Data Science, Pop. Processes & Health  
(SOC 538)

Class Schedule

Lecture:  
TU 12:00 - 2:50 PM
Room:  
SAV 409
Zoom:  
[Link available via Canvas]
URL:  
https://canvas.uw.edu/

Professor

Name:  
Zack W. Almquist
Office:  
Savery 231
Office Hours:  
By appointment
Email:  
zalmquist@uw.edu

Course Description

This course is meant to provide a broad foundation in data science, population processes and health. By the end of the course students should have solid grounding in key issues and research topics in the field. The course is geared to provide students with enough background that they could pursue PhD level comps in this research area. Beyond basic coverage of seminal papers and current research areas, students will get exposure to modern research in the area through teaching one skills lecture and completing a quarter long project.

Assignments in the course include: weekly readings and discussion (students will be assigned as lead to summarize the paper, as lead to form questions around the article and to be the first response to the question set); students will lead one data science lab for other class members (topics provided); students will complete a quarter long project which uses big data or advanced analytic tools (a series of projects will be provided or students may suggest their own topic).
Learning Goals

- Have a broad understanding of the state of the art research in data science, populations processes and health.
- Provide the necessary background for pursuing a PhD level comp in data science, populations processes and health.
- Experience presenting a computational topic to other researchers.
- Completing a research analysis and write up that uses data science skills.

Prerequisites

- This is a graduate class and as such there are no formal prerequisites. Students are assumed to have or be willing to quickly acquire sufficient understanding of R to complete the major project and lead one tutorial over the quarter.

Course Structure

The course will be broken into two components over the two hours and 50 minutes of class. The first 80 minutes of class will be devoted to discussion of the reading. A randomly chosen student will be selected to summarize the reading and another will be assigned to be the first response to the reading and provide a question about the reading for the class to consider. We will then break for 10 minutes. The next 80 minutes will be a student led tutorial based on the provided material. Depending on class size, the tutorials will be individually led or group led.

Assignments

Reading and Response

Students will be randomly assigned to summarize, first question or first response for each article over each week. Before each seminar if you are assigned to be,

- Summarizer - Please prepare this before class and submit a written summary to Canvas.
- First Question - Write up question and post to the list (no later than SUNDAY at Midnight): Article Discussions
- First response - Look up questions (Article Discussions) and prepare a response for discussion on Tuesday

Each role will be graded 1/0.

Tutorial

Every student will need to lead (or co-lead) one R tutorial for the class. This involves preparing any needed materials and sending out details to the class by Monday evening. All students are
expected to bring a laptop and follow the tutorial as best as they are able to. The tutorial presentation will be graded on,

- Preparation
- Presentation
- Question and answers

All students need to sign up for a tutorial to present to the class in the first week of class.

- Sign up - Tutorial Sign Up For SOC 538

If there are more than 9 students, tutorials will be group led; if there are less than 9 students all remaining tutorials will be group led and graded on the same scale.

*Recommendation* - Sign up for the tutorial that aligns with your course project. Ideally, course projects will be built from one of the tutorial data sets.

**Project**

Projects should be built on one of the 9 tutorials used in the course. However, if you have a relevant project you are currently working on you can use that for your course project. Please communicate to me in the first week of the class if you intend to do this.

The project is composed of weekly assignments (see Course Schedule). Five points for each weekly assignment and 50 points for the final assignment. Total is 100 points. Please discuss with me if you will be late on any particular piece of the project. All project assignments are due on Sunday at Midnight. Final Project is Due Thursday at midnight of Finals Week.

