

Minicurso

Análise de Sentimentos com R

Christian e Jennifer

Universidade Federal do Rio Grande do Norte
Semana da Estatística

27 de novembro de 2025

Agenda

- 1 Introdução à Análise de Sentimentos
- 2 Metodologia e Ferramentas (R)
- 3 Exemplos Práticos: Visualizando Dom Casmurro
- 4 Exemplos Práticos: Visualizando Górgias
- 5 Desafios e Conclusão

O que é Análise de Sentimentos?

Definição

É uma técnica de Processamento de Linguagem Natural (PLN) que identifica e extrai informações subjetivas de textos.

Características principais:

- Classificação de polaridade: Positivo, Negativo ou Neutro.
- Identificação de emoções (Léxico NRC): Raiva, Alegria, Confiança, etc.
- Transformação de literatura e diálogos em Séries Temporais.

Por que é Importante?

Para Empresas

- Monitorar reputação de marca.
- Entender satisfação do cliente (CX).
- Prevenção de crises de PR.

Para Humanidades Digitais

- Análise de estrutura narrativa.
- Evolução de personagens.
- Detecção de ironia e conflito em diálogos clássicos.

Níveis de Análise

- 1. Nível de Documento:** Classifica o livro/texto inteiro.
- 2. Nível de Sentença:** Analisa frase por frase (Ideal para diálogos).
- 3. Nível de Aspecto:** Identifica sentimentos sobre características específicas (ex: "O argumento é bom, mas o orador é ruim").

Como funciona a mágica?

Baseada em Léxico (Nossa Abordagem) Usa dicionários (Bing, NRC) com pesos pré-definidos.

Vantagem: Transparente e não exige treinamento.

Desvantagem: Falha com contexto e ironia.

Machine Learning Treina classificadores (Naive Bayes, SVM).

Deep Learning Redes Neurais (BERT, LLMs).

Ferramentas: O Poder do R

Pacote R	Função Principal	Uso no Curso
tidytext	unnest_tokens()	Limpeza e Tokenização
syuzhet	get_nrc_sentiment()	Emoções e Polaridade
ggplot2	geom_line()	Visualização de Séries
wordcloud	wordcloud()	Nuvens de Palavras

