Overview

This course facilitates the development of data management skills and surveys strategies for efficient, accurate, and responsible data science practice. While the course explores these topics using Stata, many of the principles featured in this course are universal. Prior experience with Stata is helpful but is not required for the course.

Course Objectives

• Develop data management skills
• Improve overall programming fluency in Stata
• Survey public-use datasets
• Create a data science workflow that maximizes accuracy, efficiency, and replicability

Course Requirements and Grading Criteria

Discussion (20%)
In addition to faithful attendance, each student is expected to complete all assigned readings and contribute to class discussions.

Problem Sets (30%)
Three problem sets will reinforce key concepts and techniques covered in class and/or course materials. Students will submit a single .do file and relevant data files (if applicable) for each problem set. Problem sets that are well-commented, cleanly-written, and demonstrate skill mastery will receive full credit. All problem sets must be submitted electronically before 7:00pm on the corresponding date listed in the syllabus.

Data Presentation (15%)
Each student will prepare a very brief presentation about the policy topic and dataset he/she intends to use for completing the blog post assignment (see below). Presentations that concisely and creatively address all of the requirements listed in the assignment prompt (to be distributed and reviewed later) will receive full credit.

Data Blog (35%)
Students will conduct and communicate the results of a descriptive data analysis on a topic of their choice in the form of a blog post not to exceed 1,000 words in length. Students will submit all .do files and data files used in their analysis along with the text of their posts. Late posts will not be accepted. The blog assignment must be submitted electronically before 7:00pm on the corresponding date listed in the syllabus.
Required Materials

There are no required texts for this course. All materials will be made available online and/or placed on reserve at Alderman Library.

All students must have access to Stata 14 or 15 to successfully complete this course. All UVA students can access Stata virtually (for free!) using The UVA Hive. Alternatively, students may choose to purchase a limited-term Stata license from Stata's corporate website (Stata/IC recommended; $45 as of January 17, 2018).

Recommended Materials

The following materials are useful for data wrangling, particularly with respect to Stata:

- Christopher F. Baum (2016). *An Introduction to Stata Programming*. College Station, TX: Stata Press
- UCLA’s Institute for Digital Research and Education’s Stata repository
- UNC Carolina Population Center’s stata tutorial
- User-written resources on Stata’s corporate website
- The Stata Blog

Students should also follow high-quality data blogs in order to keep abreast of current data releases, data visualizations, and policy-relevant contributions to data journalism. The following are some great resources:

- The Upshot
- FiveThirtyEight
- Wonkblog
- StatChatVA

Other Policies

Computer Use
Computers are to be used *solely* for the purpose of taking notes and actively participating in class discussion. I reserve the right to ask you not to use your computer if I have reasons to suspect that you are using it to surf the web, check email or engage in other non-course-related activities.

Email Communications
I will respond to all student questions and comments via email and/or on Collab within 48 hours.
Accordingly, please plan to initiate any important correspondence with me well in advance of key deadlines for assignments, projects, etc.

**Frank Batten School of Leadership and Public Policy Honor Statement**

The Frank Batten School of Leadership and Public Policy embraces and upholds the University’s Honor Code principles that mandate that students will not lie, cheat, or steal, nor tolerate the actions of those who do. Acting in a manner consistent with the principles of Honor benefits every member of the community while enrolled in the Batten School and in the future.

We expect every student to comply fully with all provisions of the UVa Honor System. By enrolling in this course, you agree to abide by and uphold the Honor Code System of the University of Virginia As applied to your Batten course work and requirements, and unless otherwise specified by your instructors:

- All graded assignments must be pledged.
- Students may not access any notes, study outlines, problem sets, old exams, answer keys, or collaborate with other students without explicit permission.
- When given permission to collaborate with others, students will not copy answers from another student.
- Students should always cite any resources or individuals they have consulted to complete an assignment. If in doubt, sources should be cited.
- Suspected violations will be forwarded to the Honor Committee, and, at the discretion of the instructor, students may receive “no credit” the assignment in question, independent of the actions taken by the Honor Committee.
- Any questions about what is or is not permitted on an assignment, should be clarified by students with their instructors prior to the completion of their work.

If you believe you may have committed an Honor Offense, you may wish to file a Conscientious Retraction (“CR”) by calling the Honor Offices at (434) 924-7602. According to Honor guidelines, for your retraction to be considered valid, it must, among other things, be filed with the Honor Committee before you are aware that the act in question has come under suspicion by anyone. More information can be found at [www.virginia.edu/honor](http://www.virginia.edu/honor). If you have questions regarding the course honor policy, please contact your honor representatives.

**Frank Batten School of Leadership and Public Policy Wellbeing Statement**

If you are feeling overwhelmed, stressed, or isolated, there are many individuals here who are ready and wanting to help. Both Amanda Crombie, Director of Academic Programs and Jill Rockwell, Assistant Dean for Student Services are available to help all Batten Students. They are readily accessible during walk in hours or by setting up an appointment.

Alternatively, there are also other University of Virginia resources available. The Student Health Center offers Counseling and Psychological Services (CAPS) for its students.

Call 434-243-5150 (or 434-972-7004 for after hours and weekend crisis assistance) to get started and schedule an appointment. If you prefer to speak anonymously and confidentially over the
phone, call Madison House’s HELP Line at any hour of any day: 434-295-8255.

If you or someone you know is struggling with gender, sexual, or domestic violence, there are many community and University of Virginia resources available. The Office of the Dean of Students, Sexual Assault Resource Agency (SARA), Shelter for Help in Emergency (SHE), and UVA Women’s Center are ready and eager to help.

Adaptability
Please note that this syllabus is a working document. The course schedule and reading list are subject to change.

Course Schedule and Reading List

Wednesday, January 24: Intro to Workflows, Stata, and Data Diagnostics

• J. Scott Long (2009). The Workflow of Data Analysis Using Stata. College Station, TX: Stata Press, pages 1-8; 197-241

Wednesday, January 31: Macros and Loops; Wrangling I

• Long 2009, pages 241-279
• Christopher F. Baum (2016). An Introduction to Stata Programming. College Station, TX: Stata Press, pages 47-57, 57-66
• Problem Set 1 Due

Wednesday, February 7: Wrangling II & Dataset Examples

• Baum 2016, pages 104-121
• UCLA-IDRE Post on collapse command
• Problem Set 2 Due

Wednesday, February 14: Other Stata Hacks

• Baum 2016, pages 83-103
• Possible reading on writing custom Stata programs (if time and/or interest)
• Problem Set 3 Due

Wednesday, February 21: Student Presentations

Wednesday, February 28: Data Blog DUE

• NO CLASS MEETING (our last in-class session is 02/21)