SOC 388: Social Data Science

Draft Syllabus

Professor Jeremiah Bohr
Department of Sociology
University of Wisconsin Oshkosh

UWO Bulletin Description

This course introduces students to techniques for collecting large amounts of digital data ("Big Data") for social analysis. The course will primarily focus on analyzing text-based data (such as social media) and data visualization, and students will become proficient in the R programming language. The course will provide practical, hands-on experience and requires no prior computer programming knowledge. Prerequisites: SOC 101, or instructor's consent. Recommended: SOC 281 or similar statistics course.

Additional Course Description

Our personal and professional lives increasingly exist online, which means that studying human behavior requires understanding how to access and analyze the data we leave behind in our digital lives. In this class, I will introduce you to the new field of computational social science—an interdisciplinary intersection of the social sciences and computer science. While you will not exit this class an expert in data science or computer programming, you will learn basic principles and foundational skills that you may expand on through other coursework, and/or you will have the confidence to begin teaching yourself more advanced skills by engaging the online R community.

Learning Goals

- Become familiar with key ethical issues and articulate positions regarding ethical use of digital data in analysis and reporting
- Become proficient in the basic syntax of the R programming language
- Develop analytical skills and quantitative literacy
- Gain experience collecting digital data (e.g., social media)
- Gain experience visualizing data for public presentation

Required Texts

Matt Salganik. 2018. *Bit by Bit: Social Research in the Digital Age*. Princeton University Press.

Other readings will be made available on the course website.

Boot Camps

We will go through a series of "boot camp" tutorials that will teach you technical skills in the R language. Boot camps are tiered such that each tutorial builds on skills that you learned in the previous one, and conclude with homework assignments that let you test out what you have learned. The boot camps include:

- Boot Camp 1: Getting to know base R (data types, variables, vectors, data frames, tables)
- Boot Camp 2: Mathematical operations and descriptive statistics
- Boot Camp 3: Libraries and package documentation
- Boot Camp 4: Visualizing data in the tidyverse (ggplot2)
- Boot Camp 5: Introduction to text analysis

Individual Projects

In addition to the boot camps, there are two projects that each student will individually complete (although you will begin one of them in a group). For both projects, I will help guide you through your data collection that you will use for analysis.

Group Project

Students will form groups to complete the final assignment in this class. The main goal of this assignment is to collect, analyze, and visualize data for public consumption. The group will submit a final report that includes a written analysis integrated with visualization, and any code and data necessary to replicate the analysis. Additionally, each group must formally present their findings to the class during the final week of the semester.

Grading

Boot Camps	20%
Individual Projects (2)	40%
Group Project	20%
Group Presentation	10%
Attendance/Participation	10%

Late work will be subject to penalties.

Academic misconduct will not be tolerated. You must do your own work and properly cite ideas that belong to other people. More information on plagiarism can be found here: https://wts.indiana.edu/writing-guides/plagiarism.html.

Campus Resources
All UWO students are eligible for one-to-one conferencing at the Writing Center. Their free, confidential tutoring is designed to help students work through assignments and gain additional

writing skills. The Writing Center is located in the academic support suite of the Student Success Center and here: http://www.uwosh.edu/wcenter.

The Center for Academic Resources (CAR) provides free, confidential tutoring for students in most classes on campus. CAR is located in the Student Success Center, Suite 102. Check the Content Tutoring page on CAR's website (http://car.uwosh.edu/) for a list of tutors. If your course is not listed, click on a link to request one, or call 424-2290. To schedule a tutoring session, simply email the tutor, let him/her know what class you are seeking assistance in, and schedule a time to meet.

UWO is committed to providing reasonable accommodation for students with disabilities. Please contact the Disability Services [Dean of Students Office, 125 Dempsey Hall, 424-3100 (Voice), 424-1319 (TTY)] for the University's accommodation request form and documentation requirements or visit their website at: https://www.uwosh.edu/deanofstudents/Accessibility-Center.

Please visit this page to read about all the campus services available to support your success: https://uwosh.edu/resources/).

Your final grade will be computed on the following percentage scale:

A	92.5-100%
A-	90.0-92.49
B+	87.5-89.99
B	82.5-87.49
B-	80.0-82.49
C+	77.5-79.99
C	72.5-77.49
C-	70.0-72.49
D+	67.5-69.99
D	62.5-67.49
D-	60.00-62.49
F	0-59.99

Grade Interpretation

• A – Indicates truly superior work and active participation in the learning environment. Students at this level demonstrate an ability to think critically in oral and written form, integrate concepts and theories, level of knowledge relevant to the course that is beyond expectations and far above average, and the ability to use this knowledge in unfamiliar situations.

