Programa¸ca˜o e an´alise de dados com R

Da introdu¸ca˜o a` autonomia intermedi´aria

Victor Gabriel Alcantara [victorgalcantara@usp.br](mailto:victorgalcantara@usp.br) PPGS/USP

O curso aborda o funcionamento da linguagem de programa¸c˜ao R e do software de desenvolvimento RStudio/Posit, visando oferecer no¸c˜oes gerais e opera¸c˜oes b´asicas para o dom´ınio e a autonomia na an´alise de dados com pro- grama¸c˜ao. O curso envolve opera¸c˜oes de importac˜ao, an´alise, visualiza¸c˜ao, exporta¸c˜ao e comunica¸cao de dados, e est´a estruturado seguindo os eixos do livro de [Wickham e Grolemund (2018)](https://r4ds.had.co.nz/)[**wickham2017**]. Os t´opicos mes- clam com o inglˆes para estimular a adapta¸c˜ao ao software, a` comunidade de programadores internacionais e aos manuais, que tˆem a l´ıngua inglesa como padr˜ao.

**Objetivo geral**: oferecer condi¸c˜oes para a autonomia na programa¸c˜ao

com R e no uso do software RStudio.

E´ esperado que se saiba, ao final do

curso, trabalhar com R para operar com o instrumental b´asico da Ciˆencia de Dados, que s˜ao aplica¸c˜oes da estat´ıstica e econometria.

**Dinˆamica**: online com aulas s´ıncronas e materiais disponibilizados em minha p´agina do GitHub. Ser˜ao recomendados exerc´ıcios para praticar o uso da linguagem, verificar du´vidas e exercitar a autonomia.

**Pr´e-requisitos te´oricos**: interesse em aprender e trabalhar com an´alise de dados com programa¸c˜ao. Na˜o ´e exigido nenhum conhecimento em R, matem´atica, estat´ıstica muito menos inglˆes.

**Pr´e-requisitos t´ecnicos**: E´ necess´ario um computador com configura¸c˜oes padr˜ao (4GB RAM e estrutura 64bits) e uma conex˜ao est´avel com a internet.

E´ importante ter instalado o [R](https://brieger.esalq.usp.br/CRAN/) e o [RStudio](https://www.rstudio.com/products/rstudio/download) no computador. E´ recomendado

a integra¸c˜ao na plataforma [GitHub](https://github.com/), atualmente a mais utilizada por progra- madores (n˜ao se preocupe, ser´a apresentada na aula inicial). Os materiais para o curso estar˜ao dispon´ıveis em minha p´agina do [GitHub](https://github.com/victorgalcantara?tab=repositories).

**Inscri¸c˜oes**: at´e o dia 30 de abril pelo formul´ario

# Explore: basic knowledge and workflow No¸c˜oes gerais: ´e de comer?

Histo´ria R e RStudio [**ihaka1996**][**ihaka1998**]

Input, processamento e output

Fluxo de trabalho no RStudio: janelas, linguagens suportadas e con- ven¸c˜oes para o desenvolvimento de c´odigos

Pr´aticas comunit´arias: help, Stack Overflow e GitHub

# Explore: logic operations and basic R M˜ao na massa

Opera¸c˜oes b´asicas, comandos e l´ogica

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

Subset: navega¸c˜ao em objetos

Fun¸c˜oes e pacotes: tidyverse e ggplot

# Import and tidy data Dados estruturados

Import: importa¸c˜ao de dados

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

# Analyse Estat´ısticas descritivas

Vari´aveis categ´oricas nominais e ordinais

* + Tabela de frequˆencia

Vari´aveis categ´oricas tipo likert

* + M´edia

Vari´aveis m´etricas discretas e cont´ınuas

* + M´edia, Mediana e Moda

# Visualize Gr´aficos

GGplot: A Grammar of Graphics [**wickham2016**]

Vari´aveis categ´oricas nominais e ordinais

* + Barras e Pizza

Vari´aveis m´etricas discretas e cont´ınuas

* + Barras, boxplot, histogramas e densidade

# An´alise Bivariada

Defini¸c˜ao conceitual: o que significa analisar duas vari´aveis?

Categ´orica e Categ´orica

* + Tabela de contingˆencia ou cruzada
  + Estat´ıstica Qui-Quadrado e V de Cramer

Categ´orica e M´etrica

* + Boxplot, histogramas, densidade e Diferen¸ca entre m´edias

M´etrica e M´etrica

* + Dispers˜ao
  + Correla¸c˜ao de Pearson

# Model

**An´alise Multivariada**

Regress˜ao linear simples pelo M´etodo dos M´ınimos Quadrados (MQO)

Regress˜ao linear mu´ltipla pelo M´etodo dos M´ınimos Quadrados (MQO)

# Communicate Rmarkdown