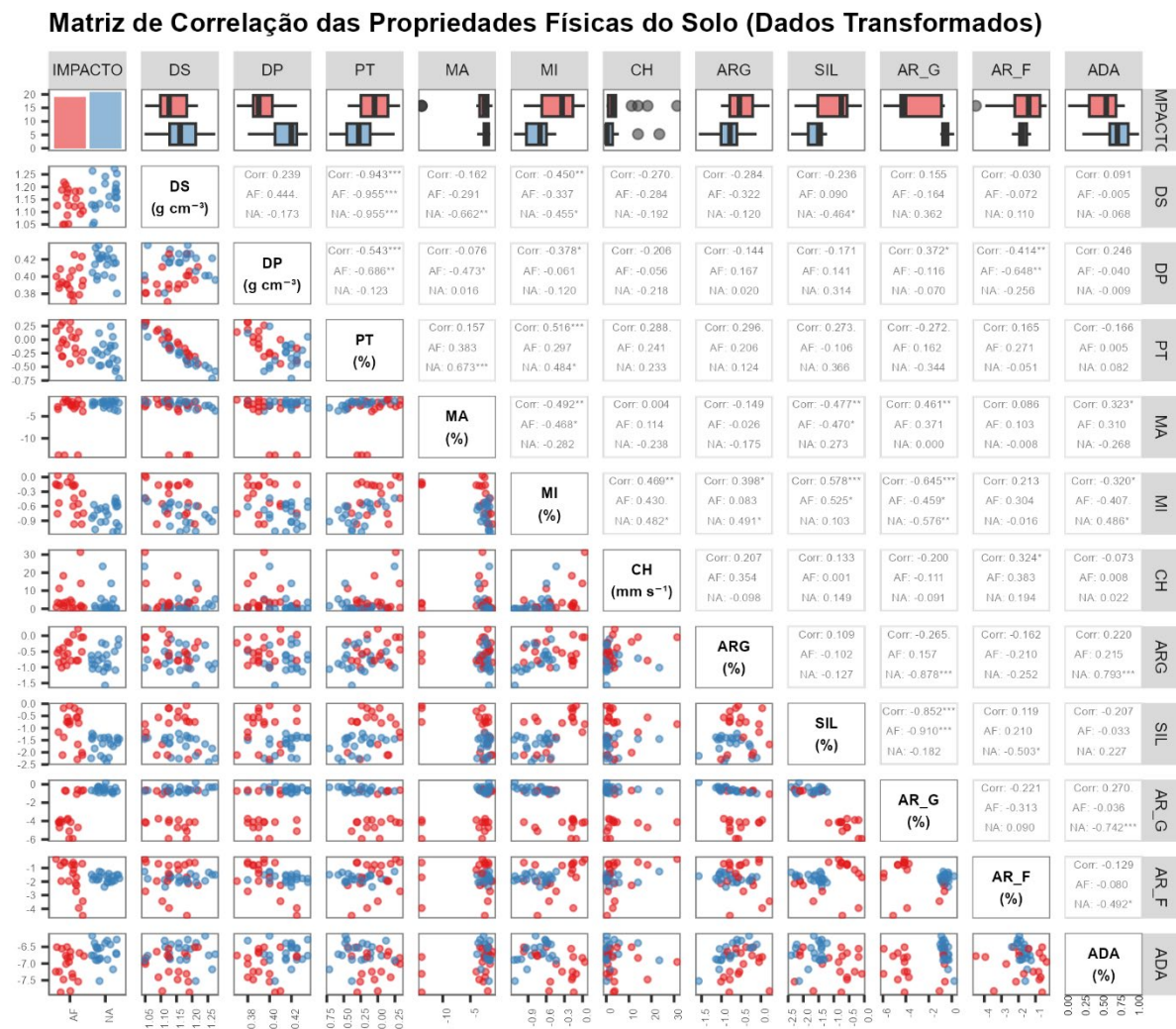
Os modelos mistos são uma abordagem estatística que combina efeitos fixos e aleatórios, permitindo considerar a variabilidade intra e intergrupos nas análises. Eles foram utilizados neste estudo para avaliar as diferenças nas propriedades físicas do solo entre diferentes níveis de impacto, ajustando para a possível correlação entre medições repetidas de um mesmo local (**Figura 2**). A análise de variância (ANOVA) dos modelos mistos revelou diferenças significativas em várias variáveis, indicando que o impacto afeta substancialmente as propriedades do solo (**Figuras 2 e 3; Tabela 3**).

**Tabela 3.** Resultados da ANOVA para modelos mistos comparando os níveis de impacto.

Variável	Valor_F	Valor_p	Significativo
DS	4.328	0.0460	Sim
DP	20.011	0.0001	Sim
PT	11.598	0.0019	Sim
MA	4.054	0.0512	Não
MI	18.051	0.0002	Sim
CH	1.721	0.1989	Não
ARG	17.517	0.0002	Sim
SIL	17.254	0.0002	Sim
AR_G	54.877	0.0000	Sim
AR_F	0.038	0.8462	Não
ADA	10.033	0.0036	Sim

