

## **eSoc 414 / Info 514: Computational Social Science**

Tuesdays and Thursdays, 12:30pm – 1:45pm

### **Professor Yotam Shmargad**

Email: [yotam@email.arizona.edu](mailto:yotam@email.arizona.edu)

Office Location: Harvill Building, Room 437B

Office Hours: Tuesdays and Thursdays, 2pm

Course Homepage: D2L

### **Course Description:**

This course will guide students through advanced applications of computational methods for social science research. Students will be encouraged to consider social problems from across sectors, including health science, environmental policy, education, and business. Particular attention will be given to the collection and analysis of data to study social networks, online communities, electronic commerce, and digital marketing. Students will consider the many research designs used in contemporary social research, including “Big” data, online surveys, and virtual experimental labs, and will think critically about claims of causality, mechanisms, and generalization.

### **Upon completion of this course, students should be able to:**

1. Comprehend the importance of sound social science research
2. Understand how to carry-out a research project, including statistical data analysis
3. Critically evaluate assumptions and methods used in published research in the social sciences
4. Understand the ethical considerations involved in computational social science research

### **A Few Words on Technology:**

1. You will have access to and will be required to retrieve all course materials from the course page in D2L. Please prepare now for this experience by familiarizing yourself with D2L, the web-based courseware supporting this course. Training for D2L can be found online at: <http://help.d2l.arizona.edu/students>.
2. We will be using R and Stata to analyze and visualize data in this course. If you do not have previous experience using R or Stata, or at least some programming experience and basic knowledge of statistics, then you will face a steep learning curve. This is not to say that you will find the course impossible, just that you should go in knowing that you will have to commit much more time to the course than what might usually be expected.
3. Finally, always have a back-up plan. If your ‘default’ or most preferred computing location fails, be prepared to find a computer to use when you need a backup machine. It would be to your benefit to familiarize yourself with services offered at the university libraries and the ILC on campus.

### **Readings:**

There are no books required for *purchase* in this class. However, we will be reading from two books, both of which are available for free online. The first is, “Bit By Bit: Social Research in the Digital Age,” by Matthew J. Salganik, a Sociology professor at Princeton University. You can find it here: <http://www.bitbybitbook.com/>. The second book will be reading in class is, “Mostly Harmless Econometrics: An Empiricist’s Companion,” by Joshua D. Angrist, an Economics professor at MIT, and Jörn-Steffen Pischke, an Economics professor at London School of Economics and Political Science. Find it here: [https://www.researchgate.net/profile/Joshua\\_Angrist/publication/51992844\\_Mostly\\_Harmless\\_Econometrics\\_An\\_Empiricist's\\_Companion/links/00b4953344a9a0cb13000000.pdf?origin=publication\\_detail](https://www.researchgate.net/profile/Joshua_Angrist/publication/51992844_Mostly_Harmless_Econometrics_An_Empiricist's_Companion/links/00b4953344a9a0cb13000000.pdf?origin=publication_detail). I will post additional readings for you on D2L. Some come from academic journals, and others may be from books or news articles. It is *crucial* that you read all of the assigned readings to do well in this class. Anyone who has not done the reading will simply not be able to participate. Needless to say, this will not do good things for your grade. Come to class well-prepared and you will be well-rewarded.

**Complete List of Assignments with Grade Breakdown and Due Dates:**

- Participation (10%)
- Paper Presentation (10%): Due on the day you present at 12:30pm (in class)
- Research Ideas (10%): Due February 21 at 12:30pm (in class)
- Assignments (5% each, 20% total):
  - Assignment 1: Due February 28 at 9:00am (D2L)
  - Assignment 2: Due March 7 at 9:00am (D2L)
  - Assignment 3: Due March 21 at 9:00am (D2L)
  - Assignment 4: Due March 28 at 9:00am (D2L)
- Research Proposal (10%): Due April 4 at 12:30pm (in class)
- Final Presentation (15%): Due April 25 at 9:00am (D2L)
- Final Paper (25%): Due May 2 at 9:00am (D2L)
  - 12-15 pages, double-spaced (for non-honors undergraduate students)
  - 15-20 pages, double-spaced (for graduate/honors students)
  - Includes: Introduction, Literature Review, Method, Results, Discussion

