

R para análise de dados - MQuinho de Inverno

Rômulo Damasceno

5/22/2021

Apresentação

Bom dia/tarde/noite! Meu nome é Rômulo Damasceno, serei o professor do curso de R para análise de dados, e te agradeço por escolher o MQuinho!

Saiba que ao nos dar essa chance você está contribuindo com um dos projetos mais bonitos do curso de Ciências Sociais (e da UFMG), que busca espalhar conhecimento de qualidade à comunidade por um valor bastante acessível :)

O projeto do curso:

Se você já fez um curso de Introdução ao R (se não fez, não se preocupe), é possível que tenha se sentido perdido e sobrecarregado com a quantidade de informações, funções e comandos que parecem vir do nada. Isso também aconteceu comigo, e é por isso que nesse curso decidi apostar numa nova abordagem.

Meu plano é que você saia deste curso sabendo caminhar no processo de análise de dados com R, e acredito que para que isso seja possível, você precisa adquirir três competências fundamentais:

1. Fazer boas perguntas para os dados, tendo clareza sobre aquilo que você quer descobrir antes de partir para a programação de fato;
2. Através dos procedimentos técnicos (a programação em si), ser capaz de obter as respostas que deseja;
3. Produzir visualizações e interpretações adequadas dos dados, que sejam capazes de construir uma narrativa que faça sentido ao leitor.

O esquema da aula consistirá em uma primeira parte demonstrativa das funções, caminhando passo a passo no que diz respeito às suas origens e usos, e uma segunda parte de aplicação destas num projeto real. A ideia é que o final do curso tenhamos um projeto completo, escrito em formato de relatório que será o compilado de tudo o que foi construído ao longo do processo.

O que aprenderemos nessa capacitação?



O Tidyverse é o conjunto de pacotes mais popular do R. Nele, várias (se não todas) as funcionalidades necessárias para análise de dados estão condensadas. Quase a totalidade do curso será feita tendo o Tidyverse como referência.

Introdução e questões fundamentais

1. Conhecendo o R e o RStudio
2. As possibilidades do R
3. Quem são as mentes por trás do programa?
4. Por quê o R e não o Excel, SPSS, SAS, Stata, etc?
5. A iniciativa Tidyverse
6. A programação e a necessidade de estruturação linear do pensamento para fazê-la.

Trazendo o dado para o R

1. Formatos de dados
2. Importação de dados para o ambiente R com os pacotes **readr** e **readxl**
3. Checagem de consistência e características gerais do banco

Análise exploratória (EDA)

Queries e transformação do banco de dados - Dplyr

1. Verbos do Dplyr - `select()` , `filter()` , `mutate()`, `arrange ()` e `summarise()`

Variáveis numéricas

1. Medidas de posição: Quartis, médias e medianas
2. Medidas de dispersão: Variância, desvio padrão e amplitude

Variáveis categóricas

1. Tabelas de frequência: contagens e percentuais
2. Análise exploratória com o pacote DataExplorer

Estética de tabelas - Kable e KableExtra

1. Introdução ao pacote KableExtra

Gráficos com o ggplot2

1. A Gramática de Gráficos
2. Quais os componentes de um ggplot? Por que isso importa?
3. Construindo gráficos: `geom_bar()`, `geom_col()`, `geom_histogram()`, `geom_boxplot` e `geom_point`
4. Estética gráfica.

Construção de relatórios com o RMarkdown