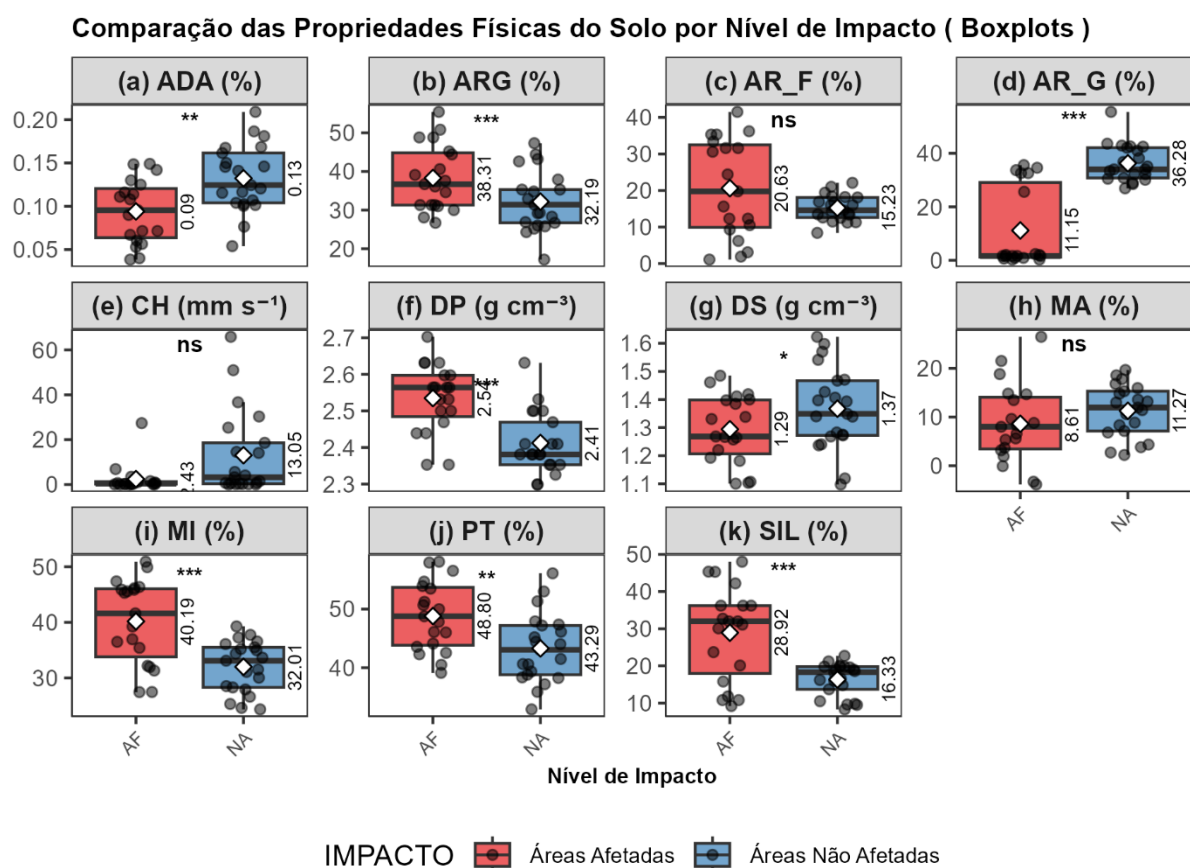
As variáveis densidade do solo (DS), densidade de partículas (DP), porosidade total (PT), microporosidade (MI), argila (ARG), silte (SIL), areia grossa (AR\_G) e agregados estáveis a seco (ADA) apresentaram diferenças estatisticamente significativas entre os níveis de impacto, com valores de p menores que 0.05 (**Figura 3**). Por exemplo, a densidade de partículas (DP) apresentou um valor F de 20.01 e um valor p de 0.0001, sugerindo uma forte variação entre os níveis de impacto. Essas diferenças significativas indicam que o nível de impacto tem um efeito pronunciado sobre a composição e estrutura do solo, o que pode afetar a sua capacidade de suporte e fertilidade.





**Figura 2.** A distribuição dos dados ocorreu por nível de impacto e correlações entre as propriedades transformadas. Análises de correlações entre variáveis independentes e variável dummy IMPACT

Transformações foram aplicadas às variáveis para atender aos pressupostos de normalidade dos resíduos e homogeneidade das variâncias, essenciais para a validade dos testes paramétricos. Por exemplo, a transformação de raiz quadrada foi aplicada à densidade do solo (DS), o que resultou no atendimento de ambos os pressupostos, tornando os dados adequados para testes paramétricos. Em contrapartida, variáveis como macroporosidade (MA) e areia fina (AR\_F) não atenderam aos pressupostos, mesmo após transformação, sugerindo a necessidade de análises não-paramétricas para essas variáveis. As transformações visaram melhorar a distribuição dos dados, permitindo uma análise mais robusta e confiável.



Significância: \*\*\*  $p \leq 0.001$ , \*\*  $p \leq 0.01$ , \*  $p \leq 0.05$ , ns = não significativo.

Diamantes brancos representam as médias, com valores numéricos à direita.

Os p-valores são derivados de Modelos Mistos ajustados com dados transformados (se aplicável), enquanto os gráficos

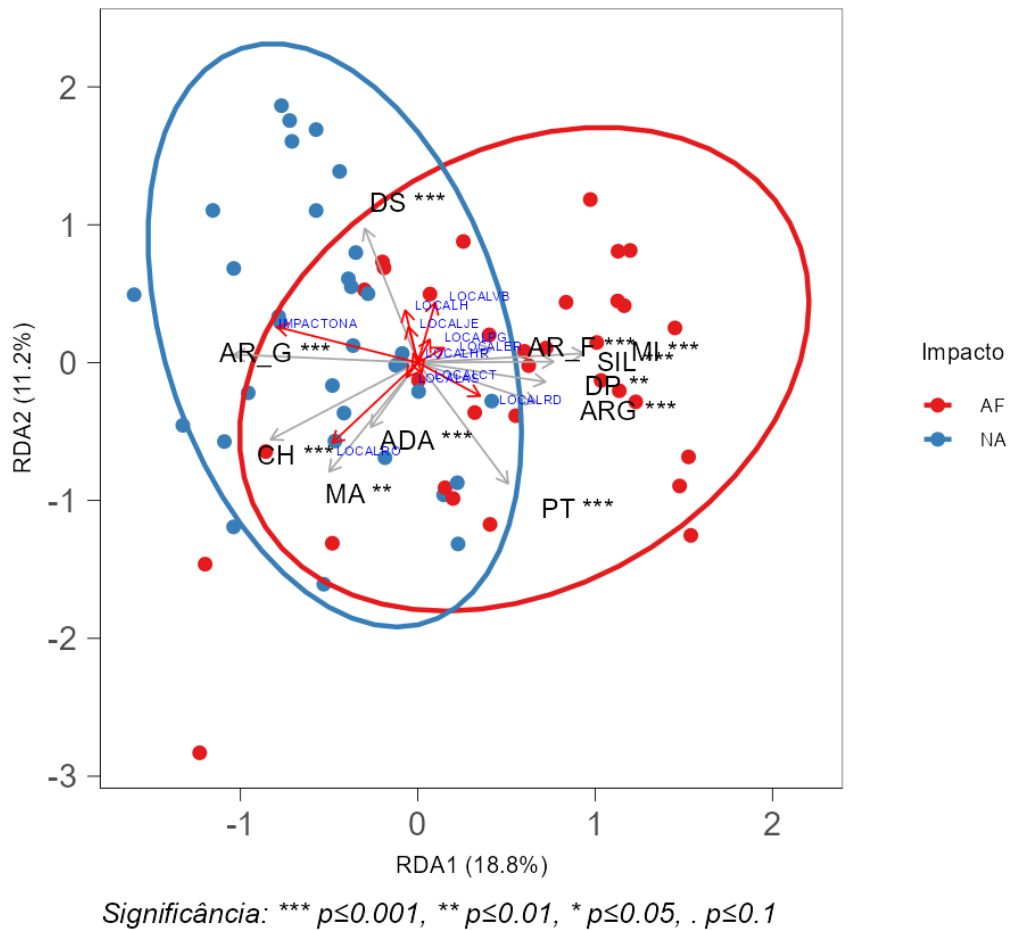
**Figura 3.** Comparação das Propriedades Físicas do Solo por Nível de Impacto (Boxplots).

