

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

% Points Earned	Number grade	Letter Grade
100-97	4.0-3.9	A
96-90	3.8-3.5	A-
87-89	3.4-3.2	B+
86-84	3.1-2.9	B
83-80	2.8-2.5	B-

79-77	2.4-2.2	C+
76-74	2.1-1.9	C
73-70	1.8-1.5	C-
69-67	1.4-1.2	D+
66-64	1.1-0.9	D
63-60	0.8-0.7	D-
59-0	0	F

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 • tidycensus](#)
- [Week 7] IRS Migration Data - [GitHub - mathewhauer/IRS-migration-data](#)
- [Week 8] FB Short term Migration Data - <https://github.com/MJAlexander/fb-migration-hurricane-maria>
- [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
 - Francesco C. Billari and Emilio Zagheni (2017). “Big Data and Population Processes: A Revolution?” In: Petrucci, Alessandra, and Rosanna Verde. SIS 2017. Statistics and Data Science: new challenges, new generations. Florence: Firenze University Press.
 - Lazer, D. M., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. Science 343(6176):1203-1205
 - Computational social science and big data
 - Ruths, D., Pfeffer, J. (2014). Social Media for Large Studies of Behavior. Science, 346(6213), 1063-1064.
 - Lazer, D. and Radford, J. (2017) Data ex Machina: Introduction to Big Data. Annual Review of Sociology (43):19-39.

- Lazer, D. M., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., ... & Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science*, 369(6507), 1060-1062.
 - Predictability of social science data
 - Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., ... & McLanahan, S. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences*, 117(15), 8398-8403.
 - **Week 2 - Data Types: Administrative and mobile phone/sensor/app data**
 - Administrative data
 - Ruggles, S. (2014). Big microdata for population research. *Demography*, 51(1), 287-297.
 - Walker, K. E. (2016). *tigris*: An R package to access and work with geographic data from the US Census Bureau.
 - Sullivan, T. A. (2020). *Census 2020: Understanding the issues*. Springer Nature.
 - Chapters 1, 2, 3, 4 and 5
 - Mobile phone/sensor
 - Blumenstock, J.E. (2012). Inferring Patterns of Internal Migration from Mobile Phone Call Records: Evidence from Rwanda. *Information Technology for Development*, 18(2):107-125.
 - Blumenstock, J.E., Cadamuro, G and On, R. (2015) Predicting Poverty and Wealth from Mobile Phone Metadata. *Science*, 350:1073- 1076.
 - Palmer, J.R.B., Espenshade, T.J., Bartumeus, F., Chung, C.Y., Ozgencil, N.E., and Li K. (2012). New Approaches to Human Mobility: Using Mobile Phones for Demographic Research. *Demography* (50):1105-1128.
 - Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F.R., Gaughan, A.E., Blondel, V.D. and Tatem, A.J. (2014) Dynamic Population Mapping Using Mobile Phone Data. *Proceedings of the National Academy of Sciences* 111(45):15888-15893.
 - Liang, Y., Sengupta, D., Campmier, M. J., Lunderberg, D. M., Apte, J. S., & Goldstein, A. H. (2021). Wildfire smoke impacts on indoor air quality assessed using crowdsourced data in California. *Proceedings of the National Academy of Sciences*, 118(36).
 - **Week 3 - Data Types: Social media and survey data**
 - Survey data
 - Japac, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., ... & Usher, A. (2015). Big data in survey research: AAPOR task force report. *Public Opinion Quarterly*, 79(4), 839-880.
 - Feehan, D. M., & Mahmud, A. S. (2021). Quantifying population contact patterns in the United States during the COVID-19 pandemic. *Nature communications*, 12(1), 1-9.
 - Social media

