PS 553: Introduction to Statistical Computing for Political Science

**Instructor:** Emily Sellars (esellars@wisc.edu)
**Time and Place:** Tuesdays, 8:45-10:45 am in Social Science 3218
**Office Hours:** TBA

This course provides an introduction to statistical programming techniques for cleaning, analyzing, and graphically representing empirical data. We will focus on learning the basics of two programs in particular, Stata and R, which are used extensively by social scientists. 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.

**Course Format**

The course will be primarily conducted in a “flipped-classroom” format. In advance of the class each week, I will post slides and other resources on Learn@UW that cover the week's material. I will present a brief overview of the material at the beginning of class, but you will spend most of the class time applying the techniques to empirical data. I will provide suggested datasets to work with each week, and you may also choose to work with your own data. I will be there to answer questions, check your code, and provide assistance.

**Assessment**

This is a one-credit class, and it will be graded on a pass/fail basis. To pass the course, you must complete a satisfactory final project. I will assign homework/practice problems roughly weekly, but they will not be graded. I will post solutions to these problems on Learn@UW. I would be happy to answer questions about the practice problems in office hours or during class work time. Keep in mind that this is a self-directed course. I will not be taking attendance or collecting homework, so it is up to you to keep up with the material and complete your final project on time. You will get out of the class exactly what you put into it.

**Final Project**

Your final project is to apply the methods from this class to a research project. You may combine this with a paper for another course with the consent of that instructor. Your final project should have three components:
1. A short paper (roughly five to seven double-spaced pages) with the following information:
   - A statement of your research question
   - A description of your data, including the sources of your data and the transformations necessary to complete your analysis
   - A description of the methods you used to analyze your data and why they are appropriate given your research question
   - A presentation of your results using the necessary tables and figures

2. Computer code (usually a Stata do file or an R script file) that transforms your raw data into the form used in your analysis

3. Computer code that executes your analysis and produces the figures and tables.

Please come and talk to me if you have questions about the final project or have specific circumstances that would require you to deviate considerably from the suggested project format. The final project will be due on the last day of the exam period (May 17th).

Topics and Preliminary Schedule

1 Programming Basics and Best Practices

January 21st and 28th:
- Introduction to R and Stata
- Stata do files and R script files
- Commenting your code
- Pseudocode
- Replication
- Some simple programming techniques

2 Data Management

February 4th, 11th and 18th:
- Inputting data into Stata and R
- Data types
- Cleaning, merging, and appending datasets
- Recoding variables and generating new variables
- Indexing
- Reshaping datasets
- Collapsing data
3 Graphics and Visualization

February 25th and March 4th:
- Scatterplots, histograms, box plots, stem-and-leaf plots, and density plots
- Editing graphs in R and Stata
- Using graphics for data exploration/diagnostics

4 Basic Data Analysis in R and Stata

March 11th and 25th:
- Summary statistics in R and Stata
- Statistics by groups in R and Stata
- Using survey weights in R and Stata
- Linear regression in R and Stata
- Exporting results from R and Stata
- Sources to learn other methods for R and Stata

5 Spatial Data in ArcGIS, R, and Stata

April 1st and 8th:
- Introduction to ArcGIS
- Types of spatial data
- Basic choropleth maps
- Spatial statistics
- Extracting spatial data

6 Text and String Processing

April 15th and 22nd:
- Dealing with string variables
- Regular expressions
- The tm package for R

7 Wild Card Weeks

April 29th and May 6th:
- Possible topics include writing your own programs in R and/or Mata, macros, more ArcGIS, simulation and Monte Carlo methods, or something else by request