Syllabus

SMPA 2152: Data Analysis for Journalism and Political Communication (Spring 2023)

Meeting Times:	Mondays, 7:10-9:40pm
Classroom:	MPA B01
Professor:	Nicholas Bell, Ph.D. (he/him/his) nicholasbell@gwu.edu
Office Hours (on Zoom):	Wednesdays 7:30-8:30pm and by appointment (Zoom link for office hours: https://gwu-edu.zoom.us/j/91601292476) I prefer to meet during office hours or by appointment. However, I am available by email, and I try to respond to emails by the end of the next business day (M-F).

Course Description

Data has been democratized. More data is available to the ordinary person than ever before, and leaders in every industry – including journalism and political communication – are seeking to take advantage of data to advance their missions. However, most of us lack the data literacy skills to make good use of these resources, and this can lead to the misapplication and misuse of data. To fully leverage the promise of big data, we must become familiar with the basic challenges inherent in data analysis and how to overcome them. This course is an introduction to the principles and practices of data analysis. The goal is for students to become responsible consumers and producers of data. Students will learn how to critically evaluate claims derived from data. Students will also learn how to ethically present data in compelling and persuasive ways to non-expert audiences. This class includes a special discussion of political polling, which is widely used in journalism and political communication but has come under increasing scrutiny in recent years. Students require only a basic aptitude in numeracy (e.g. percentages and averages) for this course.

In addition to developing data literacy, students will be introduced to the R programming language. There are many advantages to learning R: it is free and open-source, meaning that developers are continually releasing new tools to make coding easier; it is widely used by news organizations and researchers around the world; and R is one of the most powerful programming languages for statistical analysis. Students will

learn data literacy by applying the same tools and techniques used by professional data scientists.

Learning Objectives

1. You will be able to assess the pragmatic and ethical issues in collecting, manipulating, and analyzing data, known as "data literacy."
2. You will be able to obtain publicly-available data and perform basic manipulations on that data using the programming language R.
3. You will be able to visualize and present data in accurate and persuasive ways.
4. You will be familiar with the statistical concepts of sampling, uncertainty, hypothesis testing, and linear regression and how to conduct basic statistical analyses in R.
5. You will be able to develop, produce, and share an original data storytelling project.

Problem Sets

Each week, I will assign a problem set for you to complete in R. You may complete these problem sets on your own or in collaboration with other students. If you work with other students, please indicate their names at the top of your submission. You should expect to spend 3-4 hours on each problem set.

Problem sets are due each week by 11:59pm on Sunday via Blackboard. Late problem sets will be deducted 50%, and no problem sets will be accepted after the start of class on Monday. We will review the solutions to the problem set at the start of our class meeting on Monday. There will be 10 problem sets in total.

Final Project

Your final project is a data storytelling exercise that is due at the conclusion of the final exam block assigned to this course (there is no in-person final exam). More information will be provided during the course.

Attendance

Attendance is mandatory unless you email me to let me know that you will be missing class, preferably before class or as soon as possible. I have a no-questions-asked excuse policy, meaning that your absence is automatically considered excused – as long as you email me to let me know. Every class session is video recorded and made available through the course Blackboard site.

Your attendance grade is the percentage of class meetings with an unexcused absence deducted from 100 (rounded up). For example, if you have two unexcused absences, your attendance grade is $100 - ((2/14) * 100) = 86$

University Policy on Observance of Religious Holidays

Students must notify faculty during the first week of the semester in which they are enrolled in the course, or as early as possible, but no later than three weeks prior to the absence, of their intention to be absent from class on their day(s) of religious observance. If the holiday falls within the first three weeks of class, the student must inform faculty in the first week of the semester. For details and policy, see "Religious Holidays" at provost.gwu.edu/policies-procedures-and-guidelines.

Grading

Your course grade is calculated as your grade on each of the following course components weighted by:

Problem Sets	40%
Final Project	40%
Attendance	20%

Course grades are converted into letter grades according to the following rubric:

93-100 = A (4.0 GPA points)
90-92 = A- (3.7 GPA points)
87-89 = B+ (3.3 GPA points)
83-86 = B (3.0 GPA points)
80-82 = B- (2.7 GPA points)
77-79 = C+ (2.3 GPA points)
73-76 = C (2.0 GPA points)
70-72 = C- (1.7 GPA points)
67-69 = D+ (1.3 GPA points)
63-66 = D (1.0 GPA points)
60-62 = D- (0.7 GPA points)

Support for Students with Disabilities

Any student who may need an accommodation based on the impact of a disability should contact the Office of Disability Support Services (DSS) to inquire about the documentation necessary to establish eligibility and to coordinate a plan of reasonable and appropriate accommodations. DSS is located in Rome Hall, Suite 102. For additional information, please call DSS at 202-994-8250, or consult disabilitysupport.gwu.edu/.

Academic Integrity

Academic integrity is an essential part of the educational process, and all members of the GW community take these matters very seriously. As the instructor for this course, my role is to provide clear expectations and uphold them in all assessments. Violations of academic integrity occur when students fail to cite research sources properly, engage in unauthorized collaboration, falsify data, and otherwise violate the Code of Academic Integrity. If you have any questions about whether or not particular academic practices or resources are permitted, you should ask me for clarification. If you are reported for an academic integrity violation, you should contact the Office of Student Rights and Responsibilities (SRR) to learn more about your rights and options in the process. Consequences can range from failure of assignment to expulsion from the university and may include a transcript notation. For more information, please refer to the SSR website, email rights@gwu.edu, or call 202-994-6757.

Class Recordings and Use of Electronic Course Materials

Class meetings will be audio/video recorded and made available to other students in this course. As part of your participation in this course, you may be recorded. If you do not wish to be recorded, please contact me during the first week of class to discuss alternative arrangements.

