MACS 30000 - Perspectives on Computational Analysis

	Dr. Richard Evans	Dr. Benjamin Soltoff	Ms. Ging Cee Ng (TA)
Email	rwevans@uchicago.edu	soltoffbc@uchicago.edu	gingcee@uchicago.edu
Office	250 Saieh Hall	249 Saieh Hall	251 Saieh Hall
Office Hours	W 2:30-4:30pm	Th 2-4pm	TBD
GitHub	rickecon	bensoltoff	TBD

- **Meeting day/time**: MW 11:30-12:50pm, 247 Saieh Hall for Economics
- **Lab session**: T 5-5:50pm, location TBD

Course description

Massive digital traces of human behavior and ubiquitous computation have both extended and altered classical social science inquiry. This course surveys successful social science applications of computational approaches to the representation of complex data, information visualization, and model construction and estimation. We will reexamine the scientific method in the social sciences in context of both theory development and testing, exploring how computation and digital data enables new answers to classic investigations, the posing of novel questions, and new ethical challenges and opportunities. Students will review fundamental research designs such as observational studies and experiments, statistical summaries, visualization of data, and how computational opportunities can enhance them. The focus of the course is on exploring the wide range of contemporary approaches to computational social science, with practical programming assignments to train with these approaches.

Course objectives

Required textbooks

- Salganik, Matthew J. *Bit by Bit: Social Research in the Digital Age*, Princeton University Press, Open review edition.
- Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization.* John Wiley & Sons.

Evaluation

Assignment	Quantity	Points	Total Points
Short paper	4	15	60
Problem set	4	10	40
Final exam	1	20	20

Disability services

If you need any special accommodations, please provide us with a copy of your Accommodation Determination Letter (provided to you by the Student Disability Services office) as soon as possible so that you may discuss with me how your accommodations may be implemented in this course.

Course schedule (lite)

Date	Topic	Lab	Assignment
Mon, Sep. 26	Intro to Computational Social Science		
Tue, Sep. 27		TBD	
Wed, Sep. 28	Scientific method		
Mon, Oct. 3	No class (conference)		
Tue, Oct. 4		TBD	
Wed, Oct. 5	Ethics		
Mon, Oct. 10	Observational data		Short Paper 1 due
Tue, Oct. 11		TBD	
Wed, Oct. 12	Observational data		
Mon, Oct. 17	Observational data		
Tue, Oct. 18		TBD	
Wed, Oct. 19	Observational data		
Mon, Oct. 24	Collecting your own data		Short Paper 2 due
Tue, Oct. 25		TBD	
Wed, Oct. 26	Collecting your own data		
Mon, Oct. 31	Experiments		Short Paper 3 due
Tue, Nov. 1		TBD	
Wed, Nov. 2	Simulated data		
Mon, Nov. 7	Simulated data		Short Paper 4 due
Tue, Nov. 8		TBD	
Wed, Nov. 9	Data visualization and description		Problem Set 1 due
Mon, Nov. 14	Data visualization and description		
Tue, Nov. 15		TBD	
Wed, Nov. 16	Data visualization and description		
Mon, Nov. 21	Data visualization and description		Problem Set 2 due
Tue, Nov. 22		TBD	
Wed, Nov. 23	Data visualization and description		
Mon, Nov. 28	Collaboration		Problem Set 3 due
Tue, Nov. 29		TBD	
Wed, Nov. 30	Collaboration		
Fri, Dec. 2			Problem Set 4 due
Wed, Dec. 7	Final exam [10:30am-12:30pm]		

Course schedule (detailed)

1. Introduction to computational social science
    - Objectives
        -
    - Readings
        - Watts, D. J. (2007). A twenty-first century science. *Nature*, 445(7127), 489-489.
        - Lazer et. al. (2009) Computational Social Science. *Science*, 323, 721-723.
2. The scientific method
    - Objectives
        - Identify the major characteristics of the scientific method that make something "science"
        - Compare inductive and deductive theory-building
        - Assess how computational social science conforms to and differs from the traditional scientific method in both scope and process
        - Discuss the problematic aspects of quantitative research and best practices for conducting quality research

- Readings
  - Bhattacherjee, A. (2012). Social science research: principles, methods, and practices. Chapters 1-4. Skim/review as needed.
  - Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*.
  - Einav, L., & Levin, J. (2014). The Data Revolution and Economic Analysis. *Innovation Policy and the Economy*, 14(1), 1-24.
  - Schrodt, P. A. (2014). Seven deadly sins of contemporary quantitative political analysis. *Journal of Peace Research*, 51(2), 287-300.