**Grading**

Reading and Response: 40%

Tutorial: 40%

Project: 20%

**Letter grade assignment**

<table>
<thead>
<tr>
<th>% Points Earned</th>
<th>Number grade</th>
<th>Letter Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>100-97</td>
<td>4.0-3.9</td>
<td>A</td>
</tr>
<tr>
<td>96-90</td>
<td>3.8-3.5</td>
<td>A-</td>
</tr>
<tr>
<td>87-89</td>
<td>3.4-3.2</td>
<td>B+</td>
</tr>
<tr>
<td>86-84</td>
<td>3.1-2.9</td>
<td>B</td>
</tr>
<tr>
<td>83-80</td>
<td>2.8-2.5</td>
<td>B-</td>
</tr>
<tr>
<td>Range</td>
<td>Grade</td>
<td>Grade Description</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------------------</td>
</tr>
<tr>
<td>79-77</td>
<td>2.4-2.2</td>
<td>C+</td>
</tr>
<tr>
<td>76-74</td>
<td>2.1-1.9</td>
<td>C</td>
</tr>
<tr>
<td>73-70</td>
<td>1.8-1.5</td>
<td>C-</td>
</tr>
<tr>
<td>69-67</td>
<td>1.4-1.2</td>
<td>D+</td>
</tr>
<tr>
<td>66-64</td>
<td>1.1-0.9</td>
<td>D</td>
</tr>
<tr>
<td>63-60</td>
<td>0.8-0.7</td>
<td>D-</td>
</tr>
<tr>
<td>59-0</td>
<td>0</td>
<td>F</td>
</tr>
</tbody>
</table>

**Tutorials**

- [Self Study] Introduction to R - An Introduction to R
- [Self Study] CSSS Introduction to R - CSSS 508 | UW CSSS508
- [Week 2] Purple Air Sensor - Process and Display Data from Air Quality Sensors • AirSensor
- [Week 3] GSS - US General Social Survey (GSS) Data for R • gssr
- [Week 4] Twitter Workshop - Welcome! | Social Media Data for Population Research
- [Week 5-6] US Census Data - Load US Census Boundary and Attribute Data as tidyverse and sf-Ready Data Frames • tidyverse
d- [Week 7] IRS Migration Data - GitHub - mathewhauer/IRS-migration-data
- [Week 9] FB/ACS Migration Data and Model - GitHub - MJAlexander/fb-migration-bayes: Code and materials for Facebook migration modeling project

**Readings**

- **Week 1 - Introductions and overview**
  - Data Science and population processes
  - Computational social science and big data

  - Predictability of social science data

- Week 2 - Data Types: Administrative and mobile phone/sensor/app data
  - Administrative data
      - Chapters 1, 2, 3, 4 and 5
  - Mobile phone/sensor

- Week 3 - Data Types: Social media and survey data
  - Survey data
  - Social media


Social media and survey data


Week 4 - Data Types: Experiments and quasi-experiments using online data


Week 5 - Data Bias: Statistical, population and mitigation strategies


Week 6 - Data ethical issues, privacy and reproducible research


**Week 7 - Demography: Mortality, fertility & life course**

- **Fertility**

- **Mortality**

**Week 8 - Demography: Migration and mobility**

- **Big data for migration processes**

- **World migration data construction/estimation**

**Week 9 - Demography: Migration, displacement and response to disasters**

- **Credit card and IRS data**

  - **Twitter**

  - **Facebook**

- **Week 10 - Population forecasting and COVID-19**
  - **Population Forecasting**
  - **COVID-19**
## Course Schedule

*Note that the dates listed below are subject to change. Changes will be announced in class.*

