

# Categorização numérica e gráfica de palavras no Conto A Cartomante

Raiane dos Anjos Pereira  
Sabrina Evelyn Cruz Oliveira

Novembro, 2025

# Introdução

- Ferramentas computacionais de modelagem <-> análise linguística;
- Como a análise computacional pode revelar aspectos estilísticos e estruturais do conto?

# Objetivo geral

- Explorar técnicas de processamento computacional aplicadas a um texto literário;
- Demonstrar como estatísticas e visualizações podem auxiliar na análise linguística;
- Aplicar conceitos da disciplina utilizando o R, com auxílio do github.

# Metodologia

- Leitura e pré-processamento do texto;
- Padronização: minúsculas, remoção de pontuação, separação em palavras;
- Remoção de *stopwords* em português;
- Cálculo de estatísticas básicas;
- Geração de visualizações em ggplot2 e wordcloud2.

# Estatísticas básicas do texto

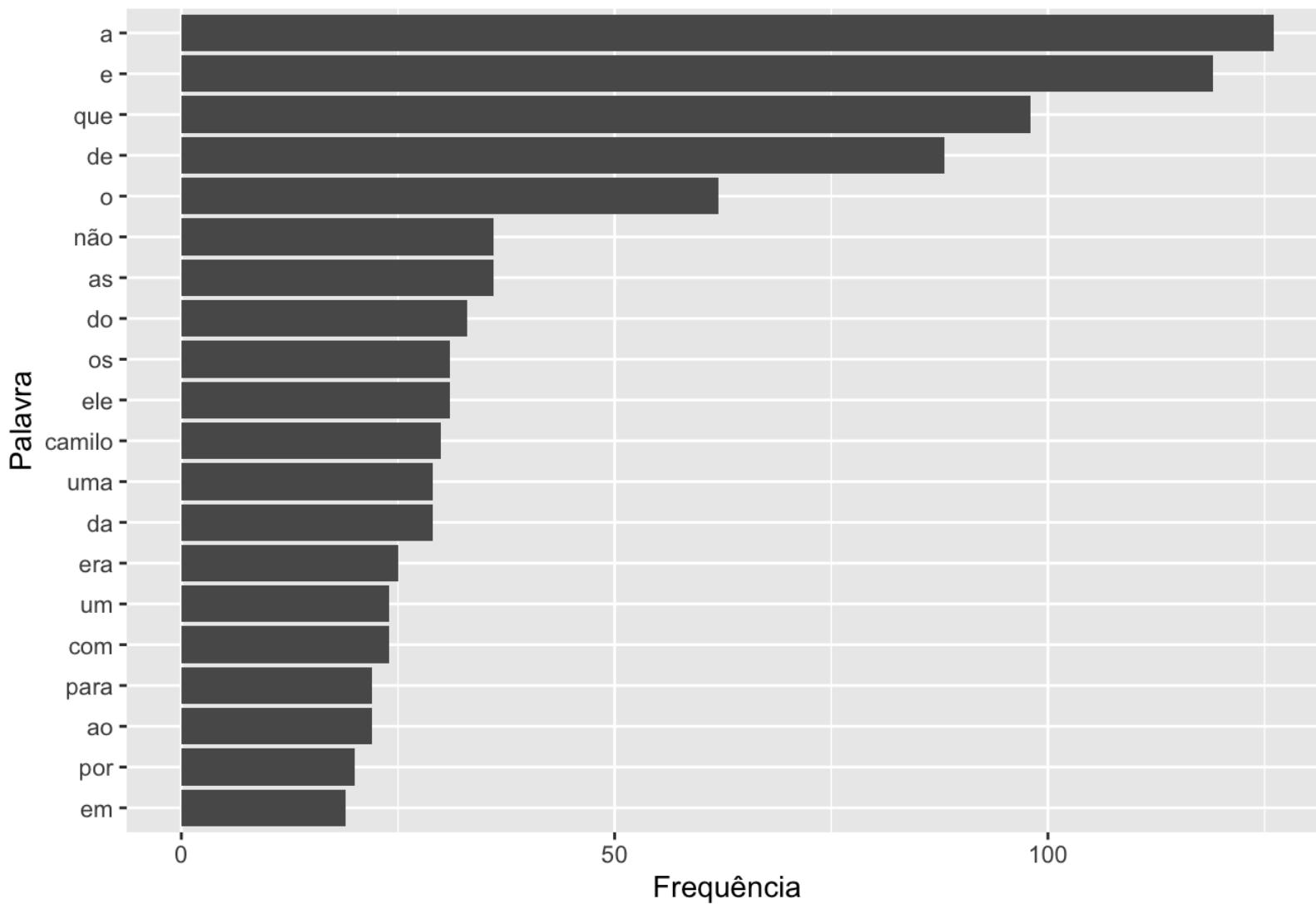
- Número de linhas;
- Número total de palavras;
- Número de caracteres;
- Média do tamanho das palavras;
- Mediana do tamanho das palavras;
- Distribuição dos tamanhos (histograma).

# Frequência de palavras

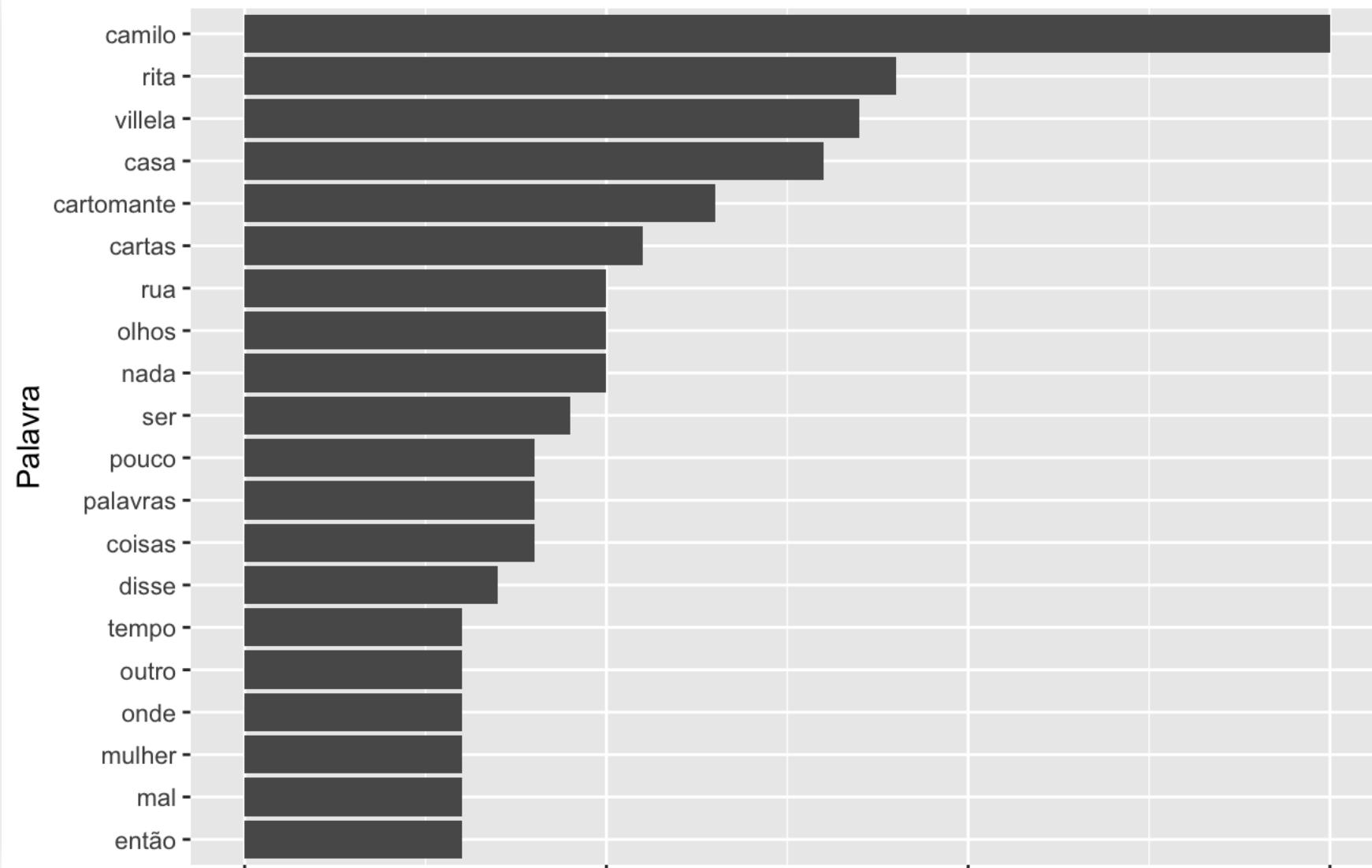
- Tabela de palavras;
- Palavras mais recorrentes no texto;
- Remoção das palavras funcionais – foco nas palavras semânticas;
- Aplicação de stopwords e filtros.