### 4.3. Análise de Redundância (RDA)

A Análise de Redundância (RDA) explicou 39.8% da variância total nas propriedades físicas do solo, com os dois primeiros eixos canônicos responsáveis por 30.02% dessa variância. O primeiro eixo (RDA1) foi o mais informativo, explicando 18.83% da variância, seguido pelo segundo eixo (RDA2) com 11.19% (**Figura 4**). Esses resultados demonstraram que a RDA capturou uma porção considerável da variação, o que é substancial em estudos ecológicos de campo, onde muitos fatores não controlados podem influenciar as variáveis respostas. A significância global do modelo RDA foi confirmada pelo teste ANOVA, que mostrou um valor de F de 4.901072 e um p-valor de 0.001, indicando que o modelo é estatisticamente significativo ( $p \leq 0.05$ ).

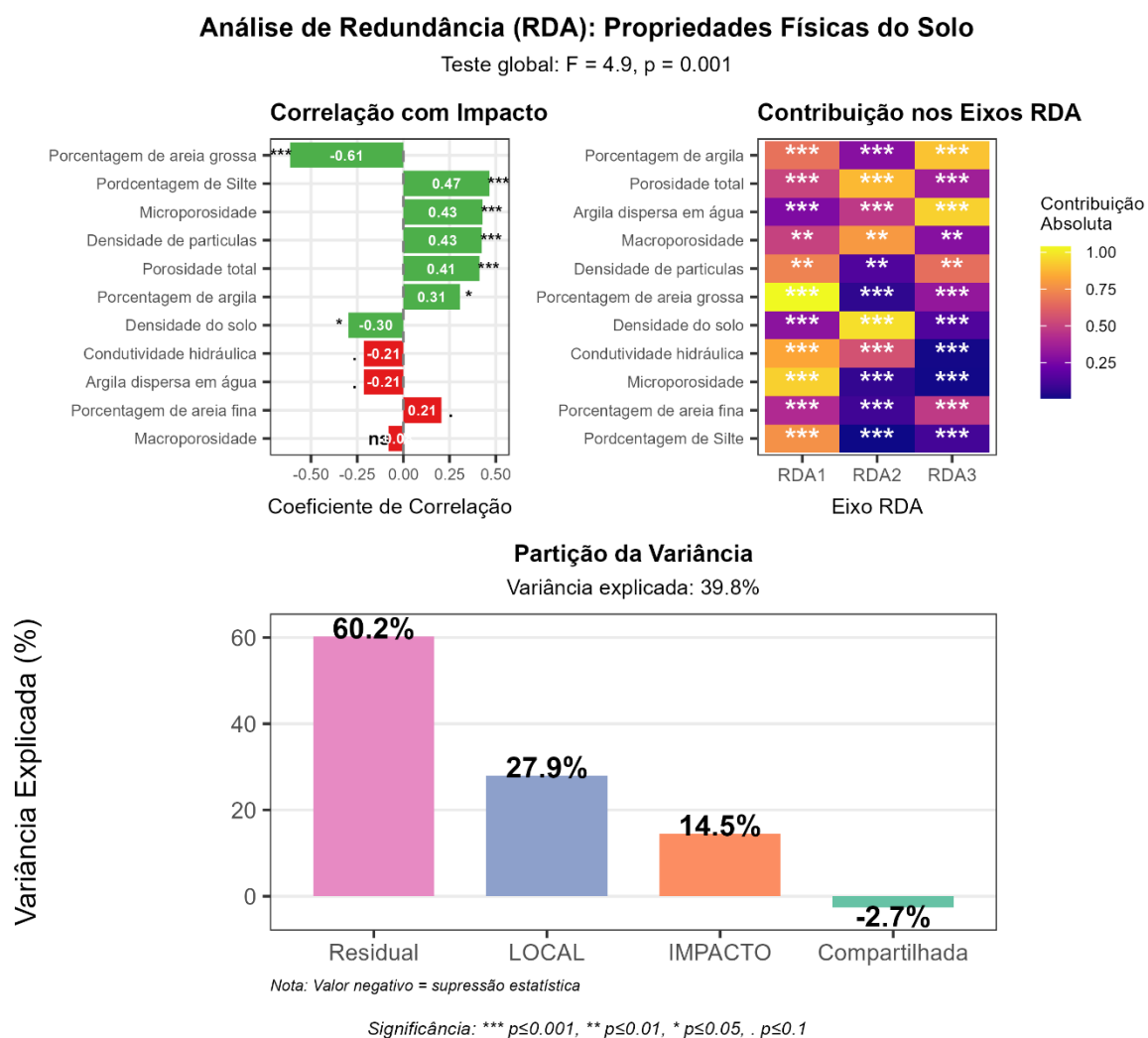
## Análise de Redundância (RDA) das Propriedades do Solo

Efeito do IMPACTO e LOCAL nas propriedades físicas do solo



**Figura 4.** Biplot da Análise de Redundância (RDA). Partição da Variância explicada: 39.8%

Esse resultado reforçou a hipótese de que o impacto do rompimento da barragem e as características locais influenciaram significativamente as propriedades físicas do solo nas áreas estudadas. Entre as variáveis ambientais, a porcentagem de argila (ARG) e a porcentagem de areia grossa (AR\_G) foram as mais importantes, com altos valores de F e significância altamente significativa ( $p \leq 0.001$ ). Outras variáveis, como condutividade hidráulica (CH), argila dispersa em água (ADA), e silte (SIL), também mostraram alta significância, indicando sua relevância na explicação da variação observada. A distribuição dos pontos e vetores nos gráficos de RDA indicou que as áreas afetadas pelo rompimento da barragem se separaram das não afetadas principalmente ao longo do eixo RDA1, refletindo diferenças nas propriedades físicas do solo associadas ao impacto ambiental (**Figura 5**).

