

Ciência de Dados para RI: Análise e Visualização

20 a 24 de julho de 2020 – 18h-23h

27 a 31 de julho de 2020 – 18h-23h

Ementa

O curso aborda técnicas exploratórias e inferenciais para análise de dados a partir de linguagens de programação. A análise exploratória é necessária para o refinamento das hipóteses a serem investigadas através dos dados. Por sua vez, a inferência estatística é o processo pelo qual são tiradas conclusões e realizadas predições a partir dos dados. Através de bases de dados reais, serão abordadas técnicas estatísticas multivariadas e modelos de regressão linear usuais na ciência de dados. Somadas às habilidades de análise de dados, serão conhecidas as melhores práticas para apresentação de resultados tanto em formato gráfico, cartográfico e de relatórios automatizados.

Objetivos

Apresentar o fluxo de trabalho do cientista de dados a partir de exemplos práticos aplicados às ciências sociais e relações internacionais. Ao final, o aluno será capaz de criar projetos, fazer seu versionamento e apresentar resultados por meio de relatórios automatizados.

Pré-requisito

Não há.

Professores

Davi Cordeiro Moreira: Professor Adjunto da Escola de Relações Internacionais da Fundação Getulio Vargas (FGV-RI). Vencedor do prêmio CAPES de melhor tese de doutorado em Ciência Política e Relações Internacionais em 2017. Suas pesquisas focam nas seguintes áreas: Metodologia em Ciência Política, Ciência de Dados, Text as Data e Política Comparada.

Rafael Nunes Magalhães: Cientista de dados e consultor. Doutor em Ciência Política pela Universidade de São Paulo e pesquisador do Caeni/USP. Possui ampla experiência profissional em projetos que utilizam ciência de dados para a tomada de decisão. Entre as instituições nas quais atuou estão: Unicef, Ibope e IBRE/FGV.

> PROFESSORES

Davi Moreira

Professor Adjunto da Escola de Relações Internacionais - FGV

Webpage: www.davimoreira.com

Email: davi.moreira@fgv.br

Rafael Magalhães

Cientista de dados

Webpage: rnunesmagalhaes.github.io

Email: rnunesmagalhaes@gmail.com

> OFFICE HOURS

Davi Moreira: [Agendar aqui](#)

Rafael Magalhães: [Agendar aqui](#)

Avaliação

Serão solicitadas tarefas específicas e uma tarefa final, totalizando, respectivamente, 70% e 30% dos pontos totais (100%). A aprovação final está condicionada a ao menos 80% do total de pontos e presença mínima em 75% das aulas.

Código de honra

Os alunos são incentivados a discutir estratégias de resolução dos exercícios, mas espera-se que os resultados representem seus esforços individuais.

Bibliografia

[ML] BISHOP, Christopher M. Pattern Recognition and Machine Learning. springer, 2006.

BLEI, D. M. Topic Modeling and Digital Humanities. Disponível em: <http://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei/>

GRIMMER, J.; STEWART, B. M. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis, 2013.

[DV] HEALY, K. Data Visualization: A Practical Introduction. Princeton University Press, 2018.

[QSS] IMAI, K. Quantitative Social Science: An Introduction. Princeton University Press, 2018.

IZUMI, M. Y.; MOREIRA, D. C. O Texto como Dado: Desafios e Oportunidades para as Ciências Sociais. REVISTA BRASILEIRA DE INFORMAÇÃO BIBLIOGRÁFICA EM CIÊNCIAS SOCIAIS - BIB, v. 2, n. 86, p. 138–174, 2018.

[SWD] KNAFLIC, Cole Nussbaumer. Storytelling with Data. John Wiley & Sons, 2015.

MCDOWELL, G. L. Cracking the Coding Interview: 189 Programming Questions and Solutions. Edição: 6th ed. Palo Alto, CA: CareerCup, 2015.

MOREIRA, D. C. (2019). Text as Data for Social Sciences. Independent Publication. Disponível em: https://bookdown.org/davi_moreira/txt4cs/

ROBERTS, M.; STEWART, B.; TINGLEY, D. stm: R Package for Structural Topic Models. Journal of Statistical Software, 2019.

SLAPIN, J. B.; PROKSCH, S.-O. A Scaling Model for Estimating Time-Series Party Positions from Texts. American Journal of Political Science, v. 52, n. 3, p. 705–722, 2008.