• B – Demonstrates a level of knowledge relevant to the course that is beyond expectations and above average. Will be able to use and extend this knowledge in some situations.

• C – Demonstrates an acceptable and average level of knowledge relevant to the course in terms of familiarity with readings, integration of concepts and theories, level of critical thinking, and active participation. Students at this level should be able to continue learning in this field of study.

• D – Demonstrates less than satisfactory performance in the course requirements and barely adequate level of knowledge relevant to the course. Unlikely to be able to apply knowledge, nor successfully continue studies in this direction.

• F – Demonstrates insufficient knowledge to be given credit for this course, and/or dismal performance and no class participation.

SCHEDULE

Week 1 INTRODUCTION

Day 1 Course Overview

Day 2 Introduction to RStudio

- Guide students through installation
- Explain basics of independent development environment (IDE)
- *Assignment:* Baseline skills survey

Week 2 RESEARCH DESIGN IN THE DIGITAL AGE

Day 1 Approaches to Digital Data

- *Reading:* Salganik, 1-11
- *Activity*: Group presentation of hypothetical "readymade" and "custommade" data projects

Day 2 Basic Skills

- *Discuss:* Baseline skills survey
- *Assign and begin Boot Camp #1*

Week 3 ETHICS IN DATA SCIENCE

Day 1 Principles of Ethical Research

- *Reading*: Salganik, 281-301
- *Due*: Boot Camp #1
- *Discuss:* Boot Camp #1 experience

Day 2 Ethical Frameworks

- *Reading*: Salganik, 301-325

- *Activity*: Group presentation on ethical research dilemma
    - Identify problem from real-world research example
    - Propose an ethical course of action
    - Identify and explain this course of action within an ethical framework
- *Assign and begin Boot Camp #2*

Week 4 BIG DATA

Day 1 Observing Behavior with Digital Trace Data

- *Discuss:* Boot Camp #2 experience
- *Reading:* Salganik, 13-62

Day 2 Observing Behavior with Digital Trace Data (continued)

- *Assign and begin Boot Camp #3*
- *Optional:* Perdue et al, "Can Big Data Predict the Rise of Novel Drug Abuse?"
- *Optional:* Lazer et al, "The Parable of Google Flu: Traps in Big Data Analysis."

Week 5 SUPERVISED LEARNING IN TEXT ANALYSIS

Day 1 Supervised vs. Unsupervised Learning

- *Discuss:* Boot Camp #3 experience
- *Assign and begin Boot Camp #4*

Day 2 Dictionaries in Text Analysis

- *Discuss:* Boot Camp #4 experience
- *Assign and begin Boot Camp #5*
- *Reading:* Flores, "Do Anti-Immigrant Laws Shape Public Sentiment? A Study of Arizona's SB 1070 Using Twitter Data"

Week 6 READYMADE DATA COLLECTION

Day 1 Individual Project #1 Overview ("readymade" text analysis)

- *Reading:* De Vaus, "What is Research Design?"
- *Discuss:* Boot Camp #5 experience
- Access Twitter API

Day 2 "Datathon" #1

Week 7 READYMADE & CUSTOMMADE DATA COLLECTION

Day 1 Accessing Census Data

Day 2 Asking Questions
- *DUE:* Individual Project #1
- Share results in class through informal presentations
- *Reading*: Salganik, 85-99

Week 8 CUSTOMMADE DATA COLLECTION

Day 1 Individual Project #2 Overview ("custommade" text analysis)
- Form data collection groups
- *Reading*: Salganik, 99-108

Day 2 Wiki surveys demonstration
- *Reading*: Salganik, 111-115
- *Goal*: be prepared to launch group survey by end of class

Week 9 CUSTOMMADE DATA COLLECTION (continued)

Day 1 *Provide updates on data collection*

Day 2 "Datathon" #2

Week 10 UNSUPERVISED LEARNING

Day 1 Topic Models
- *DUE:* Individual Project #2

Day 2 Topic Models (continued)
- Group practice with corpus provided to you

Week 11 REVIEW

Day 1 Review text analysis techniques
- Methods & Topics "speed dating" exercise

Day 2 Group Project Assignment overview
- Form groups based on shared interests

Week 12 GROUP WORK

Day 1 *Work in-class on final project*

Day 2 *Work in-class on final project*

Week 13 GROUP WORK

Day 1 *Work in-class on final project*

Day 2 *Work in-class on final project*

Week 14 GROUP PROJECT PRESENTATION

Day 1 Group presentations

- *DUE:* final group project

Day 2 Group presentations