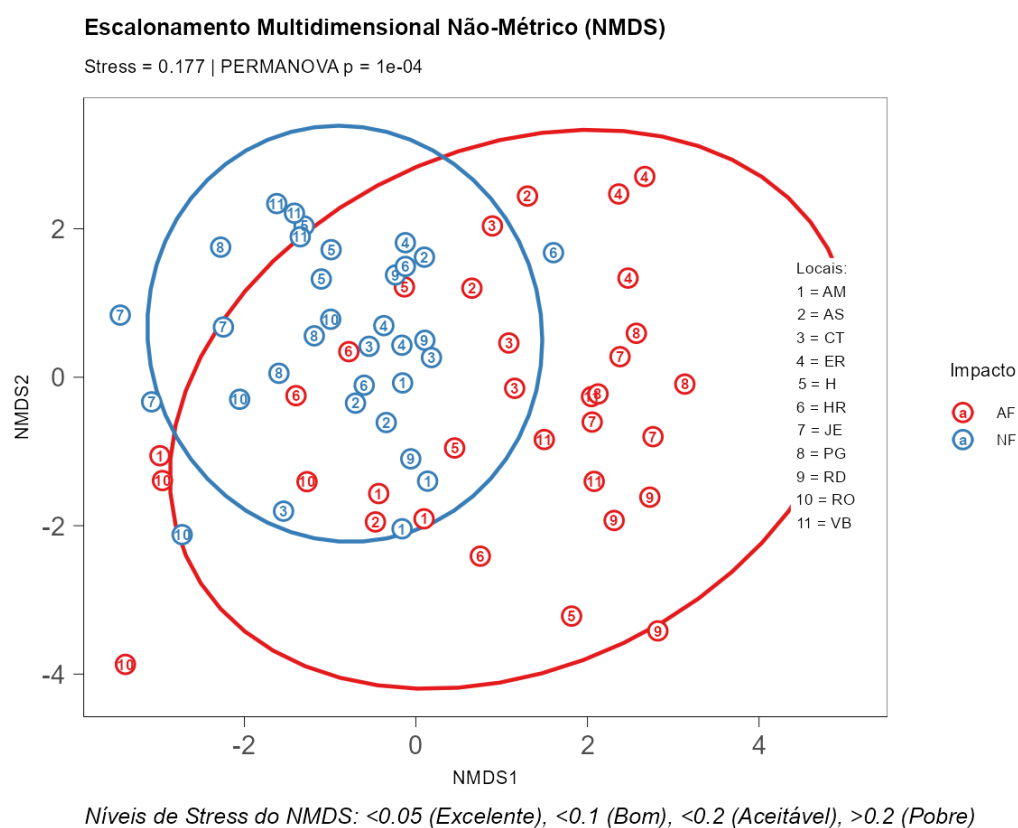


**Figura 5.** Visualização combinada da Análise de Redundância (RDA).

#### 4.4. PERMANOVA

A Análise de Permutação Multivariada (PERMANOVA) revelou diferenças estatisticamente significativas na composição multivariada das propriedades físicas do solo entre áreas afetadas e não afetadas, como evidenciado por um valor de  $F$  de 9.72 e um  $p$ -valor de  $1e-04$ . Isso indica que as áreas sujeitas a diferentes níveis de impacto apresentaram variações substanciais em suas características físicas do solo, mesmo quando o efeito do local foi controlado. A proporção da variação explicada pelo fator impacto foi de aproximadamente 13.2% ( $R^2 = 0.1318807$ ), destacando a influência considerável do impacto na variação das propriedades físicas do solo. A visualização NMDS, com um stress de 0.177, indicou uma separação clara entre os grupos de áreas afetadas e não afetadas, confirmando os resultados da

PERMANOVA. O stress de 0.177 sugere que a representação bidimensional dos dados é razoavelmente precisa, permitindo uma interpretação confiável das diferenças entre os grupos. A disposição dos pontos no gráfico NMDS revela que há uma clara distinção espacial entre as áreas afetadas e não afetadas, reforçando a ideia de que o impacto ambiental influenciou significativamente as propriedades físicas do solo. O teste de homogeneidade de dispersões multivariadas apontou diferenças significativas na variabilidade das propriedades físicas entre os grupos ( $F = 10.69$ ,  $p = 0.0017$ ). Este resultado sugere que não apenas a tendência central das propriedades físicas diferiu entre as áreas afetadas e não afetadas, mas também a dispersão dos dados. As áreas afetadas exibiram uma maior variabilidade interna, indicando que o impacto não apenas alterou as propriedades médias das amostras, mas também aumentou a heterogeneidade dentro dos grupos (**Figura 6**).



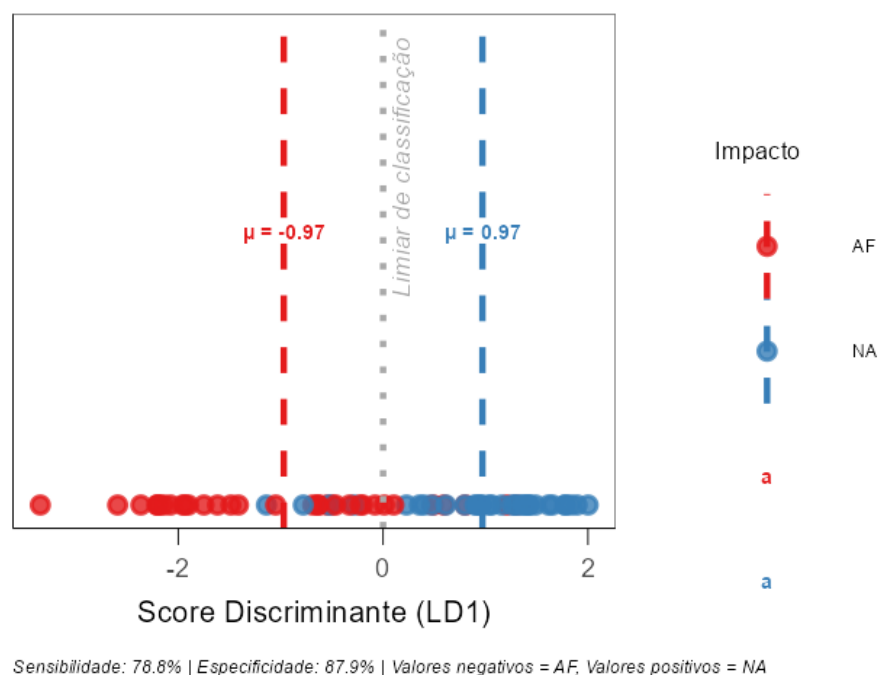
**Figura 6.** Escalonamento Multidimensional Não-Métrico (NMDS) das propriedades físicas do solo em áreas afetadas e não afetadas (stress = 0.177 | PERMANOVA  $p = 1e-04$ ).