# Visualizações criadas com o R

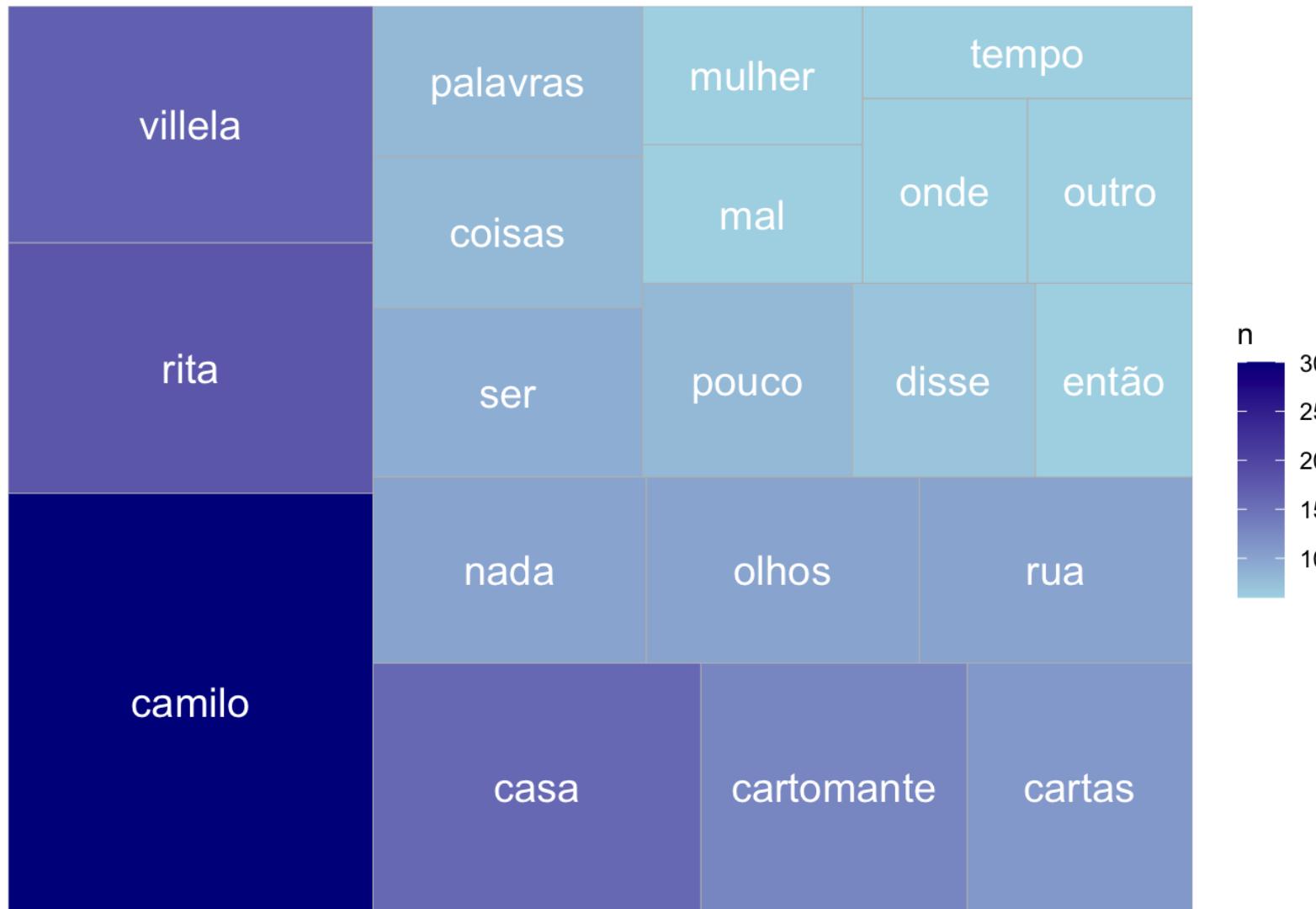
## 20 Palavras mais Frequentes em 'A Cartomante'



