

Introdução à Programação em R

Ementa

O curso pretende oferecer uma iniciação à lógica de programação em linguagem R para os alunos(os) e comunidade UNIFESP. Em específico, propõe-se apresentá-las(los) e treiná-las(los) no uso de ferramentas computacionais para (i) coleta, (ii) organização, (iii) preparação e (iv) apresentação de dados. Neste sentido, não se trata de um curso de estatística, nem de metodologia, mas sim, de iniciação em uma linguagem de programação por meio do aprendizado de suas potencialidades na análise exploratória de dados.

Dentro desta perspectiva, o curso abordará desde o elementar da linguagem (instalação do software e reconhecimento do layout e suas opções) até aplicações mais complexas como a raspagem de dados da web e elaboração de mapas. De toda maneira, o foco está na “alfabetização” do participante na linguagem R (operações matemáticas, tipos de dados e objetos, operadores relacionais e lógicos, loops e funções) a partir da manipulação de dados em data-frames.

Objetivo

O curso pretende oferecer uma iniciação à lógica de programação em linguagem R para os alunos(os) e comunidade UNIFESP. Em específico, propõe-se apresentá-las(los) e treiná-las(los) no uso de ferramentas computacionais para (i) coleta, (ii) organização, (iii) preparação e (iv) apresentação de dados.

Justificativa

Programação é uma habilidade destacada na atualidade, seja para o exercício profissional, seja para a pesquisa acadêmica. O conhecimento de ferramentas computacionais facilita e potencializa a análise e a manipulação de dados em razão da escalabilidade e reprodutibilidade características deste conhecimento. Assim sendo, são processos especialmente valorados, tanto pelo mercado, quanto pela produção científica.

Neste diapasão, a linguagem R destaca-se por cinco razões principais: (i) Trata-se de uma iniciativa de código livre - sem a exigência de pagamento por licenças ou patentes; (ii) Embora já possua uma ampla e ativa comunidade, a iniciativa continua a crescer. Recentemente, inclusive, *Microsoft* e *Oracle* começaram a desenvolver projetos nesta linguagem, ampliando a oferta de pacotes e funcionalidades; (iii) Embora as funcionalidades que mais se destacam são a modelagem, visualização e análise de

dados, o R não é somente um software estatístico especializado (como é o caso do *Stata-SAS*), mas sim, uma linguagem versátil de alto nível; e (iv) Promove integração com outras ferramentas e linguagens de maneira descomplicada; (v) a comunidade se preocupa em dar respostas a problemas sociais - projetos que adotam a linguagem costumam ter foco diverso e inclusivo.

Forma de avaliação

As(os) participantes serão avaliadas(os) por meio de exercícios de fixação de conteúdo e um projeto final a ser entregue em até 10 dias após o último encontro presencial.

Metodologia de ensino

Os encontros serão compostos de uma parte teórica introdutória, seguida da resolução de tutoriais pelos participantes. Durante esta segunda parte, as dúvidas serão debatidas e resolvidas.

A relação ensino-aprendizagem é fundamentada na necessidade da prática desde o primeiro encontro. As listas de verificação e os tutoriais se destacam como formas incrementais de fixação e aplicação dos conteúdos estudados.

Conteúdo Programático

1. Preparação do ambiente computacional: R e RStudio.
2. Gramática R: Operações matemáticas, tipos de dados, vetores, matrizes, fatores, data-frames e listas.
3. Alfabetização em R base: Operadores relacionais e lógicos, condicionais, loops e funções. (I)
4. Alfabetização em R base: Operadores relacionais e lógicos, condicionais, loops e funções. (II)
5. O pacote *Tidyverse*: Manipulação de data-frames em R, tabelas e gráficos(I)
6. O pacote *Tidyverse*: Visualização de dados e o pacote *ggplot2* (II)
7. Aplicações: Raspagem de dados na internet
8. Aplicações: Soluções via API de terceiros (basedosdados)
9. Aplicações: Mineração de textos com R
10. Aplicações: Mapas e GIS

Cronograma

- 20.02.2024 – **Preparação do ambiente computacional** (ministrante: Vinícius)
21.02.2024 – **Gramática R** (ministrante: Vinícius)
22.02.2024 – **Alfabetização R base** (ministrante: Vinícius)
23.02.2024 – **Alfabetização R base** (ministrante: Vinícius)

24.02.2024 – **Pacote *Tidyverse*** (ministrante: Vinícius)
26.02.2024 – **Pacote *Tidyverse*** (ministrante: Vinícius)
27.02.2024 – **Aplicações: Raspagem** (ministrante: Vinícius)
28.03.2024 – **Aplicações: Soluções via API** (ministrante: Vinícius)
29.03.2024 – **Aplicações: Mineração de textos** (ministrante: Vinícius)
01.03.2024 – **Aplicações: Mapas e GIS** (ministrante: Vinícius)

Estratégia de divulgação

Pretende-se divulgar o curso mediante a fixação *folder* no site da universidade, disparo de e-mails para os discentes e pôsteres afixados na universidade.

Bibliografia

- Aquino, Jakson Alves de (2014). R para cientistas sociais. Editus.
- Navarro, Danielle; Pedersen, Thomas Lin e Wickham, Hadley (2009). *ggplot2: Elegant Graphics for Data Analysis*. Ed: Springer.
- Guerra, Saulo; Oliveira, Paulo Felipe de e McDonnell, Robert (2018). *Ciência de dados com R: Uma Introdução*. Disponível em: <https://cdr.ibpad.com.br/cdr-intro.pdf>
- Grolemund, Garrett (2014). *Hands-On Programming with R*. Ed: O'Reilly Media.
- Imai, Kosuke (2017). *Quantitative Social Science: An Introduction*. Ed: Princeton University Press
- Imai, Kosuke e Laudete, Elena (2023). *Data analysis for social sciences: A friendly and practical approach*. Ed: Princeton University Press.
- Silge, Julia e Robinson, David (2017). *Text Mining with R: A Tidy Approach*. Ed: O'Reilly Media.
- Wickham, Hadley e Grolemund, Garrett (2016). *R for Data Science*. Ed: O'Reilly Media.
- Wickham, Hadley (2014). *Advanced R*. Ed: Chapman and Hall/CRC