HARVARD Kennedy School
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

Theodore Svoronos
Lecturer in Public Policy
617.495.7587
theodore_svoronos@hks.harvard.edu

January 2, 2020

Dear students,

First, thanks for expressing interest in API-202 Z! I am the instructor of this section and am excited for the coming semester.

I am writing to ask that you spend some time before the Spring semester becoming familiar with R, the statistical software package we will use in API-202 Z. If you have taken API-201 Z, you already have some familiarity with the language; if not, the language may be totally new to you. Regardless of which bucket you fall into, this document contains advice for preparation before the semester depending on your background, information on an assignment due before the semester starts, and some guidance on using R.

Contents of this document:

(1) Pre-semester work
(2) Guidance on pre-semester work
(3) Training sessions
(4) Accessing R
(5) Why we are using R
(6) Learning R

---

(1) Pre-semester work

Tutorials: The tutorials to learn R before the Spring consist of the following 5 primers:

1. Visualization Basics
2. Programming Basics
3. Working with Tibbles
4. Isolating Data with dplyr
5. Creating Variables with dataframes

R winter assignment: Regardless of your prior background with R, you will be expected to complete and submit this assignment, which will help you practice the skills that you need to master *before* starting the course. You will be asked to submit your assignment on Canvas, and will receive instructions on how to do so in the coming weeks. The due date for this assignment is the first day of classes, January 27th.

Shiro Kuriwaki: Shiro is a Ph.D. student who will be conducting the R math camp sessions. He has experience helping students from a range of programs (undergraduate, masters, and PhDs) learn R, and will be a valuable resource for you. If you have questions about the tutorials or the R winter assignment, please feel free to reach out to him (kuriwaki@g.harvard.edu).

(2) Guidance on pre-semester work

Below is some specific guidance on how to engage with R before API-202 Z gets underway, which varies by your background in R. Please note that this is the minimum we expect you to do before you begin the course, taking into account that you probably have limited time before you come. If you have more time available, I highly encourage you to explore other R resources (such as the other primers from RStudio).

Note: I would expect most students who you have taken API-201 Z to fall into the "Intermediate" category.

- **Beginner**: Have never used R or do not feel comfortable importing data, calculating basic summary statistics, or generating new variables from an existing dataset.
  - Complete the 5 R primers
  - Complete the R winter assignment
  - Estimated time: 12-15 hours (depending on your prior programming skills)
- **Intermediate**: Have some familiarity with the R interface, through a course, online tutorial, or self-learning, but have not used it for long problem sets or research projects.
  - Complete the subset of the R primers that suits your needs
  - Complete the R winter assignment
  - Estimated time: 8-12 hours
- **Advanced**: Have significant applied experience using R in real-world contexts, comfortable with advanced commands and their options.
  - Make sure you are familiar with the `tidyverse` suite of packages (especially `dplyr` and `ggplot`).
  - Complete the R winter assignment
  - Estimated time: 3-5 hours

(3) Training sessions

We have developed several in-person training sessions for early in the semester to build off the work that you do in these tutorials. Signups for these sessions will take place during the first week of class.

(4) Accessing R

R is a free programming language. At the beginning, we will be using RStudio Cloud, which will enable you to learn R with a light setup. Part of the winter assignment asks you to set up an RStudio Cloud account and get started. You can use RStudio Cloud from any computer, but requires a decent internet connection. At some point after you arrive to HKS, you will be asked to transition from RStudio Cloud to RStudio Desktop.[1]
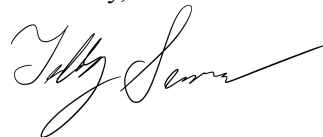
(5) Why we are using R

This is the first year the Z-section statistics sequence will be using R as the default programming language. Up until this year, we had been using Stata. While this decision involved some tradeoffs, we ended up going with R because we believe that it will allow our graduates to be better prepared for a wider range of jobs, it is better for areas of data analysis that are becoming increasingly popular (machine learning, data visualization, etc.), it is free, and it is better for analyzing a broader range of data (text, GIS, etc.).[2]

(6) Learning R

Learning R can be exciting and empowering, but it also involves some challenges. The language is open-source, which means that anyone in the world can develop code (which become packages) that others can download and use. While this has led to rapid innovation and versatility, it means that there are often many packages available for doing the same task, and choosing which one is best can be a challenge (this is especially true for Stata users, who are used to having a single company behind the code). To help you learn R in the most efficient way possible, have developed a "standardized" toolkit for coding in R that will provide you with the default HKS way of doing certain common tasks in R. We will share a cheat sheet related to this toolkit once the semester starts.[3] You should feel free to deviate from this version, but we encourage you to adopt it when you are first learning R.

I hope you will make use of these resources to develop your R skills before you arrive on campus, so you can spend more time in API-202 Z learning about how to conduct statistical analyses in the policy world and less time on mastering the software needed to conduct such analyses. Please do not hesitate to contact me if you have any questions or need further information. I look forward to working with you in the coming semester!

Sincerely,

Teddy Svoronos
Lecturer in Public Policy
617.495.7587 | theodore_svoronos@hks.harvard.edu

---

[1] You may download these beforehand if you wish. To install R, select the Windows or Mac links here. To install RStudio, R's Interactive Developer Environment, select the appropriate operating system under "Installers for Supported Platforms" here.

[2] Also, unlike Stata, SAS, or SPSS, R is a generalized programming language, meaning that it handles various data types and can build other software. And unlike Python and other programming languages one might use in Computer Science, R is a language that is built around data analysis as one of its main goals.

[3] In the meantime, you might want to develop your own cheat sheet as you go through the tutorials, so you can keep track of the commands you are learning and more easily recall them when you need them in the assignment.