

SOC 365 Spring 2022
Data Management for Social Science Research
mid-semester update

General Information

Time Monday 9:55 AM – 11:50 AM
Location Sewell 3218
Mode In-person (class meetings) with online evaluation
Class Page All class materials will be posted on Canvas
Instructor Griffin JM Bur
Email gjmbteaching@gmail.com (please do not use my UW address as it is much easier for me to miss important e-mails there).
Office Hours: Mondays: 12:30 PM – 1:30 PM
Fridays: 11:00 AM – 12:00 PM
I'm also happy to meet with you by appointment.

I am happy to meet in person (my office is SSSB 3475) or via Zoom.
The Zoom link for office hours is here: <https://tinyurl.com/365OfficeHours>.

COURSE OVERVIEW

Course Description

This is a course in how to manage social scientific data. Statistics courses in sociology (and other social sciences) might provide students with a solid theoretical understanding of data analysis, but they 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, concrete work will be supplemented throughout the semester with discussion of the bigger picture – why is careful and effective data management and preparation so essential? For your final project, 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 a prior or concurrently taken statistics course—e.g., Soc. 360, or an equivalent—is required so that everyone has a baseline understanding of the statistics we will work with. Knowing how and when to use basic descriptive statistics such as two-way tables, means, and correlations will be reviewed and is a required part of the course. 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 listed on our Canvas page as well as at <http://www.stata.com/links/resources-for-learning-stata/>.

Learning Objectives

Beyond the specific substantive and methodological content that I will cover in this course, I have designed this course to achieve the following instructional objectives designated as priorities by the Department of Sociology:

Conduct Research and Analyze Data. Although professional-quality research requires graduate-level training, we expect that all undergraduate majors will be able to conduct small-scale research in which they formulate a research question, collect data, analyze results, and draw conclusions.

Communicate Skillfully. Sociology majors write papers and make oral presentations that build arguments and assess evidence in a clear and effective manner.

Prepare for Graduate School and the Job Market. Students use their social research skills to identify opportunities for employment or further study, assess their qualifications for these opportunities, and identify strategies for gaining the necessary knowledge and experience to improve their qualifications. Students are encouraged to develop and maintain portfolios of their written work and educational experiences to aid them in preparing applications.

Improve Project Management Skills. Students will improve their skills in time management, ordering and executing a series of complex and inter-related tasks, and integrating distinct components of a project into a final product.

Required Texts

Michael N. Mitchell, *Data Management Using Stata: A Practical Handbook*. (Available at UBS)
All other readings will be available as PDFs in Canvas.

The data sets used in this book can be downloaded from:

<http://www.stata-press.com/data/dmus.html>.

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.

COURSE LOGISTICS

Course Website

All course materials will be centralized on the Canvas site. Each week in the semester has its own page, where all readings, assignments, and other materials will be posted. All assignments should be submitted through Canvas.

Course Communication

My main mode of communicating class announcements will be in-class and through Canvas. Please make sure you that have turned on notifications so that you receive these announcements as an e-mail. Announcements are also visible in the “Announcements” tab in Canvas.

Individual communication should take place over e-mail (gjmbteaching@gmail.com) or in-person during office hours. Please allow me 24 hours during the business week to respond to you. If you still don’t get a response after 24 hours, please send a reminder e-mail to make sure I didn’t miss your e-mail.

Computing

All assignments will require manipulation of sample survey data, using the statistical package Stata. 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 to you through two modes; you are welcome to choose whichever your prefer:

- You may download Stata to your personal computer. Stata is available to you for free from the campus software library (<https://www.doit.wisc.edu/services/software/>). We'll go over this on the first day of class.
- You may use Stata via Winstat, meaning that you can do your work from any computer as long as you have an internet connection. An introduction to Winstat will also be provided on the first day of class.

Help with Stata

Even experienced Stata users frequently use the help manuals. These are available both in Stata (type `help [your command name or topic]`) or for free online as PDFs. Another quick way to find help (and a surprisingly reliable option) is to search for your Stata question via a search engine (e.g., "how to tabulate a variable by values of another variable 'Stata'").

