COURSE INFORMATION

Instructors:  Hannah Chapman, Ph.D. candidate (hschapman@wisc.edu)
             Michael DeCrescenzo, Ph.D. candidate (decrescenzo@wisc.edu)

Office Hours:  By appointment

Time and Place:  Fridays 8–10 a.m., Social Science 3218

Credits:  1 unit, pass/fail

Credit hour policy standards:  Per new university policy (https://teachlearn.provost.wisc.edu/course-syllabi/), syllabi must specify how courses meet credit hour policy requirements. We employ the “45 hours per credit” definition. Language recommended by the university (with slight edits) follows:

The credit standard for this course is met by an expectation of a total of 45 hours of student engagement with the courses learning activities, which include regularly scheduled in-class instruction (2 hours weekly, for 7 weeks) in addition to required exercises, reading, consulting online resources, office-hour meetings with instructors, and other work described in the syllabus such as the final project.

Course designations/attributes:  Graduate or professional standing. Undergrads may enroll with instructor consent

Online resources:  Canvas: https://canvas.wisc.edu/courses/86298
                  Online R materials: https://mikedecr.github.io/811
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 goal of the course is to provide a solid foundation in the fundamentals of these two programs so that you can use them in your own research and learn advanced methods on your own.

The course will meet for the first half of the Spring 2018 semester and will be divided between the two instructors. Hannah Chapman will focus on Stata for three weeks, and Michael DeCrescenzo will focus on R for three weeks. Week 7 will be an open work and question session with both instructors for questions about course materials and final projects.

OFFICE HOURS AND COMMUNICATION

Both instructors are available to you for questions during class and office hours. Because of the small class size, the individual research interests of each student, and the shortened class schedule, office hours will be held by appointment only. To benefit from each instructor’s specialized knowledge, please direct questions about Stata to Hannah Chapman and questions about R to Michael DeCrescenzo. For broader administrative issues and questions, please contact both instructors by email.

ASSIGNMENTS AND GRADING

This is a one-credit class graded on a pass/fail basis. Assignments include a final memo and weekly coding exercises.

Each week, we will distribute exercises that complement the material for each session. These exercises should be submitted to the appropriate Dropbox folder on Canvas no later than 5pm on the Thursday following the class session. Solutions will be posted to the course website, and instructors will be available to answer questions about exercises during class and office hours.

To pass the course, you must also complete a satisfactory final memo. This 4–5 page (double-spaced) memo must include:

1. A research question of interest to you.

2. A description of data used to address this question, including sources and methods you used to clean and organize the data.

3. A brief analysis of the data using the tools learned in class.
4. An accompanying script file (a .do file, a .R file, or both) 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. Code is (obviously) not counted toward the 4–5 pages of writing.

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 provide comments on your code and analysis, although you will not receive a letter grade for the assignment. Please consult instructors if you have questions or encounter specific circumstances that require you to deviate from this project format.

**Final memos are due before spring break, on March 24**

This course is intended to supplement your statistical training in political science and to facilitate your own research. It is not, however, an end-all training in these softwares. It is important to recognize that each researcher has individual needs and develops individual competencies with their software tools of choice. When this class ends, your paths with these softwares will diverge significantly. As such—although we are happy to help with difficult problems—we encourage students to practice their skills in reading software documentation and seeking online help.

The material will be most useful if you keep up with course material and complete your final memo on time. Further, we will not take attendance.

**CLASS CONTENT**

Courses will cover the following main themes: data manipulation, graphics, statistical analysis, and some special topics. If there is any specific content that you would like us to cover (e.g. survey data, panel data, spatial analysis, Bayesian analysis, etc.), **please notify us ASAP**.

Notes and slides will be posted online and should be reviewed *before* coming to class. We will spend class time on a few examples of the concepts in the notes, leaving lots of time to work on exercises and ask questions.

Here is an abbreviated list of lecture dates and topics. Detailed descriptions of each session can be found in the following section. **Do check the detailed descriptions for preparatory readings and tasks!** It is very important that you complete the “before class” material for each lesson.