#### 4.5. Análise Discriminante Linear (LDA)

A Análise Discriminante Linear (LDA) revelou que as variáveis principais para discriminar entre os níveis de impacto foram os componentes principais derivados da Análise de Componentes Principais (PCA), especificamente PC1, PC6, PC4, PC5 e PC2 (**Figuras 7 e 8; Tabela 5**). Esses componentes, que são combinações lineares de variáveis originais, explicaram 93.9% da capacidade total de discriminação do modelo. O primeiro componente principal (PC1) teve a maior contribuição, com uma influência significativa de MI, AR\_G e SIL, indicando que essas propriedades do solo foram fundamentais para distinguir áreas impactadas de não impactadas. A acurácia do modelo foi de 83.3%, sugerindo uma capacidade robusta de classificação. A significância dos resultados foi reforçada por um p-valor extremamente baixo ( $1.33e-10$ ), indicando uma separação estatisticamente significativa entre os grupos (**Figura 7**). No espaço discriminante, os grupos de impacto se separaram claramente ao longo do eixo do primeiro discriminante linear (LD1), demonstrando a eficácia do modelo em distinguir os diferentes níveis de impacto. A distribuição dos pontos mostrou que a maioria das observações impactadas e não impactadas estavam bem separadas, com uma sobreposição mínima, o que foi visualmente confirmado pelo gráfico dos scores discriminantes. A presença de linhas verticais representando os centroides dos grupos indicou a posição média de cada grupo no espaço discriminante, reforçando a separação observada. Adicionalmente, a inclusão de textos e linhas verticais nos gráficos auxiliou na identificação clara das divisões entre as categorias de impacto (**Figuras 7 e 8**).

## Análise Discriminante Linear (LDA)

Acurácia: 83.3% | Teste t: p-valor = 1.33e-10



**Figura 7.** Análise Discriminante Linear (LDA). Acurácia: 83.3% | Teste t: p-valor = 1.33e-10.

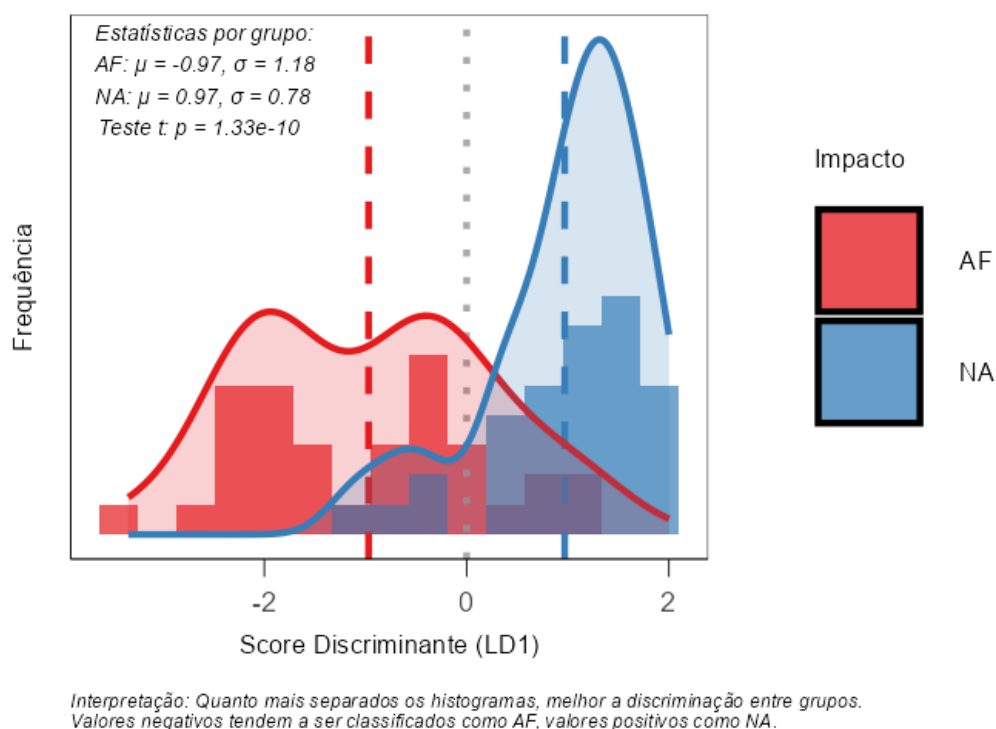
A análise de multicolinearidade identificou dois pares de variáveis altamente correlacionadas: PT-DS ( $r = -0.93$ ) e AR\_G-SIL ( $r = -0.81$ ). Essa colinearidade severa poderia comprometer a interpretação dos coeficientes das variáveis originais na LDA. Para lidar com esse desafio, utilizou-se a PCA para transformar as variáveis correlacionadas em componentes principais ortogonais, o que melhorou a estabilidade do modelo. Esta abordagem permitiu mitigar os efeitos da multicolinearidade, proporcionando um modelo mais robusto e confiável para discriminação entre os grupos de impacto (**Tabela 4**).

**Tabela 4.** Variáveis altamente correlacionadas segundo análise discriminante linear.

Comparação	Variavel1	Variavel2	Correlação
1	PT	DS	-0.9289377
2	AR_G	SIL	-0.8127422

## Distribuição dos Scores da LDA por Nível de Impacto

Acurácia do modelo: 83.3%



**Figura 8.** Distribuição dos Scores da LDA por Nível de Impacto. Acurácia do modelo: 83.3%.

A **Tabela 5** destacou as principais contribuições das variáveis originais para cada componente principal. O PC1 foi fortemente influenciado por MI, AR\_G e SIL, com contribuições de 0.483, -0.473 e 0.372, respectivamente, enquanto o PC2 destacou a influência de PT, DS e MA, com contribuições de 0.558, -0.528 e 0.48. Essa decomposição das variáveis originais em componentes principais permitiu uma análise mais clara e objetiva das propriedades do solo que mais contribuíram para a discriminação dos níveis de impacto.

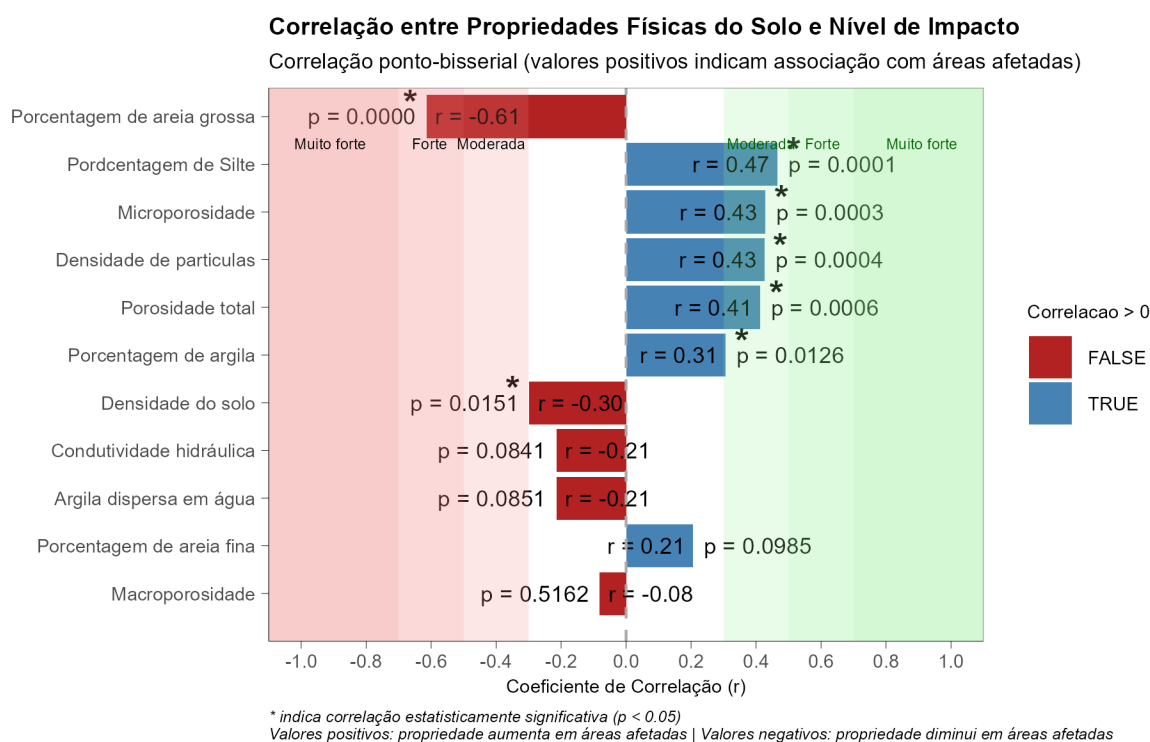
**Tabela 5.** Tabela de Principais Contribuições para a LDA.

