

Lego I

Bruno Marques Schaefer

Atualizada em: 19 de março de 2025

Ementa

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:

- 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.
- Entender a utilidade dos fundamentos da estatística, incluindo noções de amostragem (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

- Interpretar e avaliar criticamente análises quantitativas de dados, como as publicadas em periódicos acadêmicos.
- 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 e/ou híbridas 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. Monitoria: Carlos Freitas

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:

- LLAUDET, Elena; IMAI, Kosuke. Data analysis for social science: A friendly and practical introduction. Princeton University Press, 2022.
- KELLSTEDT, Paul M.; WHITTEN, Guy D. Fundamentos da pesquisa em ciência política. Editora Blucher, 2021.
- WICKHAM, Hadley; ÇETINKAYA-RUNDEL, Mine; GROLEMUND, Garrett. R for data science. "O'Reilly Media, Inc.", 2023.
- **Usando R: Um Guia para Cientistas Sociais**

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 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)

0.0.1 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.

0.0.2 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/>
- **Tutorial**

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 (brunoschaefer@iesp.uerj.br) 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 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.

GPT e outros LLM

O uso de LLMs (como o ChatGPT) 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 não for suficiente ou possível, aí, sim, use LLM como uma espécie de tutor. Importante notar que, mesmo assim, LLMs não são livres de erros e podem fornecer informações incorretas.

Parte 1 - Nivelamento

Leituras obrigatórias:

- Dias 10 à 14 de Março de 2025
- Dia 1 Usando R, Básico, Cap. 1
- Dia 2 Usando R, Básico, Cap. 2
- Dia 3 Usando R, Básico, Cap. 3
- Dia 4 Usando R, Básico, Cap. 4
- Dia 5 Usando R, Básico, Cap. 5

Leituras sugeridas:

- Wickham, H., Çetinkaya-Rundel, M., e Golemund, G. (2023). R for data science. O'Reilly Media, Inc. <https://r4ds.had.co.nz/>, The whole game, Caps. 1–8
- Caps 1 - 8
- Llaudet, E. and Imai, K. (2022). *Data analysis for social science: A friendly and practical introduction*. Princeton University Press
- Introdução

Parte 2 - Desenho de Pesquisa e Programação

Aula 1 – Fundamentos da pesquisa quantitativa 18/03

Leituras obrigatórias:

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

Leituras e Materiais Complementares:

- Chen, Yunsong, et al. "Social prediction: a new research paradigm based on machine learning." *The Journal of Chinese Sociology* 8 (2021): 1-21.
- Silva, G. P. d. (2018). *Desenho de pesquisa*
- Capítulo 2 - Panorama Geral das Ciências sociais
- Discriminação esperada e busca por emprego - Causalidade
- Forecasting - Predição
- Desigualdade racial e acesso a creches - descrição
- Refining your research topic
- Como definir um problema de pesquisa

Aula 2 – Desenho de pesquisa e mensuração 25/03

Leituras obrigatórias:

- Kellstedt, P. M.,; Whitten, G. D. (2018). *The fundamentals of political science research*. Cambridge University Press., Caps. 3 e 4
- King, G., Keohane, R. O., Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton university press., Cap. 1

Leituras e Materiais Complementares:

- Abdulai, R. T., e 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.
- Figueiredo Filho, D. B., Paranhos, R., Rocha, E. C., Silva Jr, J., e Santos, M. D. (2012). Levando Gary King a sério: desenhos de pesquisa em Ciência Política. *Revista Eletrônica de Ciência Política*, 3(1-2), 86-117.
- Nicolau, J. (2013). Breve roteiro para redação de um projeto de pesquisa. *Revista Estudos Políticos*, 4(7), 345–353.
- Problema de Pesquisa

Aula 3 – Importação, combinação e tratamento de dados 01/04

Leituras obrigatórias:

- Wickham, H., Çetinkaya-Rundel, M., and Grolemund, G. (2023). *R for data science*. O'Reilly Media, Inc
- Caps. 7 a 11

Aula 4 – Visualização de dados 08/04

Leituras obrigatórias:

- Usando R, Básico, Cap. 4
- KASTELLEK, Jonathan P.; LEONI, Eduardo L. Using graphs instead of tables in political science. *Perspectives on politics*, v. 5, n. 4, p. 755-771, 2007.

Leituras e Materiais Complementares:

- Wickham, H. (2009). *ggplot2: elegant graphics for data analysis* new york. NY: Springer
- Caps. 13 a 20
- Modern Data Visualization with R
- Galeria de Gráficos

Aula 5 – Indicadores e estatísticas descritivas 15/04

Leituras obrigatórias:

- Kellstedt, P. M. and Whitten, G. D. (2018). *The fundamentals of political science research*. Cambridge University Press
- Cap. 5

Leituras e Materiais Complementares:

- Conceito e mensuração de democracia
- Goertz, G. (2008). Concepts, theories, and numbers: A checklist for constructing, evaluating, and using concepts or quantitative measures.

