How I Use R

David Keyes // R for the Rest of Us

Contents

Chapter 1

Introduction

Since 2018, I've been teaching people to use R through my company, R for the Rest of Us. It's an incredibly rewarding experience to see people learn to use this powerful piece of software, but it can also be frustrating.

One of the hardest parts of learning R (or any language) is taking knowledge from exercises and applying it to an actual project you're working on. Concepts that make sense in the classroom suddenly become muddled when you're back at your desk trying to use R to write a report.

One of the biggest challenges I've had as a teacher is helping people in this stage right after they learn some fundamental R skills. After I teach folks R fundamentals through my online courses, workshops, or custom training sessions, I struggle to find resources for them to help them apply their new knowledge to their daily work.

There are many great learning resources at the beginner stage and some incredible tutorials to master complex tasks in R. But, drawing from a concept in urban planning, there are far fewer resources in the middle.

Stretching the metaphor perhaps to its breaking point, new R users at the "detached single-family home" stage can't get to the advanced "mid-rise" level without going through the middle stage. The "missing middle" in the R neighborhood is the lack of resources to that answer the types of nuts and bolts questions that new R users often have. Things like:

- How should I organize my file structure when creating a new project?
- Should I do data cleaning in an RMarkdown file or an R script file?
- How do I find packages? How do I know if the packages I find are high quality?

This book is my attempt to provide answers to these types of questions. It will be an opinionated look at how one person (me) uses R. I'm writing it not

Figure 1.1: Source: Missing Middle Housing

because I think my approach is the best and that everyone should use it. I'm writing it because I believe that I can offer some ideas that can help the students I work with – and perhaps you as well – to go from learning R to using it in their daily practice. R has been transformative for my work and I want it to be the same for you.

Note

What follows is an in-progress outline of the rest of this book. It will change over the next few months as I write the book.

Chapter 2

Setting Myself up for Success

2.1 Software

2.1.1 Why I Don't Work Just in R

2.1.2 Why I Use RStudo, not something like R Commander

2.1.3 RStudio Options

Uncheck load data on start new session

2.2 Code Style

2.2.1 Examples of other style guides

- http://jef.works/R-style-guide/
- https://style.tidyverse.org/ + https://styler.r-lib.org/index.html
- https://google.github.io/styleguide/Rguide.xml
- https://ropensci.github.io/dev_guide/
- https://csgillespie.github.io/efficientR/

2.2.2   My style preferences

- Spaces between things
- Line breaks
- Put all packages at top of script

2.3   Working with Files

2.3.1   Directory organization

2.3.1.1   Use data/data-raw structure

2.3.1.2   Naming files

https://speakerdeck.com/jennybc/how-to-name-files?slide=3)

2.4   Packages

2.4.1   Where do I find packages?

2.4.2   How do I evaluate packages?

2.4.3   Why I Use the Tidyverse

- https://joss.theoj.org/papers/10.21105/joss.01686

2.5   Workflow

2.5.1   When to use .R vs .Rmd

Break up data cleaning (R script) and reporting (RMarkdown)

2.5.2   Load all data at top

2.5.3   Add sections in R scripts to enable TOC

2.6   Git/GitHub

Pluses

- Hard to get set up

Minuses

- Multiple people can work at a time
- Version control
- Branching

2.6.1   As compared to Google Drive, Dropbox, etc.

Pluses

- Easy

Minuses

- Only one person can work at a time
- No version control

Chapter 3

Working with Data

3.1 Importing Data

3.1.1 Why to use read_csv not read.csv

3.1.2 clean_names() function

3.1.3 Variable and value labels

3.1.3.1 Codebooks

https://rubenarslan.github.io/codebook/   https://cran.r-project.org/web/packages/vtable/index.html   https://cran.r-project.org/web/packages/sjlabelled/index.html

Include discussion of R vs SPSS

http://josiahparry.com/post/2019-12-14-spss-haven/

3.1.4 Examining Data

- Skimr
- Naniar
- DataMaid

3.2   Data Wrangling and Analysis

3.3   General Practices

- Restart session often

- Create as few objects as possible

- Load all data at top of code

Chapter 4

Reporting Results with RMarkdown

Explain how I didn't use RMarkdown for a long time and really missed out

4.1 General Practices

4.1.1 Naming Code Chunks

4.1.2 Use TOC

4.1.3 YAML

- https://github.com/r-lib/ymlthis

4.1.4 Parameterized reports

4.1.5 Misc

- https://github.com/ThinkR-open/remedy

4.1.6   Tables

4.2   What Format I Knit To When

4.2.1   Word

Using Word reference documents for style

4.2.2   PDF

I basically don't because I'm scared off by LaTeX

4.2.3   HTML

4.2.3.1   Distill

4.2.3.2   Bookdown

https://alison.rbind.io/talk/2019-rsc-bookdown/

4.2.3.3   Pagedown

4.2.3.4   Dashboards

4.2.3.5   Flexdashboard

4.2.3.5.1   crosstalk

4.2.3.5.2   Shiny

Chapter 5

Visualizing Data

5.1 Themes

5.2 Using Fonts in Plots

extrafont package

5.3 ggplot extensions

5.4 Formatting numbers etc

- Twitter tip to not use scientific notation: https://twitter.com/ecologyofgavin/status/1188865515059585025
- scales package

5.5 Mapping

5.5.1 General packages

- tmap
- leaflet
- ggplot
- mapview

5.5.2   Geocoding

- ggmap
- https://rdrr.io/cran/tmaptools/man/geocode_OSM.html
- opencage

5.5.3   Tables

See blog post

Chapter 6

What I Do When Things Go Wrong (and They Always Do)

6.1 Guides to getting help in R

- https://socviz.co/appendix.html#a-little-more-about-r
- https://sctyner.github.io/rhelp.html

Chapter 7

Read More

- https://whattheyforgot.org/