Applied R in the Classroom

JD Long & Dusty Turner

Abstract

Introduction - Dusty

The Father of Modern Econometrics, Adam Smith, once frustratingly stated:

> The discipline of colleges and universities is in general contrived, not for the benefit of the students, but for the... ease of the masters.[1]

Adam Smith is making the claim that methods of instruction used by professors are those which are easiest for the instructor, but not necessarily what is best for the student. This is understandable, as research and other administive demands force instructors to rely on tried teaching techniques which are familiar to the teacher and not what will equip the student to be most successful in their career.

The authors of this article will make the case that the R open sourse statistical programming languge can bridge this gap between a Smith's proverbial teacher's ease and a student's benefit. As R continues to be one of the more popular coding languages in industry with ever increasing technical support, the barriar for entry keeps falling. There are many tools available in R which can aide the teaching process to get students loading and exploring data quickly requiring low overhead for proficiency to teach. With ample open source support, a wide acceptance in industry and many additional features to explore and present data, teaching with R is both satisfies the ease for the instructor and has long term benefit for the students.

This is something we are certain Adam Smith would be proud of.

The R Ecosystem - JD

CRAN

Describe Base R install, CRAN pacakges, and RStudio

RStudio

R Studio & Rstudio.cloud - how to share a project for the class

Tidyverse

library(tidyverse) explain each package purpose

Installation

an aside on how to do this on your personal computer - link to R Cookbook

Projects

Directory structure in the project.

[1] Smith and Krueger, "The Wealth of Nations."

Tutorial begins - Dusty will start

Loading Data (Gapminder)

```
library(tidyverse)

gapminder <- read_csv("01_data/gapminder.csv")
```

Or directly from a URL: will be a 404 until the repo is public

```
gapminder <- read_csv("https://github.com/CerebralMastication/r_for_the_student/blob/master/01_data/gapm
```

Identifying data types

```
summary(gapminder)
```

```
##    country           continent              year         lifeExp
##  Length:1704        Length:1704        Min.   :1952   Min.   :23.60
##  Class :character   Class :character   1st Qu.:1966   1st Qu.:48.20
##  Mode  :character   Mode  :character   Median :1980   Median :60.71
##                                        Mean   :1980   Mean   :59.47
##                                        3rd Qu.:1993   3rd Qu.:70.85
##                                        Max.   :2007   Max.   :82.60
##       pop              gdpPercap
##  Min.   :6.001e+04   Min.   :   241.2
##  1st Qu.:2.794e+06   1st Qu.:  1202.1
##  Median :7.024e+06   Median :  3531.8
##  Mean   :2.960e+07   Mean   :  7215.3
##  3rd Qu.:1.959e+07   3rd Qu.:  9325.5
##  Max.   :1.319e+09   Max.   :113523.1
```

Or using the skimr package

```
library(skimr)
```

```
##
## Attaching package: 'skimr'

## The following object is masked from 'package:stats':
##
##     filter
```

```
skim(gapminder)
```

```
## Skim summary statistics
##  n obs: 1704
##  n variables: 6
##
```

```
## -- Variable type:character ------------------------------------------------------------------------
##   variable missing complete    n min max empty n_unique
##  continent       0     1704 1704   4   8     0        5
##    country       0     1704 1704   4  24     0      142
##
## -- Variable type:numeric --------------------------------------------------------------------------
##   variable missing complete    n     mean         sd       p0        p25
##  gdpPercap       0     1704 1704  7215.33 9857.45      241.17    1202.06
##    lifeExp       0     1704 1704    59.47   12.92       23.6       48.2
##        pop       0     1704 1704 3e+07       1.1e+08 60011    2793664
##       year       0     1704 1704  1979.5    17.27     1952       1965.75
##       p50      p75         p100     hist
##   3531.85  9325.46 113523.13    <U+2587><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581>
##     60.71    70.85     82.6     <U+2581><U+2582><U+2585><U+2585><U+2585><U+2585><U+2587><U+2583>
##  7e+06    2e+07         1.3e+09 <U+2587><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581>
##   1979.5   1993.25   2007       <U+2587><U+2583><U+2587><U+2583><U+2583><U+2587><U+2583><U+2587>
```

`dplyr` Verbs

(select, filter, mutate, group_by, summarize, arrange)

Plotting/EDA

Esquisse? - https://cran.r-project.org/web/packages/esquisse/readme/README.html

Regression

parsnip/tidymodels

text from prior draft left below for salvaging

An example of how to cite in R

With more than 275 practical recipes, this expanded edition[2] helps you perform data analysis with R quickly and efficiently. The R language[3] provides everything you need to do statistical work, but its structure can be difficult to master. These task-oriented recipes make you productive with R immediately. Solutions range from basic tasks to input and output, general statistics, graphics, and linear regression.

Each recipe addresses a specific problem and includes a discussion that explains the solution and provides insight into how it works. If you're a beginner, R Cookbook will help get you started. If you're an intermediate user, this book will jog your memory and expand your horizons Grolemund and Wickham.[4] You'll get the job done faster and learn more about R in the process.

R Studio Layout

One of the most common mistakes I run into in class on day one is for beginners to try to run R directly from the console. They have R Studio downloaded, but do not understand the difference between R and R studio. R is the program that executes all commands you provide. It is the engine of the car, if you will. R

[2] Teetor and Long, "R Cookbook, 2nd Edition."
[3] Allaire et al., *Rmarkdown*; Xie, *Knitr*.
[4] "R for Data Science."

Studio is the shell that keeps the user organized. R Studio is the shell around the car. To continue, it is the steering wheel, accelerator, brakes, and seatbelts.

To ensure you have opened R Studio, you screen should be broken into 3 to 4 panels.

1) Top Right: The Environment
2) Bottom Right: The Files Viewer (what do you call this quadrants?)
3) Bottom Right: The Console
4) Top Left: The Script

Upon first opening R Studio, you may not have panel 4. In order to open up a script, you will need to select `file` -> `New File` -> `R Script`.

Many new users also make the mistke of not realizing that they have not opened up a script. They begin executing in the console again and do not take advantage of the ease that scripts provide in writing reproducable code.

Possibly note on how to execute lines of code. Both code chuncks and specific lines of code.

Tidyverse

Installation

paraphrase what's in R Cookbook?

Installing R

Installing R Studio

Installing the Tidyverse

Using R

Loading Data

Plotting

quick intro to ggplot

That graphical tool for ggplot?

Plotting Multivariate Data

ggpairs?

Bibliography

Allaire, JJ, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. *Rmarkdown: Dynamic Documents for R*, 2019. https://CRAN.R-project.org/package=rmarkdown.

Grolemund, Garrett, and Hadley Wickham. "R for Data Science." *R for Data Science*, n.d. https://r4ds.had.co.nz/.

Smith, Adam, and Alan B. Krueger. "The Wealth of Nations." Bantam Classics, 2003.

Teetor, Paul, and JD Long. "R Cookbook, 2nd Edition," n.d. http://www.cookbook-r.com/.

Xie, Yihui. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2019. https://CRAN.R-project.org/package=knitr.