

Gov385L: Making Big Data

—FIND, CURATE, CREATE HIGH-QUALITY SOCIAL SCIENCE DATA—
Graduate Level, Spring 2026
Time/place TBD at UT Austin

Course Description

In the social sciences, quantitative analyses depend critically on the availability of high-quality data, but these data are rarely readily available at the click of a mouse. Instead, they are often messy, in difficult formats, or in forms not immediately usable in statistical software programs. In this class, we will learn how to find, curate, and create high-quality social science data. We will balance the teaching of practical tools—such as web scraping, crowd-sourcing techniques, online surveys, massive database management, and dataset merging—with the examination of theoretical perspectives on the role of measurement

error in the analysis of causality. Code will also be treated as a form of data, and some emphasis will be placed on how to make software packages that improve reproducibility and enhance research impact.

Projects completed during the course will give students an opportunity to create new, high-quality data contributing not only to their own research but also to the broader cumulation of knowledge in the social sciences. Geared toward graduate students with some coding background in R or Python; undergraduates with appropriate background are welcome (instructor permission needed).

Course Details

- **Instructor:** Connor T. Jerzak
 - **Level:** Graduate (Undergraduates welcome with permission)
 - **Term:** Spring 2026
 - **Schedule:** TBD
 - **Location:** TBD
 - **Syllabus:** ConnorJerzak.com/MakingBigDataSyllabus
 - **Feedback:** admonymous.co/cjerzak
-