1. Introdução

O objetivo do curso é dar uma abordagem geral da análise da POF no R. Para tanto, é preciso que haja um básico de conhecimento tanto do R quanto da pesquisa em si, para que seja possível combinar os dois. Por um lado, é preciso saber diversas ações no R, como criar colunas, juntar tabelas e importar dados. Por outro, é necessário entender a estrutura da pesquisa, como a divisão em unidades de consumo (e não domicílios), a coexistência de diversos registros e o uso dos fatores de expansão.

É irreal pensar que um curso é capaz de esgotar todas as informações relativas ao R e à POF. Por isso creio fundamental dar ferramentas para que seja possível buscar soluções com independência. Nesse caso, entra o stack overflow, a documentação de pacotes e funções e o google (no R) e, de outro, os scripts, resultados e manuais que o IBGE fornece para a POF.

Ao mesmo tempo, o fato de ser um curso abre a possibilidade de boas práticas de programação e de abordagem de pesquisa quantitativa. Coisas como projetos, foco no script, nomeação adequada de funções e arquivos, organização de pastas, funções, iteração, controle de versão (se der tempo) etc. serverm para qualquer trabalho no R. E muitas das informações sobre a POF são similares para outras pesquisas (principalmente amostrais e do IBGE).

Dito isso, penso que a conclusão dessa parte inicial é que, dado o escopo da POF e do R, o curso deve ter um norte extremamente prático, focando nas informações necessárias para analisar a POF no R e, a partir daí, cada um aprofundar os conhecimentos na medida de sua necessidade.

2. Introdução ao R

Levando em conta as considerações acima, eu dividiria a introdução ao R em dois blocos: uma abordagem direta a dataframes por meio do tidyverse (sem gastar muito tempo com outras estruturas como listas e vetores) e práticas de programação que garantam fluidez, reprodutibilidade e segurança no código. Pretendo estruturar para que o segundo venha naturalmente, para que fique evidente os motivos de seguir esses hábitos.

2.1. Para começar: R, RStudio e tidyverse

O R é uma linguagem de programação com raízes na análise estatística. Por isso seu uso disseminado para análise de dados. O RStudio é um IDE (“Integrated Development Environment”) – é um programa que fornece um ambiente para rodar o R. Por fim, o tidyverse é um conjunto de pacotes que seguem uma mesma lógica e são extremamente úteis para uma melhor performance do R.

Para instalar o R, é preciso entrar em <https://cran.r-project.org/mirrors.html> e escolher o site de preferência. Eu sempre baixo o da Fiocruz, mas só por bairrismo mesmo.

O RStudio está em: <https://www.rstudio.com/products/rstudio/download/#download>

Não é preciso nunca abrir o R – pode usar direto o RStudio. A primeira coisa que eu recomendo fazer é mudar para um tema escuro. Trabalhar com código é cansativo para os olhos e usar um tema escuro (qualquer um) ajuda a mitigar isso. Para fazer isso, basta clicar em Tools > Global Options > Appearance > Editor theme. Eu gosto do idle fingers e do merbivore soft.

Falar das janelas do RStudio.

O R vem com funções embutidas. Exemplos: sum, date.

Pacotes são “unidades compartilháveis de código”, juntando dados, funções e testes. O R possui o CRAN, um repositório de pacotes que seguem critérios de manutenção e documentação. O CRAN é periodicamente checado, e diversos pacotes podem ser removidos (um exemplo é o *ecoseries*, pacote que baixava dados do BCB e do Sidra). Nem todo pacote precisa estar no CRAN – qualquer um pode criar um pacote e disponibilizar no github (eu mesmo tenho alguns).

O tidyverse não é, de fato, um pacote. Ele é uma coleção de pacotes voltados para data science com a mesma estrutura e filosofia. O tidyverse melhora muito a performance e a legibilidade do R, e por isso vamos usá-lo diretamente. Muitos pacotes podem não fazer parte do núcleo do tidyverse mas seguem a mesma filosofia (exemplos notáveis são o glue e o lubridate). O símbolo de que um pactoe é compatível com o tidyverse é que seu símbolo é um hexágono.

Para usar um pacote, é necessário instalá-lo (somente uma vez) e, a cada novo uso do R, carregá-lo. O comando para instalar é install.packages(“nome\_do\_pacote”) – no caso, install.packages(“tidyverse”). Para carregar, library(nome\_do\_pacote) – no caso, library(tidyverse).

2.2. Básico de mexer em um dataframe

Operações: visualizar (View), primeiras linhas (head), últimas linhas (tail).

Selecionar colunas (select), criar colunas (mutate), filtrar dados (filter), ordenar (arrange), agrupar (group\_by), resumir (summarise) e juntar (binds e joins).

<https://www.tidyverse.org/blog/2017/12/workflow-vs-script/>

<https://martinctc.github.io/blog/rstudio-projects-and-working-directories-a-beginner%27s-guide/>

<https://r4ds.had.co.nz/workflow-projects.html>

<https://rstats.wtf>

<https://stat545.com/index.html>

<https://stat545.stat.ubc.ca>