SPA 096: Conduct of Inquiry Lab

Instructor: Wali Reheman
School of Public Affairs, American University

Fall 2024

Course Information

- **Meeting Time & Location:** Tuesdays 4:30pm-5:30pm, Zoom
- **Instructor:** Wali Reheman
- **Email:** rw8143a@american.edu
- **Phone:** 202-391-3091

Course Description

This course is a lab supplement to the **SPA 612: Conduct of Inquiry I**. This lab is an introduction to R programming, focusing on data wrangling and visualization. The course aims to equip students with essential skills to manipulate, analyze, and visualize data using R and Stata.

Course Objectives

- Understand the basics of R programming and Stata.
- Learn data manipulation techniques using Stata and dplyr and tidyverse in R.
- Create various types of data visualizations using Stata and ggplot2 in R.
- Develop skills to clean and prepare datasets for analysis in R and Stata.

Evaluation

The evaluation of this course will be entirely based on participation.

Prerequisites

- Please install R and R Studio before the first class. Professor Ryan Moore has a short guide to R and R Studio. https://www.ryantmoore.org/files/class/introPolResearch/intro_R_short.pdf
- We will start using R Markdown from Week 2. If you have time, please See this introduction about R Markdown. https://rmarkdown.rstudio.com/lesson-1.html

Textbook and Materials

- R for Data Science (2e) by Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Grolemund. https://r4ds.hadley.nz/
- Introduction to Stata by CSCAR (University of Michigan). https://cscar.github.io/workshop-stata-intro/01-the-basics-of-stata.html

Software

R and RStudio
These are the two required applications for R users in this course. Both are free to download and use!

- To download R, you must use a CRAN (Comprehensive R Archive Network) mirror: https://cran.r-project.org/mirrors.html
- To download RStudio, you can go here: https://posit.co/download/rstudio-desktop/

Stata
Stata is not a free application, which makes it more difficult for students to use on personal computers. There are a few options:

1. With a good Internet connection, Stata can also be accessed remotely through AU's Virtual Applications: https://apps.american.edu/
2. You can also use Stata in some of our on-campus labs: https://edspace.american.edu/ctrl/stata/

Chat GPT

Chat GPT is our friend in this class. The benefit is obvious: it gives you answers and lets you ask follow-up questions! However, I suggest you try to use Chat GPT only for debugging. When you run into errors in your R code or Stata code, describe the problem and share the relevant code snippets. Remember that it's best used to support your learning.
If you have to let Chat GPT write R code or Stata code for you, please add the following to your prompt: Explain to me why you do it this way, and add comments to explain each step of your code.

Weekly Schedule

1. **Week 1 (8.27): Introduction to R**
    - Overview of R and RStudio.
    - R Markdown for reporting and reproducibility.
    - Basic R syntax and commands.
    - Installing and loading packages.
2. **Week 2 (9.03): Data and Basics**
    - Importing data from various sources.
    - Exporting data to different formats.
    - Understanding vectors, data frames, and lists.
    - Manipulating and exploring data structures.
3. **Week 3 (9.10): Data Wrangling with `dplyr`**
    - Introduction to the `dplyr` package.

- Selecting, filtering, and mutating data.

4. **Week 4 (9.17): Stata**
    - Understanding previous content in Stata.
5. **Week 5 (9.24): Review**
    - Troubleshooting questions and problems
6. **Week 6 (10.01): Data Summarization**
    - Grouping data.
    - Calculating summary statistics.
    - Creating descriptive tables.
7. **Week 7 (10.08): Introduction to `ggplot2`**
    - Basics of the `ggplot2` package.
    - Creating simple plots.
8. **Week 8 (10.15): Advanced `ggplot2`**
    - Customizing plots.
    - Using facets and themes.
9. **Week 9 (10.22): Stata**
    - Understanding previous content in Stata.
10. **Week 10 (10.29): Review**
    - Troubleshooting questions and problems
11. **Week 11 (11.05): Election Day- no classes.**
12. **Week 12 (11.12) : Integration with with `dplyr` and `ggplot2`**
    - Integration with `dplyr` and `ggplot2`.
13. **Week 13 (11.19): Output and report your results**
    - `stargazer` in R
    - Output tables in Stata
14. **Week 14 (11.26): Thanksgiving holiday - no classes**
15. **Week 15 (12.03): Review**
    - Troubleshooting questions and problems
16. **Week 16 (12.10): Final Week - no Classes**

Office Hours

By appointment (please email to schedule).

Policies

- **Attendance:** Regular attendance is expected. Notify the instructor in advance if you need to miss a class.
- **Academic Integrity:** Adherence to the university's academic integrity policy is mandatory.
- **Intellectual Property:** Course content and developed ideas is the intellectual property of the instructor or student who created it, and may not be distributed without consent.

Additional Resources

- **AU Quantitative Support:** American University offers quantitative support including Math & Stats Lab and Online Statistical Software Support. https://www.american.edu/provost/academic-access/quantitative.cfm
- **Swirl:** An interactive platform for learning R directly within the R console. https://swirlstats.com/
- **DataCamp:** Offers interactive courses on R programming and data science. https://www.datacamp.com/