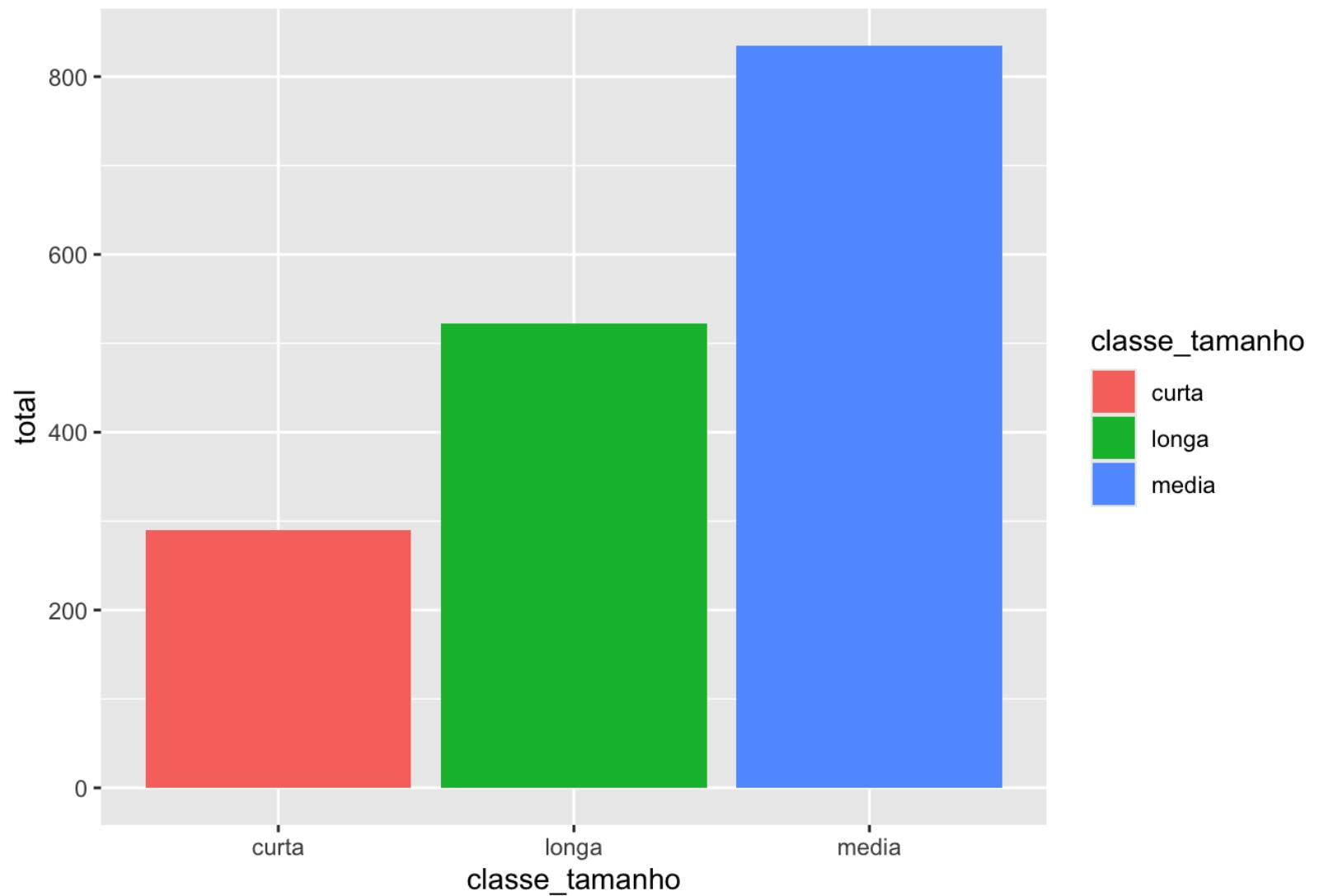
## Top 20 palavras semânticas em 'A Cartomante'



