

Gov 395L: MAKING BIG DATA

Connor T. Jerzak

Department of Government

University of Texas at Austin

Spring 2023



British museum object #92687 showing a 6th century BCE Babylonian map of the world. With better data gained via exploration, understanding of the earth would have been improved.

URL: britishmuseum.org/collection/object/W_1882-0714-509

Why You Should Take This Course

Data contributions are increasingly important in social science

Why?

NEW DATA \leadsto BROADER EMPIRICAL CLAIMS ON CURRENT THEORY

- * Broader by place (e.g., more countries)
- * Broader by time (e.g., more time slices)

NEW DATA \leadsto OPENS UP NEW AREAS OF INQUIRY

- * Improved unit resolution: Regional \leadsto city \leadsto neighborhood \leadsto individual level theories + data
- * New units of analysis: Analyze political dynamics in hitherto unexplored venues (e.g. bureaucratic, online, sports, etc.)

Problem Most methods courses focus on methods for analyzing existing data

Solution This course introduces students to the theory & practice of creating new data

The task of science is both to
extend our experience (this class)
and reduce it to order. (Stat 1-2, ML, etc.)

—Niels Bohr, *Atomic Theory and the Description of Nature, Introductory Survey*
(p. 1), Cambridge University Press. 1934.

What We Mean by Theory & Practice of Making Data

- PRACTICE**
- Learn tools for extracting information & converting it to data
 - * Web scraping, crawling, processing scrapped pages
 - Learn tools for processing massive databases
 - * Merging, tabulating via SQL large data
 - Learn how to make data collaboratively
 - * Crowd sourcing, online experiments, expert coding projects, etc.
 - Learn how to distribute code as data via **R** packages

- THEORY**
- Learn how causal inference changes when variables are measured imperfectly
 - Learn principles of good measurement
 - Learn how to articulate the strengths, weaknesses & new possibilities of a new data product

In the course, practice will precede theory to give you project tools ASAP

What Mentality We Will Have in the Class

A hacker mentality will help us hone the art of making big data:*

- “No matter what, I will innovate solutions to secure the target data.”
- “I will eagerly learn new tools to assist me in accomplishing my research goals.”
- “Fermi gave me the wonderful method of doing things quickly and easily. Any problem can be solved. Any problem can be solved by sitting down for 20 minutes and thinking about it starting from first principles.” – Freeman Dyson, [youtube.com/watch?v=G_hWfGOMh8A](https://www.youtube.com/watch?v=G_hWfGOMh8A)

* Term used with the meaning described in

Levy, S., 1984. *Hackers: Heroes of the computer revolution*. Garden City, NY: Anchor Press/Doubleday.

Course Details

Instructional Staff

Instructor

NAME: Connor T. Jerzak

EMAIL: connor.jerzak@austin.utexas.edu

URL: ConnorJerzak.com

OFFICE: BAT 3.108

VIRTUAL OFFICE: utexas.zoom.us/my/jerzak

OFFICE HOURS: Wednesdays 10:45 am–12 noon and by appointment

Course Details

- CLASSES (14)
- * *Part 1*: Lecture introducing tools + theory
 - * *Part 2*: Tutorial applying tools
 - * *Part 3*: Student (individ.) presentation on data product
- PRESENT (1)
- * Select social science paper whose major contribution is new data
 - * Ideally, relevant to both class topic of the day and your research (or one of the two)
 - * Discuss: what, why, how, strengths, limits
- PSETS (3-4)
- * Done in pairs; emphasis on learning/applying tools
 - * One question about your (individ./group) project
- PROJECT (1)
- * Ideally: add a data contribution to your dissertation
 - * Can be actual data product or software package
 - * To be released for replication purposes on Harvard Dataverse (unless prior permission)

Syllabus review: Latest version at
ConnorJerzak.com/teaching

Pre-semester Course Feedback

Pre-semester Course Feedback

- 1 **STUDENTS VOTE:** What topics should we consider in the “Students Vote” week?
- 2 **External speakers:** Would you prefer more external speakers or more tutorial time?
 - *Speaker pro:* Interesting, motivational, new perspectives
 - *Speaker con:* Less time to hone practical tools together
- 3 **PROBLEM SETS:** Should we craft to hone learning on particular skills or generate to help launch projects?
 - *Instructor-directed pro:* More collective help via collaboration; higher learning floor
 - *Project-directed pro:* Tailored to your projects; higher learning ceiling

Other Groundrules

- **LIMIT COMPUTER USE** during lecture component
- **EATING/ETC. OKAY**: Feel free to maximize your efficient use of time by having lunch/etc. in class
- **STANDING, ETC. OKAY**: Feel free to stand or move around (within reason!) during class
 - It's hard to focus sitting still for so long
 - Neural-imaging studies show increased frontal cortex activation during movement

Introducing some of the tools
we'll be using (brief tutorial)

Things to Set Up for Next Time

- 1 Windows users: Download terminal rdr-it.com/en/howto/windows-10-install-windows-terminal/
- 2 Download R and RStudio if you need to
- 3 Try out the first Code Academy HTML module: codecademy.com/learn/learn-html



For more information, ConnorJerzak.com/teaching