

# Lego I

Fernando Meireles

Monitor: Carlos Freitas

📍 Sala Olavo Brasil

📅 Quartas-feiras, 9h-12h

🔗 [fmeireles.com/lego](https://fmeireles.com/lego)

## Sobre

**Lego I** é uma introdução à Ciência Social quantitativa. O curso apresenta os fundamentos da análise de dados, incluindo noções sobre desenho de pesquisa, probabilidade e inferência, testes de hipóteses e modelos de regressão linear. Complementarmente, o curso também introduz o uso de programação em R, um ambiente amplamente utilizado para análise de dados e *data science* dentro e fora da academia. A dinâmica do curso combinará tutoriais e sessões guiadas de monitoria nas quais resolveremos problemas aplicados de pesquisa social de forma reprodutível.

## Objetivos

O objetivo central do curso é equipar alunos e alunas para a realização de análises quantitativas de dados – o que, de quebra, significa capacitá-los para acessar o crescente mercado de trabalho para cientistas de dados e para cientistas sociais quantitativos fora da academia. Até o final do curso, espera-se que estudantes consigam:

- 1) Usar programação para carregar ou extrair, manipular, analisar e visualizar dados de diferentes tipos e fontes, de pequenas planilhas de Excel a bases com milhares ou milhões de entradas.
- 2) Entender a utilidade dos fundamentos da estatística, incluindo noções de amostragem (i.e., como entrevistar 2 mil pessoas nos permite inferir sobre uma população de milhões), mensuração de fenômenos, resumo de dados, teste de hipóteses e relações entre variáveis.
- 3) Interpretar e avaliar criticamente análises quantitativas de dados, como as publicadas em periódicos acadêmicos.

4) Produzir e apresentar de forma reprodutível os resultados de suas próprias análises quantitativas utilizando ferramentas como **Zotero**, **quarto** e **LaTeX**, todos *softwares* de código aberto que nos ajudarão a automatizar diferentes etapas da escrita acadêmica.

## Público-alvo

O curso é *obrigatório* para alunos e alunas de pós-graduação ingressantes na Sociologia ou na Ciência Política do IESP-UERJ, mas também é aberto a pessoas de outras instituições – dentro da disponibilidade de vagas remanescentes. O curso não prevê pré-requisitos, embora espera-se que estudantes tenham feito o *nivelamento* em programação em R oferecido na primeira semana de aulas.

## Atividades complementares

### Nivelamento

Aprender programação requer prática frequente, especialmente no início. É por esta razão que o curso oferece um *nivelamento* em programação em R na primeira semana de aulas composto por sessões expositivas *remotas*, de cerca de 3 horas de duração, para além de atividades complementares. Vale notar: *o nivelamento é essencial para o aproveitamento do curso*, especialmente para ingressantes no IESP-UERJ. Ao final do nivelamento, teremos uma lista de exercícios para prática do conteúdo visto que poderá ser entregue para obtenção de pontos extras na nota final do curso.

### Sessões de monitoria

Para além das aulas presenciais, **Lego I** contará com sessões de monitoria semanais *remotas* para tirar dúvidas sobre o conteúdo visto em aula. A participação nas monitorias é obrigatória e será computada como parte da frequência formal no curso.

## Leituras

Em cada aula teremos *leituras obrigatórias*, que devem ser realizadas *previamente* aos nossos encontros; e *materiais complementares*, úteis para aprofundamento e para apresentar aplicações práticas em R. Neste semestre, adotaremos principalmente o seguintes livros como base:

### *Fundamentos de pesquisa e estatística*

- King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton university press.

- Kellstedt, P. M., Whitten, G. D., & Tuch, S. A. (2022). *The fundamentals of social research*. Cambridge University Press.

### *Análise de dados em R*

- **Usando R: Um Guia para Cientistas Sociais**
- Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science*. O'Reilly Media, Inc. <https://r4ds.had.co.nz/>

## **Avaliação**

**Lego I** adota um sistema de avaliação contínuo, que premia o esforço e a dedicação dos alunos e alunas ao longo do curso. Temos dois grandes conjuntos de avaliação: i) *realização de, pelo menos, 3 listas de exercícios*; e, ii), um *pequeno projeto final de análise de dados*. Para além destes, o curso prevê duas formas de obter *pontos extras*, isto é, pontos que podem compensar eventuais dificuldades em alguma das avaliações.

### **Listas de exercícios (60 pontos)**

Para praticar o conteúdo visto, teremos *4 listas de exercícios*. A entrega das dos exercícios realizados deverá ser feita por meio de um modelo que será disponibilizado no início da disciplina que alternará código e explicações por escrito. A avaliação levará em conta a capacidade de implementar o conhecimento visto no curso e o esforço aplicado na tarefa – e não a obtenção de um resultado específico.