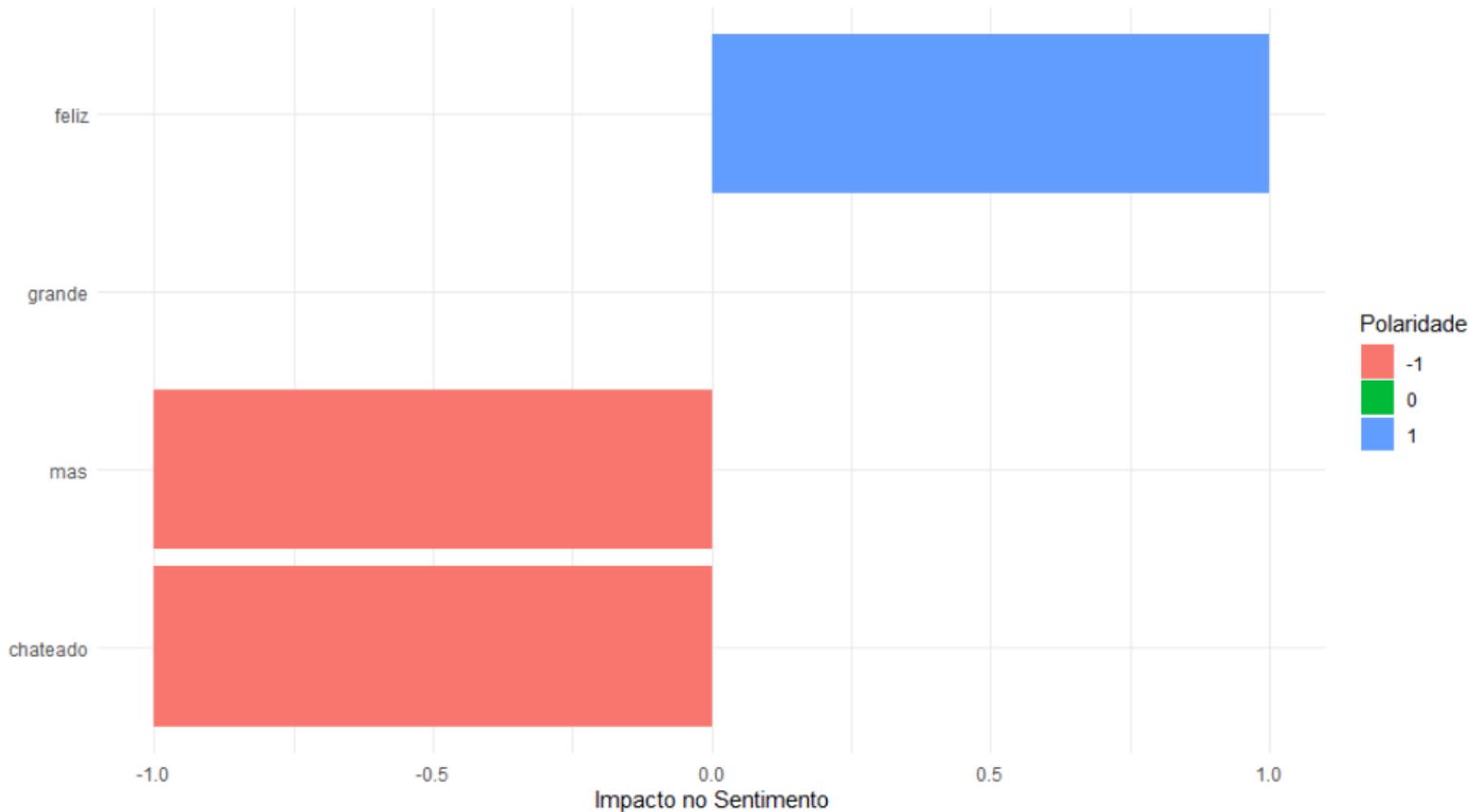
Abordagem Tidytext e OpLexicon

```
1 library(tidyverse); library(tidytext); library(lexiconPT)
2
3 # 1. Dados (Trecho de Eclesiastes)
4 texto <- texto <- c(
5   "Estou muito feliz com o resultado, foi um sucesso!",
6   "Mas o atraso foi um grande problema e estou bem chateado"
7 )
8 df_texto <- tibble(id = 1, texto = texto)
9
10 # 2. Tokeniza o (Quebra em palavras)
11 palavras <- df_texto |>
12   unnest_tokens(word, texto)
13
14 # 3. Cruzamento com o Lexico (Inner Join)
15 sentimentos <- palavras |>
16   inner_join(oplexicon_v3.0, by = c("word" = "term"))
```

Visualizando a Contribuição no Sentimento

```
1 # Contribuição de cada palavra para o saldo final  
2 sentimentos |>  
3   count(word, polarity, sort = TRUE) |>  
4   mutate(word = reorder(word, n*polarity)) |>  
5   ggplot(aes(word, n*polarity, fill = factor(polarity))) +  
6   geom_col() +  
7   coord_flip() +  
8   labs(y = "Impacto no Sentimento", x = NULL, fill = "Polaridade") +  
9   theme_minimal()
```

