

Plano do curso

Introdução ao R

Fernando Meireles

Universidade Federal de Minas Gerais

<fmeireles@ufmg.br>

1/2017

Este curso oferece uma introdução a uma das ferramentas mais versáteis e potentes para a análise dados: o ambiente de programação R. Ao longo dele, veremos como utilizá-lo para contar histórias convincentes com base sólida nos dados de forma rápida, eficiente e reproduzível, o que envolve realizar tarefas comuns numa pesquisa científica, tais como: coletar, limpar, manipular, documentar, visualizar e produzir inferências a partir dos mais diversos tipos de dados. Após uma introdução ao básico da linguagem R, realizaremos atividades práticas que cobrirão desde o cálculo de estatísticas descritivas básicas até a exportação dos resultados de forma automatizada para *Word*, \LaTeX , *.pdf*, *HTML* ou qualquer outro meio. Por fim, trabalharemos num mini-projeto de pesquisa onde será aplicado todo o conteúdo visto.

Deve ficar claro, entretanto, que não esgotaremos as possibilidades do R (na verdade, seria irrealista prometer isto). Procurarei apenas mostrar que o R é uma espécie de ambiente onde sempre há uma forma mais simples de se fazer algo – e que, como tal, ele pode ser eficientemente usado com qualquer finalidade, acadêmica ou não. Ao fim do curso, é esperado que os(as) alunos(as) estejam minimamente preparados para incorporar o R as suas rotinas.

Pré-requisitos

Não é requerido nenhum conhecimento prévio de programação para acompanhar o curso. Conhecimentos básicos de estatística (i.e. descritiva e teste de hipóteses) são desejáveis, mas não são necessários.

Materiais

A princípio, realizaremos o curso num laboratório de informática. De qualquer modo, sendo possível, é recomendado que os(as) alunos(as) tragam *notebooks*; isto facilita o

aprendizado e o uso posterior do R (já que me permite ajudar presencialmente a contornar quaisquer erros causados por incompatibilidade de *hardware* ou *software*, estes mais comuns em Linux e OS X). No último tópico deste plano, indico os procedimentos necessários para ter a última versão funcional do R devidamente instalada.

Todos os os *scripts* e demais materiais usados no curso serão disponibilizados antes da realização das atividades. A quem interessar, uma boa fonte introdutória é o curso *online* do *Datacamp*. Basicamente, ele simula um *console* do R e ensina o básico de forma prática, com dicas e ajudas eventuais, um corretor que notifica erros e um sistema de avaliação progressivo. O curso é gratuito (mas em inglês) e pode ser acessado em:

- www.datacamp.com

Por fim, a melhor fonte em português para solucionar dúvidas pontuais é o *Stackoverflow*, um fórum de programação dedicado ao R (mas não só a ele). Ali, não só é possível encontrar respostas a diversas questões já feitas como também postar outras novas.

- pt.stackoverflow.com

Logística do curso

As aulas serão todas práticas, precedidas ou sucedidas de breves explicações acompanhadas de *slides* (que ficarão à disposição após o término do curso). Como a parte principal do curso será orientada, auxílio e esclarecimento de dúvidas serão feitos principalmente em aula (outras questões podem ser discutidas antes ou depois das aulas, ou por e-mail).

Datas, horários e locais no curso: Laboratório de Informática PGP, 22 a 24 de março, às 14.h.

Avaliação

Realizaremos várias tarefas durante o curso, mas elas não serão avaliadas. A única tarefa que será avaliada é a atividade final, que poderá ser realizada em duplas. Esta consistirá no teste de uma hipótese previamente formulada com base em alguns dados que serão disponibilizados. Para tanto, deverá ser entregue um *script* documentado e reproduzível com a análise. Não serão dadas notas, apenas será avaliado o nível de desenvolvimento das análises - a ideia principal aqui, portanto, é incentivar que cada um procure suas próprias soluções, e não receitas prontas.

Atendimento a necessidades especiais

Alunos(as) com quaisquer necessidades ou solicitações individuais podem me procurar diretamente, por e-mail ou pessoalmente, para obterem auxílio. Todos os pedidos serão mantidos em sigilo.