Para cálculo da nota final, serão consideradas as 3 melhores notas obtidas nas listas de exercícios, isto é, *estudantes que realizarem os quatro exercícios poderão descartar aquele no qual obtiverem a menor nota*. Cada lista de exercícios considerada corresponderá a *20 pontos da nota final do curso*.

### **Projeto final (40 pontos)**

O projeto final consistirá em um pequeno exercício prático de análise de dados. Para ele, alunos e alunas, individualmente ou em duplas, deverão escolher um tema de pesquisa e um base de dados de sua preferência que permita responder uma pergunta de pesquisa social. Há várias possibilidades aqui: explorar uma base de dados original por meio de visualizações e estatísticas descritivas; usar uma base de dados de outra pesquisadora para replicar seus achados; ou, ainda, aproveitar a oportunidade para fazer a parte empírica de um trabalho de outra disciplina.

O projeto deverá ser escrito usando o *template* que veremos no início do curso e deverá ter, *no máximo*, 15 páginas, incluindo tabelas, gráficos e bibliografia. Na avaliação, serão considerados a capacidade de aplicar o conteúdo do curso; a organização e estrutura do documento; e a qualidade das visualizações, tabelas e estatísticas utilizadas.

## Pontos extras (10 pontos extras)

### Participação em aulas e monitorias (5 pontos extras)

Este curso recompensará com pontos extras as pessoas que frequentarem as aulas com assiduidade e que se engajarem nas atividades propostas, especialmente sessões de monitoria e nivelamento, cuja frequência é obrigatória. Vale dizer: o esforço dedicado para aprender e para ajudar colegas a aprender será valorizado.

### Lista de exercícios do nivelamento (5 pontos extras)

Ao final do *nivelamento*, será entregue uma lista de exercícios que também valerá 5 *pontos extras para a nota final do curso*. A lista será composta por exercícios de programação em R.

## Software

Como dito, adotaremos programação em R neste curso. Trata-se de uma das principais linguagens de programação usadas na academia e **uma das 10 mais utilizadas no mundo** e que, além disso, conta com inúmeras extensões (pacotes, como veremos) que implementam diferentes funcionalidades e ferramentas estatísticas, incluindo várias úteis para análise de dados em Ciências Sociais. Fora o R, também usaremos o **Rstudio**, que é uma interface que facilitará o nosso trabalho. Ambos podem ser instalados em:

- **R e RStudio Desktop**

Para um tutorial mais detalhado, em vídeo, sobre como instalar o R e o **RStudio** no *Windows*, **está disponível neste vídeo**. Seja pela razão que for, caso tenha problemas ao instalar o R e/ou RStudio uma alternativa será usar o RStudio Cloud, uma versão online do RStudio que dispensa instalação e que pode ser acessada pelo navegador. Para acessar o RStudio Cloud, basta criar uma conta em

- <https://posit.cloud/>

Para um tutorial em vídeo sobre como usar o RStudio Cloud, **veja este vídeo**.

## Políticas

### Política de Gênero

Em aulas de metodologia, homens frequentemente monopolizam a participação. Para evitar isso, seguiremos dois protocolos neste curso: para intervir, é necessário estender a mão; quando mulheres falam, colegas não as interrompem.

## Atendimento a Necessidades Especiais

Se você, por qualquer razão que seja, tenha necessidades ou solicitações individuais, não hesite em procurar auxílio, tanto por **e-mail** quanto pessoalmente. Faremos o melhor possível para tornar este curso um ambiente de aprendizado acessível e inclusivo.

## Plágios

Apresentar texto ou código de outra pessoa como se fosse próprio, inclusive em trabalhos finais de disciplina, viola **normativas da nossa universidade**, o que pode resultar em sanções. Embora a colaboração entre colegas seja estimulada neste curso, evite copiar códigos inteiros de outra pessoa – listas de exercícios serão checadas automaticamente para identificar similaridades.

## Plano das Aulas

### Parte I – Nivelamento

*Leituras obrigatórias:*

- **Dia 1** – Usando R, Básico, Cap. 1
- **Dia 2** – Usando R, Importação de dados, Cap. 2
- **Dia 3** – Usando R, Manipulação de dados, Cap. 3
- **Dia 4** – Usando R, Visualizações, Cap. 4
- **Dia 5** – Usando R, Estatísticas descritivas e regressão, Cap. 5

*Leituras sugeridas:*

- Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science*. O'Reilly Media, Inc. <https://r4ds.had.co.nz/>, *The whole game*, Caps. 1–8

### Parte II – Programação

#### Aula 1 – Fundamentos da pesquisa quantitativa

*Leituras obrigatórias:*

- Kellstedt, P. M., & Whitten, G. D. (2018). *The fundamentals of political science research*. Cambridge University Press., Cap. 1, 2 & 3

#### Aula 2 – Desenho de pesquisa e mensuração

*Leituras obrigatórias:*

- King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton university press., Cap. 1

- Kellstedt, P. M., & Whitten, G. D. (2018). *The fundamentals of political science research*. Cambridge University Press., Cap. 4

*Leituras sugeridas:*

- Abdulai, R. T., & Owusu-Ansah, A. (2014). Essential ingredients of a good research proposal for undergraduate and postgraduate students in the social sciences. *Sage Open*, 4(3), 2158244014548178.
- Nicolau, J. (2013). Breve roteiro para redação de um projeto de pesquisa. *Revista Estudos Políticos*, 4(7), 345–353.

