In this course, we will study how traditional methods used in social sciences can help us make sense of new data sources, and how these new data sources may require new approaches and research design. There will be a mix of lectures, student-led discussions and demos, and hands-on computational activities with various tools. The course covers substantive topics relevant to demographic research as well as a selection of data science tools to extract Internet data, manage large data sets, analyze them, and visualize them.

**Diversity of Student Backgrounds:** Students in this class have different backgrounds. Some students are pursuing a PhD or an MA, some others are undergraduates. Some students may have strong computational and statistical skills, some others may not. Some students may be familiar with population studies, some others not. To accommodate the range of backgrounds, I emphasize substance, and key statistical and computational concepts. There will also be different types of homework assignments. Some of them will involve computing and coding. Some others will be critical reflections about the readings. In short, I facilitate and encourage the participation of students who do not have extensive background in statistics, or computational methods, but are eager to learn.

**Goals:** In this course, we will discuss a number of substantive topics related to the emergence of (big) data-driven discovery in social sciences, with emphasis on population processes. By the end of the course, students will be familiar with relevant literature at the intersection of demographic research and computational social science. The main goals of the course are i) to develop critical thinking about the emergent field of big data analysis, ii) to learn some of the methods, approaches and tools of data science in the context of big data analysis, iii) to identify research questions in your own area of interest that could be addressed with innovative data sources and to devise an appropriate research plan.
Course Requirements and Grading

<table>
<thead>
<tr>
<th>Participation &amp; student-led discussion</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework assignments</td>
<td>50%</td>
</tr>
<tr>
<td>Term paper</td>
<td>30%</td>
</tr>
<tr>
<td>Final presentation</td>
<td>10%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
</tr>
</tbody>
</table>

**Class Participation & student-led discussions or demos:** Class participation will count towards your final grade. Please help create a constructive learning environment. Different people have different ways in which they participate best, all of which are valid: thoughtful preparation, sharing a well-formulated idea after a long pause, stimulating discussion through questions, helping a classmate understand a concept, discussing ideas and challenges during office hours, sharing news articles with the class, etc. I strongly encourage you to interact with me and the other students. Even if you feel uncertain about how to express something, I would rather have you speak up than say nothing at all. Listen to your peers, wait for your turn to speak, and refrain from using discriminatory language. If you are a talker, make sure that your quieter peers get a chance to speak. If you are shy, remember that if you have a question, most likely there is at least one other person with the same question who would be happy to listen to the answer.

During the quarter, you will be asked to choose a paper of your choice that is relevant for the general topic of the course, and share it with your classmates in advance of our meeting. In class, you would briefly present the main results, explain why you have chosen the paper and what is relevant about it, and lead a short discussion about the paper. Alternatively, you will have the opportunity to share a data science tool that is of interest to you in the form of a mini-demo or tutorial.

**Homework assignments:** There will be homework assignments almost every week. For some assignments, you will be expected to work on some technical problems related to statistical concepts or computational tasks discussed in class. These assignments would require coding, mostly in R. For other assignments, you may be expected to write a short commentary about assigned readings or topics. You may work in small groups (2-3 people) on the assignments, but each person of the group must submit a copy of the assignment and report the names of all the group participants.

**Term paper:** The term project is an empirical research brief on a relevant topic of your choice. You could replicate existing studies, test new ideas, develop a new visualization that is relevant for your own research, etc.. You could use existing data or collect your own. I encourage you to be adventurous. The style and sophistication of analysis depend on the student's background. In terms of format and length, you should follow the guidelines for submission to the *Descriptive Findings* series of *Demographic Research*:
http://www.demographic-research.org/info/general_information.htm. Early in the term, you should discuss your ideas with me, so that we can define a feasible plan.

**Final presentation:** During the last class meeting, students will present their term projects. This is your chance to practice your communication skills and to receive constructive feedback from your peers. You will be asked to submit the final papers a few days after your presentation. That way, you can incorporate the feedback that you received after your presentation.

**Class Conduct**

The class atmosphere will be quite relaxed. Just a few guidelines to make sure:

- Arriving a bit late is tolerated as long as you make an effort to minimize the disturbance for other students.

- Eating and drinking in class is allowed, but please make sure that you are not disturbing others.

- Please turn off your cellphone or put it on silent mode.

- If you cannot make it to class for whatever reason, make sure that you know what happened during the lecture and lab that you missed.

- If you are having trouble with the course material or personal problems that are hindering your performance in the class, please come and talk to me so that we can solve the problem before it is too late. It is better to bring up any concerns as early as they arise.

- Please always show respect to your fellow classmates.

**Students with Disabilities**

Please inform me as soon as possible of special needs that you may have. The sooner you notify me, the better I will be able to make appropriate arrangements.

**Academic Integrity**

A fundamental tenet of all educational institutions is academic honesty. Students must do all their work within the boundaries of acceptable academic norms. See the UW statement about student academic responsibility prepared by Committee on Academic Conduct in the College of Arts and Sciences (https://depts.washington.edu/grading/pdf/AcademicResponsibility.pdf). Students found guilty of plagiarism or academic dishonesty will be subject to appropriate disciplinary actions.
Course schedule, format and reading list

Each session will be a mix of lecture, discussion, and lab (hands-on computational activities). I will provide source code and material for the lab on a weekly basis. Most of the examples will be provided in R (and in some cases, in Python). Familiarity with R is useful, but not a pre-requisite.

Below is the course schedule and list of readings. The reading list may change. Additional readings, including news reports, demos and tutorials may be added during the course of the quarter, depending on students' interests and time availability.

Week 1 **Mon, Oct 3rd - Introduction: Challenges and opportunities for “Big Data” research**

Lab: Data manipulation with R: Baby names data


Week 2 **Mon, Oct 10th - Twitter data and difference-in-differences estimation**

Lab: Collect and analyze Twitter data


Week 3 Mon, Oct 17th – Guest lecture by Kivan Polimis: Introduction to Python and related tools for Web-scraping

Week 4 Mon, Oct 24th – Addressing selection bias in `digital breadcrumbs’
Lab: Estimating demographic attributes of social media users using the Face++ API


Week 5 Mon, Oct 31st - Mobile phones, demography and development
Lab: Managing large data sets and scalability


Week 6  **Mon, Nov 7th: Ethical issues, privacy and reproducible research**  
**Lab: Managing large data sets and scalability (continued)**


Week 7  **Mon, Nov 14th – Socinfo workshop on “Web data, Social Media Data and Demographic Research”, in Bellevue, WA**

Alternative activities for those who cannot attend the workshop.

Individual meetings with Emilio about term projects

Week 8  **Mon, Nov 21st: Web experiments**  
**Lab: Interactive Data Visualization with Rshiny**


Week 9  **Mon, Nov 28th: Demographic microsimulation and agent-based models**  
**Lab: Applications of demographic and stable population theory to online populations**


---

Week 10  **Mon, Dec 5th: Conclusions and students' presentations**