The Social Science Computing Cooperative (SSCC) also provides statistical consulting to students in this course. Information on contacting the SSCC, housed in Sewell, is here: <https://www.sccc.wisc.edu/get-help/>. You can, of course, always email me or stop by my office hours as well, and I will be thrilled to meet with you. If you have a Stata question, don’t beat your head against a brick wall for too long, but also please attempt to figure things out on your own before seeking help.

Credits: 3-credits. The credit standard for this course is met by an expectation of a total of 135 hours of student engagement with the course learning activities (at least 45 hours per credit), which include regularly scheduled instructor-student meeting times during class of 115 hours per week. The additional time is spent on required readings, problem sets, a final project, significant independent engagement with a data set of the students’ choice for the final project, and other student work as described in the syllabus.

Prerequisites: This course is open to graduate students and upper-level undergraduates. Sociology 360 or the equivalent, or Sociology 357 or the equivalent are required. Concurrent enrollment is permitted. Exceptions may be made with instructor consent.

Course Designation and Attributes: Honors optional

ASSIGNMENTS AND GRADING

Course Requirements

- **Attendance & Participation:** Lectures and in-class work will assume that you have read the assigned materials *before class*. Every semester that I've taught, there is a significant fraction of the class that realizes this only very belatedly, to both their detriment and mine. My style of lecture is to go deeper and unpack the material in an intuitive but rigorous way. Your job *before lecture* is to do basic fact-finding. That will be especially easy in this course since we don't have a lot of theoretically-challenging material, but I still want to stress that you *must read* and you *must read before class*. Attendance and participation will not factor directly into your grade, but you should assume that it will have a *large* indirect impact (via your ability to complete the homework assignments and the quality of your final project and exam). Lecture slides will be posted on the course web-page but may not contain all the important information covered in class. I will likely record my lectures, but if you anticipate any problems in attending class regularly, please talk to me: even though I do not require attendance, I like to try to accommodate students where possible and to know about challenges they are facing. I am happy to help you out to the extent that I can.
- **Reading Notes:** A small portion of your grade will be based on the submission of your reading notes for the first several weeks of the semester. Each chapter from the main textbook leads you through hands on exercises in Stata. You are required to submit the .do files that accompany your reading (your "reading notes") for the first few weeks. The grading is described in detail below. I encourage you to continue taking these notes throughout the whole class, but I will not require that you turn them in to me after the first few weeks. Reading notes (your do-files) are due prior to the start of each class (Mondays at 9:50 AM).
- **Weekly Exercises:** Exercises are due at 9:50 AM the day of class. I will post model answers and allocate some class time to discuss questions about the assignments.

Late work loses five percent per day of the total possible credit. Students will not get credit for assignment that are turned in that have substantial problems or are incomplete. Because I understand that sometimes unanticipated events occur, *your lowest homework score will be dropped*.

These assignments will count for 25% of your final grade so it is in your best interest to complete and submit these on time. I generally grade these on a five point scale: {0, 0.25, 0.5, .75, and 1}. Assignments that are turned in on time, are complete, and demonstrate thought and effort will be given full credit; I do not expect complete accuracy when I assign perfect scores, only substantial effort and accuracy. Putting effort into the homework and learning from the feedback you receive is the best way to do well on the exam and final project.

I encourage you to work together to complete the assignments, but you must turn in individual assignments. Everyone should put effort into answering all the questions, but discussion and collaboration with others is acceptable and, indeed, very useful.

Please submit all assignments via the Canvas “Assignments” page by 9:50 AM (right before class) on the date that they are due. If you have a mental or physical health emergency or ongoing condition, please reach out to me, but if you do not have a true extenuating circumstance, please accept the (fairly generous) five percent per day late penalty.

