

Syllabus

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

Meeting Times: Mondays and Wednesdays, 5:45-7:00pm

Classroom: MPA B01

Professor: Nicholas Bell, Ph.D. (he/him/his)
nicholasbell@gwu.edu

Office Hours (on Zoom): Wednesdays 8:00-9:00pm and by appointment
(Zoom link for office hours:
<https://gwu-edu.zoom.us/j/4852885019>)

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.

Course Materials

Students are required to purchase a [Posit Cloud](#) student access plan for the duration of the course. Posit Cloud is a web-based version of the popular R development environment RStudio. Even if you are familiar with the desktop version of RStudio, you will need to purchase a Posit Cloud membership. All of the code we work on during this class will be available on Posit Cloud. At the time of this writing, a membership costs \$5 per month. If the course materials pose a financial burden, please contact me and we will work something out.

Assignments

There will be five assignments in this course that will ask you to apply the R skills we learn in class to typical issues we encounter when working with data. You may complete these assignments 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 assignment.

Assignments are due by 11:59pm Eastern of the following Tuesday. This means that you have 6 days to complete each assignment. Late assignments will be deducted 50%, and no assignments will be accepted after the start of class on Wednesday. We will review the solutions to the assignment at the start of our class meeting on Wednesday.

Mid-term and Final Exams

The mid-term exam is scheduled for October 25 and it is an in-class exam. The mid-term exam will cover material from both the data analysis and coding tracks through October 18.

The final exam will take place during the final exam period. The final exam will cover material from the data analysis track beginning October 23, but will include material from the coding track from the entire course. This is because the skills we learn in the first half of the course are the “building blocks” for the skills we will learn in the second half of the course, so it is not possible to separate them.

Attendance

Attendance is mandatory. If you miss a class due to illness, family emergencies, University-scheduled events, and other unusual circumstances, you must email me and let me know. You do not need to provide proof of your reason for missing class, but misrepresentation of your reason for excusal is a violation of the [Code of Student Conduct](#).

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 $100 - ((2/24) * 100) = 92$.

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:

Assignments	25%
Mid-term exam	25%
Final exam	40%
Attendance	10%

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

This course consists of two parallel tracks, a Data Analysis track that will be taught on Mondays and a Coding track that will be taught on Wednesdays. The Data Analysis track will introduce you to key principles of data analysis and core concepts in statistics. The coding track will teach you how to conduct data analyses using the statistical programming language R.

The readings assigned for each class should be completed by the following meeting of that track, e.g., the readings assigned in the Data Analysis track on August 28 are to be completed by September 11, while the readings assigned in the Coding track on August 30 are to be completed by September 6. 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 Modern Dive 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>.

Week	Data Analysis Track (Mondays)	Coding Track (Wednesdays)
1	<p>August 28: Introduction</p> <p>Homework</p> <ul style="list-style-type: none"> • Ismay & Kim, Preface: Introduction for Students • Excerpt of Berinato (2016), Good Charts (on Blackboard) • Fry (2021), "When Graphs Are a Matter of Life and Death" (The New Yorker) (on Blackboard) 	<p>August 30: Introduction</p> <p>Homework</p> <ul style="list-style-type: none"> • Ismay & Kim, Ch. 1: Getting Started with Data in R
2	<p>September 4: No Class (Labor Day)</p>	<p>September 6: Base R</p> <p>Homework</p> <ul style="list-style-type: none"> • Ismay & Kim, Ch. 2: Data Visualization • Assignment #1 (due September 12)
3	<p>September 11: Data Visualization</p> <p>Homework</p> <ul style="list-style-type: none"> • 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) 	<p>September 13: Data Visualization I</p> <p>Homework</p> <ul style="list-style-type: none"> • No homework
4	<p>September 18: No Class (professor at conference)</p>	<p>September 20: Data Visualization II</p> <p>Homework</p> <ul style="list-style-type: none"> • Wickham, Çetinkaya-Rundel, and Grolemund, R for Data Science, 2nd ed., Ch. 29: Quarto
5	<p>September 25: No Class (Yom Kippur)</p>	<p>September 27: Quarto</p> <p>Homework</p> <ul style="list-style-type: none"> • Ismay & Kim, Ch. 3: Data Wrangling • Assignment #2 due October 3

6	<p>October 2: Researcher Choices and Bias</p> <p><u>Homework</u></p> <ul style="list-style-type: none"> • Bergstrom & West (2020), <u>Calling Bullshit</u>, Ch. 4: "Causality" (on Blackboard) 	<p>October 4: Data Wrangling I</p> <p><u>Homework</u></p> <ul style="list-style-type: none"> • Ismay & Kim, <u>Ch. 4: Data Importing and Tidy Data</u>
7	<p>October 9: Correlation vs. Causation</p> <p><u>Homework</u></p> <ul style="list-style-type: none"> • O'Neil (2017) "<u>The era of blind faith in big data must end</u>" (video) • Diakopoulous (2016), "<u>BuzzFeed's pro tennis investigation displays ethical dilemmas of data journalism</u>" (Columbia Journalism Review) • Bouie (2021), "<u>Quantifying the Pain of Slavery</u>" (New York Times) 	<p>October 11: Data Wrangling II</p> <p><u>Homework</u></p> <ul style="list-style-type: none"> • Wickham, Çetinkaya-Rundel, and Golemund, <u>R for Data Science</u>, 2nd ed., <u>Chs. 15, 17, & 18: Strings, Factors, and Dates and times</u> • Assignment #3 (due October 17)
8	<p>October 16: Data Ethics & Responsibilities</p> <p><u>Homework</u></p> <ul style="list-style-type: none"> • Ismay & Kim, <u>Ch. 7: Sampling</u> 	<p>October 18: Data Wrangling III</p> <p><u>Homework</u></p> <p>No homework</p>
9	<p>October 23: Sampling</p> <p><u>Homework</u></p> <ul style="list-style-type: none"> • Excerpt of Morris (2022), <u>Strength in Numbers</u> (on Blackboard) • Keeter, Kennedy, and Deane (2020), "<u>Understanding how 2020 election polls performed and what it might mean for other kinds of survey work</u>" (Pew Research Center) 	<p>October 25: Mid-term exam</p> <p><u>Homework</u></p> <p>No homework</p>
10	<p>October 30: Political Polling</p> <p><u>Homework</u></p> <p>No homework</p>	<p>November 1: Analyzing Polls in R</p> <p><u>Homework</u></p> <ul style="list-style-type: none"> • Assignment #4 (due November 7)
11	<p>November 6: Guest Speaker: Emily Guskin, Polling Analyst at The Washington Post</p> <p><u>Homework</u></p> <p>No homework</p>	<p>November 8: Election Results in R</p> <p><u>Homework</u></p> <p>No homework</p>

12	November 13: Machine Learning/AI <u>Homework</u> • Ismay & Kim, Ch. 9: Hypothesis Testing	November 15: No Class (professor away)
13	November 27: Hypothesis testing <u>Homework</u> • Excerpt of Gonick & Smith (1993), <u>The Cartoon Guide to Statistics</u> (on Blackboard)	November 29: Hypothesis testing <u>Homework</u> • Ismay & Kim, Ch. 5: Basic Regression • Ismay & Kim, Ch. 6: Multiple Regression
14	December 4: Regression <u>Homework</u> No homework	December 6: Regression <u>Homework</u> • Assignment #5 (due December 10)
15	December 11: Review session	
Final exam date TBA		

Version: 2
Last Updated: August 16, 2023
Subject to change.