PS 553: Introduction to Statistical Computing for Political Science  
Spring 2017

Instructors: Sarah Bouchat (bouchat@wisc.edu)  
Hannah Chapman (hschapman@wisc.edu)  
Time and Place: Fridays 8–10 am in Social Science 3218  
Office Hours: By appointment

Course Overview
This course provides an introduction to statistical programming techniques for cleaning, analyzing, 
and graphically representing empirical data. We will learn basic techniques in both R and Stata, 
which are used extensively in social science research. The primary goal of the course is to provide 
a solid foundation in basic programming skills and in the fundamentals of these two programs 
so that you will have the necessary background to use these techniques in your own research and 
learn more advanced methods on your own.

The course will be divided into two parts and will meet for the first half of the Spring 2017 semester. 
Part I of this course will focus on R. Part I will last the first 3 weeks of the semester. Part II, focusing 
on Stata, will be in weeks 4–6 of the semester. Week 7 will be an open work and question session 
with both instructors so that you can ask questions about course material and your final projects.

Communication
Both instructors are available to you for questions during class and office hours. Because of the 
small class size and shortened class schedule, office hours will be by appointment. To fully benefit 
from the instructors resources and knowledge, please direct questions specific to material from Part 
I or Part II of the course to the relevant instructor. For administrative issues and questions, please 
contact both instructors by email.

Assignments
This is a one-credit class graded on a pass/fail basis. We will distribute exercises that align with the 
material for each weekly session, but these will not be collected or graded. Solutions will be posted 
to the course website, and we will be available to answer questions during class or office hours.

This course is self-directed and We will not be taking attendance. To pass the course, you must 
complete a satisfactory final memo. This 4–5 page, double-spaced memo must include:

(a) A research question of interest to you

(b) A description of data used to address this question, including sources and methods you used 
to clean and organize the data

(c) Code (a Stata .do file or R script file) that you used to do the analysis in the memo, including 
the transformation of your raw data as well as code to produce figures and tables

This memo can aid you in a final research project for another class this semester or in a research 
project of your own interest. We will be providing comments on your code and analysis, although 
you will not receive a letter grade for the assignment. This course is intended to supplement your 
statistical training in political science and to facilitate your own research. The material will be most 
useful if you keep up with course material and complete your final memo on time. Please consult 
instructors if you have questions or have specific circumstances that require you to deviate from 
this project format.

Final memos are due before Spring Break, on 17 March 2017.
Part I: R Preliminary Schedule

1  Introduction and Data Management (20 January)

Topics
- Introduction to R
- R script files
- Commenting your code
- Pseudocode
- Replication
- Simple programming techniques
- Inputting data
- Data types
- Data cleaning, merging, appending
- Recoding variables
- Indexing
- Reshaping
- Collapsing data

Readings & Resources
- Why You Should Master R (even if it might eventually become obsolete)
- Download R
  - Base R
  - RStudio (desktop)
- R for Data Science: Basics link
- R for Data Science: Scripts link
- Optional: Advanced R: Data types link

2  Graphics and Visualization & Basic Data Analysis (27 January)

Topics
- Scatterplots, histograms, box plots, density plots
- Editing graphics in R
- Using graphics for diagnostics
- ggplot2
- Summary statistics
- Statistics by groups
- Linear regression
- Exporting results
- Additional resources for learning other methods in R

Readings & Resources
- R for Data Science: ggplot2 link
R for Data Science: Exploratory Data Analysis link
Rice, Ken. “How to Visualize Summaries.” (Slides) link

3 Advanced Topics (3 February)

Topics
- tidyverse
- dplyr
- Regular expressions
- Text analysis*
- ggmap and Google location data*
- Scraping html with R*

Readings & Resources
- R for Data Science: Wrangle data with tidyverse (sections 9–21) link
- Regular Expression Tester
- dplyr vignettes link

Part II: Stata Preliminary Schedule

1 Introduction and Data Management in Stata (10 February)

Topics
- What is Stata good for?
- Introduction to Stata
- Stata best practices
- Stata Do files
- Inputting data
- Data cleaning, merging, appending
- Recoding variables
- Installing packages

Readings & Resources
- TBD

2 Graphics and Visualization & Basic Data Analysis (17 February)

Topics
- Editing graphics in Stata
- Using graphics for diagnostics
- Summary statistics
- Linear regression
- Post-estimation tests
Readings & Resources
TBD

3 Advanced Topics (24 February)

Topics
- Exporting results
- Troubleshooting
- Panel data
- Survey data
- Student choice!

Readings & Resources
TBD