- Exams: There will be two open-book exams, during which you will be asked to conduct a series of data manipulations similar to those covered in the text and the weekly assignments. The second test will technically be non-cumulative, although much of the knowledge gained in the first half of the class will be necessary to carry out the tasks in the second half. The tests will count for 40 percent of your final grade (20 percent * 2).
- Final Project: The course will culminate in a final project that showcases the technical and analytic skills developed in class. You will write a research memo on a topic of your choosing. This project will involve choosing a publicly available data set (in consultation with the instructor), carefully describing those data, addressing missing data, conducting consistency checks, recoding variables, and preparing basic descriptive results. This project will count for 30 percent of your final grade in the course. I will provide detailed instructions for the project early in the semester. The same late penalty as above applies with the caveat that I must have your final project before grades are due.

Grading

Your final grade will be allocated as follows:

- Reading Notes: 5%
- Weekly Assignments: 25%
- Exam 1: 20%
- Exam 2: 20%
- Final Project: 30%

I assign grades based on the following cut points:

A	93-100
AB	87-93
B	82-87
BC	77-82
C	70-77
D	60-70
F	< 60

Late Work: Deadlines in this class are set entirely for your benefit. The course assignments are designed to build on each other over the course of the semester. Falling behind on these assignments will hurt your ability to do well in the course. Please do your best to stick to the deadlines I have assigned, as this will put you on the best path for success in the course.

That said, I understand that things happen. If you need an extension for an assignment, please communicate with me as soon as you become aware of your need for an extension; I'm happy to help where I can. My goal is to be as generous as possible, with an eye towards equity — with respect to you but also your fellow students — and also keeping you on track for "hard deadlines" (those imposed by the UW which I cannot modify beyond a certain point). The five percent default late penalty is meant to be generous and to account for "routine" lateness; if you foresee long-term difficulty completing the course given what's on your plate, contact me. More on this follows below.

COURSE POLICIES

Accommodations

We know students have many issues in their lives that may affect classroom performance, including hunger, homelessness, money problems, family responsibilities or crises, death of a family member or close friend, illness, or disability. If you are having difficulty securing food or housing or are facing financial, family, crime victimization, or life stresses, you can get support from the Dean of Students <https://doso.students.wisc.edu/student-assistance/> and are invited, at your option, to inform your professor of your circumstances. If you have life circumstances beyond your control, you have the right to request accommodation. Accommodation adjusts deadlines or the details of assignments to give you a chance to do your work and demonstrate your knowledge. Accommodation does not get you out of work nor does it give you a higher grade. You can receive accommodation for circumstances beyond your control, disabilities, and religious observances. Requests for accommodation for disability, religious observance, or chronic problems beyond your control should be requested during the first two weeks of class. Requests for unexpected or emergent problems should be requested as soon as possible. If you have a disability that affects your ability to meet these requirements make sure that you have worked with the McBurney Center (<http://www.mcburney.wisc.edu/>) to determine the best accommodations for you and that those accommodations have been communicated early in the semester.

Sexual harassment and misconduct. Professional conduct and appropriate behavior are critical to create a safe learning environment for students and instructors alike. Here is a statement about sexual harassment from the University:

What is Sexual Harassment?

Unwelcome sexual advances, requests for sexual favors, and verbal or physical conduct of a sexual nature constitute sexual harassment when:

- submission to such conduct is a condition of employment, academic progress, or participation in a university program; or
- submission to or rejection of such conduct influences employment, academic or university program decisions; or
- the conduct interferes with an employee's work or a student's academic career, or creates an intimidating, hostile or offensive work, learning, or program environment.

Tangible Action or Quid Pro Quo (This for That) Sexual Harassment and Hostile Environment Sexual Harassment are both illegal and unacceptable.

One practical implication is that instructors (faculty or graduate students) may not date students to whom they will be assigning grades. If you believe that you have been harassed, contact your instructor or the chair of the Department of Sociology (socchair@ssc.wisc.edu).

