




Universidade Federal da Paraíba
Centro de Ciências Sociais Aplicadas
Departamento de Economia

PROGRAMA DA DISCIPLINA
INTRODUÇÃO À CIÊNCIA DE DADOS NO R
CALENDÁRIO 2024.2

Código:	PSAE000174
C.H./Créditos:	60h/04
Local:	Sala CCSA 209
Horários:	Terça-feira e Quinta-feira, 10h
Professor:	Antonio Vinícius B. Barbosa
Email:	antonio Barbosa@ccsa.ufpb.br

© **Objetivo:** Esta disciplina introduz as principais técnicas computacionais em R para a leitura, manipulação, visualização e reprodução de dados econômicos. Com o advento da tecnologia da informação e o rápido crescimento da quantidade de dados disponíveis, é indispensável o uso de ferramentas estatísticas para uma análise econômica robusta e eficiente. Utilizando-se do fato de o R ser uma ferramenta estatística poderosa, gratuita, extensível e extremamente flexível, o curso busca desenvolver habilidades que permitam a estruturação de modelos teóricos e empíricos em Economia de uma maneira unificada. Ao final do curso, espera-se que o aluno seja capaz de:

1. Ler e manipular dados no R;
2. Computar estatísticas descritivas dos dados;
3. Entender como visualizar dados e apresentar informações estatísticas;
4. Criar gráficos específicos através do pacote ggplot2

 **Ementa:** Os principais tópicos do curso incluem:

- Introdução e Conceitos Básicos
- Introdução ao R
- Estruturas de dados
- Leitura e exportação de dados
- Manipulação de dados
- Visualização de dados
- Programação básica em R

📁 **Conteúdo Programático:** Os principais tópicos do curso incluem:

■ **Introdução e Conceitos Básicos**

■ Conceito de computação com dados; Novas competências e paradigmas; O uso de linguagens de programação para Economia; Por que utilizar o R?; Exemplos de análise de dados.

■ **Introdução ao R**

■ O ambiente estatístico R; Instalação do R; Ambientes de desenvolvimento integrado (IDE); RStudio; Instalação e gerenciamento de pacotes; Obtendo ajuda; Operações básicas.

■ **Estruturas de dados**

■ Trabalhando com dados; Tipos de dados no R; Vetores: criação e manipulação; Entendendo e manipulando matrizes; Arrays; Entendendo e criando Data Frames; Listas no R; Operações lógicas e condicionais; Manuseando dados faltantes; Aplicações: Bolsa de Valores; Índices de Preços ao Consumidor.

■ **Leitura e exportação de dados**

■ Geradores de números aleatórios; Leitura de dados em diferentes formatos; Salvando objetos em R; Leitura de objetos no formato do R; Salvando o ambiente de trabalho; Exemplos e aplicações de leitura de dados.

■ **Manipulação de dados**

■ Criação de variáveis; Conversão de classes; Ordenação dos dados; Junção (merge) de dados; Manipulação de variáveis contendo datas; Manipulando de variáveis contendo strings; *Tidying data*: utilizando os pacotes `dplyr` e `tidyr`.

■ **Visualização de dados**

■ Gráficos básicos no R: gráfico de barras, gráfico de pizza, histogramas, densidades, box plots, gráfico de dispersão, gráfico de linhas, correlogramas, mosaicos. Introdução ao pacote `ggplot2`; Gráficos interativos; Visualizando dados em Mapas.

■ **Programação básica em R**

■ Introdução aos *loops*; Expressões e condições; Introdução a funções; Criando funções básicas no R; Exemplos e aplicações.

📖 **Metodologia:** A metodologia adotada é baseada em aulas expositivas do conteúdo previsto, na apresentação e discussão de textos complementares, e na resolução de listas de exercícios. As aulas serão interativas, com a participação dos alunos expressando opiniões e apresentando dúvidas durante sua realização. A principal linguagem utilizada nesta disciplina será o R (r-project.org), juntamente com o ambiente de desenvolvimento integrado RStudio (<https://posit.co/>). As atividades práticas de laboratório cobrirão os assuntos discutidos durante a exposição do conteúdo e introduzirá tópicos adicionais.

Literatura sugerida:

- BRAUN, W. John; MURDOCH, Duncan J. A First Course in Statistical Programming with R. Cambridge University Press, 2016.
- DAYAL, Vikram. An Introduction to R for Quantitative Economics: Graphing, Simulating and Computing, Springer Briefs in Economics, Springer (India), 2015.
- ROCHA, M.; FERREIRA, F. G. Análise e Exploração de Dados com R. FCA, 1ª Edição. 2017.
- TEETOR, Paul. R cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics. O'Reilly Media, Inc., 2011.
- WICKHAM, Hadley; GROLEMUND, Garrett. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media, Inc., 2016.
- CHAMBERS, John. Software for Data Analysis: Programming with R. Springer Science Business Media, 2008.
- KABACOFF, R. I.; ACTION, R. In. Data Analysis and Graphics with R. 2011.
- MATLOFF, Norman. The Art of R programming: A Tour of Statistical Software Design. No Starch Press, 2011.
- XIE, Yihui et al. Tufte: Tufte's styles for R Markdown documents. In: R Markdown: The Definitive Guide. Vienna, Austria: Chapman and Hall/CRC, 2019. p. 1-45.
- RStudio Cheat Sheets (<https://www.rstudio.com/resources/cheatsheets/>)

✓ Os alunos serão informados ao longo do curso em relação à referências complementares.

Material: Todo o material utilizado ao longo do curso (slides, scripts, tutoriais e listas de exercícios) serão disponibilizados na página do curso.

Comunicação: Toda a comunicação entre alunos e professor sobre os elementos do curso (atendimento, dúvidas, listas, soluções e notícias) será realizada exclusivamente através da página da disciplina no **Sigaa** (<https://sigaa.ufpb.br/sigaa/>).

Avaliação: A avaliação na disciplina será dada através de listas de exercícios e de uma avaliação final, de acordo com a seguinte função:

$$NF = f(\text{Listas de Exercício}, \text{Projeto}).$$

A nota final será uma média ponderada, de acordo com os seguintes pesos:

Descrição	Peso
Listas de Exercício	40%
Projeto	60%

Atendimento Extraclasse: Horário de atendimento às **quartas-feiras**, 09h-11h, e às **sextas-feiras**, 14h-16h, na **Sala 05** dos professores do CCSA. O agendamento deverá ser feito exclusivamente através da página https://calendly.com/antonio_barbosa