Math Camp - R Section
Summer 2019

Instructor:
Shiro Kuriwaki (kuriwaki@g.harvard.edu)

Course Assistants:
Jake Schneider (jake_schneider@student.hks.harvard.edu)

Course Website:
Canvas (https://canvas.harvard.edu/courses/62068)
Rstudio Cloud (https://rstudio.cloud/spaces/18236/projects)

Overview

This is a 5-day workshop for putting the MPA-ID cohort up and running on the same speed for data analysis in R. Each day includes a brief lecture, demonstration, followed by group work and discussion.

The distinct features of this program compared to other introductions to R are:

- Material tailored to social science and international development
- Coverage of the entire pipeline from data import to cleaning to visualization
- Worked examples and exercises
- Groupwork and learning from each other

Each session is accompanied by some exercises that put into practice the concepts from the day's lecture. The final assignment is due the last day, on August 27th, and some groups will be asked to present or walk through their report in the class.

Schedule

Homework by August 5th: Pre- Math Camp Assignment

Day 1: Mastering the Basics of R in tidyverse
(Thursday August 15, 3 - 5p.m.[1])

- Overview
- dplyr verbs, grouping by a variable, count

[1] We will try to end the formal sessions earlier than the stipulated time, and leave time for individual students or groups to ask the instructors any questions about exercises they may have.

- ggplot aesthetics and layers
- function anatomy: arguments, defaults, pipes

Homework before day 2: simple visualization assignment and dplyr factual questions

Day 2: Mastering the Basics of R beyond the tidyverse
(Monday August 19, 3 - 5p.m.)

- base-R classes: vectors, matrices, summary statistics, 2D arrays, factors, lists
- formula syntax
- object assignment

Homework before day 3: Start visualization assignments

Day 3: Perfecting Graphs
(Wednesday August 23, 3 - 5p.m.)

- How our eyes see data: data-to-ink ratio, position, length, area, density, color
- ggplot technicalities: axes, legends, annotations, and saving to PNG/PDF
- formula syntax and factor reordering in the case of graphics
- (Discuss professional visualizations)

Homework before day 4: Continue visualization assignments

Day 4 : Proper Workflow, R functions recap
(Monday August 26, 1:30 - 3:30p.m.)

- Proper workflow: order of code in a script, restarting and re-running for replicability, input and output, file paths
- Review of core concepts and under-discussed topics in cheatsheet
- Discussion of R and RStudio Desktop

Homework before Day 5: (1) install R and RStudio on Desktop, (2) with your group, prepare a memo on the climate and growth data

Day 5: Presentations, Preparing for Classes
(Tuesday August 27, 1 - 3:30p.m.)

- Discussion of submissions to final exercise 1
- Presentation of climate data analysis memos (three groups)
- Transition to RDesktop
- reproducible examples (reprex)

Exercises and Group Presentation

Students will work in the orientation groups assigned to them on Thursday, August 15th. Students do not need to submit their work (though they are welcome to if they want feedback), other than the following are two exercises for the last day.

Both are based on the dataset from the Dell, Jones, and Olken article[2] we will discuss in class.

1. *Final exercise 1* (due on Canvas by August 27th, 9:30am)*:* A common stylized fact is that countries with warmer climates tend to be poorer than those with colder ones. Using the data you have, make *one* self-contained, well- labelled, and visually pleasant figure or table that is a fair empirical assessment of this stylized fact. Try to utilize the principles and techniques for good graphics you learned as much as possible.

2. *Final exercise 2* (presentations on August 27th, submissions due end of the day): Using the climate data (but not limited to it), create a memo that analyzes a particular aspect of the relationship between climate and political economy.

   The memo can be focused on a specific region, a specific country (or countries), a specific sector, or a specific time period. Examples of ambitious (and by no means necessary) extensions include adding new data from more recent years, or collecting more outcome variables from other data source. Dividing up tasks between group members is encouraged, but group members should learn from each other as much as possible. *The point is to do independent data analysis on an aspect of the data that interests you and present it well.*

   The memo can be made in a google doc or word doc, but it has to be presentable as-is (e.g., in a PDF). Using Rmarkdown is optional.

Grades and Assessment

The submitted problem set will be graded but mainly to provide personal feedback. No grades are recorded, and they will in no way impact your other courses.

Reference Texts

Each day will have slides and a Rstudio cloud space to work on code. We will be basing most of the concepts of the following online (mostly open-source) resources:

---

[2] Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. 2012. "Temperature Shocks and Economic Growth: Evidence from the Last Half Century." *American Economic Journal: Macroeconomics*, 4 (3): 66-95. A cleaned dataset is available on the Canvas assignment page.

- Hadley Wickham, The tidyverse style guide
- Garrett Grolemund and Hadley Wickham, *R for Data Science*
- Garett Grolemund, *Hands-on Programming for R*
- Claus Wilke, *Fundamentals of Data Visualization*.
- John Rauser, "How Humans See Data" (youtube)
- Edward Tufte, *The Visual Display of Quantitative Information*
- Jennifer Bryan and Jim Hester, *What they Forgot to Teach You about R*
- Matthew Gentzkow and Jacob Shapiro, *Code and Data for the Social Sciences: A Practitioner's Guide*

Announcements and Asking Questions

Watch out for any announcements about the course by periodically checking the Canvas page, or by the Announcements we make on the website (these are accompanied by email notifications by default).

We will use the "Discussion" feature on Canvas to collect questions you might have as you prepare for math camp and work on the exercises. Please consider asking a question there first *before emailing me* in person -- This allows all students to benefit from each other's comments although I'm happy to answer in person too.

Some ways to ask questions are better than others. A question should have enough information for me to help you -- this typically involves a reproducible example (a reprex) for coding errors or screenshots for non-code issues.[3]

Getting in touch

The best way to contact Shiro is through email (kuriwaki@g.harvard.edu). My regular office is not in the HKS building (CGIS Knafel 428) so it is harder for me to meet in HKS outside scheduled hours in short notice, but I am happy to schedule meetings by email. For information about the stats sequence, please contact Professor Dan Levy (dan_levy@hks.harvard.edu).

[3] From the website: "Imagine that you've made a cake, and for some reason it's turned out absolutely awful - we're talking completely inedible. Asking a question *without* a reprex is like asking, "Why didn't my cake turn out right?" – there are hundreds of possible answers to that question, and it's going to take a long time to narrow in on the exact cause for your inedible cake creation."