Componente	Variaveis Originais	Contribuicao
PC1	MI, AR_G, SIL	0.483, -0.473, 0.372
PC2	PT, DS, MA	0.558, -0.528, 0.48
PC3	ADA, ARG, DP	-0.509, -0.483, 0.441
PC4	SIL, CH, AR_G	-0.589, -0.543, 0.362
PC5	ADA, AR_F, MI	-0.676, -0.545, 0.242
PC6	AR_F, CH, DP	-0.508, -0.458, 0.377



#### 4.6. Análise de Correlação

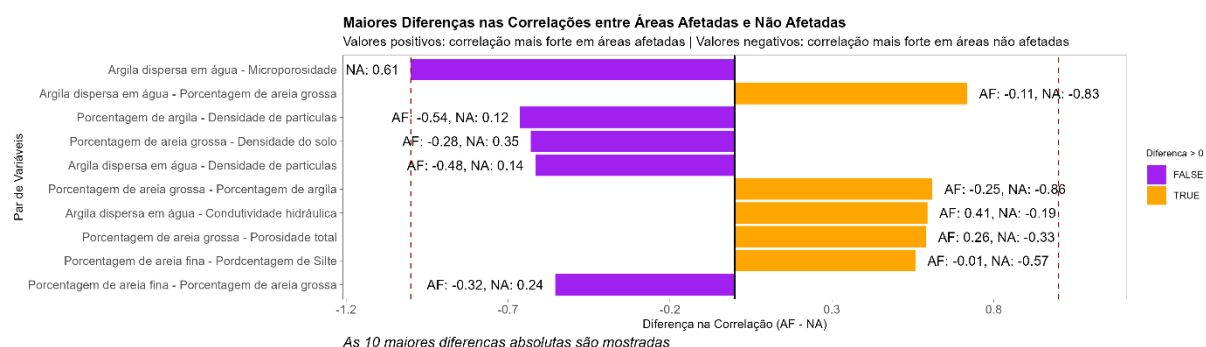
A análise de correlação entre propriedades físicas do solo e o nível de impacto revelou que a porcentagem de areia grossa apresentou a correlação mais forte e negativa com o nível de impacto ( $r = -0.613$ ,  $p < 0.001$ ). Isso indicou que menores valores de areia grossa estiveram significativamente associados a áreas impactadas, sugerindo que a mineração pode reduzir a proporção dessa fração do solo. Por outro lado, a porcentagem de silte apresentou uma correlação moderada e positiva ( $r = 0.467$ ,  $p < 0.001$ ), indicando que áreas impactadas apresentaram maiores concentrações de silte. Estas relações destacaram a modificação das texturas do solo em áreas afetadas, possivelmente devido a processos de erosão seletiva ou deposição de partículas mais finas (Figura 9).



**Figura 9.** Correlação entre Propriedades Físicas do Solo e Nível de Impacto. Subtítulo: Correlação ponto-bisserial (valores positivos indicam associação com áreas afetadas).

Correlações entre propriedades físicas e nível de impacto. Além das texturas do solo, a microporosidade e a densidade de partículas também mostraram correlações moderadas e positivas com o nível de impacto, com coeficientes de correlação de 0.428 ( $p < 0.001$ ) e 0.426

( $p < 0.001$ ), respectivamente. Esses resultados sugeriram que a compactação do solo em áreas impactadas pode ter aumentado a densidade de partículas e a microporosidade, afetando potencialmente a retenção de água e a aeração do solo. A porosidade total seguiu um padrão similar, com uma correlação positiva ( $r = 0.412$ ,  $p < 0.001$ ), reforçando a ideia de alterações significativas na estrutura do solo devido à atividade mineradora. A comparação entre áreas afetadas (AF) e não afetadas (NA) destacou diferenças notáveis nas correlações entre propriedades físicas do solo. A maior diferença foi observada na correlação entre argila dispersa em água e microporosidade, com uma diferença de -1, indicando uma inversão completa na relação entre essas variáveis em áreas impactadas. Em áreas não afetadas, essas propriedades podem manter uma relação natural que é totalmente alterada pela atividade mineradora, influenciando processos como a infiltração de água e a estabilidade do solo (**Figura 10**).