**Grade Distribution:**

90-100% = A “exemplary, far beyond reqs/expectations”

80-89% = B “exceeds requirements/expectations”

70-79% = C “meets requirements/expectations”

60-69% = D “falls short of requirements/expectations”

< 60% = E “repeat of course needed”

All written work will be evaluated for format, organization, style, grammar, and punctuation as well as content and argument. Written work turned in for this course is expected to be formatted in accordance with the American Psychological Association. I will post the instructions for course assignments on D2L well before the due dates.

**Attendance, Due Dates, and Missing Work:**

1. The UA’s policy concerning Class Attendance and Administrative Drops is available at:  
<http://catalog.arizona.edu/2015-16/policies/classatten.htm>
2. Attendance is “voluntary,” but there is no way you will pass the class without showing up for face-to-face meetings. If you have a lot going on in your personal or professional life that will conflict with your attendance, this may not be a good time to enroll in this course! Attendance is critical.
3. All holidays or special events observed by organized religions will be honored for those students who show affiliation with that particular religion.
4. Absences pre-approved by the UA Dean of Students (or Dean designee) will be honored.
5. Arriving late and leaving early is extremely disruptive to others in the class. Please avoid this kind of disruption.
6. Missed class assignments or exams cannot be made up without a well-documented, verifiable, excuse (for example, a physician’s medical excuse). The validity of such excuses will be assessed by the professor during a face-to-face meeting. Indeed, *due dates are firm*, and late work will be accepted only with a verifiable and valid excuse.

**Course Conduct and Campus Policies (be familiar with all campus policies):**

1. Food and technologies are issues in classrooms. Cellular telephones are distracting, so please put them away. Laptops can be utilized, but only for note-taking purposes. Please follow classroom rules regarding food and beverages in the classroom.
2. This is a safe environment for sharing and generating unique ideas. Please try to be “open” to diverse perspectives and learn from others who may pose views that differ from your own.

3. Students are encouraged to share intellectual views and discuss freely the principles and applications of course materials. However, graded work/exercises must be the product of independent effort unless otherwise instructed. Students are expected to adhere to the UA Code of Academic Integrity as described in the UA General Catalog. See: <http://deanofstudents.arizona.edu/academic-integrity/students/academic-integrity>.
4. It is the University's goal that learning experiences be as accessible as possible. If you anticipate or experience physical or academic barriers based on disability or pregnancy, please let me know immediately so that we can discuss options. You are also welcome to contact Disability Resources (520-621-3268) to establish reasonable accommodations. For additional information on Disability Resources and reasonable accommodations, please visit <http://drc.arizona.edu/>.
5. The Arizona Board of Regents' Student Code of Conduct, ABOR Policy 5-308, prohibits threats of physical harm to any member of the University community, including to one's self. See: <http://policy.arizona.edu/threatening-behavior-students>.
6. All student records will be managed and held confidentially. <http://www.registrar.arizona.edu/ferpa/default.htm>
7. Information contained in the course syllabus, other than the grade and absence policy, may be subject to change with advance notice, as deemed appropriate by the instructor.
8. UA Non-discrimination and Anti-harassment policy: <http://policy.arizona.edu/human-resources/nondiscrimination-and-anti-harassment-policy>.
9. Confidentiality of Student Records: <http://www.registrar.arizona.edu/ferpa/default.htm>.
10. Information contained in the course syllabus, other than the grade and absence policy, may be subject to change with advance notice, as deemed appropriate by the instructor.

#### **Requirements for the Course:**

To succeed in this course, 2-3 hours of study time per hour of formal class time (or per unit) are required. This means that in addition to our three hours of formal class meeting time, 6-9 hours a week of study time are needed in order to meet course expectations. These hours should be spent on reading texts, writing papers, researching for new information, or thinking about course content. College-level reading and writing abilities are assumed.

