

STRANGE-R METHODOLOGIES: Ensinando R de trás-pra-frente no serviço público

Carolina Musso, Hudson G. V. Fontenele, Paola F. de Oliveira, Taynná V. R. Almeida and Janaina Sallas

palavras-chave: educação; epidemiologia; vigilância à saúde; SUS

INTRODUÇÃO

O uso do R na rotina do serviço público de saúde é limitado pelo tempo necessário para o aprendizado da linguagem. Por outro lado, os profissionais da área de saúde se veem cada vez mais compelidos a dominar ferramentas poderosas de análise como esta, com a necessidade de aplicabilidade imediata. Essa realidade é particularmente marcante no caso de epidemiologistas que trabalham na vigilância em saúde, cujo trabalho está intimamente atrelado à análise de grandes volumes de dados. Assim, há necessidade de profissionais capacitados para a manipulação de grandes bases de dados provenientes de diferentes fontes de forma concomitante e rápida, situação evidenciada pela pandemia de COVID-19.

O curso proposto teve o objetivo de tornar um profissional de saúde capaz de reconhecer os potenciais da ferramenta R/RStudio e saber buscar as informações nas fontes necessárias. Com metodologia diferenciada, esse curso foi voltado especificamente para profissionais da vigilância em saúde do Ministério da Saúde e da Secretaria de Saúde do Distrito Federal, Brasil.

METODOLOGIA

Esse curso, se propôs a abordar a aplicação de R de um ponto de vista prático, partindo da aplicação direta de códigos de análise automatizadas prontos e, a partir daí, solidificar a base teórica necessária para a autonomia dos profissionais. Portanto, para o desenvolvimento do curso, foi requisitado que os alunos possuísem um contato prévio mínimo com a ferramenta.

A primeira parte do curso (módulos 1 e 2) consistiu na execução de códigos previamente implementados. Nesta etapa, era esperado que os alunos adquirissem a capacidade de reconhecer possíveis erros que pudessem surgir ao executar códigos que outros colegas (mais fluentes na linguagem) houvessem escrito. O aluno foi estimulado a identificar esses erros comuns e a trabalhar diferentes maneiras de como solucioná-los. Ele também foi estimulado, neste momento, a intuitivamente identificar potenciais locais no código para customizações específicas, de modo a ter um contato com a linguagem de forma mais empírica. Na fase intermediária (módulos de 3 a 6), o curso se propôs a fornecer as ferramentas necessárias para que, em seu projeto final (módulo 7), os alunos fossem capazes de criar manipulações e visualizações próprias com os bancos de interesse, resultando em um documento reproduzível útil ao serviço.

O curso teve uma duração de seis meses com uma carga horária de 96h. A principal bibliografia utilizada foi o livro online Batra (2021). As aulas foram síncronas, em sua maioria presenciais, de duas horas cada, ministradas duas vezes por semana e foi estruturado em sete módulos:

Módulo 1 “Como rodar aquele código que alguém fez pra mim”: Executar rotinas já estabelecidas. Uso do GitHub. Compreender e identificar possíveis *bugs* e como corrigi-los. Introdução ao RMarkdown. Conceitos básicos (objetos, vetores, console, data frame) foram abordados de forma "imersiva" e manipulados de forma intuitiva sem apresentação formal.

Módulo 2 “Mas é que eu queria esse gráfico de outra cor”: Ciclo da análise (Transformar, Visualizar, Comunicar). Adaptar os códigos já existentes de gráficos em pacote *ggplot2*, “brincando” com os parâmetros. Foram apresentados os conceitos de Tidy Data, gramática dos gráficos (camadas, estética, mapeamento), os gráficos de Box-plot e Gráfico de Barras, edição de títulos/eixos/legendas, e facetamento.

Módulo 3 “Às vezes devemos dar um passo para trás antes de continuar”: Introdução clássica à programação em R: Partes do R-Studio, Funções, Objetos, *tidy data* importação de dados e exportação de objetos.

Módulo 4 “Nada se cria, tudo se transforma”: Manipulação de dados com *tidyverse*, funções *select()*, *arrange()*, *mutate()*, *group_by()*, *summarise()*, *pivot_longer()*, *pivot_wider()* dentre outras além dos pacotes *lubridate*, *stringr*, *forcats* e outros.

Módulo 5 “Servidor público gosta de tabela”: Confeção e customização de tabelas com o pacote *flextable* a extensões.

Módulo 6 “Módulo de tópicos avançados”: Confeção de mapas e arquivos *Rmarkdown* avançado, confeção de *dashboards* com o pacote e *flexdashboard*.

Módulo 7 “Tudo junto misturado: Trabalho final”: Criar um relatório simples, mas reproduzível, do zero com uma base de interesse para a rotina de trabalho do aluno.

Todas as aulas expositivas foram permeadas de exercícios práticos que foram compartilhados via GitHub. Cada módulo foi finalizado com sessões de tutoria individualizadas. Eventualmente foram realizados *quizzes* com a plataforma *Kahoot* como ferramenta lúdica de teste e fixação de conteúdo. Para auxiliar na resolução de exercícios e tarefas, o curso contou com o auxílio de dois monitores da Universidade de Brasília.

RESULTADOS E CONCLUSÕES

A apresentação dos códigos implementados no início do curso surtiu o efeito esperado e muitas pessoas foram capazes de manipular trechos de código intuitivamente e se sentirem mais motivadas. Seis pessoas foram capazes de finalizar o curso, sendo que quatro delas possuíam um conhecimento prévio quase nulo na programação. Essas pessoas foram capazes de progredir e confeccionar trabalhos aplicáveis e úteis em suas unidades de trabalho como, por exemplo: rotinas de qualificação do banco de dados de COVID-19, boletim informativo de consultas médicas na população indígena, e análise do banco de dados de doenças de notificação compulsória.

Das atividades apresentadas, o uso do GitHub foi a ferramenta mais desafiadora e a menos recomendada para aprendizado e compartilhamento das atividades. Os módulos necessitam de redimensionamento, com aumento de tempo direcionado ao módulo de manipulação. Uma das aulas consideradas mais interessantes e importantes foram as aulas destinadas à identificação de erros.

A boa receptividade dos alunos concluintes e os produtos gerados nos levam a crer que a metodologia, com as devidas adaptações, é viável para ser aplicada no serviço público e mesmo em outras áreas, como em universidades. Espera-se que cursos como este estimulem a implementação de novas práticas no serviço público para que os profissionais despendam mais tempo em tarefas como a interpretação dos dados, ao passo que as tarefas mais repetitivas e automatizáveis possam ser delegadas às máquinas.

REFERÊNCIAS

BATRA, N., et al (2021) *The Epidemiologist R Handbook*. <https://epirhandbook.com/>