Resultado



Dom Casmurro

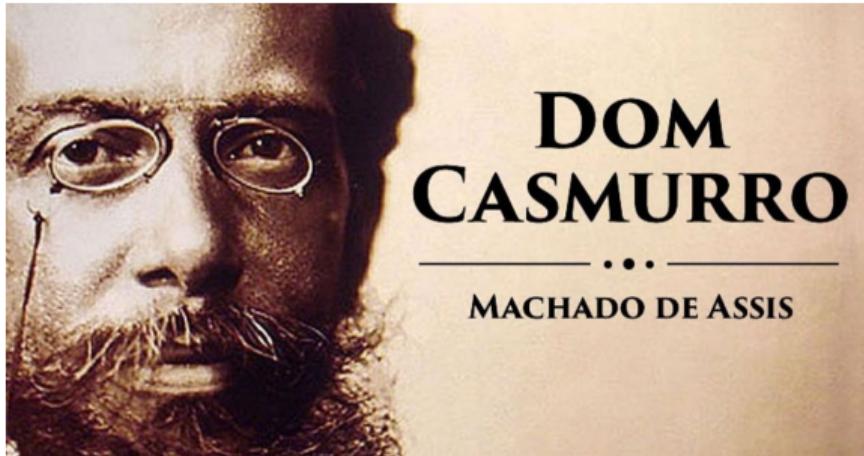
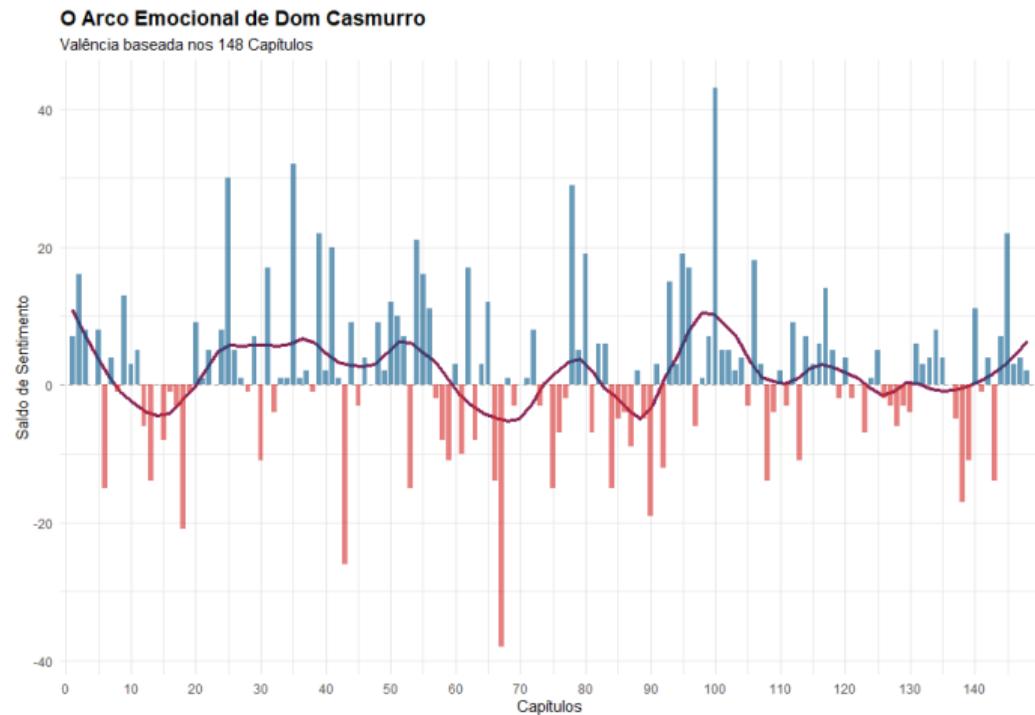


Figura: Machado de Assis, 1899

O Arco Emocional da Narrativa

Visão Macroscópica

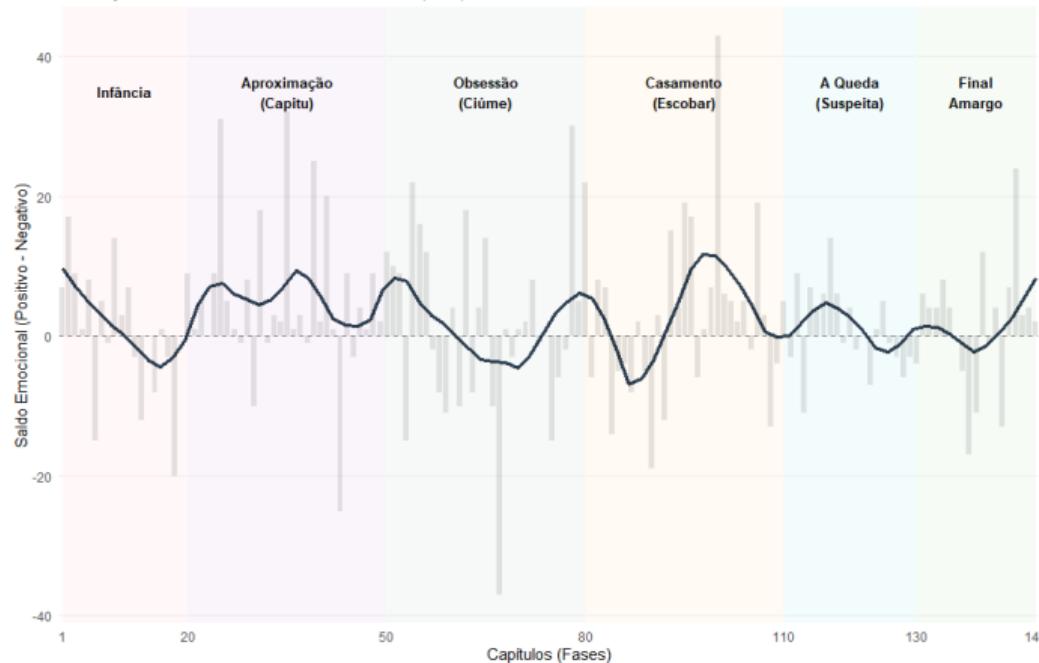
- Visualização dos 148 capítulos.
- Barras representam o saldo emocional (Positivo - Negativo).
- A linha de tendência suaviza o ruído, revelando a "respiração" do livro.
- Note a oscilação constante entre euforia e melancolia.