Academic honesty

By enrolling in this course, each student assumes the responsibilities of an active participant in UW-Madison's community of scholars in which everyone's academic work and behavior are held to the highest academic integrity standards. Academic misconduct compromises the integrity of the university. Cheating, fabrication, plagiarism, unauthorized collaboration, and helping others commit these acts are examples of academic misconduct, which can result in disciplinary action. This includes but is not limited to failure on the assignment/course, disciplinary probation, or suspension. Substantial or repeated cases of misconduct will be forwarded to the Office of Student Conduct & Community Standards for additional review. For more information, refer to studentconduct.wiscweb.wisc.edu/academic-integrity/. According to UWS 14, academic misconduct is defined as:

- Seeks to claim credit for the work or efforts of another without authorization or citation;
- uses unauthorized materials or fabricated data in any academic exercise;
- forges or falsifies academic documents or records;
- intentionally impedes or damages the academic work of others;
- engages in conduct aimed at making false representation of a student's academic performance;
- assists other students in any of these acts.

For a complete description of behaviors that violate the University's standards as well the disciplinary penalties and procedures, please see the Dean of Students website. If you have questions about the rules for any of the assignments or exams, please ask your instructor.

Departmental Notice of Grievance and Appeal Rights

The Department of Sociology regularly conducts student evaluations of all professors and teaching assistants near the end of the semester. Students who have more immediate concerns about this course should report them to the instructor or to the chair, 8128 Social Science (socchair@ssc.wisc.edu).

COURSE SCHEDULE

Week 1 (Jan 31): Introduction. Getting onto Winstat. The importance of good data management. Basics of data structure and do files. Organizing your files.

Read before class: None

Due: Nothing

Week 2 (Feb 7): Reading data into Stata.

Read before class: Mitchell, Chapters 1 and 2 (notes required)

Due: Reading notes (Mitchell, Chapters 1 and 2) (9:50 AM the day of lecture).

Assign: Exercise 1

Week 3 (Feb 14): Data cleaning

Read before class: Mitchell, Chapter 3 (notes required)

Due: Exercise 1, reading notes (Mitchell, Ch. 3)

Assign: Exercise 2

- Week 4 (Feb 21): Labeling, codebook, documentation, and fact-checking
Read before class: Mitchell, Chapter 4 (notes required); Long Chapter 2
Due: Exercise 2, Reading notes (Mitchell, Chapter 4)
Assign: Exercise 3
- Week 5 (Feb 28): Recoding, creating new variables
Read before class: Mitchell, Chapter 5
Due: Exercise 3, Reading notes (Mitchell, Chapter 5)
Project proposal (Feb 28, 11:59 PM)
Assign: Exercise 4
- Week 6 (Mar 7): Basic descriptive statistics (tabulation and summarization)
Read before class: Kohler and Kreuter Chapter 7
Due: Exercise 4
Assign: Exercise 5
- Week 7 (Mar 14): No class owing to Spring Break
- Week 8 (Mar 21): Missing data
Read before class: Allison Chapter 1
Due: Exercise 5
Assign: Exercise 6
- Week 9 (Mar 28): Exam I
- Week 10 (Apr 4): Combining data-sets
Read before class: Mitchell, Chapter 6
Due: Exercise 6
Assign: Exercise 7
- Week 11 (Apr 11): Processing observations across subgroups
Read before class: Mitchell, Chapter 7
Due: Exercise 7, Preliminary descriptive statistics for final project (11:59 PM)
Assign: Exercise 8
- Week 12 (Apr 18): Reshaping data
Read before class: Mitchell, Chapter 8
Due: Exercise 8
Assign: Exercise 9
- Week 13 (Apr 25): Programming for data management
Read before class: Mitchell, Chapter 9
Due: Exercise 9
Assign: Exercise 10

Week 14 (May 2): Paper preparation, review

Read before class: None

Due: Exercise 10

Monday, May 9: Final draft of term papers due to Canvas assignment folder at 11:59 PM.

Final exam: May 12, 2:45 pm. The UW has not currently published locations for these exams.