[R4DS] WICKHAM, H.; GROLEMUND, G. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 1 edition ed. Sebastopol, CA: O'Reilly Media, 2017. Disponível em: <https://r4ds.had.co.nz/>

XIE, Y. bookdown: Authoring Books and Technical Documents with R Markdown. Disponível em: <https://bookdown.org/yihui/bookdown/>

Programa de aulas

Aula	Tema	Bibliografia
1	<p>Introdução</p> <p>Apresentação. Projetos e versionamento. Scripts, objetos, operadores lógicos. Boas práticas de programação.</p> <p><i>Avaliação de projetos do Banco Mundial</i></p>	[R4DS] Caps. 1, 2 e 27
2	<p>Processamento de dados</p> <p>Tipos de dados. Estruturas de dados. Importação de bases de dados. Operações com variáveis. Resumindo e agrupando dados. União de bases de dados.</p> <p><i>Quality of Government: padrões de desenvolvimento econômico em diferentes regimes</i></p>	[R4DS] Caps. 5, 11 e 12
3	<p>Análise exploratória</p> <p>Medidas de tendência central e de dispersão. Tabelas de contingência. Resumos gráficos. Interpretação de estatísticas descritivas.</p> <p><i>Explorando a vitimização de civis no Afeganistão</i></p>	[R4DS] Cap. 7

4	<p>Visualização de dados I: Gráficos</p> <p>Gramática dos gráficos. Gráficos básicos. Personalização e temas.</p> <p><i>Gapminder: a evolução internacional de indicadores de bem-estar</i></p>	<p>[DV] Caps. 1 e 3 [R4DS] Caps. 3 e 28</p>
5	<p>Visualização de dados II: Mapas</p> <p>Dados geoespaciais. Elaboração de mapas. Gráficos dinâmicos.</p> <p><i>A campanha russa de Napoleão Bonaparte</i></p>	<p>[DV] Caps. 7 e 8</p>
6	<p>Inferência, Estimação e Modelos de Regressão</p> <p>Descrição e inferência. Princípios de amostragem. Comunicação de incerteza. Intervalos de confiança e margens de erro.</p> <p>Associação e correlação. Regressão linear. Inferência e predição em modelos de regressão. Regressão e causalidade.</p> <p><i>World Values Survey: analisando o apoio à democracia no mundo</i></p> <p><i>Determinantes das eleições de 1932 na Alemanha</i></p>	<p>[QSS] Cap. 7 [ML] Cap. 3.1</p>

7	<p>Análise Multivariada e Modelos Preditivos</p> <p>Análise de clusters. Análise de componentes principais. Análise fatorial. Análise de correspondência</p> <p>Previsão e classificação. Algoritmos supervisionados e não supervisionados. Diagnóstico e calibração de modelos preditivos. Tidymodels.</p> <p><i>ONU: padrões de votação na Assembleia Geral</i></p> <p><i>Análise de resultados eleitorais nos EUA</i></p>	<p>[ML] Caps. 9.1 e 12.1</p> <p>[R4DS] Caps. 22, 23 e 24</p>
8	<p>Análise de Redes</p> <p>Descrição de redes. Medidas de centralidade. Grafos dirigidos e não dirigidos.</p> <p><i>Redes sociais: análise de dados do Twitter</i></p>	<p>[QSS] Cap. 5.2</p>
9	<p>Text as Data</p> <p><i>Impeachment: análise de votos proferidos</i></p>	<p>GRIMMER, J.; STEWART, B. M. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis, p. mps028, 22 jan. 2013.</p>
10	<p>Amarrando os elementos para comunicação efetiva</p> <p>Criação de apresentações. Princípios de visualização de dados. Storytelling with data. Técnicas de solução de problemas</p> <p><i>Análise da evolução do Covid-19</i></p>	<p>[DV] Cap. 1</p> <p>[SWD] Caps. 4, 5, 12, 13 e 14</p>

Bancos de dados para Relações Internacionais

AidData: <http://aiddata.org/>

Banco Mundial (World Bank). <http://data.worldbank.org/>

CIA World Factbook. <https://www.cia.gov/library/publications/the-world-factbook/rankorder/rankorderguide.html>

Comparative Study of Electoral Systems (CSES). <http://www.cses.org/>

Correlates of War. http://www.correlatesofwar.org/data-sets/folder_listing

FMI. <http://data.imf.org/>

Global Health Observatory (OMS). <http://apps.who.int/gho/data/?theme=home>

Human Development Reports (ONU) (IDH, Inequality, Gender Development, Gender Inequality, Poverty). <http://hdr.undp.org/en/data-explorer>

Latinobarómetro. <http://www.latinobarometro.org/latContents.jsp>

OCDE. <https://data.oecd.org/>

OMC. https://www.wto.org/english/res_e/statistics_e/statistics_e.htm

ONU. <http://data.un.org/Explorer.aspx>

Paul Hensel's International Relations Data Site. <http://paulhensel.org/data.html>

Penn World Tables. <https://pwt.sas.upenn.edu/>

Quality of Government Institute. <http://qog.pol.gu.se/data/datadownloads>

Transparency International (Corruption, Bribe Payer). <https://www.transparency.org/research/>

World Values Survey. <http://www.worldvaluessurvey.org/WVSContents.jsp>