Syllabus
Soc. 365
Computing in Sociological Research
Fall 2014

General Information:
Time: Friday 8:50-10:45
Room: 3218 Sewell Social Science Building
Instructor: Jim Raymo
Office: 4452 Social Science Building
Tel: 262-2783
Email: jraymo@ssc.wisc.edu (Julia Goldberg: jsgoldbe@ssc.wisc.edu)
Office Hours: 4-5 Monday (Julia’s office hours TBA)
Class Page: class materials will be posted on Learn@UW

Required Textbooks:
Michael N. Mitchell, Data Management Using Stata: A Practical Handbook. (Available at UBS)

The data sets used in this book can be downloaded from:
These files will allow you to replicate all of the examples in the book. We will download these files on the first day of class.

Introduction
This is a course in how to manage social scientific data. Statistics courses in Sociology (and other social sciences) provide students with a solid theoretical understanding of data analysis but typically do not provide sufficient training in how to actually prepare and work with real data to apply those analytical tools. Because the large majority of research time (in any research job) is spent on data management, this is an important shortcoming in your training. By the end of this course you will understand the structure of different types of social scientific data, how to clean messy data, how to effectively document data, how to merge data from multiple sources, and how to restructure data for analysis. You will also learn techniques for visual display of data (i.e., graphing) to identify patterns and problems and to effectively convey information to consumers of your research. To this end, we will work with a widely-used software package for the management and analysis of social scientific data (Stata). Hands-on, nuts and bolts, work will be supplemented throughout the semester with discussion of the bigger picture – why is careful and effective data management and preparation so essential? You will also have a chance to work with a publicly available data set of your choosing to examine a question that interests you. This is not a statistics course but you will find that a prior statistics course—e.g., Soc. 360, or the equivalent—will be very helpful. I am also assuming that you have all had an introductory
research methods course such as Soc 357. No previous experience with Stata is necessary, but again, any familiarity you have with the program or similar programs (SPSS, SAS, etc.) will certainly be a plus. There are many on-line resources for learning Stata and for troubleshooting – you may want to explore some of the sites at http://www.stata.com/links/resources-for-learning-stata/.

Course Requirements:

Class participation: You are expected to complete the assigned reading before each class. Lectures and in-class work will assume that you have read the required readings. Valuable class time is reserved for hands-on work, examples, discussion, and clarification. Class participation will not factor into your grade directly but you should assume that it will have an indirect impact (via your ability to complete the homework assignments and the quality of your final project and exam). If unable to attend for some reason, you should arrange to get class notes from another student.

Weekly Exercises: Exercises are due before class the week after they are assigned. I will not read and grade each assignment but the class TA (Julia Goldberg) and I will provide feedback on a few randomly chosen assignments each week, will post model answers on Learn@UW, and will allocate some class time the following week to discuss questions about the assignments. Late exercises will not be accepted. If for some reason you do not complete your assignment on time, I encourage you to complete it on your own, but I will not accept it for credit. Most of these assignments will be based on simulated data provided by the instructor. Some assignments will also make use of data from large publicly-available surveys such as the General Social Survey. These assignments will count for 33% of your final grade so it is in your best interest to complete and submit these on time (I will keep track of whether assignments are done completely and on time). Please submit all assignments via Learn@UW by 8:30 (right before class) on the date that they are due.

Test: On the last day of class (12/12), we will have an open-book exam on which you will be asked to work with a prepared data set to conduct a series of data manipulations similar to those covered in the text and the weekly assignments. The test will count for 33% of your final grade.

Project: The course will culminate in a research project in which you will use the technical and analytic skills developed in class to address a research question of your choosing. This project will involve choosing a publicly available data set (in consultation with the professor), carefully describing those data, addressing missing data, conducting consistency checks, recoding variables, and preparing basic descriptive results to answer your question. This project will count for 34% of your final grade in the course. Detailed instructions for the project are on the Learn@UW page.

Computing: All assignments will require manipulation of sample survey data, using the statistical package Stata, Version 13. There are many other similar statistical packages but I think that Stata offers the best combination of power, flexibility, and ease of use. For those of you thinking about graduate school in Sociology or another social science, this course will be an excellent opportunity to master the leading software package. Stata is available via Winstat, meaning that you can do your work from any computer as long as you have an internet connection. An
An introduction to Winstat will be provided on the first day of class. You are certainly welcome to buy a copy of Stata for your own computer if you like but it is not cheap (around $200). If you are interested in this, please let me know.