3. No class (Big Questions, Big Data, and Big Computation (B^3): Frontiers of Computational Social Science conference)
4. Ethics
   - Objectives
     - Review historical examples of ethical violations in pre-computational era
     - Identify the need to practice ethics in research
     - Review the Belmont Report's principles for human subjects research
     - Assess the ethical dilemmas involved in examples of computational social science research
     - Identify guiding ethical principles researchers should follow
   - Readings
     - "Chapter 6: Ethics." *Bit by Bit*.
     - Facebook emotional contagion study
       * Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *PNAS*, 111(24), 8788-8790.
       * Editorial Expression of Concern: Experimental evidence of massive-scale emotional contagion through social networks. (2014) *PNAS*, 111(29), 10779.
       * Watts, D. J. (2014). Stop complaining about the Facebook study. It's a golden age for research. *The Guardian*.
       * Rosen, J. (2014). Facebook's controversial study is business as usual for tech companies but corrosive for universities. *The Washington Post*.
       * Vertesi, J. (2014). The Real Reason You Should Be Worried About That Facebook Experiment. *Time*.
     - Parry, M. (2011). Harvard Researchers Accused of Breaching Students' Privacy. *Chronicle of Higher Education*.
     - Zimmer, M. (2016). OkCupid Study Reveals the Perils of Big-Data Science. *Wired*.
     - UChicago Social & Behavioral Sciences Institutional Review Board
       * Skim site
       * Specifically read "Does My Research Need IRB Review?"
5. Observational data (counting)
   - Objectives
     - Define big data and how it differs from traditional data
     - Identify main characteristics of big data
     - Evaluate studies which use observational data to count things
     - Assess weaknesses in research designs using observational data
   - Readings
     - "Chapter 2: Observing Behavior." *Bit by Bit*. Sections 2.1-2.4.1.3.
     - King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review*, 107(02), 326-343.
     - Kossinets, G., & Watts, D. J. (2006). Empirical analysis of an evolving social network. *Science*, 311(5757), 88-90.
     - Edelman, B. G., & Luca, M. (2014). Digital discrimination: The case of airbnb.com. *Harvard Business School NOM Unit Working Paper*, (14-054).
     - Chetty, R., Hendren, N., Kline, P., Saez, E., & Turner, N. (2014). Is the United States still a land of opportunity? Recent trends in intergenerational mobility. *The American Economic*

Review, 104(5), 141-147.