A Anatomia do Ciúme

A Anatomia do Ciúme: Fases de Dom Casmurro

Correlação entre a narrativa e a valéncia emocional (NRC)



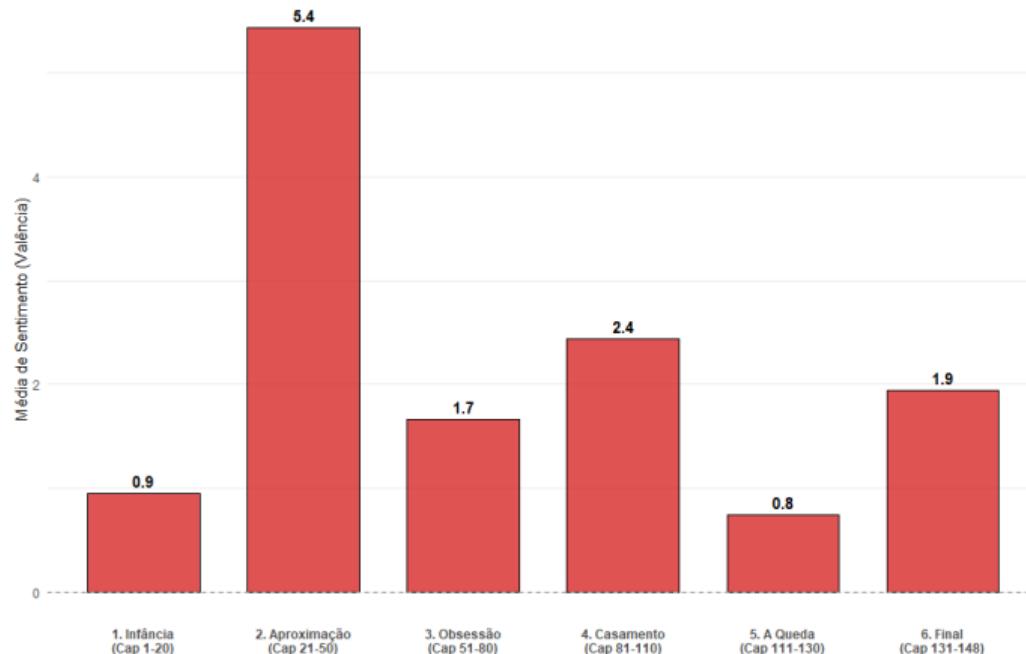
Segmentação Temporal

- **Infância/Aproximação:** Predomínio de picos positivos.
- **Obsessão:** O início da queda e instabilidade.
- **A Queda:** O vale profundo no final do livro (morte de Escobar e exílio de Capitu).

O Declínio do Sentimento

Média Emocional por Fase Narrativa

Dom Casmurro: O declínio do sentimento ao longo da trama



Queda dramática de **5.4** (Aproximação) para **0.8** (A Queda).

Górgias



Platão, c. 380 a.C.