Students are encouraged to use electronic course materials, including recorded class sessions, for private personal use in connection with their academic program of study. Electronic course materials and recorded class sessions should not be shared or used for non-course related purposes unless express permission has been granted by the instructor. Students who impermissibly share any electronic course materials are subject to discipline under the Student Code of Conduct. Please contact the instructor if you have questions regarding what constitutes permissible or impermissible use of electronic course materials and/or recorded class sessions.

Additional Resources for Students

- **Counseling and Psychological Services**
  202-994-5300
  GW's Colonial Health Center offers counseling and psychological services, supporting mental health and personal development by collaborating directly with students to overcome challenges and difficulties that may interfere with academic, emotional, and personal success. healthcenter.gwu.edu/counseling-and-psychological-services

- **Writing Center**
  GW's Writing Center cultivates confident writers in the University community by facilitating collaborative, critical, and inclusive conversations at all stages of the writing process. Working alongside peer mentors, writers develop strategies

to write independently in academic and public settings. Appointments can be booked online at gwu.mywconline.com.

- **Statistical Consulting**
  Academic Commons provides GW students with access to statistical consulting (including in R) through Penji. Students can connect with a statistical consultant at academiccommons.gwu.edu/statistical-consulting.

Safety and Security

- Monitor GW Alerts and Campus Advisories to Stay Informed before and during an emergency event or situation
- In an emergency: call GWPD/EMeRG at 202-994-6111 or 911
- For situation-specific actions: refer to GW's Emergency Response Handbook and Emergency Operations Plan
- In the event of an armed intruder: **Run. Hide. Fight.**

Course Outline

You are expected to read the assigned texts **BEFORE** class. These readings introduce the concepts that we will discuss in class. Readings with an embedded link can be accessed online. All other readings are available on Blackboard. References to Ismay & Kim refer to Statistical Inference via Data Science: A ModernDive into R and the Tidyverse by Chester Ismay and Albert Y. Kim, an open-source online textbook that is available at https://www.moderndive.com.

Date	**Topics and Readings**
Data Analysis in R	
January 23	Introduction to Data Analysis, R, and RStudio • Chalabi (2017), "Three Ways to Spot a Bad Statistic" (video) • Bouie (2021), "Quantifying the Pain of Slavery" (New York Times) • Ismay & Kim, Preface: Introduction for Students • Ismay & Kim, Ch. 1: Getting Started with Data in R
January 30	Class Cancelled (Professor Sick)
February 6	Data Visualization I • Excerpt of Berinato (2016), Good Charts (on Blackboard) • Fry (2021), "When Graphs Are a Matter of Life and Death" (The New Yorker) (on Blackboard) • Ismay & Kim, Ch. 2: Data Visualization

February 13	Data Visualization II & R Markdown • Wickham and Grolemund, R for Data Science, Ch. 27: R Markdown Note: Homework due February 26 due to Martin Luther King, Jr. Day
February 20	**President's Day (no class)**
February 27	Researcher Choices and Bias & Data Wrangling I • Retro Report (2021), "What's in a Number? (video) • Reinhart (2015), Statistics Done Wrong, Ch. 9: "Researcher Freedom: Good Vibrations?" (on Blackboard) • Aschwanden (2015), "Science Isn't Broken" (FiveThirtyEight) • Ismay & Kim, Ch. 3: Data Wrangling
March 6	Correlation vs. Causation & Data Wrangling II • Spiegelhalter (2019), The Art of Statistics, Ch. 4: "What Causes What?" (on Blackboard) • Bergstrom & West (2020), Calling Bullshit, Ch. 4: "Causality" (on Blackboard) • Ismay & Kim, Ch. 4: Data Importing and Tidy Data
March 13	**Spring Break (no class)**
March 20	Data Ethics & Exploring the tidyverse • O'Neil (2017) "The era of blind faith in big data must end" (video) • Diakopolous (2016), "BuzzFeed's pro tennis investigation displays ethical dilemmas of data journalism" (Columbia Journalism Review) • Willis (2014), "Professors' Research Project Stirs Political Outrage in Montana" (New York Times) • Wickham and Grolemund, R for Data Science, Chs. 14 - 16: Strings, Factors, & Dates and Times
Statistics in R	
March 27	Sampling & Exploratory Data Analysis • Ismay & Kim, Ch. 7: Sampling • Wickham and Grolemund, R for Data Science, Ch. 7: Exploratory Data Analysis • Schulson (2020), "In Fallout Over Polls, 'Margin of Error' Gets New Scrutiny" (Undark) • The Economist (2021), "How data detectives spotted fake numbers in a widely cited paper" (on Blackboard)
April 3	Guest Speaker: Andrew Van Dam (The Washington Post)

April 10	Political Polling • Excerpt of Morris (2022), Strength in Numbers (on Blackboard) • Keeter, Kennedy, and Deane (2020), "Understanding how 2020 election polls performed and what it might mean for other kinds of survey work" (Pew Research Center) • Rutenberg, Bensinger, and Eder (2022), The 'Red Wave' Washout: How Skewed Polls Fed a False Election Narrative (New York Times)
April 27	Hypothesis Testing & Uncertainty • Ismay & Kim, Ch. 9: Hypothesis Testing • Kupferschmidt (2018), More and more scientists are preregistering their studies. Should you? (Science magazine)
April 24	Simple Regression • Ismay & Kim, Ch. 5: Basic Regression • Ismay & Kim, Ch. 6: Multiple Regression
Other Topics in R	
May 1	Mapping in R • No assigned readings.
May 3	TBD • No assigned readings.

Version: 3
Last Updated: February 15, 2023
Subject to change.