#### **Graduate/Honors Students' Requirements:**

Graduate students and undergraduate students earning credit with the University of Arizona Honors College will be held to the following enhancements:

1. Graduate and honors students will be required to read additional research articles not expected of non-honors undergraduate students.
2. Graduate and honors students will be expected to 'journal' about the readings each week. Each week, that is, students will be required to write a paragraph about each reading, and to submit their summaries to D2L.
3. For graduate and honors students, the final paper is required to be 15-20 pages, which is longer than the 12-15 page limit for non-honors undergraduate students. The additional pages should be used to write a deeper, more thorough review of the relevant literature, run additional and/or more sophisticated analyses, or both.

Students wishing to contract this course for Honors Credit should email me to set up an appointment to discuss the terms of the contract and to sign the Honors Course Contract Request Form. The form is available at <http://www.honors.arizona.edu/documents/students/ContractRequestForm.pdf>.

## **Section 1. Social Research: Past, Present, and Future**

In this first section, we will take a very close look at studies that combine ideas from traditional social science research with concepts, tools, and methods that have more recently been popularized due to the wide proliferation of computers and digital media. We will discuss the ethical considerations that arise when data about large numbers of people are readily available and as experimentation becomes the new normal for companies with a technological bent. We will discuss various social science methods, including observational data analysis, surveys, and experiments. By the end of the first section, you should start thinking about a few research ideas that you might explore for your final paper.

### **1. Thursday, January 12: Welcome and Course Overview**

#### Assignment for Next Class:

- Read: Lazer et al. 2009. "Computational Social Science." *Science*.
- Read: Blumenstock et al. 2015. "Predicting Poverty and Wealth from Mobile Phone Metadata." *Science*.
- Read: Provost and Fawcett 2013. "Data Science and its Relationship to Big Data and Data-Driven Decision Making." *Big Data*.
- Read: Preface and Chapter 1 of "Bit By Bit: Social Research in the Digital Age."

#### **Graduate/Honors Students**

- Read: Donoho 2015. "50 years of Data Science." *Tukey Centennial Workshop*.

### **2. Tuesday, January 17: Social Science vs. Data Science**

### **3. Thursday, January 19: High-Performance Computing**

#### Assignment for Next Class:

- Read: Kramer et al. 2014. "Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks." *Proceedings of the National Academy of Sciences*.
- Read: Soeller et al. 2016. "MapWatch: Detecting and Monitoring International Border Personalization on Online Maps." *Proceedings of the 25<sup>th</sup> International World Wide Web Conference*.
- Read: Chapter 6 of "Bit By Bit: Social Research in the Digital Age."

### **4. Tuesday, January 24: Ethics**

### **5. Thursday, January 26: Ethics**

#### Assignments for Next Class:

- Read: Lazer 2015. "Issues of Construct Validity and Reliability in Massive, Passive Data Collections." *The City Papers: An Essay Collection from The Decent City Initiative*.
- Read: Goel et al. 2010. "Predicting Consumer Behavior with Web Search." *Proceedings of the National Academy of Sciences*.
- Read: Chapter 2 of "Bit By Bit: Social Research in the Digital Age."

#### **Graduate/Honors Students**

- Read: Farber 2015. "Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers." *Quarterly Journal of Economics*.

### **6. Tuesday, January 31: Big Data**

### **7. Thursday, February 2: Prediction vs. Causality**

#### Assignments for Next Class:

- Burke and Kraut 2014. "Growing Closer on Facebook: Changes in Tie Strength Through Social Network Site Use." *Proceedings of the 32<sup>nd</sup> Annual ACM Conference on Human Factors in Computing Systems*.
- Jean et al 2016. "Combining Satellite Imagery and Machine Learning to Predict Poverty." *Science*.
- Read: Chapter 3 of "Bit By Bit: Social Research in the Digital Age."

#### **Graduate/Honors Students**

- Read: Salganik and Levy 2015. "Wiki Survey: Open and Quantifiable Social Data Collection." *PLoS ONE*.

