

PROPOSTA DE LIVRO

# Usando R:

## Um Guia Para Cientistas Políticos

FERNANDO MEIRELES

Universidade Federal de Minas Gerais

<[www.fmeireles.com](http://www.fmeireles.com)>

DENISSON SILVA

Universidade Federal de Minas Gerais

<[www.denissonsilva.com](http://www.denissonsilva.com)>

### 1 Proposta

O uso de ferramentas computacionais está mudando rapidamente a forma com que fazemos pesquisa. Se no início da década termos como *webscraping* e *big data* sequer eram usados, hoje eles são recorrentes nas melhores revistas acadêmicas: estão presentes em trabalhos que usam informações de redes sociais para prever resultados eleitorais (Burnap et al., 2016; MacWilliams, 2015; Vepsäläinen et al., 2017); ou em análises que usam dados de Censo de toda uma população para investigar o perfil daqueles que decidem entrar para a política (Dal Bó et al., 2017). Mais importante, pesquisas como essas podem ser realizadas em uma fração de tempo, sem ter de se restringir à disponibilidade de material empírico já compilado, de auxiliares de pesquisa ou de orçamentos vultuosos (Silva and Meireles, 2016). Nas mãos de alguém com treinamento adequado, um simples *notebook* substitui tudo isso.

O livro que propomos desenvolver visa apresentar uma das mais potentes e versáteis dessas ferramentas computacionais: o ambiente de programação R. Crescentemente usado na academia e no setor privado, o R permite desde extrair e sistematizar bases massivas de dados até produzir relatórios replicáveis – um formato novo de documento que combina programação, dados e texto para gerar livros, artigos e *websites*. Além disso, o R oferece um leque completo de técnicas para análise quantitativa, é gratuito e possui uma rede mundial de usuários e desenvolvedores que trabalham continuamente para o seu aperfeiçoamento.<sup>1</sup>

Nosso objetivo ao oferecer um novo livro é cobrir a demanda acadêmica mais recente por R para a pesquisa científica na área de Ciência Política e similares.

---

<sup>1</sup>Para essas e outras vantagens do uso do R, ver: <https://www.nature.com/news/programming-tools-adventures-with-r-1.16609>

Diferentemente de outro que é popular (ver Seção 3), o nosso oferecia três principais inovações. Primeiro, o de ser direcionado à pesquisa acadêmica de forma prática, com exemplos reais, cobrindo etapas normalmente não discutidas em livros-texto. Segundo, e relacionado ao primeiro ponto, nosso foco será apresentar as ferramentas mais usadas na pesquisa política, o que nos permite explorar cada tópico em maior profundidade – como um guia deve ser. Finalmente, nosso livro não ensinaria receitas fechadas, mas formas de resolver problemas com R. Essa mudança pode parecer pequena, mas indica que esperamos que nossos potenciais leitores implementem em suas rotinas o que aprenderam com o livro.

No que segue, discutimos brevemente a diferença do livro proposto para outro similar. Depois, apresentamos nossas credenciais e um esquema do conteúdo planejado para o texto.

## 2 Público-alvo

O livro seria endereçado principalmente para a comunidade acadêmica da Ciência Política, o que envolve professores universitários, alunos de pós-graduação e graduação da área. De qualquer forma, o uso de R desperta grande interesse em diversos outros públicos. A título de exemplo, um *software* que desenvolvemos em R para extrair dados eleitorais – algo notadamente específico – já foi instalado mais de 6.500 vezes desde seu lançamento oficial.<sup>2</sup> Pelo livro pretender cobrir, etapa por etapa, como se realiza pesquisa acadêmica usando R, acreditamos que isso favoreça a sua inclusão em ementas de cursos.