- Gerring, J. (1999). What makes a concept good? A criterial framework for understanding concept formation in the social sciences. *Polity*, 31(3), 357-393.
- Sartori, G. (1970). Concept misformation in comparative politics. *American political science review*, 64(4), 1033-1053.
- COLLIER, David; MAHON JR, James E. Conceptual “stretching” revisited: Adapting categories in comparative analysis. *American political science review*, v. 87, n. 4, p. 845-855, 1993.
- **Estatísticas descritivas via modelsummary**

Parte 3 - Estatística

Aula 6 – Efeitos causais em experimentos 22/04

Leituras obrigatórias:

- Llaudet, E. and Imai, K. (2022). *Data analysis for social science: A friendly and practical introduction*. Princeton University Press
- Cap. 2 “Estimating causal...”

Leituras e Materiais Complementares:

- **Como evidência empírica afeta decisões sobre políticas públicas Auditorias Públicas Discriminação esperada e busca por emprego**
- **Experimentos em políticas públicas**

Aula 7 – Inferência 29/04

Leituras obrigatórias:

- Llaudet, E. and Imai, K. (2022). *Data analysis for social science: A friendly and practical introduction*. Princeton University Press
- Cap. 3 “Inferring...”
- Kellstedt, P. M., e Whitten, G. D. (2018). *The fundamentals of political science research*. Cambridge University Press., Cap. 6

Leituras e Materiais Complementares:

- Meireles, F., e Russo, G. (2022). Pesquisas eleitorais no Brasil: tendências e desempenho. *Estudos Avançados*, 36, 117-131.
- **Pesquisas Eleitorais**
- **PNAD e Censo Calculadora de amostra**
- 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., and Tourangeau, R. (2009). *Survey methodology*, volume 561. John Wiley & Sons
- Cap. 4

Aula 8 – Predição 06/05

Leituras obrigatórias:

- Llaudet, E. and Imai, K. (2022). *Data analysis for social science: A friendly and practical introduction*. Princeton University Press
- Cap. 4 "Predicting Outcomes..."

Leituras e Materiais Complementares:

- Chen, Yunsong, et al. "Social prediction: a new research paradigm based on machine learning." *The Journal of Chinese Sociology* 8 (2021): 1-21.
- **Predizendo eleições no segundo turno**
- **Escolaridade e renda**
- Verhagen, M. D. (2022). A pragmatist's guide to using prediction in the social sciences. *Socius*, 8, 23780231221081702.

Aula 9 – Estimando efeitos causais com dados observacionais 13/05

Leituras obrigatórias:

- Llaudet, E. and Imai, K. (2022). *Data analysis for social science: A friendly and practical introduction*. Princeton University Press

- Cap. 5 "Estimating causal effects with..."

Leituras e Materiais Complementares:

- Desafios da avaliação de políticas
- DAGs
- DAGs 2.0
- Literatura em inferência causal

Aula 10 – Probabilidade 20/05

Leituras obrigatórias:

- Llaudet, E. and Imai, K. (2022). *Data analysis for social science: A friendly and practical introduction*. Princeton University Press
- Cap. 6 "Probability"

Leituras e Materiais Complementares:

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

Aula 11 – Quantificando a incerteza 27/05

Leituras obrigatórias:

- Llaudet, E. and Imai, K. (2022). *Data analysis for social science: A friendly and practical introduction*. Princeton University Press
- Cap. 7 "Quantifying uncertainty"

Leituras e Materiais Complementares:

- Shmueli, G. (2010). *To explain or to predict?*

Aula 12 – Teste de hipótese 03/06

Leituras obrigatórias:

- Kellstedt, P. M., e Whitten, G. D. (2018). The fundamentals of political science research. Cambridge University Press., Cap. 7

Leituras complementares

- Çetinkaya-Rundel, M., e Hardin, J. (2021). Introduction to modern statistics. OpenIntro. <https://openintro-ims.netlify.app/>, Cap. 11
- Mesquita, E. B. de, e Fowler, A. (2021). Thinking clearly with data: A guide to quantitative reasoning and analysis. Princeton University Press., Cap. 2

Aula 13 – Regressão de novo 10/06

Leituras obrigatórias:

- Kellstedt, P. M., e Whitten, G. D. (2018). The fundamentals of political science research. Cambridge University Press., Cap. 8
- Mesquita, E. B. de, e Fowler, A. (2021). Thinking clearly with data: A guide to quantitative reasoning and analysis. Princeton University Press., Cap. 5

Aula 14 – Regressão com mais de uma variável 17/06

Leituras obrigatórias:

- Kellstedt, P. M., e Whitten, G. D. (2018). The fundamentals of political science research. Cambridge University Press., Caps. 9-11

Leituras e Materiais Complementares:

- Shmueli, G. (2010). To explain or to predict?.

Aula 15 – Auxílio nos projetos finais 24/06

Materiais complementares

- LITMAP - Encontre evidências científicas
- Ferramentas de IA para pesquisa

- 16 sites de pesquisa acadêmica que farão você esquecer do Google
- Mapa da literatura
- Mapa de citações das Ciências Sociais
- Gerenciamento de Referências Bibliográficas
- SciSpace - IA para revisões
- Berkeley Initiative for Transparency in the Social Sciences
- O que são pré-prints?
- Ética em IA
- Editoria de Replicabilidade
- Replication Crisis
- Ética em pesquisa
- Pré-Análise
- Transparência e replicabilidade - PodMétodos
- Breve Manual de Escrita e Comunicação
- 25 dicas para revisar textos acadêmicos
- Escrevendo um resumo
- Um guia para escrever profissionalmente na Ciência Política
- Scientific Writing Made Easy
- Exemplos de resumos