- January 26, NO CLASS
- February 2, Introduction and Stata (Data Manipulation)
- February 9, Stata (Graphics and Data Analysis)
- February 16, R (Data Manipulation)
- February 23, R (Graphics)
- March 2, R (Data Analysis and Advanced Topics)
- March 9, Stata (Advanced Topics)
**Detailed Class Schedule**

*January 26, No Class*

*February 2, Introduction and Stata (Data Manipulation)*

- Before class:
  - Introduction and additional resources presentation
- Topics include:
  - What is Stata good for?
  - Introduction to Stata
  - Best practices
  - Stata .do files
  - Inputting data
  - Data cleaning, merging, appending
  - Recoding variables
  - Installing packages

*February 9, Stata (Graphics and Data Analysis)*

- Before class:
  - Introduction and additional resources presentation
- Topics include:
  - Editing graphics in Stata
  - Using graphics for diagnostics
  - Summary statistics
  - Linear regression
  - Post-estimation tests

*February 16, R (Data Manipulation)*

- Before class:
  - Review online introduction and both the “R Basics” and “Data” lessons ([https://mikedecr.github.io/811](https://mikedecr.github.io/811))
- Topics include:
– Introduction to R
– R script files
– Object-oriented programming
– Directories
– Data frames, data entry, and external data
– Cleaning and recoding existing data (dplyr)
– Merging and appending new data
– Summarizing data (including aggregating and collapsing)
– Reshaping data (wide, long and “tidy” data, tidyr)
– Piping data (magrittr)

• Additional reading and resources:
  – R for Data Science (http://r4ds.had.co.nz/)
  – Intro to piping, tidyr, and dplyr
    (https://rpubs.com/bradleyboehmke/data_wrangling)
  – Short style guide (http://adv-r.had.co.nz/Style.html)
  – Tidyverse style guide (http://style.tidyverse.org/)
  – Download Rstudio (https://www.rstudio.com/products/rstudio/download/)
  – Rstudio blog (https://blog.rstudio.com/)
  – R-Bloggers (https://www.r-bloggers.com/)
  – RPubs (https://rpubs.com/)
  – David Robinson (http://varianceexplained.org/)

February 23, R (Graphics)

• Before class:
  – Review online “Graphics” lesson
  – Comfort with dplyr and tidyr verbs

• Topics include:
  – Intro to ggplot2
  – ggplot2 vs base graphics
  – Geoms for various graphics
  – Aesthetic mappings from data (color, fill, size, shape)
  – Facets and scales
  – Themes and theme elements
  – Saving graphics
- Prepping data (tidy data, factors and strings using forcats and stringr)
- Extensions for ggplot

- Additional reading and resources:
  - R4DS chapter on visualization (http://r4ds.had.co.nz/data-visualisation.html)
  - Data Visualization for Social Science by Kieran Healy (http://socviz.co/index.html)

March 2, R (Data Analysis and Advanced Topics)

- Before class:
  - Review online “Analysis” lesson
  - Comfort with setting and navigating directory pathways (when loading data and saving graphics)

- Topics include:
  - How regression works in R
  - Linear and generalized linear models
  - Summarizing model estimation: tables and graphics
  - Exporting model output (e.g. to \LaTeX)
  - Post-estimation analysis
  - Generating model predictions
  - Resources for other estimation tools (arm and rms, survey weights, time series and panel analysis, hierarchical models, Bayesian estimation software, etc.)
  - Advanced R topics: lists, apply() functions, nesting and mapping
  - R for Replication
  - Student choice!

- Additional reading and resources:
  - Advanced R by Hadley Wickham (http://adv-r.had.co.nz/)

March 9, Stata (Advanced Topics)

- Before class:
– Introduction and additional resources presentation

• Topics include:
  – Exporting results
  – Troubleshooting
  – Panel data
  – Survey data
  – *Student choice!*