6. Observational data (measuring)
   - Objectives
     - Assess the use of big data for measurement
     - Explain ideal point estimation
     - Evaluate the use of non-representative data to draw inferences about populations
     - Review an example of incorrect interpretation of big data and discuss how to avoid making these errors
   - Readings
     - Bonica, A. (2014). Mapping the ideological marketplace. *American Journal of Political Science*, 58(2), 367-386.
     - Wojcik, S. P., Hovasapian, A., Graham, J., Motyl, M., & Ditto, P. H. (2015). Conservatives report, but liberals display, greater happiness. *Science*, 347(6227), 1243-1246.
     - Emotional timeline of September 11, 2001
       * Back, M. D., Küfner, A. C., & Egloff, B. (2010). The emotional timeline of September 11, 2001. *Psychological Science*, 21(10), 1417-1419.
       * Pury, C. L. (2011). Automation can lead to confounds in text analysis Back, Küfner, and Egloff (2010) and the Not-So-Angry Americans. *Psychological Science*, 22(6), 835-836.
       * Back, M. D., Küfner, A. C., & Egloff, B. (2011). "Automatic or the people?" Anger on September 11, 2001, and lessons learned for the analysis of large digital data sets. *Psychological Science*, 22(6), 837-838.
7. Observational data (forecasting)
   - Objectives
     - Define forecasting
     - Compare and contrast the goals of explanation and forecasting
     - Identify appropriate benchmarks for evaluating prediction model accuracy
     - Examine the use of Google search queries to predict flu trends
     - Identify why Google Flu Trends (GFT) failed in the long-run
   - Readings
     - 2.4.2 Forecasting and nowcasting. *Bit by Bit*.
     - Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *PNAS*, 107(41), 17486-17490.
     - Google Flu Trends
       * Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012-1014.
       * Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google flu: traps in big data analysis. *Science*, 343(6176), 1203-1205.
     - Forecasting terrorism
       * Schrodt, P. A., Yonamine, J., & Bagozzi, B. E. (2013). Data-based computational approaches to forecasting political violence. In *Handbook of computational approaches to counterterrorism* (pp. 129-162). Springer New York.
       * Brandt, P. T., Freeman, J. R., & Schrodt, P. A. (2011). Real time, time series forecasting of inter-and intra-state political conflict. *Conflict Management and Peace Science*, 28(1), 41-64.
8. Observational data (approximating experiments)
   - Objectives
     - Identify the importance of randomness to experiments
     - Explain why approximating experiments is better than simple observational studies
     - Evaluate the importance of finding good comparison groups
     - Define quasi-experimental design strategies
       * Natural experiment
       * Matching
     - Regression discontinuity

    - Assess the value-added of approximate experiments in examples of computational social science
  - Readings
    - 2.4.3 Approximating experiments. *Bit by Bit.*.
    - Phan, T. Q., & Airoldi, E. M. (2015). A natural experiment of social network formation and dynamics. *PNAS*, 112(21), 6595-6600.
    - Hersh, E. D. (2013). Long-term effect of September 11 on the political behavior of victims' families and neighbors. *PNAS*, 110(52), 20959-20963.
    - Cohen, P., et al. (2016). Using Big Data to Estimate Consumer Surplus: The Case of Uber. Working paper.
9. Collecting data (fundamentals)
  - Objectives
    - Discuss why asking is still an essential research method in a digital age full of observing
    - Identify the major sources of survey error (selection and measurement)
    - Review the major eras of survey research (face-to-face, random digit dialing, internet)
    - Review probability sampling and survey weighting
    - Evaluate the benefits of non-probability sampling in survey design
  - Readings
    - "Chapter 3: Asking Questions." *Bit by Bit.* Sections 3.1-3.4.
    - Schuldt, J. P., Konrath, S. H., & Schwarz, N. (2011). "Global warming" or "climate change"? Whether the planet is warming depends on question wording. *Public Opinion Quarterly*, 75(1): 115-124.
    - Pew Research Center Survey Methodology
      * U.S. Survey Research: Our survey methodology in detail
      * U.S. Survey Research: Sampling
    - Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, 31(3), 980-991.
    - FiveThirtyEight: Is A 50-State Poll As Good As 50 State Polls?
10. Collecting data (digitally-enriched)
  - Objectives
    - Examine the usage of multilevel regression and poststratification (MRP) to estimate parameters of interest
    - Define amplified/enriched asking and assess its benefits and drawbacks
    - Explain wiki surveys and how they differ from traditional open/close-ended questions
    - Identify and assess the usefulness of new survey research techniques for the digital age (enriched asking, linking survey responses to digital traces, always-on surveys, wiki surveys, etc.)
  - Readings
    - "Chapter 3: Asking Questions." *Bit by Bit.* Sections 3.5-3.7.
    - Lax, J. R., & Phillips, J. H. (2009). How should we estimate public opinion in the states?. *American Journal of Political Science*, 53(1), 107-121.
    - Ansolabehere, S., & Hersh, E. (2012). Validation: What big data reveal about survey misreporting and the real electorate. *Political Analysis*, 20(4): 437-459.
    - Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264), 1073-1076.
    - Salganik, M. J., & Levy, K. E. (2015). Wiki surveys: Open and quantifiable social data collection. *PLoS One*, 10(5), e0123483.
    - Sugie, N. F. (2016). Utilizing Smartphones to Study Disadvantaged and Hard-to-Reach Groups. *Sociological Methods & Research*, 0049124115626176.
11. Experiments
  - Objectives
    - Define causal inference
    - Identify why experiments can be used to determine causal inference
    - Explain the difference between internal and external validity, and how massive experiments can be used to overcome external validity
    - Compare between- vs. within-subjects designs

