
Nome da Disciplina: R para Ciência de Dados

Código: ECN932

Turma: D e E

Classificação: Optativa

Carga Horária: 60 horas/aula

Créditos: 4 créditos

Curso: Economia

Pré-requisito: Nenhum

Período Letivo: 1º Semestre de 2023

Horários: 4ª – 09h20 às 12h50

Professor: Lucas Carvalho, Alexandre Stein e Rafael Ribeiro

EMENTA

Introdução ao R, criação de projetos e organização de scripts, estrutura de dados, tipos de vetores. Manipulação, limpeza, visualização e análise descritiva de dados. Ferramentas básicas de programação. Introdução a conceitos e ferramentas de geoprocessamento e visualização de dados espaciais. Aplicações diversas de R para ciência de dados.

OBJETIVOS

O objetivo da disciplina é fornecer aos alunos ferramentas para utilização da linguagem R para ciência de dados. Mais especificamente, busca-se promover habilidades na manipulação, limpeza e visualização de dados, realização de análise descritiva de dados e introdução a ferramentas de geoprocessamento. Além disso, serão introduzidas aplicações diversas utilizando R para trabalhos quantitativos.

CONTEÚDO PROGRAMÁTICO

- I. Introdução ao R.
- II. Manipulação, limpeza e visualização de dados.
- III. Análise descritiva de dados.
- IV. Programação básica.
- V. Ferramentas de geoprocessamento e manipulação de dados espaciais.
- VI. Aplicações diversas.

ESTRATÉGIAS E PROCEDIMENTOS DE AVALIAÇÃO

Quatro exercícios básicos: 10 pontos cada

Um exercício completo: 20 pontos

Trabalho final de análise de dados: 40 pontos

CRONOGRAMA DE AULAS

Data	Tema	Pacotes
08/mar	Abertura da disciplina, introdução ao R, plots básicos.	ggplot2
15/mar	Criação de projetos e organização de scripts. Workflow basics, estrutura de dados, tipos de vetores. Manipulação de dados com Rbase.	Rbase
22/mar	Importação de dados e geração de dados aleatórios. Manipulação de dados dplyr 1.	readxl, tibble, dplyr, data.table
29/mar	Manipulação de dados com dplyr 2.	dplyr
05/abr	Visualização e apresentação de dados.	ggplot2, ggside, ggrepel, ggbetweenstats, patchwork
12/abr	Limpeza de dados, expressões regulares, strings e variáveis de tempo.	dplyr, lubridate
19/abr	Manipulação de grandes bases de dados.	data.table
26/abr	Extra - Trabalho e exercícios	-
03/mai	Programação básica 1	R base
10/mai	Programação básica 2	purrr
17/mai	Introdução a conceitos e ferramentas de geoprocessamento.	sf, geobr
24/mai	Ferramentas de visualização de dados espaciais.	tmap, ggplot, esquisser
31/mai	Aplicação de ferramentas de geolocalização.	-
07/jun	<i>Application Programming Interface (APIs)</i>	-
14/jun	<i>Webscrapping</i>	-
21/jun	Automatização de coleta de dados via FTP.	-
28/jun	Análise de dados qualitativos	-
05/jul	Extra - Trabalho e exercícios	-

BIBLIOGRAFIA

✓ Wickham, Hadley, and Garrett Golemund. *R for data science: import, tidy, transform, visualize, and model data.* " O'Reilly Media, Inc.", 2016.

Lovelace, Robin, Jakub Nowosad, and Jannes Muenchow. *Geocomputation with R.* Chapman and Hall/CRC, 2019.

Pebesma, Edzer, and Roger Bivand. "Spatial Data Science: With Applications in R." (2022).