- Zagheni, E., Weber, I., & Gummadi, K. (2017). Leveraging Facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*, 721-734.
 - Cesare, N., Lee, H., McCormick, T., Spiro, E., & Zagheni, E. (2018). Promises and pitfalls of using digital traces for demographic research. *Demography*, 55(5), 1979-1999.
 - Social media and survey data
 - Salomon, J. A., Reinhart, A., Bilinski, A., Chua, E. J., La Motte-Kerr, W., Rönn, M. M., ... & Tibshirani, R. J. (2021). The US COVID-19 Trends and Impact Survey: Continuous real-time measurement of COVID-19 symptoms, risks, protective behaviors, testing, and vaccination. *Proceedings of the National Academy of Sciences*, 118(51).
 - Alekseev, G., Amer, S., Gopal, M., Kuchler, T., Schneider, J. W., Stroebel, J., & Wernerfelt, N. C. (2020). The Effects of COVID-19 on US small businesses: Evidence from owners, managers, and employees (No. w27833). National Bureau of Economic Research.
- **Week 4 - Data Types: Experiments and quasi-experiments using online data**
 - Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *science*, 311(5762), 854-856.
 - King, G., Pan, J. and Roberts, M.E. (2014) Reverse-engineering censorship in China: Randomized Experimentation and Participant Observation, *Science* 345(6199).
 - Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111(24):8788-8790.
- **Week 5 - Data Bias: Statistical, population and mitigation strategies**
 - Feehan, D. M., & Cobb, C. (2019). Using an online sample to estimate the size of an offline population. *Demography*, 56(6), 2377-2392.
 - Chatterjee, N., Chen, Y. H., Maas, P., & Carroll, R. J. (2016). Constrained maximum likelihood estimation for model calibration using summary-level information from external big data sources. *Journal of the American Statistical Association*, 111(513), 107-117.
 - Bradley, V. C., Kuriwaki, S., Isakov, M., Sejdinovic, D., Meng, X. L., & Flaxman, S. (2021). Unrepresentative big surveys significantly overestimated US vaccine uptake. *Nature*, 600(7890), 695-700.
- **Week 6 - Data ethical issues, privacy and reproducible research**
 - Zimmer, M. (2010). But the Data is Already Public: On the Ethics of Research in Facebook. *Ethics and information technology*, 12(4), 313-325.
 - Sweeney, L. (2000). Simple Demographics Often Identify People Uniquely. Carnegie Mellon University Data Privacy Working Paper 3. Pittsburgh 2000.

- Van Riper, D., Kugler, T., Schroeder, J., & Ruggles, S. (2020, November). Differential Privacy and Racial Residential Segregation. In 2020 APPAM Fall Research Conference. APPAM.
- **Week 7 - Demography: Mortality, fertility & life course**
 - Fertility
 - Cacioppo, J. T., Cacioppo, S., Gonzaga, G. C., Ogburn, E. L., & VanderWeele, T. J. (2013). Marital satisfaction and break-ups differ across on-line and off-line meeting venues. *Proceedings of the National Academy of Sciences*, 110(25), 10135-10140.
 - Bellou, A. (2015) The Impact of Internet Diffusion on Marriage Rates: Evidence from the Broadband Market. *Journal of Population Economics* 28, 265-297.
 - Mortality
 - Chae, D.H., Clouston, S. et al. (2015) Association between an Internet-based Measure of Area Racism and Black Mortality. *PloS ONE* 10(4).
 - Vos, T., Lim, S. S., Abbafati, C., Abbas, K. M., Abbasi, M., Abbasifard, M., ... & Bhutta, Z. A. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, 396(10258), 1204-1222.
- **Week 8 - Demography: Migration and mobility**
 - Big data for migration processes
 - Fiorio, L., Abel, G., Cai, J., Zagheni, E., Weber, I., & Vinué, G. (2017, June). Using Twitter data to estimate the relationship between short-term mobility and long-term migration. In *Proceedings of the 2017 ACM on web science conference* (pp. 103-110).
 - DeWaard, J., Johnson, J., & Whitaker, S. (2019). Internal migration in the United States: A comprehensive comparative assessment of the Consumer Credit Panel. *Demographic research*, 41, 953.
 - Hauer, M., & Byars, J. (2019). IRS county-to-county migration data, 1990–2010. *Demographic Research*, 40, 1153-1166.
 - World migration data construction/estimation
 - Azose, J. J., & Raftery, A. E. (2019). Estimation of emigration, return migration, and transit migration between all pairs of countries. *Proceedings of the National Academy of Sciences*, 116(1), 116-122.
 - Abel GJ, Sander N (2014) Quantifying global international migration flows. *Science* 343:1520–1522.
- **Week 9 - Demography: Migration, displacement and response to disasters**
 - Credit card and IRS data
 - DeWaard, J., Johnson, J. E., & Whitaker, S. D. (2020). Out-migration from and return migration to Puerto Rico after Hurricane Maria: Evidence from the consumer credit panel. *Population and Environment*, 42(1), 28-42.