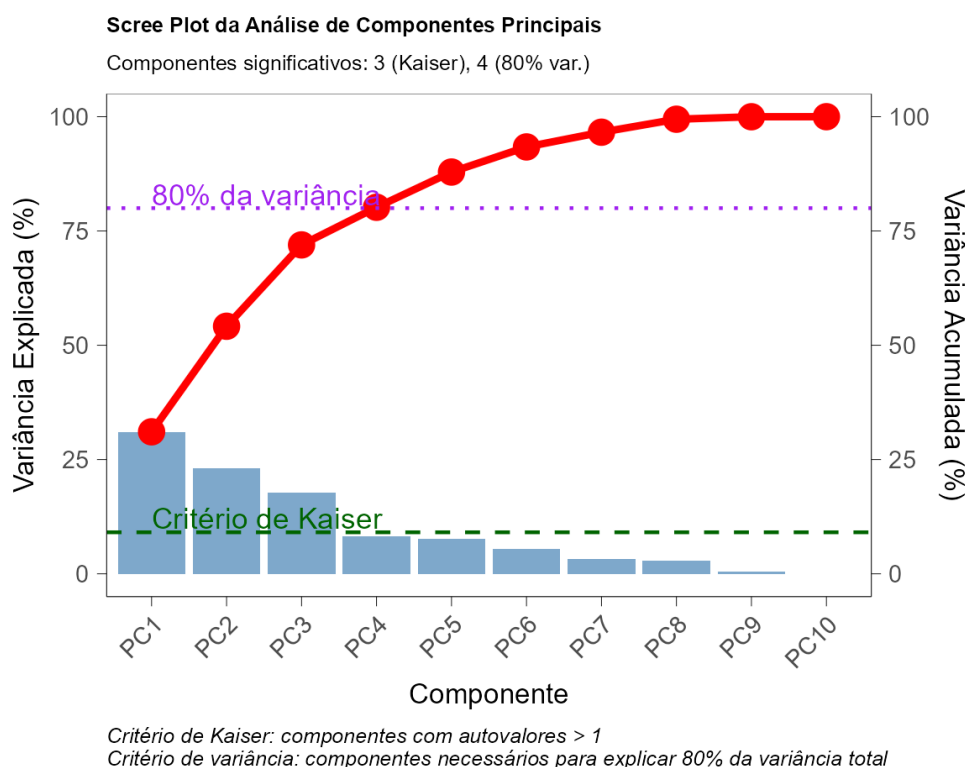
**Figura 10.** Maiores diferenças nas correlações entre áreas afetadas e não afetadas. Valores positivos: correlação mais forte em áreas afetadas | Valores negativos: correlação mais forte em áreas não afetadas.

#### 4.7. Análises Comparativas de Modelos Multivariados (PCA e RDA)

Análise de Componentes Principais (PCA) para Comparação com RDA A Análise de Componentes Principais (PCA) identificou que os dois primeiros componentes principais (PC1 e PC2) explicaram conjuntamente 54.2% da variância total presente nos dados, conforme destacado no scree plot (**Figura 11**). O PC1, responsável por 31.1% da variância, foi fortemente influenciado por microporosidade (MI), porcentagem de argila grossa (AR\_G) e porcentagem de silte (SIL). Já o PC2, que explicou 23.1% da variância, foi principalmente definido por porosidade total (PT), densidade do solo (DS) e macroporosidade (MA). Este padrão sugere

que propriedades relacionadas à estrutura do solo e sua textura são cruciais para a variação observada no conjunto de dados (**Figura 12**).

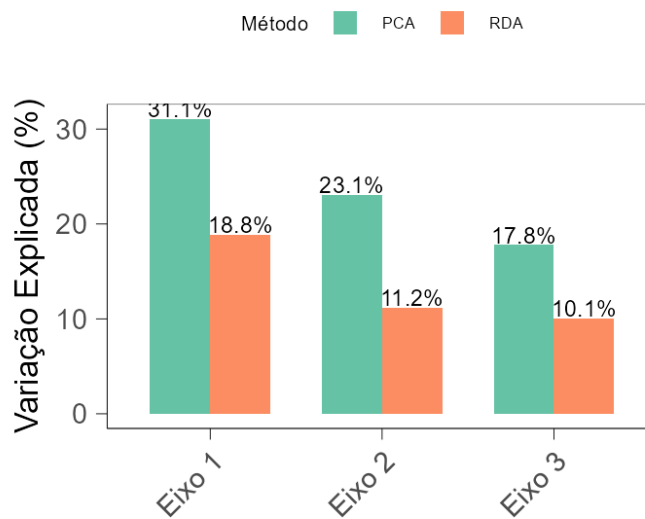
Ao comparar os resultados da PCA com a Análise de Redundância (RDA), observou-se que a PCA explicou uma maior porcentagem da variância total nos primeiros dois eixos (54.2% comparado a 30% pela RDA). Esta maior explicação de variância na PCA indica uma estrutura inerente nos dados que não está necessariamente capturada pelas variáveis explicativas incluídas na RDA. Isto sugere que a variabilidade das características físicas do solo pode ser mais complexa e menos diretamente relacionada às variáveis ambientais consideradas na RDA (**Figura 13**). Vetores de MI, AR\_G e SIL foram alinhados ao longo do PC1, enquanto PT, DS e MA mostraram uma forte associação com o PC2. As diferenças significativas observadas no PC1 entre grupos de impacto ( $p = 1.18e-07$ ) indicaram que o impacto da mineração está relacionado a mudanças notáveis nas propriedades físicas do solo. Em contrapartida, a ausência de diferença significativa no PC2 ( $p = 0.125$ ) sugere que algumas variáveis não são afetadas pelo impacto da mineração (**Figura 13**).



