

# EPPS6302 Methods of Data Collection and Production

Fall 2023 Tuesday: 4:00 pm - 6:45 pm GR3.402

**Instructor:** Dr. Karl Ho; 972-883-2017; GR3.203; [kyho@utdallas.edu](mailto:kyho@utdallas.edu)  
**Office Hours:** Wednesday, Thursday: 2 pm – 4 pm, by appointment.  
**Teaching Assistant:** Kelvin Chan; GR.3.314; [KelvinChunMan.Chan@utdallas.edu](mailto:KelvinChunMan.Chan@utdallas.edu)  
**Office Hours:**

**Overview:** This is a data method course introducing the theory, practices and issues of data collection and production. It aims at providing a comprehensive framework in understanding data, and how social scientists conduct research starting from the *data generation process*. We will cover data methods, data management, big data trends and how to prepare data for next phases of research including modeling and reporting. This survey course is designed to equip social scientists with data generation concepts, tools and best practices. Topics on new developments and tools of big data in social science research will also be covered.

## Learning Objectives:

At the completion of this course, students will be able to:

1. Understand Data theory
2. Be familiar with data methods by tasks and modes
3. Develop instruments and schemes for data collection

## Required Text:

Foster, Ian, Rayid Ghani, Ron S. Jarmin, Frauke Kreuter, and Julia Lane (editors). 2016. [Big Data And Social Science: A Practical Guide to Methods and Tools](#), Chapman and Hall/CRC Press. ‡  
Groves, R. M., et al. 2009. *Survey Methodology* Second Edition. Wiley-Interscience. ‡  
Guest, G., Namey, E.E. and Mitchell, M.L., 2012. *Collecting qualitative data: A field manual for applied research*. Sage. ‡  
Fowler, Jr., Floyd J. 2014. *Survey Research Methods*, 5th edition. Los Angeles, CA: Sage Publications. ISBN paperback 978-1-4522-5900-0. (‡)

(‡) Available electronically at McDermott Library

(†) Older version available electronically at McDermott Library

## Recommended Reading:

### Data theory:

Breiman, Leo. 2001. "Statistical modeling: The two cultures (with comments and a rejoinder by the author)." *Statistical science* 16, no. 3 199-231.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures & their Consequences*. Sage Publications.

### Survey:

Ansolabehere, Stephen and Douglas Rivers. 2013. "Cooperative Survey Research." *Annual Review of Political Science* 16: 307-329.

Biemer, Paul P. Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly*, 74(5):817–848, 2010.

Biemer, Paul P. *Latent Class Analysis of Survey Error*. John Wiley & Sons, 2011.

Biemer, Paul P. and Lars E. Lyberg. *Introduction to Survey Quality*. John Wiley & Sons, 2003.

Fowler, Jr., Floyd J. 2014. *Survey Research Methods*, 5th edition. Los Angeles, CA: Sage Publications. ISBN paperback 978-1-4522-5900-0.

Groves, Robert M. *Survey Errors and Survey Costs*. John Wiley & Sons, 2004.

Hansen, Morris H., William N. Hurwitz, and William G. Madow. *Sample Survey Methods and Theory*. John Wiley & Sons, 1993.

Saris, Willem E. & Irmchaud N. Gallhofer. 2014. *Design, Evaluation, and Analysis of Questionnaires for Survey Research, 2nd Edition*. Wiley. (‡)

Valliant, Richard, Jill A Dever, and Frauke Kreuter. *Practical Tools for Designing and Weighting Survey Samples*. Springer, 2013.

Weisberg, Herbert, *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago: University of Chicago Press, 2005

### ***Experiments***

Druckman, James N., and Donald P. Green, eds. 2021. *Advances in experimental political science*. Cambridge University Press. (‡)  
Hainmueller, Jens, Dominik Hangartner, and Teppei Yamamoto. 2015. "Validating Vignette and Conjoint Survey Experiments against Real-World Behavior." *Proceedings of the National Academy of Sciences* 112(8): 2395–2400.

### ***Record Linkage:***

Enamorado, Ted, Benjamin Fifield, and Kosuke Imai. (2019). "Using a Probabilistic Model to Assist Merging of Large-scale Administrative Records." *American Political Science Review*, Vol. 113, No. 2 (May), pp. 353-371.

### ***Scraping methods:***

Rudis, Bob. 2018. *21 Recipes for Mining Twitter Data with rtweet* (<https://rud.is/books/21-recipes/>) (*deprecated*)  
Saltz, J.S. and Stanton, J.M., 2017. *An introduction to data science*. SAGE Publications.  
Scheuren, Fritz and William E. Winkler. 1993. Regression analysis of data files that are computer matched. *Survey Methodology*, 19(1):39–58.

### **APSA Style Manual for Political Science Revised 2018 edition:**

<https://mk0apsaconnectbyy6p6.kinstacdn.com/wp-content/uploads/sites/43/2018/11/Style-Manual-for-Political-Science-2018.pdf>

### **Course GitHub:**

<https://github.com/datageneration/datamethods>

**Participation:** Full attendance of all classes is required and imperative. All class members have to actively participate in class preparation and discussion. Participation entails full preparation for class including research of class materials, completion of assignments and full involvement in the team project. Only medical emergency (with documentation) is considered excused absence. If you have a special medical condition, please contact me in first week and special arrangements can be made. Missing one class is equivalent to missing a chance of moving one letter grade up.

