

Programação e análise de dados com R

Do introdutório à autonomia intermediária

Victor Gabriel Alcantara
victorgalcantara@usp.br
PPGS/USP

Este curso aborda o funcionamento da linguagem de programação R e do software de desenvolvimento RStudio/Posit, visando oferecer noções gerais e operações básicas para o domínio e a autonomia na análise de dados com programação. O curso envolve operações de importação, análise, visualização, exportação e comunicação de dados, e está estruturado seguindo os eixos do livro de Wickham e Grolemund (2018). Por seguir o livro de referência, os tópicos mesclam com o inglês para estimular a adaptação ao software, à comunidade de programadores internacionais e aos manuais, que têm a língua inglesa como padrão.

Objetivo geral: oferecer condições para a autonomia na programação e no uso do software, por isso os fundamentos são abordados com cuidado. É esperado que se saiba, ao final do curso, trabalhar com essa linguagem de programação para operar com o instrumental básico da Ciência de Dados, que no fundo são aplicações da estatística e econometria.

Dinâmica: o curso é online com aulas síncronas. As aulas serão acompanhadas de materiais que estarão disponibilizados em minha página do GitHub. Serão recomendados exercícios para praticar o uso da linguagem, verificar dúvidas e exercitar a autonomia.

Pré-requisitos teóricos: interesse em aprender e trabalhar com análise de dados com programação. Não é exigido nenhum conhecimento em R, matemática, estatística muito menos inglês.

Pré-requisitos técnicos: É necessário um computador com configurações padrão (4GB RAM e estrutura 64bits) e uma conexão estável com a internet. É importante ter instalado o R e o RStudio no computador. É recomendado a integração na plataforma GitHub, atualmente a mais utilizada por programadores (não se preocupe, será apresentada na aula inicial). Os materiais para o curso estarão disponíveis em minha página do GitHub.

Inscrições: até o dia 30 de abril pelo formulário

1 Explore: basic knowledge and workflow

Noções gerais: é de comer?

História R e RStudio

Input, processamento e output

Fluxo de trabalho no RStudio: janelas, linguagens suportadas e convenções para o desenvolvimento de códigos

Práticas comunitárias: help, Stack Overflow e GitHub

2 Explore: logic operations and basic R

Mão na massa

Operações básicas, comandos e lógica

Classes de objetos: vetores, arrays, matrizes, listas e data.frames/tibbles

Subset: navegação em objetos

Funções e pacotes: tidyverse e ggplot

3 Import and tidy data

Dados estruturados

Import: importação de dados

Tidy and Transform: manuseio dos dados (filter, select, rename, mutate, group by etc.)

4 Analyse

Estatísticas descritivas

Variáveis categóricas nominais e ordinais

Retomando factor

Tabela de frequência

Variáveis categóricas tipo likert

Média

Variáveis métricas discretas e contínuas

Média, Mediana e Moda

5 Visualize

Gráficos

Exportação de dados e resultados de análises

Variáveis categóricas nominais e ordinais

Barras e Pizza

Variáveis métricas discretas e contínuas

Barras e histogramas

Histogramas e densidade

Análise Bivariada

Definição conceitual: o que significa analisar duas variáveis?

Categórica e Categórica

Tabela de contingência ou cruzada

Estatística Qui-Quadrado

Categórica e Métrica

Barras, boxplot e densidades

Diferença entre médias

Métrica e Métrica

Dispersão

Correlação de Pearson

Métodos para exportação dos resultados

6 Model

Análise de dados IV - Multivariada

Regressão linear simples pelo Método dos Mínimos Quadrados (MQO)

Regressão linear múltipla pelo Método dos Mínimos Quadrados (MQO)

7 Communicate

Rmarkdown

Chunks