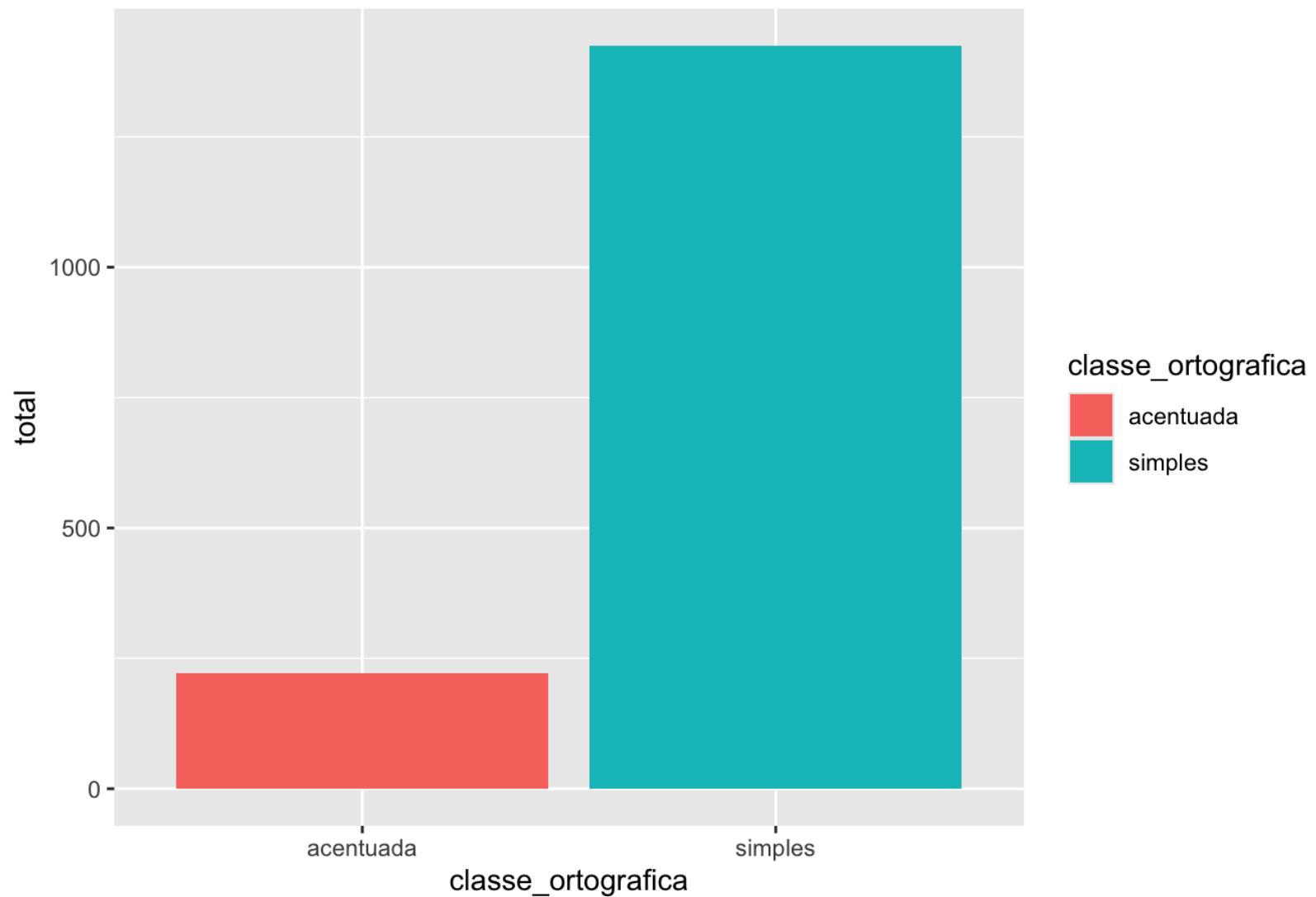
Top 20 palavras semânticas - Treemap



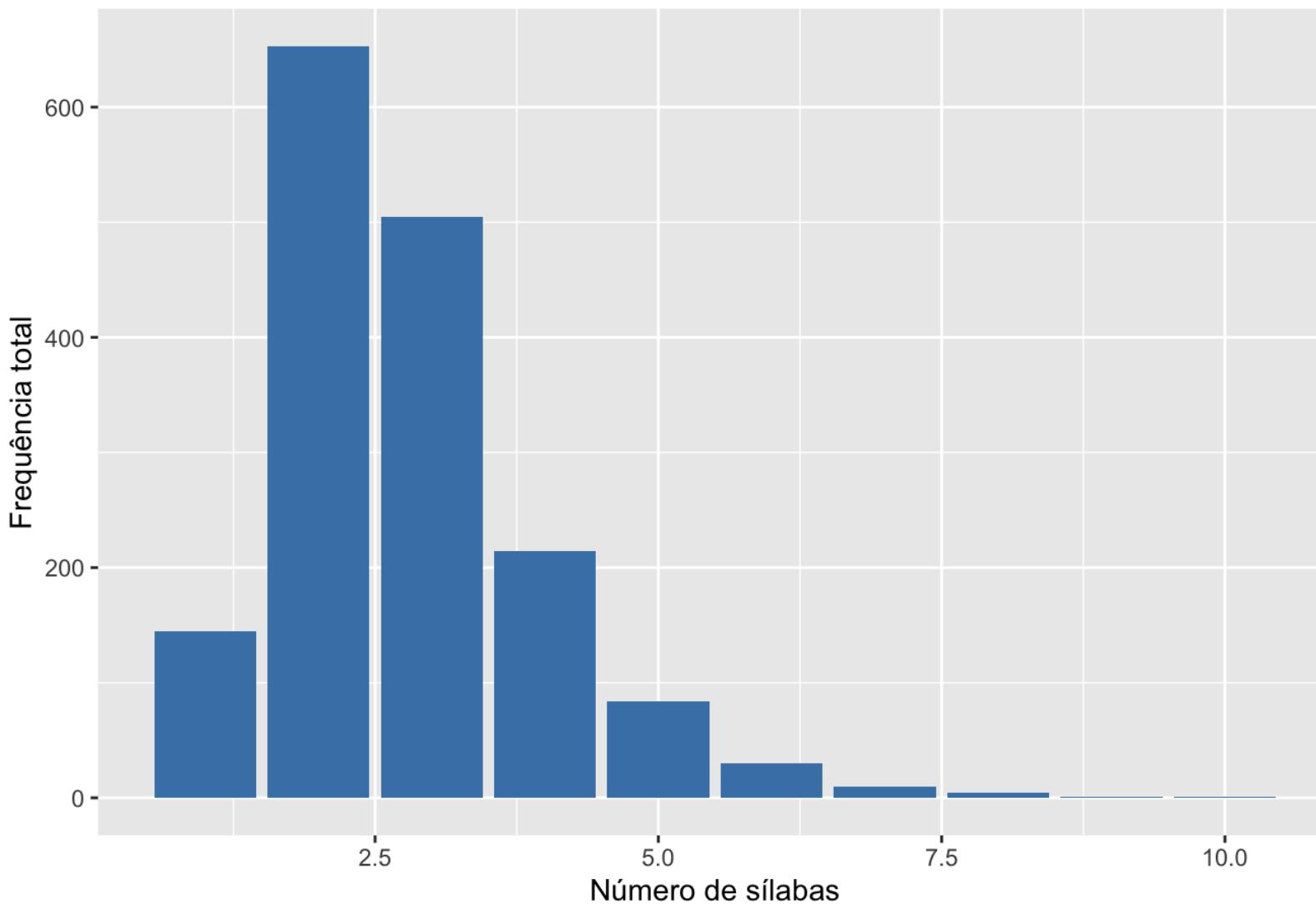
Distribuição por tamanho da palavra



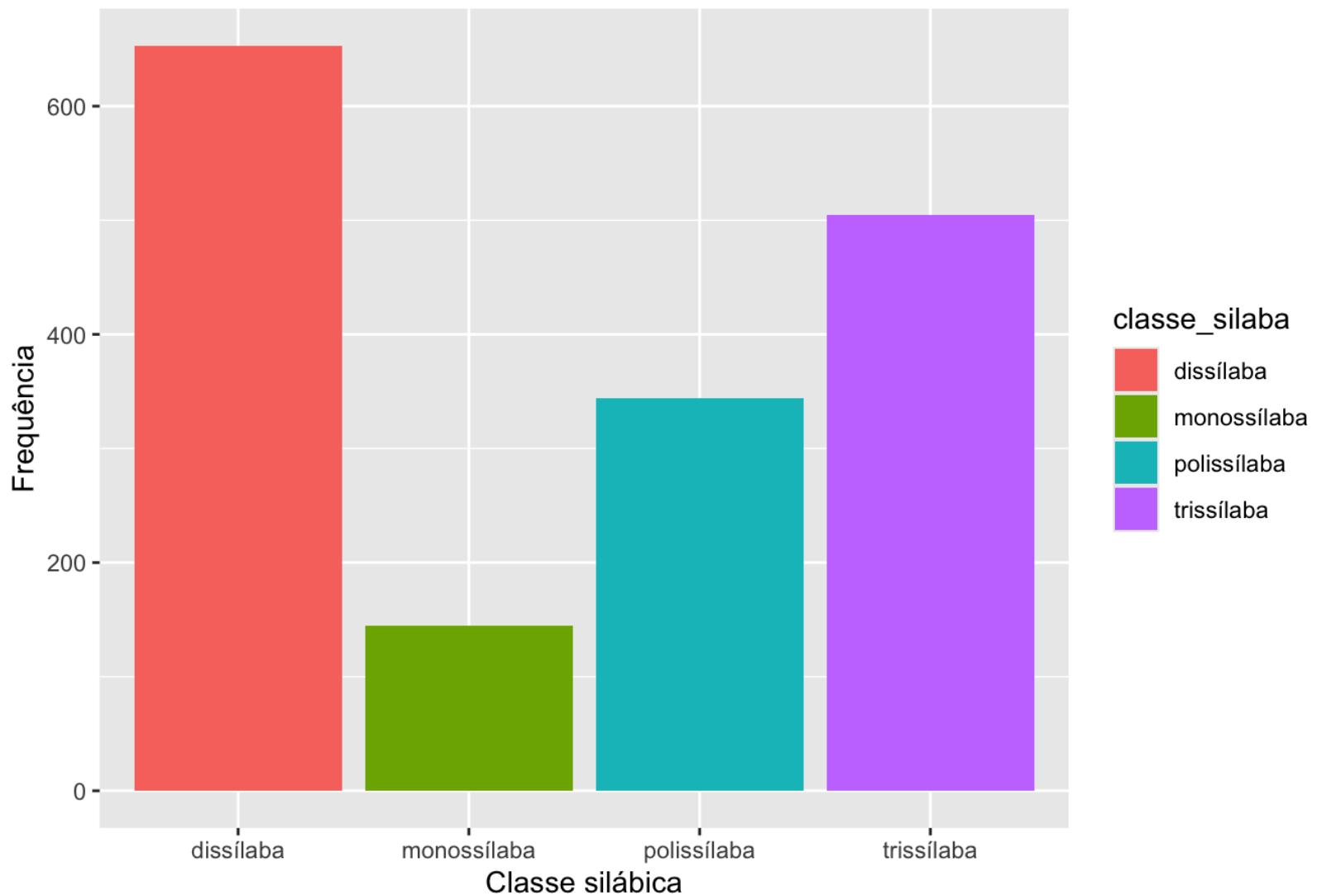
Distribuição por tipo ortográfico



Distribuição das palavras por número de sílabas



