

# Graduação & Pós-Graduação em Ciência Política

## INTRODUÇÃO À ANÁLISE DE TEXTOS COM A LINGUAGEM DE PROGRAMAÇÃO R.

**Carga Horária: 03 Horas**

**Coordenadores: Felipe Borba e Steven Ross**

**Professor: Steven Dutt-Ross**

## PLANO DE ENSINO

### Ementa

Os métodos digitais trouxeram uma nova abordagem para os processos de pesquisa em Ciências Sociais e Humanas. Nesse minicurso, apresentaremos um dos principais novos métodos: **Análise de Texto** como dado não estruturado. Assim, vamos analisar de forma prática os textos por meio de uma abordagem quantitativa. Discutiremos as principais abordagens de análise de texto. Os(as) alunos(as) irão utilizar dados e informações reais e serão capazes de construir visualizações de dados, a partir de registros brutos, além de avaliarem as características desses textos para extrair *insights* úteis.

### Objetivo

O objetivo central do curso é promover a capacidade de criar códigos para análise de texto utilizando a linguagem de programação R.

### Programação

#### 1. Motivação (25 minutos)

1. Matérias
2. Estudos de casos

#### 2. Termos e Conceitos (35 minutos)

1. Introdução ao texto como dado
2. Conceitos Básicos: stopwords, corpus, token, lemma e stem

### 3. Aplicação no R (2 horas)

1. Introdução/revisão de R
2. Carregar dados de texto no R
3. Limpeza de texto 4. Aplicação de Conceitos Basicos
5. Análise de palavras: Frequência e Nuvem
6. Relacionamento entre palavras: Bigramas e Trigramas
7. Análise de palavras: Tf-idf
8. Análise de Emoji

#### Bibliografia básica

ROSS, Steven Dutt; CRUZ, Breno de Paula Andrade. Análise quantitativa de textos: Apresentação e operacionalização da técnica via Twitter. *Administração: Ensino e Pesquisa*, v. 22, n. 1, 2021. Disponível em <https://raep.emnuvens.com.br/raep/article/view/1859/419>

SILGE, Julia; ROBINSON, David. *Text mining with R: A tidy approach*. " O'Reilly Media, Inc.", 2017. Disponível em <https://www.tidytextmining.com/index.html>

WICKHAM, H. and GROLEMUND, G. *R for data science: import, tidy, transform, visualize, and model data*. " O'Reilly Media, Inc.", 2016.

#### Bibliografia complementar

1. SAMPAIO, Rafael Cardoso et al. Muita Bardin, pouca qualidade: uma avaliação sobre as análises de conteúdo qualitativas no Brasil. *Revista Pesquisa Qualitativa*, v. 10, n. 25, p. 464-494, 2022.
2. FILGUEIRA, L. *WordCloud no R – Estatística é com R! UFF*. Disponível em <http://www.estadisticacomr.uff.br/?p=322> Publicado em julho 2, 2016. Acesso em 12/07/2022.

Para saber mais, visite os sites:

<http://giel.uniriotec.br/> &

<https://dataunirio.github.io/>