- Fussell, E., Curtis, K. J., & DeWaard, J. (2014). Recovery migration to the City of New Orleans after Hurricane Katrina: A migration systems approach. *Population and environment*, 35(3), 305-322.
 - Curtis, K. J., Fussell, E., & DeWaard, J. (2015). Recovery migration after Hurricanes Katrina and Rita: Spatial concentration and intensification in the migration system. *Demography*, 52(4), 1269-1293.
 - Twitter
 - Martín, Y., Cutter, S. L., Li, Z., Emrich, C. T., & Mitchell, J. T. (2020). Using geotagged tweets to track population movements to and from Puerto Rico after Hurricane Maria. *Population and Environment*, 42(1), 4-27.
 - Martín, Y., Cutter, S. L., & Li, Z. (2020). Bridging twitter and survey data for evacuation assessment of Hurricane Matthew and Hurricane Irma. *Natural hazards review*, 21(2), 04020003.
 - Facebook
 - Maas, P., Iyer, S., Gros, A., Park, W., McGorman, L., Nayak, C., & Dow, P. A. (2019, May). Facebook Disaster Maps: Aggregate Insights for Crisis Response & Recovery. In *KDD (Vol. 19, p. 3173)*.
 - Alexander, M., Polimis, K., & Zagheni, E. (2019). The impact of Hurricane Maria on out-migration from Puerto Rico: Evidence from Facebook data. *Population and Development Review*, 617-630.
- **Week 10 - Population forecasting and COVID-19**
 - Population Forecasting
 - Raftery, A. E., Li, N., Ševčíková, H., Gerland, P., & Heilig, G. K. (2012). Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences*, 109(35), 13915-13921.
 - Azose, J. J., Ševčíková, H., & Raftery, A. E. (2016). Probabilistic population projections with migration uncertainty. *Proceedings of the National Academy of Sciences*, 113(23), 6460-6465.
 - COVID-19
 - Jones, J. H., Hazel, A., & Almquist, Z. (2020). Transmission dynamics models for the SARS Coronavirus 2. *American Journal of Human Biology*, 32(5).
 - Irons, N. J., & Raftery, A. E. (2021). Estimating SARS-CoV-2 infections from deaths, confirmed cases, tests, and random surveys. *Proceedings of the National Academy of Sciences*, 118(31).
 - Thomas, L. J., Huang, P., Yin, F., Luo, X. I., Almquist, Z. W., Hipp, J. R., & Butts, C. T. (2020). Spatial heterogeneity can lead to substantial local variations in COVID-19 timing and severity. *Proceedings of the National Academy of Sciences*, 117(39), 24180-24187.
 - Thomas, L. J., Huang, P., Yin, F., Xu, J., Almquist, Z. W., Hipp, J. R., & Butts, C. T. (2022). Geographical patterns of social cohesion drive disparities in early COVID infection hazard. *Proceedings of the National Academy of Sciences*, 119(12).

Course Schedule

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

	Articles	R Tutorial	Weekly Project Activity
Week 1 - Introductions and overview			
Discussion			Pick project questions and data.
Tutorial		https://clanfear.github.io/CSS508/	
Week 2 - Data Types: Administrative and mobile phone/sensor/app data			
Discussion			Produce data tables, descriptive statistics and visualization.
Tutorial		https://mazamascience.github.io/AirSensor/	
Week 3 - Data Types: Social media and survey data			
Discussion			Write a formal description of data and your research question.
Tutorial		https://kjhealy.github.io/gssr/	
Week 4 - Data Types: Experiments and quasi-experiments using online data			
Discussion			Pick an analysis method that will answer research questions.
Tutorial		https://mjalexander.github.io/social_media_workshop/	
Week 5 - Data bias: Statistical, population and mitigation strategies			
Discussion			Draft analysis results.
Tutorial		https://walker-data.com/tidycensus/	
Week 6 - Data ethical issues, privacy and reproducible research			
Discussion			Finalize analysis results.
Tutorial		https://walker-data.com/tidycensus/	
Week 7 - Demography: Mortality, fertility & life course			
Discussion	Review		Write up methods section.
Tutorial		https://github.com/MJAlexander/states-mortality	
Week 8 - Demography: Migration and mobility			
Discussion			Write up analysis

Tutorial		https://github.com/mathewhauer/IRS-migration-data	results.
Week 9 - Week 9 - Demography: Migration, displacement and response to disasters			
Discussion			First draft of project paper.
Tutorial		https://github.com/MJAlexander/fb-migration-hurricane-maria	
Week 10 - Week 10 - Population forecasting and COVID-19			
Discussion			Work on final draft
Tutorial		https://github.com/MJAlexander/fb-migration-bayes	
Week 10			
Finals			Submit final draft of project. Due Thursday Finals Week.

R Resources

Datacamp

This class is supported by [DataCamp](#), 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

- [Programming - Part 1 \(Writing code in RStudio\)](#)
- [Introduction to R](#)
- [Intermediate R](#)
- [Introduction to the Tidyverse](#)
- [Reporting with R Markdown](#)
- [Managing - Part 1 \(Projects in RStudio\)](#)

- [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](#)