## Distribuição por categoria silábica



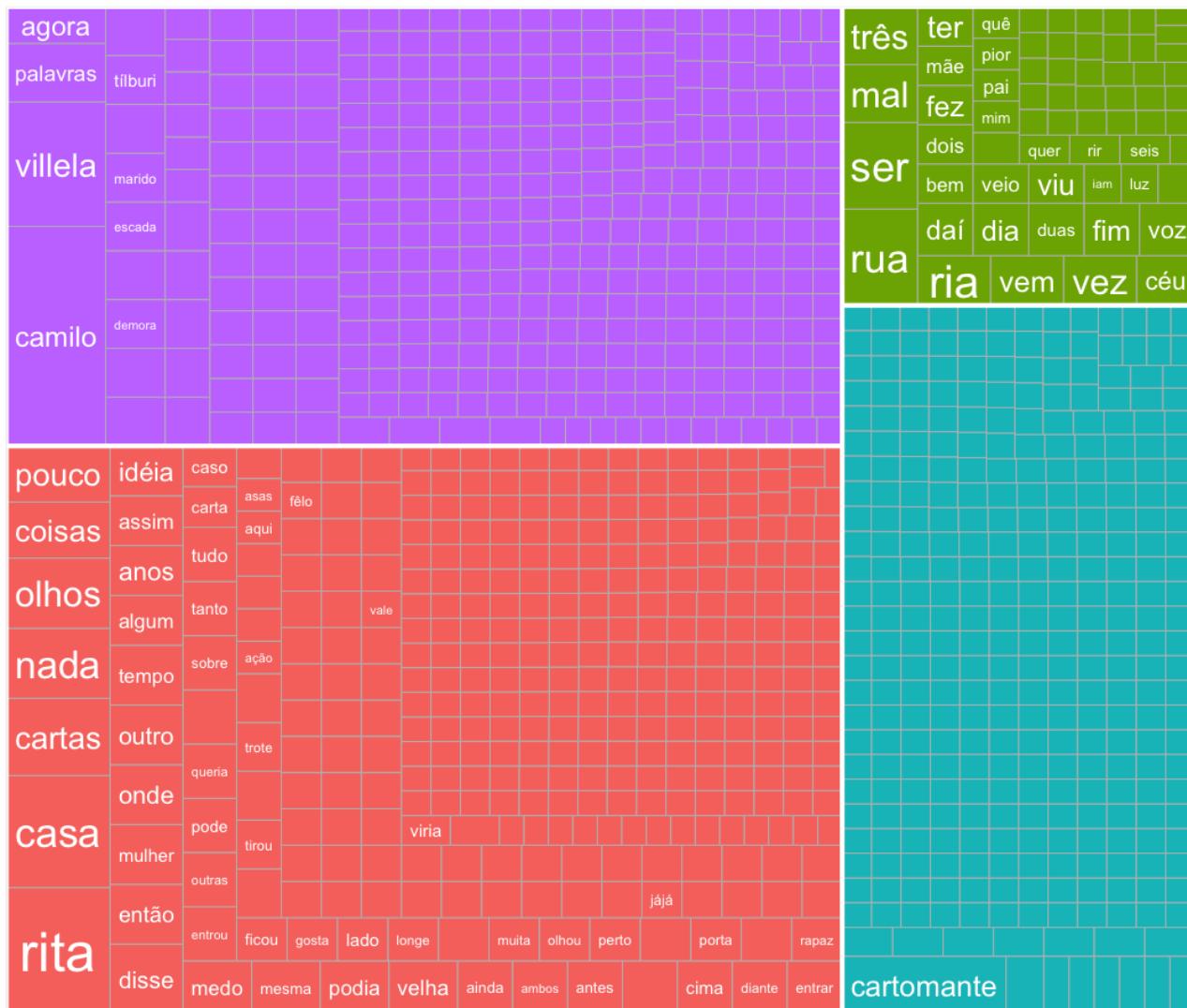
Treemap por Categoria Silábica



Classe Silábica

- dissílaba
- monossílaba
- polissílaba
- trissílaba

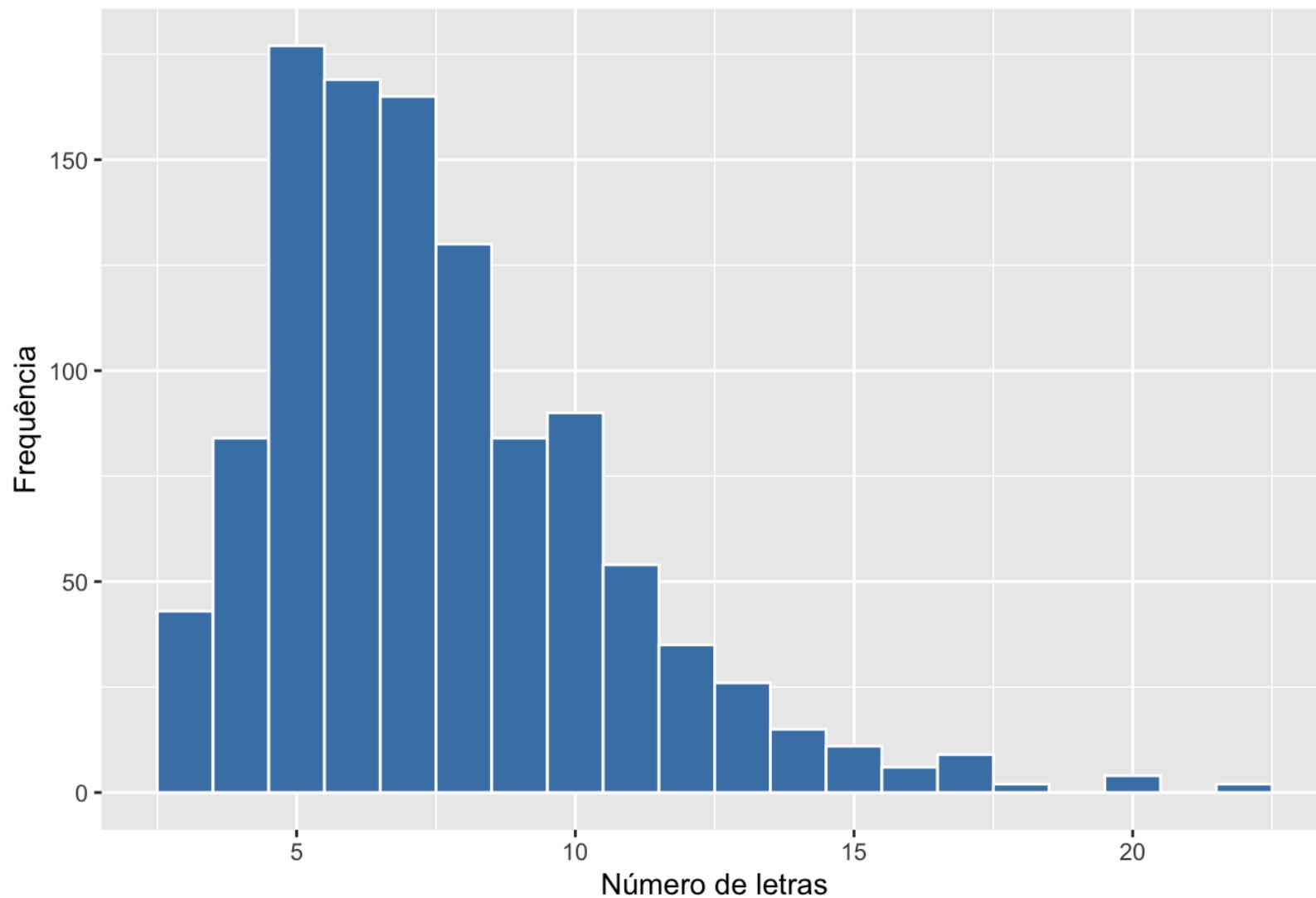
## Treemap de Palavras por Categoria Silábica