### **Aula 3 – Visualização de dados**

*Leituras obrigatórias:*

- Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science*. O'Reilly Media, Inc. <https://r4ds.had.co.nz/>, *Visualize*, Caps. 9 & 11

*Leituras sugeridas:*

- Healy, K. (2018). *Data visualization: A practical introduction*. Princeton University Press., Caps. 1, 3 & 4
- Kastellec, J. P., & Leoni, E. L. (2007). Using graphs instead of tables in political science. *Perspectives on Politics*, 5(4), 755–771.

### **Aula 4 – Manipulação de dados**

*Leituras obrigatórias:*

- Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science*. O'Reilly Media, Inc. <https://r4ds.had.co.nz/>, *Visualize*, Caps. 9 & 11

### **Aula 5 – Importação e combinação de bases de dados**

*Leituras obrigatórias:*

- Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science*. O'Reilly Media, Inc. <https://r4ds.had.co.nz/>, Caps. 19 & 20

### **Aula 6 – Indicadores e estatísticas descritivas**

*Leituras obrigatórias:*

- Kellstedt, P. M., & Whitten, G. D. (2018). *The fundamentals of political science research*. Cambridge University Press., Cap. 6

## Parte III – Estatística

### Aula 7 – Amostragem

#### *Leituras obrigatórias:*

- Babbie, E. R. (2020). *The practice of social research*. Cengage AU., Cap. 7
- Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (Vol. 561). John Wiley & Sons., Cap. 4

#### *Leituras sugeridas:*

- Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (Vol. 561). John Wiley & Sons., Caps. 1 & 2
- Çetinkaya-Rundel, M., & Hardin, J. (2021). *Introduction to modern statistics*. OpenIntro. <https://openintro-ims.netlify.app/>, Cap. 2

### Aula 8 – Inferência randomizada

#### *Leituras obrigatórias:*

- Çetinkaya-Rundel, M., & Hardin, J. (2021). *Introduction to modern statistics*. OpenIntro. <https://openintro-ims.netlify.app/>, Cap. 11–12

### Aula 9 – Distribuições de probabilidade

#### *Leituras obrigatórias:*

- Kellstedt, P. M., & Whitten, G. D. (2018). *The fundamentals of political science research*. Cambridge University Press., Cap. 7
- Çetinkaya-Rundel, M., & Hardin, J. (2021). *Introduction to modern statistics*. OpenIntro. <https://openintro-ims.netlify.app/>, Cap. 13

### Aula 10 – Teste de hipóteses

#### *Leituras obrigatórias:*

- Kellstedt, P. M., & Whitten, G. D. (2018). *The fundamentals of political science research*. Cambridge University Press., Cap. 8
- Çetinkaya-Rundel, M., & Hardin, J. (2021). *Introduction to modern statistics*. OpenIntro. <https://openintro-ims.netlify.app/>, Cap. 16–18

## **Aula 11 – Comparando grupos**

*Leituras obrigatórias:*

- Agresti, A. (2018). *Statistical methods for the social sciences*. Pearson., Cap. 7

## **Aula 12 – Relações entre variáveis**

*Leituras obrigatórias:*

- Mesquita, E. B. de, & Fowler, A. (2021). *Thinking clearly with data: A guide to quantitative reasoning and analysis*. Princeton University Press., Cap. 2
- Agresti, A. (2018). *Statistical methods for the social sciences*. Pearson., Cap. 8

## **Aula 13 – Inferência com regressão linear**

*Leituras obrigatórias:*

- Kellstedt, P. M., & Whitten, G. D. (2018). *The fundamentals of political science research*. Cambridge University Press., Cap. 9
- Mesquita, E. B. de, & Fowler, A. (2021). *Thinking clearly with data: A guide to quantitative reasoning and analysis*. Princeton University Press., Cap. 5

## **Aula 14 – Regressão linear multivariada**

*Leituras obrigatórias:*

- Kellstedt, P. M., & Whitten, G. D. (2018). *The fundamentals of political science research*. Cambridge University Press., Cap. 10–11

## **Aula 15 – Auxílio nos projetos finais**