O Vocabulário do Diálogo

profissional experimental oportunidade
médico disseste combate
compromete pergunta
preciso couro belo parte
estiveres festa certo coisa irmão
gentileza
preferir tes ano culpa falar
tes expoção arte modo novo
resto dano amigo acordo expo
vida resto mundo filho exemplo
momento atrasado pessoa cansado casa
compreendes nós
inexperiência podes nome presente
preferir

contrário alma grande idade
filosofia indivíduo forte capaz
cheio vez fraco modo
cidade lei coisa inferior
mau bom natureza tonel filho
filho cheio cidade inferior tua
relação homem feio felizes
caro vida caso justiça
relação certo certo belo

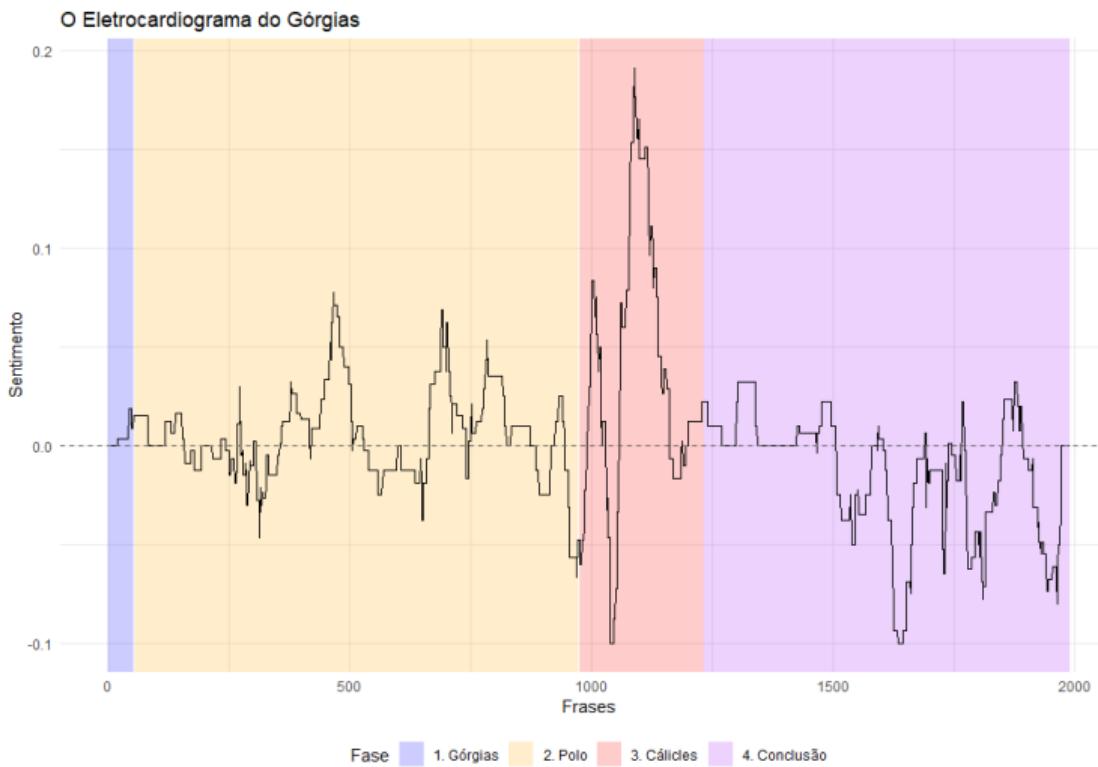
homem médico tirano
força bem feliz
fato feio alma
justo grande
dúvida mal parte
vista certo arte bom
vez coisa caso mau
mesmo belo
ponto nome cidade
orador pessoa

retórica justiça entanto contrário
respeito certo injustiça menos
ordem nome verdade acordo
poder modo orador sede
morte arte pessoa vista
modo amigável caro justo mau caso
amigo alma coisa cidade o
agradável dor bem corpo tempo
relação grande deus fato prazer cidadão
político mesmo vez vida
condição possível mal próprio
preciso maneira público indivíduo