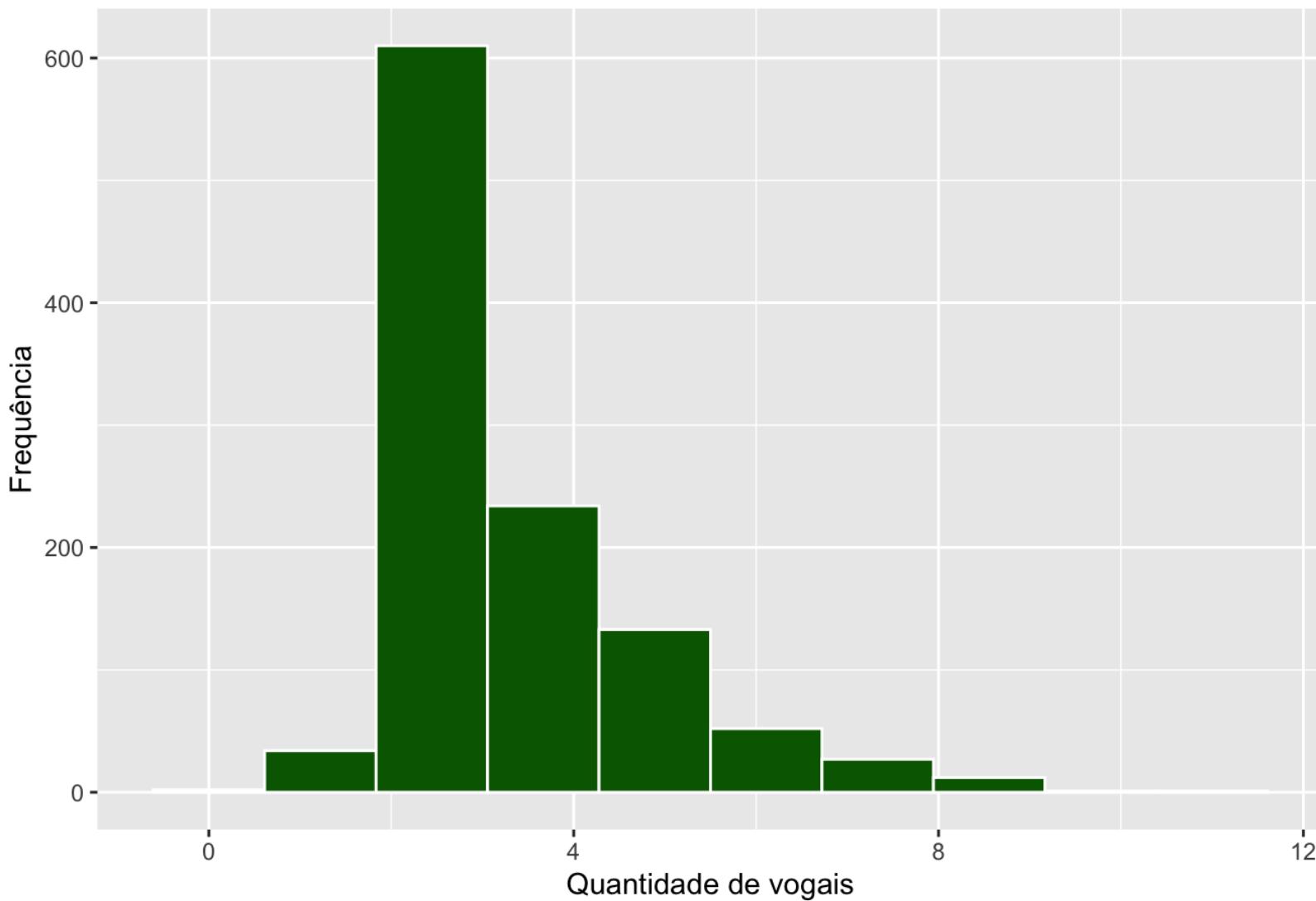
Classe Silábica

- dissílaba
- monossílaba
- polissílaba
- trissílaba

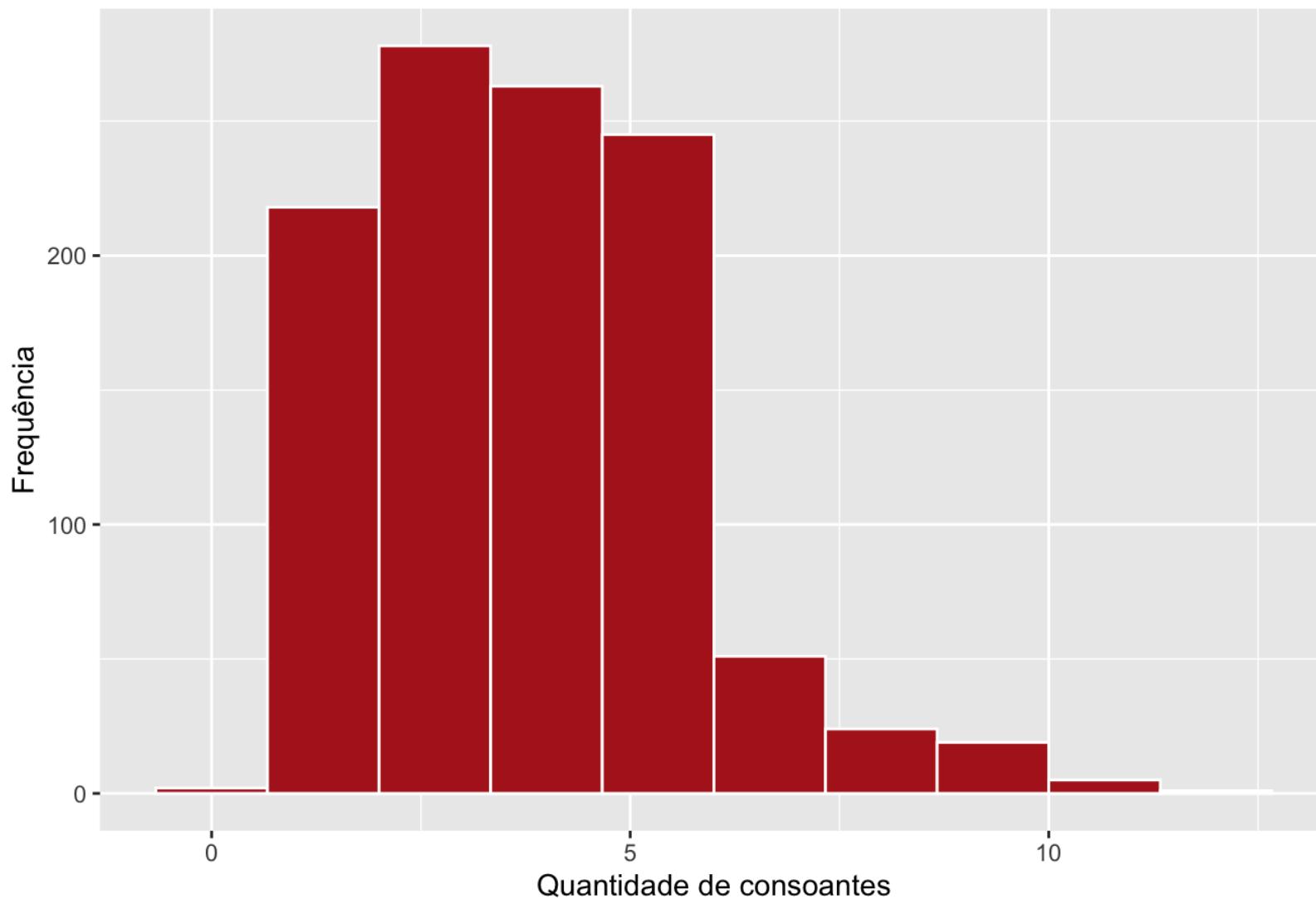
## Distribuição do Tamanho das Palavras



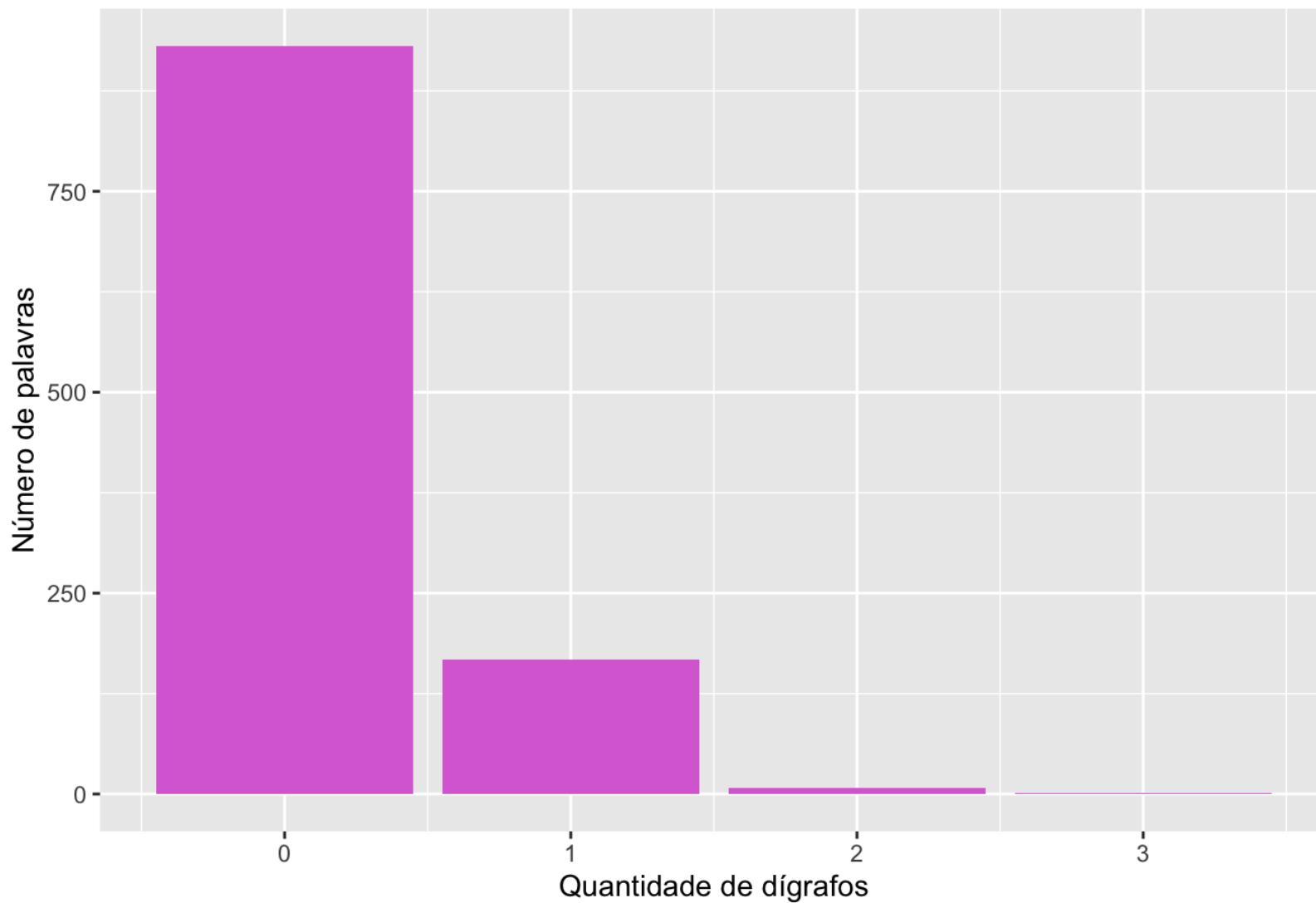
### Distribuição de Vogais nas Palavras



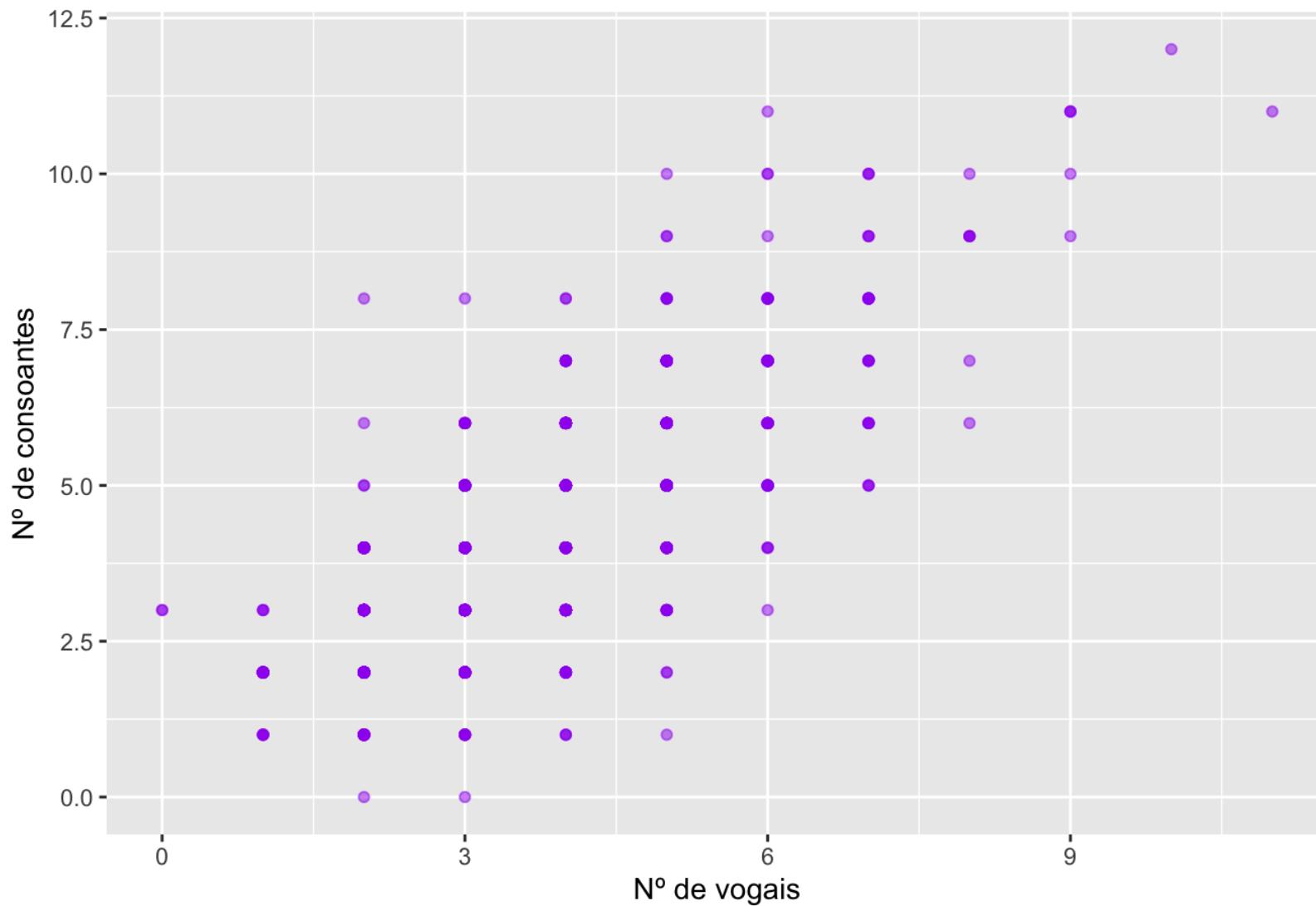
## Distribuição de Consoantes nas Palavras



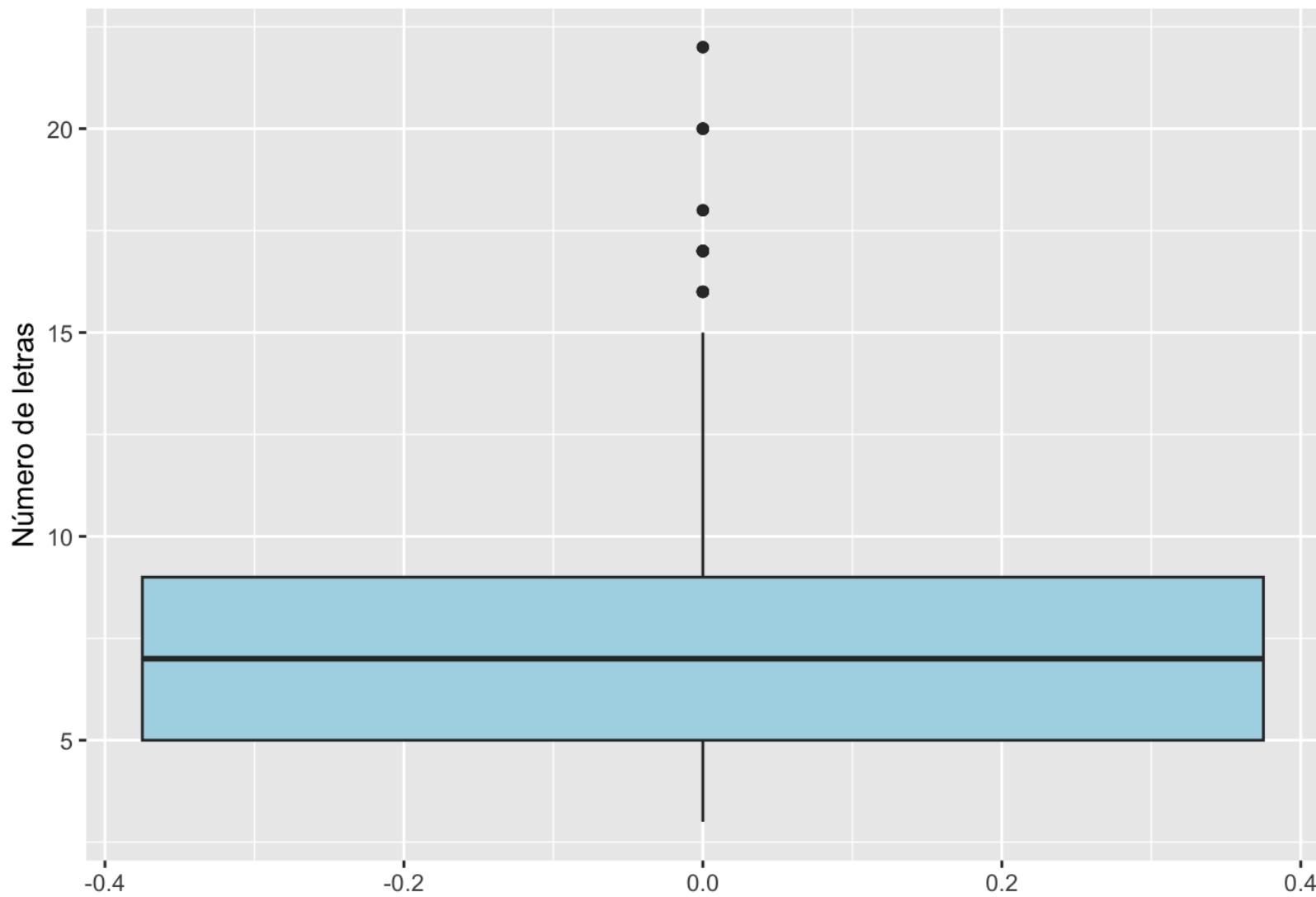
### Frequênciа de Dígrafos por Palavra



## Relação entre Vogais e Consoantes



### Boxplot do Tamanho das Palavras



# Interpretação dos resultados

- Palavras semânticas revelam temas centrais da narrativa;
- A distribuição de tamanhos mostra preferências estilísticas do autor;
- Padrões de frequência reforçam recorrências temáticas;
- Visualizações ajudam a observar tendências que passariam despercebidas na leitura intuitiva.

# Outras possibilidades de análise

- Outras categorias linguísticas (classes gramaticais, estrutura silábica) – utilizando outros pacotes do R;
- Comparação com outros contos de Machado de Assis;

# Possíveis contribuições

- Aproxima ferramentas computacionais da análise linguística;
- Exibe potencial de visualização de dados aplicados à linguística e à literatura;
- Demonstra o uso do R como ferramenta de pesquisa;
- Mostra como métodos quantitativos podem complementar a interpretação qualitativa.

# Considerações finais

- O uso do R permitiu automatizar etapas do processamento textual;
- As análises revelaram padrões relevantes no texto;
- O trabalho integra conhecimentos de linguística, estatística e programação;
- Demonstra o caráter interdisciplinar e aplicável da modelagem computacional.

Obrigada!