### **Assignments:**

Exercises will be assigned. Students are expected to finish all assignments and publish products to own website. apply techniques to own work. At beginning of each class, students will be selected to present and discuss assignment solutions.

### **Class Q&A:**

Ask class related questions on Piazza (<https://piazza.com/utdallas/fall2023/epps6302/home>). Frequently Asked Questions (Q&A) will be posted there. In other words, seek answers to class, assignment, coding questions on that site.

### **Grading and Requirements:**

#### *Team Project:*

Collaborative research is important part of this class. Each student must join a team composed of four members to work on a team project and final executive presentation. The final project must be between 25-30 pages in length. Originality is the first quality the instructor demands. Please read carefully the University's policy about cheating and plagiarism (see below for a brief version). Due dates of the proposal and progress reports are listed on the schedule below. No late submission will be accepted. Proposal of the project must be consulted with the instructor in advance and subject to the instructor's approval. It constitutes 20% of the final grade. The group project reports account for 30% of final grade. The final team project presentation is responsible for 30 percent of the final grade.

Each team must design and implement a data project using the ***systems and techniques covered in class***. Identify one original topic to investigate. Originality is the first quality the instructor demands. Original means the topic or data or method(s) must be originally designed and has never been published. Replication with new data and/or method does count but warrants strong justification (e.g. current project, dissertation research).

### **Accounts required for this class:**

Twitter Developer  
GitHub  
Qualtrics (UTD server)  
RStudio Cloud

Participation	10%
Assignment (posted on website)	10%
Project Proposal	20%
Final project presentation:	30%
Final project	30%

**Document guideline:**

All documents in this class must adhere to the following general guidelines:

- Submit electronic document to instructor's email in PDF format (NO Word or GoogleDoc):
  - With name and content on file name:
    - E.g. doe\_finalproject.pdf
  - On letter size paper
  - One inch on all margins
  - 12-point font
- NO submission using link to cloud or online drive (e.g. Google drive, box.com)

Following the APSA style and format is **mandatory**

**Schedule (Tentative):**

Week	Topic	Suggested Reading:	Due
1	<u>Introduction: Data Theory and Methods</u>		
2	<u>Survey Modes and Design</u>	Groves 1, 2; Fowler 1, 2	Team building
3	<u>Questionnaire Design</u>	Groves 7, 8, 9	
4	<u>Sampling, Survey Errors and Weighting</u>	Groves, Chapters 3, 4, 5, 6, 10; Fowler 3, 4, 5	
5	<u>Experiments</u>	Druckman, J.N., Green, D.P., Kuklinski, J.H. and Lupia, A., 2006. The growth and development of experimental research in political science. <i>American Political Science Review</i> , 100(4), pp.627-635; Hainmueller, Jens, Dominik Hangartner, and Teppei Yamamoto. 2015. "Validating Vignette and Conjoint Survey Experiments against Real-World Behavior." <i>Proceedings of the National Academy of Sciences</i> 112(8): 2395–2400.	
6	<u>Complex data</u>	Foster et al. Chapter 1, 4	
7	<u>Text data</u>	Foster et al. Chapter 7, S&S Chapter 14,15	
8	<u>Web data</u>	Foster et al. Chapter 2	Proposal & presentation 10/10
2	Social Media data	TBD	
10	<u>Spatial data:</u>	Ballas, Dimitris et al., 2018. GIS and the social sciences: theory and applications , New York: Routledge. (‡); Parker, R.N. and Asencio, E.K., 2009. GIS and spatial analysis for the social sciences: Coding, mapping, and modeling. Routledge.	
11	Record linkage	Foster et al. Chapter 3; Enamorado, Ted, Benjamin Fifield, and Kosuke Imai. (2019). "Using a Probabilistic Model to Assist Merging of Large-scale Administrative Records." <i>American Political Science Review</i> , Vol. 113, No. 2 (May), pp. 353-371.	
12	Event data	TBD	
13	<u>Qualitative Data</u>	Guest, G., Namey, E.E. and Mitchell, M.L., 2012. <i>Collecting qualitative data: A field manual for applied research</i> . Sage.	
14	Fall break		
15	Special topic		
16	Final Project Presentations		Final project 12/5



---

## **Comet Creed**

*This creed was voted on by the UT Dallas student body in 2014. It is a standard that Comets choose to live by and encourage others to do the same:*  
“As a Comet, I pledge honesty, integrity, and service in all that I do.”

---

## **Academic Support Resources**

The information contained in the following link lists the University’s academic support resources for all students. Please see <http://go.utdallas.edu/academic-support-resources>.

---

## **UT Dallas Syllabus Policies and Procedures**

The information contained in the following link constitutes the University’s policies and procedures segment of the course syllabus. Please review the catalog sections regarding the [credit/no credit](#) or [pass/fail](#) grading option and withdrawal from class.

Please go to <http://go.utdallas.edu/syllabus-policies> for these policies.

*The descriptions and timelines contained in this syllabus are subject to change at the discretion of the Professor.*