- Assess benefits and drawbacks of experiments vs. observational data
    - Review different approaches to actually running a digital experiment
- Readings
    - "Chapter 4: Running experiments." *Bit by Bit*.
    - Adams, T. G., Stewart, P. A., & Blanchar, J. C. (2014). Disgust and the politics of sex: exposure to a disgusting odorant increases politically conservative views on sex and decreases support for gay marriage. *PLoS One*, 9(5), e95572.
    - Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415), 295-298.
    - King, G., Pan, J., & Roberts, M. E. (2014). Reverse-engineering censorship in China: Randomized experimentation and participant observation. *Science*, 345(6199), 1251722.
    - Edelman, B. G., Luca, M., & Svirsky, D. (2015). Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment. *Harvard Business School NOM Unit Working Paper*, (16-069).

12. Simulated data
    - Objectives
        - Define the data generating process
        - Define simulated data and where it fits in the scientific method
        - Define indirect inference
        - Under what circumstances do we use simulated data?
        - Show examples of simulated data.
        - Have students practice creating and using simulated data.
    - Readings
        - "Indirect Inference," New Palgrave Dictionary of Economics
        - Wolpin, Kenneth I., *The Limits of Inference without Theory*, MIT Press, 2013.
        - Benoit, Kenneth, "Simulation Methodologies for Political Scientists," *The Political Methodologist*, 10:1, pp. 12-16.
        - Davidson, Russell and James G. MacKinnon, "Section 9.6: The Method of Simulated Moments," *Econometric Theory and Methods*, Oxford University Press, 2004.
13. Simulated data (cont.)
    - Objectives
        -
    - Readings
        -
14. Data visualization and description
    - Objectives
        - Teach students ways to write efficient code to visualize, slice, and dice data
        - Basic statistics are simply the first step (mean, var, cov)
        - Start with the almighty histogram
        - Teach groupby mechanisms
        - Point students to theory in order to get correct descriptives
            - Foreshadow some machine learning techniques
        - Most results from good papers can be summarized by a descriptive statistic
            - The researcher couldn't know this ex ante
        - Dynamic descriptive statistics
        - 3-D description
        - Descriptive visualization
    - Readings
        - Scott, David W., Chapters 1-4, *Multivariate Density Estimation: Theory, Practice, and Visualization*, 2nd edition, John Wiley & Sons, 2015.
15. Data visualization and description (cont.)
    - Objectives
        -

- Readings
    -

16. Data visualization and description (cont.)
    - Objectives
        -
    - Readings
        -
17. Data visualization and description (cont.)
    - Objectives
        -
    - Readings
        -
18. Data visualization and description (cont.)
    - Objectives
        -
    - Readings
        -
19. Collaboration: distributed data collection and analysis
    - Objectives
        - Introduce students to distributed collaboration methods in:
            * data collection
            * data analysis
            * coding (e.g., Git, GitHub)
                · Why git and not Dropbox or Google Docs
            * Data visualizations
    - Readings
        - "Chapter 5: Collaborating." *Bit by Bit*.
        - I will create document for git motivation and workflow
        - Chacon, Scott and Ben Straub, *Pro Git: Everything You Need to Know about Git*, 2nd edition, Apress, 2014.
        - OSPC Data Visualizations project
20. Collaboration: distributed data collection and analysis (cont.)
    - Objectives
        -
    - Readings
        -