■ Fase Górgias ■ Fase Polo

■ Fase Cálices ■ Fase Monólogo

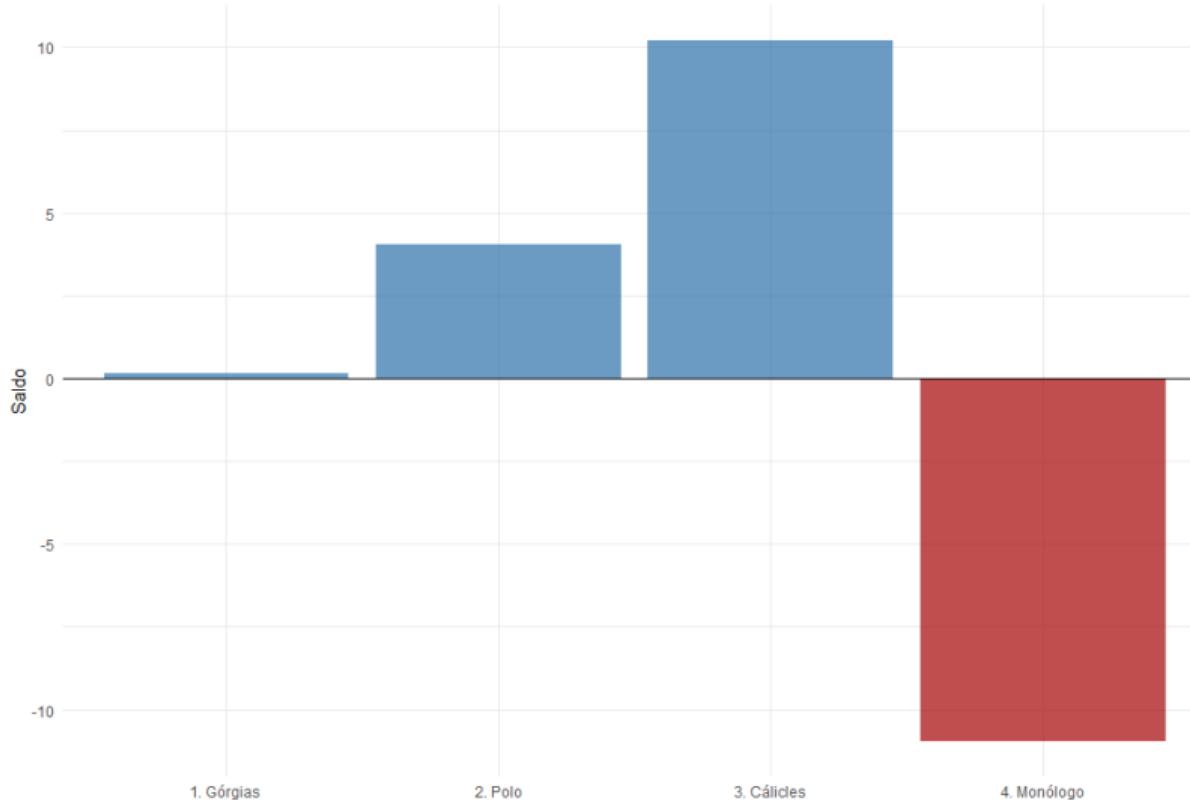
O "Eletrocardiograma" do Diálogo



Dinâmica Emocional

- **Fases Iniciais (Górgias/Polo):** Relativa estabilidade. O debate é técnico sobre a definição de retórica.
- **A Irrupção de Cálices:** O momento de maior tensão e picos emocionais. Cálices ataca a filosofia de Sócrates apaixonadamente.
- **O Monólogo Final:** Sócrates assume a fala sozinho. A volatilidade diminui e o tom se torna grave.

Saldo Emocional Total por Fase



- **Ironia e Sarcasmo:** O grande vilão da análise léxica (ex: "Marcela amou-me durante quinze meses e onze contos de réis").
- **Contexto Histórico:** Palavras mudam de sentido (ex: "Vício" e "Virtude" em Platão vs. Hoje).

Conclusão

- A estatística nos permite "ler" livros inteiros em segundos (Macroanálise).
- As visualizações revelam estruturas emocionais invisíveis à leitura comum.
- **Cuidado:** A análise quantitativa não substitui a interpretação humana.

**Dados sem contexto são apenas números.
Com contexto, são histórias.**

FIM!