<table>
<thead>
<tr>
<th>Week</th>
<th>Articles</th>
<th>R Tutorial</th>
<th>Weekly Project Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Week 1 - Introductions and overview</td>
<td></td>
<td>Pick project questions and data.</td>
</tr>
<tr>
<td></td>
<td>Discussion</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tutorial</td>
<td><a href="https://clanfear.github.io/CSSS508/">https://clanfear.github.io/CSSS508/</a></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Week 2 - Data Types: Administrative and mobile phone/sensor/app data</td>
<td></td>
<td>Produce data tables, descriptive statistics and visualization.</td>
</tr>
<tr>
<td></td>
<td>Discussion</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tutorial</td>
<td><a href="https://mazamascience.github.io/AirSensor/">https://mazamascience.github.io/AirSensor/</a></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Week 3 - Data Types: Social media and survey data</td>
<td></td>
<td>Write a formal description of data and your research question.</td>
</tr>
<tr>
<td></td>
<td>Discussion</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tutorial</td>
<td><a href="https://kjhealy.github.io/gssr/">https://kjhealy.github.io/gssr/</a></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Week 4 - Data Types: Experiments and quasi-experiments using online data</td>
<td></td>
<td>Pick an analysis method that will answer research questions.</td>
</tr>
<tr>
<td></td>
<td>Discussion</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tutorial</td>
<td><a href="https://mjalexander.github.io/social_media_workshop/">https://mjalexander.github.io/social_media_workshop/</a></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Week 5 - Data bias: Statistical, population and mitigation strategies</td>
<td></td>
<td>Draft analysis results.</td>
</tr>
<tr>
<td></td>
<td>Discussion</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tutorial</td>
<td><a href="https://walker-data.com/tidycensus/">https://walker-data.com/tidycensus/</a></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Week 6 - Data ethical issues, privacy and reproducible research</td>
<td></td>
<td>Finalize analysis results.</td>
</tr>
<tr>
<td></td>
<td>Discussion</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tutorial</td>
<td><a href="https://walker-data.com/tidycensus/">https://walker-data.com/tidycensus/</a></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Week 7 - Demography: Mortality, fertility &amp; life course</td>
<td></td>
<td>Write up methods section.</td>
</tr>
<tr>
<td></td>
<td>Discussion</td>
<td>Review</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Week 8 - Demography: Migration and mobility</td>
<td></td>
<td>Write up analysis</td>
</tr>
<tr>
<td></td>
<td>Discussion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 9 - Week 9 - Demography: Migration, displacement and response to disasters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Discussion</td>
<td></td>
<td>First draft of project paper.</td>
<td></td>
</tr>
<tr>
<td>Tutorial</td>
<td><a href="https://github.com/mathewhauer/IRS-migration-data">https://github.com/mathewhauer/IRS-migration-data</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Week 10 - Week 10 - Population forecasting and COVID-19 |
|---|---|---|
| Discussion |  | Work on final draft |
| Tutorial | https://github.com/MJAlexander/fb-migration-hurricane-maria |  |

| Week 10 |
|---|---|---|
| Finals |  | Submit final draft of project. Due Thursday Finals Week. |

**R Resources**

**Datacamp**

This class is supported by [DataCamp](https://www.datacamp.com), the most intuitive learning platform for data science and analytics. Learn any time, anywhere and become an expert in R, Python, SQL, and more. DataCamp's learn-by-doing methodology combines short expert videos and hands-on-the-keyboard exercises to help learners retain knowledge. DataCamp offers 325+ courses by expert instructors on topics such as importing data, data visualization, and machine learning. They're constantly expanding their curriculum to keep up with the latest technology trends and to provide the best learning experience for all skill levels. Join over 5 million learners around the world and close your skills gap.

**Vic’s Datacamp Recommendations**

- [Introduction to R](https://www.datacamp.com/learning/library/introduction-r)
- [Intermediate R](https://www.datacamp.com/learning/library/intermediate-r)
- [Introduction to the Tidyverse](https://www.datacamp.com/learning/library/introduction-tidyverse)
- [Reporting with R Markdown](https://www.datacamp.com/learning/library/reporting-r-markdown)
- Introduction to Writing Functions in R
- Introduction to Statistics in R (Chapters 1-3)
- Foundations of Probability in R (Chapters 1-2)

Chuck’s R Introduction to R for Social Scientists
- CSSS 508 | UW CSSS508

Chris Adolph’s Visualization Course
- Chris Adolph :: Visual