### **8. Tuesday, February 7: Surveys**

9. Thursday, February 9: Linking Data

Assignment for Next Class:

- Klar and Shmargad (forthcoming). "The Effect of Network Structure on Preference Formation." *Journal of Politics*.
- Salganik et al 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science*.
- Read: Chapter 4 of "Bit By Bit: Social Research in the Digital Age."

10. Tuesday, February 14: Experiments

11. Thursday, February 16: Experiments

Assignment for Next Class:

- Complete: Research Ideas

12. Tuesday, February 21: Research Ideas

**Due in Class Today: Research Ideas (Bring to class)**

**Section 2. Linear Modeling and Causal Inference**

In this second section of the class, we will begin to look more closely at how observational data can be used to make causal statements about the world, including how instrumental variable and panel data methods can help give estimates from linear models a causal interpretation. In four assignments, we will pursue a research project to model how political campaign funding imbalances relate to electoral results. These assignments will go through how to collect and visualize data, build models and test hypotheses, explore moderation with interactions, and improve causal inference with fixed effects. By the end of the second section, you should begin to settle on a research question and dataset that you will analyze for your final paper.

13. Thursday, February 23: Assignment 1: Data Collection and Visualization

Assignment for Next Class:

- Complete: Assignment 1
- Read: Centola 2011. "An Experimental Study of Homophily in the Adoption of Health Behavior," *Science*.
- Read: Chapter 1-2 of "Mostly Harmless Econometrics: An Empiricist's Companion."

14. Tuesday, February 28: Linear Regression

**Due Today: Assignment 1 (D2L by 9:00am)**

15. Thursday, March 2: Assignment 2: Model Building and Hypothesis Testing

Assignment for Next Class:

- Complete: Assignment 2
- Read: Eagle et al 2010. "Network Diversity and Economic Development," *Science*.
- Read: Chapter 3 of "Mostly Harmless Econometrics: An Empiricist's Companion." **(stop at section 3.3)**

16. Tuesday, March 7: Instrumental Variables

**Due Today: Assignment 2 (D2L by 9:00am)**

17. Thursday, March 9: Assignment 3: Exploring Moderation with Interactions

Assignment for Next Class:

- Complete: Assignment 3
- Read: Tucker 2008. "Identifying Formal and Informal Influence in Technology Adoption with Network Externalities." *Management Science*.
- Read: Chapter 4 of "Mostly Harmless Econometrics: An Empiricist's Companion." **(stop at section 4.3)**

18. Thursday, March 14, 16: **No Class**

19. Tuesday, March 21: Panel Data

Due Today: Assignment 3 (D2L by 9:00am)

20. Thursday, March 23: Assignment 4: Improving Inference with Fixed Effects

Assignment for Next Class:

- Complete: Assignment 4
- Read: Shmargad 2014. "Social Media Broadcasts and the Maintenance of Diverse Networks," *Proceedings of the 35<sup>th</sup> International Conference on Information Systems*.
- Read: Shmargad 2016. "The Monetization of Information Broadcasts: A Natural Experiment on an Online Social Network," *AAAI Spring Symposium*.
- Read: Chapter 5 of "Mostly Harmless Econometrics: An Empiricist's Companion."

21. Tuesday, March 28: **No Class**

Due Today: Assignment 4 (D2L by 9:00am)

22. Thursday, March 30: How to Write a Paper

Assignment for Next Class

Complete: Research Proposal

### **Section 3. Final Projects**

We conclude the class with an opportunity to discuss your proposed research project with the class. I will host several research workshops, which are intended to help you with collecting, organizing, and analyzing your data. You will then share the findings of your research project and submit your final paper to me.

23. Tuesday, April 4: Research Proposal

Due in Class Today: Research Proposal (Bring to class)

24. Thursday, April 6: Research Workshop 1

25. Tuesday, April 11: **No Class**

26. Thursday, April 13: Research Workshop 2

27. Tuesday, April 18: Research Workshop 3

28. Thursday, April 20: **No Class**

29. Tuesday, April 25: Presentations 1

Due Today: Presentation (D2L by 9:00am)

30. Thursday, April 27: Presentations 2

31. Tuesday, May 2: Course Overview

Due Today: Final paper (D2L by 9:00am)