Política de gênero

Cursos de programação e de metodologia frequentemente são o reino do clube do bolinha: homens são maioria, monopolizam a participação e os computadores. Por conta disso, seguiremos um protocolo muito simples neste curso: alunas têm preferência sobre computadores e participação – o que implica em não interromper colegas, não centralizar as atividades em grupo e priorizar o uso de *mouse* e teclado às mulheres sempre que faltarem computadores.

Plano das aulas

1 Introdução ao R

1.1 O R faz (quase) tudo; 1.2 A linguagem R; 1.3 Básico: objetos, classes e vetores; 1.4 Básico II: manipulando objetos e vetorização; 1.5 Funções: o básico; 1.6 Chamadas, argumentos e ajuda; 1.7 Gráficos básicos; 1.8 Pacotes; 1.9 Algumas estatísticas e uma regressão linear: os presidentes americanos estão ficando mais altos com o passar do tempo?

Leituras e materiais sugeridos:

- Jakson Alves de Aquino. *R para cientistas sociais*. EDITUS - Editora da UESC, 2014, caps. 1 e 2.
- Victor Lemes Landeiro. Introdução ao uso do programa r.
- Poder da linguagem R fascina analistas. (*website*).

2 Análise de Dados

2.1 Abrindo e carregando dados; 2.2 Carregando dados II; 2.3 Manipulando `data.frames` I; 2.4 Manipulando `data.frames` II: `dplyr`; 2.5 Gráfico II: `ggplot2`; 2.6 Análise: quem sobreviveu ao Titanic?.

Leituras e materiais sugeridos:

- Jakson Alves de Aquino. *R para cientistas sociais*. EDITUS - Editora da UESC, 2014, caps. 4 e 5.
- Fernando Meireles. `rsciolo` - an r package to scrape meta-data from scientific articles hosted on scielo. 2016. URL <https://github.com/meirelesff/rSciolo>
- Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer, 2015

- Como importar qualquer arquivo no R (*website*).
- Como importar dados em .csv no R (*website*).

3 Realizando um projeto

3.1 Estatística multivariada; 3.2 Reportando resultados; 3.3 Reprodutibilidade; 3.4 Projeto final (com o pacote `electionsBR`): Quem faz mais votos em Viçosa?.

Leituras e materiais sugeridos:

- C D Shikida; Rodrigo Fernandez. *Notas introdutórias em econometria aplicada usando R/Rstudio*. 2016. URL <http://wp.ufpel.edu.br/cdshikida/apostilarstudio/>
- Fernando Meireles; Denisson Silva; Beatriz Costa. `electionsbr`: R functions to download and clean brazilian electoral data. 2016. URL <http://fmeireles.com/files/electionsbr.pdf>

Carga horária total: 12h.

Para instalar o R

Para instalar o R, basta ir ao site do CRAN (*Comprehensive R Archive Network*), que é a rede de fundadores e administradores do *core* da linguagem R, e baixar o *setup* indicado para o seu sistema operacional:

- cran.r-project.org

Feito isto, já é possível usar o R – mas só via *console*, o que não é tão fácil/útil. É por isso que usaremos uma IDE (i.e. Ambiente de Desenvolvimento Integrado) neste curso: especificamente, usaremos o *RStudio*. Para baixá-lo, basta entrar no seguinte site e escolher a opção mais adequada para o seu sistema operacional:

- www.rstudio.com

Referências

- Jakson Alves de Aquino. *R para cientistas sociais*. EDITUS - Editora da UESC, 2014.
- Fernando Meireles; Denisson Silva; Beatriz Costa. electionsbr: R functions to download and clean brazilian electoral data. 2016. URL <http://fmeireles.com/files/electionsbr.pdf>.
- C D Shikida; Rodrigo Fernandez. *Notas introdutórias em econometria aplicada usando R/Rstudio*. 2016. URL <http://wp.ufpel.edu.br/cdshikida/apostilarstudio/>.
- Victor Lemes Landeiro. Introdução ao uso do programa r.
- Fernando Meireles. rscielo - an r package to scrape meta-data from scientific articles hosted on scielo. 2016. URL <https://github.com/meirelesff/rScielo>.
- Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer, 2015.