## 3 Originalidade e Relevância

Salvo melhor juízo, o único livro sobre uso de R na pesquisa social de forma ampla no Brasil é de Aquino (2014). Ele tem o mérito de ser pioneiro e de cobrir uma boa gama de aplicações, mas seu conteúdo está desatualizado em termos de novas tecnologias e práticas de programação, que ajudam iniciantes a aprender mais rápidos e, usuário com nível intermediário, a ganhar produtividade. Numa seara onde avanços são rápidos, isso importa – nesses últimos quatro anos, para mencionar apenas um exemplo, o número de *softwares* de terceiros desenvolvidos em R saltou de cerca de 6.000 para mais 11.000.<sup>3</sup>

---

<sup>2</sup>Disponível em: <<https://cranlogs.r-pkg.org/badges/grand-total/electionsBR>>.

<sup>3</sup>Disponível em: <<https://blog.revolutionanalytics.com/2017/01/cran-10000.html>>.

## 4 Sobre os autores

Usamos R desde o início de nossas formações como cientistas políticos, somando mais de uma década de experiência conjunto. Em nosso trabalho, empregamos rotineiramente essa ferramenta, como foco em técnicas quantitativas de análise, como atestam nossos currículos. Já prestamos consultoria a diversas agências e pesquisadores; desenvolvemos bancos de dados populares na área; e ministramos cursos de treinamento em R em instituições como Universidade Federal de Minas Gerais, Universidade Federal de Viçosa e Universidade Federal de Pernambuco. Cabe destacar que também desenvolvemos *software* com R. Mais importante, temos conjuntamente uma dos pacotes mais populares na área de Ciências Sociais, que extrai e limpa automaticamente dados eleitorais do Repositório do Tribunal Superior Eleitoral (TSE).<sup>4</sup>

## 5 Conteúdo do livro

O material do livro seria dividido em duas grandes partes. Na primeira (capítulos 1 a 4), ofereceríamos uma introdução ao R. Ao longo deles, mostraríamos como realizar as tarefas mais comuns numa pesquisa, tais como: coletar, ler, limpar, manipular e documentar os mais diversos tipos de dados. Complementarmente, também cobriríamos as noções básicas de *data science* e *big data* de forma aplicada, com exemplos reais sobre como criar gráficos, tabelas e realizar testes e estatísticas descritivas comuns. Na segunda parte (capítulos 5 a 11), veríamos mais detidamente como o R pode ser usado para contar histórias convincentes com base sólida nos dados de forma rápida, eficiente e reproduzível. Isso envolveria conteúdos mais avançados, como modelagem quantitativa, *webscraping* (i.e., coleta de dados digitais) e de exportação de relatórios e resultados de forma automatizada para Word,  $\text{\LaTeX}$ , HTML ou qualquer outro meio.

---

<sup>4</sup>Disponível em: <<http://electionsbr.com/>>.

Tentativamente, o sumário do livro seria o seguinte:

- Prefácio;
- Introdução;
- 1 Introdução ao R;
- 2 Bases de dados;
- 3 Importação;
- 4 Manipulação;
- 5 Gráficos;
- 6 Modelos;
- 7 Documentos;
- 8 Replicação;
- 9 Econometria;
- 10 Análise de textos;
- 11 Web Scraping.

## Referências

- Burnap, P., Gibson, R., Sloan, L., Southern, R., and Williams, M. (2016). 140 characters to victory?: Using twitter to predict the uk 2015 general election. *Electoral Studies*, 41:230–233.
- Dal Bó, E., Finan, F., Folke, O., Persson, T., and Rickne, J. (2017). Who becomes a politician? *The Quarterly Journal of Economics*, 132(4):1877–1914.
- de Aquino, J. A. (2014). R para cientistas sociais.
- MacWilliams, M. C. (2015). Forecasting congressional elections using facebook data. *PS: Political Science & Politics*, 48(4):579–583.
- Silva, D. and Meireles, F. (2016). Ciência política na era do big data: automação na coleta de dados digitais. *Revista Política Hoje-ISSN: 0104-7094*, 24(2):87–102.
- Vepsäläinen, T., Li, H., and Suomi, R. (2017). Facebook likes and public opinion: Predicting the 2015 finnish parliamentary elections. *Government Information Quarterly*, 34(3):524–532.