

# Lego I

Fernando Meireles

Monitor: Guilherme de Abreu Duque

📍 Sala Olavo Brasil

📅 Quartas-Feiras, 9h às 12h

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

## Sobre o curso

**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. Paralelamente, 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 escalável e reproduzível.

## 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 e na imprensa.

4) Produzir e apresentar de forma reprodutível os resultados de suas próprias análises quantitativas utilizando ferramentas como **Zotero**, **quarto** e **L<sup>A</sup>T<sub>E</sub>X**, 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. O curso não prevê pré-requisitos, embora se espere que estudantes tenham feito o *nivelamento* em programação em R oferecido na semana anterior ao início formal das 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 semana anterior ao início formal das aulas composto por sessões expositivas *remotas*, de cerca de 2 horas de duração às manhãs, para além de monitoria para prática de atividades no turno da noite, de cerca de 1 hora. 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. Por ser também um curso de extensão, sua conclusão valerá um certificado de 15 horas de formação complementar.

### 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 e auxiliar na resolução de exercícios e tarefas. A participação nas monitorias será computada como parte da avaliação de participação 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 os 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*

- Meireles, F., & Silva, D. (2024). *Usando r: Uma introdução para pesquisadores em humanidades digitais*. Editora FGV. Uma versão preliminar do livro pode ser vista [aqui](#).
- 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 4 listas de exercícios*; e, ii), uma *prova final* sobre o conteúdo visto ao longo do curso. Para além destes, o curso prevê duas formas de obter *pontos extras*, isto é, pontos que podem compensar eventuais dificuldades nas demais avaliações.

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

Para praticar o conteúdo visto, teremos *4 listas de exercícios*. A entrega 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*.

### **Prova (40 pontos)**

A prova final cobrirá todo o conteúdo visto no curso. Seu formato será misto: teremos questões objetivas e uma questão discursiva, que servirá para avaliar a capacidade de usar análise quantitativa de dados para estruturar um projeto de pesquisa social. Na avaliação desta questão serão considerados a capacidade de aplicar o conteúdo do curso; e a organização e estrutura do desenho de pesquisa proposto.

## 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, contribuírem com comentários e dúvidas e que se engajarem nas atividades propostas, especialmente sessões de monitoria e nivelamento. 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.

## Nota final

A nota final será calculada pela soma dos pontos obtidos nas listas de exercícios e na prova final, acrescida dos pontos extras eventualmente obtidos. A pontuação máxima possível, assim, é de *110 pontos*, o que garante o teto de *100 pontos* – dessa forma, alguém que tiver *90 pontos* na soma das listas e da prova final e *10 pontos extras*, por exemplo, ainda assim terá nota final de *100 pontos*.

## 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 Posit Cloud, uma versão online do RStudio que dispensa instalação e que pode ser acessada pelo navegador. Para acessar o Posit Cloud, basta criar uma conta em

- <https://posit.cloud/>

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

# Políticas

## Uso de LLMs (ChatGPT)

O uso de LLMs (como o **ChatGPT** ou o **Gemini**) é permitido neste curso, mas é importante usá-los apenas *para complementar o seu aprendizado em aula*. Em vez de apenas colar o enunciado de uma questão no chat e delegar totalmente a sua resolução (o que eu muito provavelmente descobrirei), tente primeiro resolvê-la por conta própria. Se a consulta a textos ou a mim à monitoria não for suficiente ou possível, aí, sim, use LLM como uma espécie de tutor. Exemplo: “Estou tentando implementar uma simulação para calcular intervalos de confiança via bootstrap. O código que fiz até agora está em anexo, mas retorna o seguinte erro: MENSAGEM DE ERRO. Você pode me explicar didaticamente qual é o problema?”. **Este guia** tem algumas outras dicas que podem ser úteis para fazer bom uso de LLMs na pós-graduação. Importante notar que, mesmo seguindo boas práticas, LLMs não são livres de erros e podem fornecer informações incorretas.

## 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 administrativas. 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.

## Política de Gênero

Em aulas de metodologia, especialmente em turmas grandes como as típicas deste curso, homens frequentemente monopolizam a participação. Para evitar isso, seguiremos dois protocolos neste curso: para intervir, é necessário estender a mão; quando alguém fala, colegas evitam fazer interrupções.

## Atendimento a Necessidades Especiais

Se você, por qualquer razão que seja, tiver 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.

# 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., & Tuch, S. A. (2022). *The fundamentals of social research*. Cambridge University Press., Cap. 1, 2 & 3

#### *Leituras sugeridas:*

- Imai, K., & Williams, N. W. (2022). *Quantitative social science: An introduction in tidyverse*. Princeton University Press., Cap. 1

### 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., & Tuch, S. A. (2022). *The fundamentals of social 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
- Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis* new york. NY: Springer. <https://ggplot2-book.org/>
- 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/>, *Whole game*, Caps. 3 & 5

### **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., & Tuch, S. A. (2022). *The fundamentals of social 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., & Tuch, S. A. (2022). *The fundamentals of social 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., & Tuch, S. A. (2022). *The fundamentals of social 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., & Tuch, S. A. (2022). *The fundamentals of social 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., & Tuch, S. A. (2022). *The fundamentals of social research*. Cambridge University Press., Cap. 10–11

## **Aula 15 – Prova final**