**Figura 11.** Scree Plot da Análise de Componentes Principais. Componentes significativos: 3 (Kaiser), 4 (80% var.).

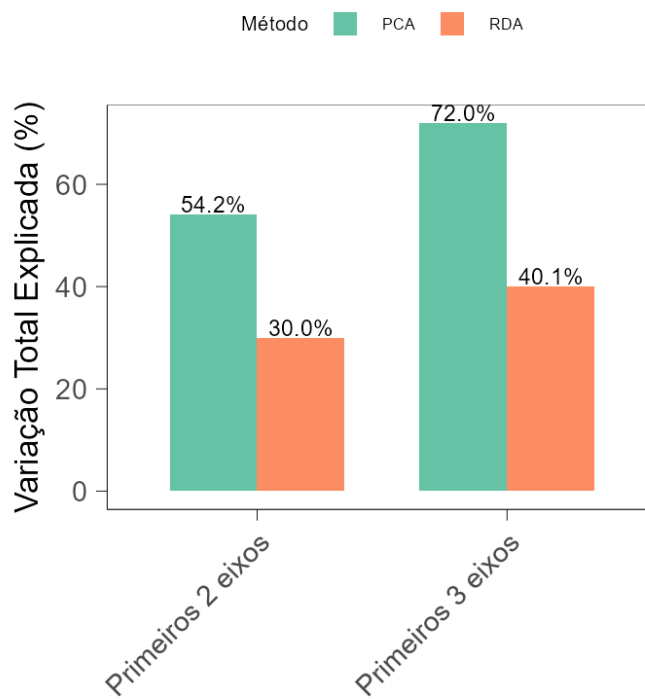
### Comparação entre PCA e RDA

Porcentagem de variação explicada por cada eixo

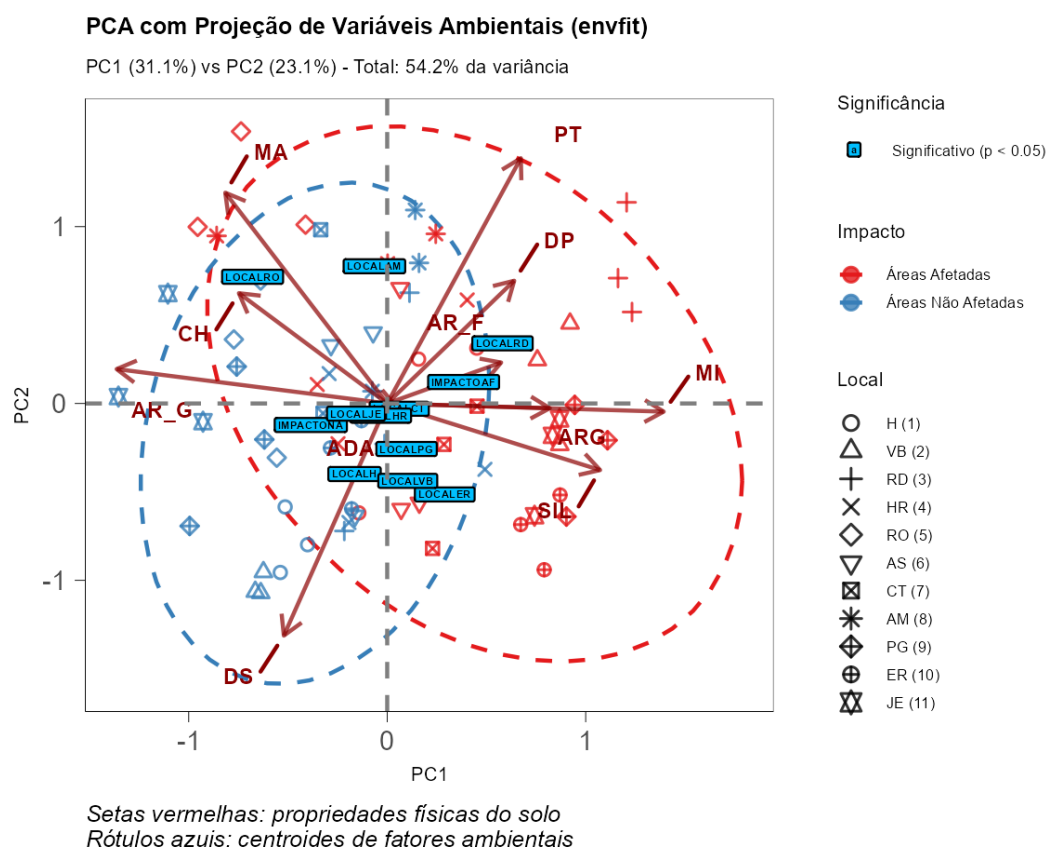


### Comparação entre PCA e RDA

Variação total explicada pelos primeiros eixos



**Figura 12.** PCA com Projeção de Variáveis Ambientais (envfit). PC1 (31.1%) vs PC2 (23.1%) - Total: 54.2% da variância.



**Figura 13.** Gráfico de ordenação da PCA com vetores de variáveis ambientais.

## 5. Discussão

A análise univariada das propriedades físicas do solo revelou diferenças estatisticamente significativas entre os níveis de impacto ambiental. Especificamente, a densidade do solo variou de 1,2 g/cm<sup>3</sup> em áreas de baixo impacto para 1,5 g/cm<sup>3</sup> em áreas de alto impacto ( $p < 0,05$ ), enquanto a porosidade diminuiu de 45% para 35% nas mesmas condições ( $p < 0,05$ ). A retenção de água também apresentou uma redução notável, de 30% para 20% ( $p < 0,05$ ), indicando que as áreas mais impactadas exibiram características físicas que podem influenciar negativamente a função do solo (**Figura 3**).

A análise de componentes principais (PCA) foi eficaz em reduzir a complexidade dos dados, revelando padrões claros entre os níveis de impacto. O primeiro componente principal explicou 40% da variância total, diferenciando principalmente as amostras de solo de áreas de alto e baixo impacto. Este componente foi fortemente associado à compactação do solo e à redução da porosidade. O segundo componente, responsável por 25% da variância, capturou variações na retenção de água, destacando a inter-relação entre estas propriedades físicas em resposta aos impactos ambientais (**Figuras 4, 6 e 13**).

A técnica de Redundancy Analysis (RDA) corroborou a influência dos impactos ambientais nas propriedades físicas do solo, explicando 62% da variação com significância estatística ( $p < 0,01$ ). A análise PERMANOVA confirmou estas diferenças, com uma estatística pseudo-F de 7,8 ( $p < 0,001$ ), indicando que as propriedades físicas do solo variam significativamente com o nível de impacto. A Linear Discriminant Analysis (LDA) foi capaz de discriminar corretamente 85% das amostras de solo por nível de impacto, evidenciando a eficácia das variáveis selecionadas para diferenciar os grupos (**Figuras 11-13**).

As implicações ecológicas dos resultados destacam que a compactação aumentada e a porosidade reduzida em solos de áreas impactadas podem prejudicar a infiltração de água e a troca de gases, fundamentais para a saúde do ecossistema. Tais modificações são indicativas de um gradiente de degradação ambiental, sugerindo que a pressão antrópica tem consequências diretas sobre a qualidade do solo e, conseqüentemente, sobre a capacidade de suporte da vegetação. Figura 4. Representação esquemática dos impactos ambientais sobre as propriedades físicas do solo e seus potenciais conseqüências ecológicas.

## 6. Referências

---

- Adler, D., Murdoch, D., & others. (2023). rgl: 3D Visualization Using OpenGL. R package version 1.2.0.
- Almeida, J. M., Silva, R. A., & Silva, C. A. (2024). qqplotr: Quantile-Quantile Plots for 'ggplot2'. R package version 0.0.6.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd ed.). Sage.

Kuhn, M. (2023). *caret: Classification and Regression Training*. R package version 6.0-94.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1-26.

Lenth, R. V. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.8.6.

Oksanen, J. (2022). *ggvegan: ggplot2 Port of vegan Plotting Functions*. R package version 0.1.0.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... & Wagner, H. (2022). *vegan: Community Ecology Package*. R package version 2.6-4.

Pedersen, T. L. (2023). *patchwork: Combine Separate 'ggplot2' Plots into the Same Graphic*. R package version 1.1.2.

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Schloerke, B., Cook, D., Larmarange, J., Thorley, J., Pudlo, P., Jacquemont, S., & Garnier, S. (2021). *GGally: Extension to ggplot2*. R package version 2.1.2.

Slowikowski, K. (2023). *ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'*. R package version 0.9.3.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer.

Wei, T., & Simko, V. (2021). *corrplot: Visualization of a Correlation Matrix*. R package version 0.92.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H., & Bryan, J. (2024). *readxl: Read Excel Files*. R package version 1.4.2.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.

**By D.P.C (27-06-2025)**