You will probably not be able to complete the course without frequent reference the Stata manuals. Fortunately, the manuals are now available electronically via Winstat. The actual manuals are quite expensive (and heavy) so it is very nice to have full access via Winstat for no cost. To access Stata manuals, you can just click on Start and Stata in Winstat – the manuals are in pdf format.

In addition to Stata (and other statistical resources available via Winstat), SSCC provides statistical consulting to students in this course. I encourage you to take advantage of these resources when you get stuck.

The course web page. I have set up a page for the course on Learn@UW, which contains this syllabus, slides from class, weekly assignments, additional readings, and data sets we will use in class and for assignments. Each Friday evening (or Saturday morning) I will put up an “Illustrative Answer” to the exercise you turn in that day, which you may also download or print. I urge you to check the web site frequently since it will contain the most up-to-date information regarding the course.

Grading summary:
Weekly assignments: 33%
Final project: 33%
Final exam: 34%

Special Needs: To make special arrangements for testing, assignments, or other aspects of the course you must qualify for disability services through the McBurney Center. Their website has detailed instructions on how to qualify: http://www.mcburney.wisc.edu/. Please notify me in the first two weeks of class if you have or anticipate having authorization from the Center and we will make the necessary arrangements.

Academic Honesty: As with all courses at the University of Wisconsin, you are expected to follow the University’s rules and regulations pertaining to academic honesty and integrity. Students are expected to know and follow the standards outlined by the Offices of the Dean of Students. See their website (http://students.wisc.edu/saja/misconduct/misconduct.html) for a complete description of behaviors that violate the University’s standards as well the disciplinary penalties and procedures.
**Course content**

Week 1 (Sept. 5): Introduction. What do data look like? Why is data management so important? 
Real world applications.

Week 2 (Sept. 12): Inputting data into Stata  
Reading: Mitchell, Chapters 1 and 2, NYT articles (“Big Data’s Impact,” “How Companies Learn Your Secrets,” and “Sizing up Big Data”)  
Assignment: Exercise 1

Week 3 (Sept. 19): Data cleaning  
Reading: Mitchell, Chapter 3  
Assignment: Exercise 2  
Due: Exercise 1

Week 4 (Sept. 26): Labeling, codebook, documentation  
Reading: Mitchell, Chapter 4, Long, Chapter 2  
Assignment: Exercise 3  
Due: Exercise 2

Week 5 (Oct. 3): Recoding, creating new variables  
Reading: Mitchell, Chapter 5  
Assignment: Exercise 4  
Due: Exercise 3

Week 6 (Oct. 10): Combining data sets  
Reading: Mitchell, Chapter 6  
Assignment: Exercise 5  
Due: Exercise 4

Week 7 (Oct. 17): Basic descriptive statistics (tabulation and summarization)  
Reading: Kohler and Kreuter Chapter 7  
Assignment: Exercise 6  
Due: Exercise 5

Week 8 (Oct. 24): Visualizing data - Graphics in Stata (Guest lecture: Russell Dimond)  
Assignment: Exercise 7  
Due: Exercise 6

Week 9 (Oct. 31): Missing data  
Reading: Allison Chapter 1  
Assignment: Exercise 8  
Due: Exercise 7
Week 10 (Nov. 7): Processing observations across subgroups  
  Reading: Mitchell, Chapter 7  
  Assignment: Exercise 9  
  Due: Exercise 8

Week 11 (Nov. 14): Reshaping data (longitudinal data)  
  Reading: Mitchell, Chapter 8  
  Assignment: Exercise 10  
  Due: Exercise 9

Week 12 (Nov. 21): Programming for data management 1 (Guest lecture: Russell Dimond)  
  Reading: Mitchell, Chapter 9  
  Assignment: Exercise 11  
  Due: Exercise 10

Week 13 (Nov. 28): No Class. Thanksgiving break

Week 14 (Dec. 5): Programming for data management 2  
  Reading: TBA (focus on use of \texttt{egen} command)  
  Due: Exercise 11

Week 15 (Dec. 12): Test, Final draft of term papers due