Syllabus

SMPA 2152: Data Analysis for Journalism and Political Communication (Spring 2021)

Meeting Times:	Tuesdays and Thursdays, 8:00-9:15am
Classroom: Virtual Classroom Zoom Link:	MPA 306 https://gwu-edu.zoom.us/j/99899478602
Professor:	Nicholas Bell (he/him/his) nicholasbell@gwu.edu
Office Hours (on Zoom):	Mondays 4:00-6:00pm and by appointment (Zoom link for office hours: https://gwu-edu.zoom.us/j/92058537984)
	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. This class is <u>not</u> a statistics class. 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 programming language R. There are many advantages to learning R: it is free and open-source, meaning that developers are continually releasing new tools to ease programming; 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, meaning that it is well-suited for confronting the problems and issues in data that we will discuss in this class. 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 able to develop, produce, and share an original data storytelling project.

COVID-19 Protocols

In accordance with University policy, masks must be worn over your nose and mouth at all times when not actively eating or drinking. You may briefly remove your mask while actively eating and drinking and then must replace it. Failure to adhere to the indoor masking policy will result in your removal from the classroom.

If you test positive for COVID-19 or are experiencing COVID-19 symptoms, **do not come to class** and contact the Colonial Health Center at 202-994-5300. Students who are unable to attend class due to COVID-19 exposure or illness or who choose not to attend class due to concerns about COVID-19 exposure (as well as students who cannot attend for any other legitimate reason) will be provided with a video recording of the class session. You do not need to provide medical documentation to be excused from class and receive a video recording of the class session.

Virtual Meeting Policies

Please use the Zoom link listed at the top of the syllabus and on Blackboard to access our virtual classroom.

- Please use your first and last name as your display name. You might also consider adding your preferred pronouns to your profile. (Instructions for changing your display name; instructions for adding and sharing your preferred pronouns.
- You are not required to turn on your video camera, but it is preferred that you do so. You are welcome to apply a virtual background to your video to protect your

privacy. A set of GW-themed virtual backgrounds and instructions for applying them are available at https://www.alumni.gwu.edu/gw-branded-video-callbackgrounds

- Please remain muted when not speaking. To ask to speak, please use the "raise hand" feature, which is accessed via the "reactions" button on the Zoom toolbar.
- Please be aware that all virtual meetings are recorded and will be made available to other students who are unable to attend.

How the Course Will Work

Each Tuesday, I will present a lecture on one or more concepts of data analysis, such as misleading uses of data, best practices for data visualization, ethical data analysis, and current debates in political polling. We will also use this time for class discussion, to complete exercises in small groups, and other activities. Please read the assigned texts before our class meeting to get the most out of these sessions.

On Thursdays, we will apply the principles from Tuesday's lecture in R. I encourage you to read the assigned texts <u>after</u> our class meeting. I find that it is easier to learn R through practice rather than reading. The assigned texts will provide a review of what was covered in class and offer a few extensions that you can try on your own.

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 other students, please indicate their names at the top of your submission. Problem sets will be graded out of four possible points:

- 4. Problem set answers are correct (or nearly correct) <u>and</u> R code successfully applies skills learned in class
- 3. One or more problem set answers are incorrect <u>or</u> R code applies skills learned in class incorrectly
- 2. Partially incomplete problem set <u>or</u> R code does not consistently apply skills learned in class
- 1. Mostly incomplete problem set <u>or</u> R code does not apply skills learned in class

Problem sets are due each week by 5:00pm on Wednesday via Blackboard. Late problem sets will be deducted 1 point, and no problem sets will be accepted after the start of class on Thursday. We will review the solutions to the problem set at the start of our class meeting on Thursday. 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. Due to the COVID-19 pandemic, 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. Students who are unable to attend class will be provided with a video recording of the class session.

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 - ((2/27) * 100) = 93

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)

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 SRR website (https://studentconduct.gwu.edu/academic-integrity), 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

• Disability Support Services (DSS)

202-994-8250

Any student who may need an accommodation based on the potential impact of a disability should contact Disability Support Services at disability support.gwu.edu to establish eligibility and to coordinate reasonable accommodations.

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-andpsychological-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.

• 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.

Course Outline

You are expected to read the assigned texts for Tuesday's class <u>before</u> our meeting. These readings introduce the concepts that we will discuss in class. On Thursdays, we will have a lesson on programming in R. You should read the texts listed under these classes <u>after</u> our meeting, since they review the skills that you have just learned.

Readings with an embedded link can be accessed online. All other readings are available on Blackboard, except for two textbooks from which we will read multiple chapters. These textbooks are available to you for free:

- <u>Programming Skills for Data Science: Start Writing Code to Wrangle, Analyze,</u> <u>and Visualize Data with R</u> (Addison-Wesley, 2019) by Michael Freeman and Joel Ross (available through GW Libraries)
- <u>R for Data Science</u> (O'Reilly, 2017) by Hadley Wickham and Garrett Grolemund (free online access)

Week	Date	Topics and Readings	
Introduction to Data Analysis			
1	January 11 (Virtual Meeting)	 Guiding Principles of Data Analysis Cepelewicz (2021), "The Hard Lessons of Modeling the Coronavirus Pandemic" (Quanta Magazine) 	
	January 13	Getting Started in R	
	(Virtual Meeting)	• Freeman and Ross, Chs. 5, 6 (except 6.4), and 7	
2	January 18	 Interrogating Data Wheelan (2014), <u>Naked Statistics</u>, Ch. 7: "The Importance of Data" Mull (2021), "The Great Shoplifting Freak-Out" (The Atlantic) The Economist (2021), "How data detectives spotted fake numbers in a widely cited paper" 	

		Creating and Importing Datasets in R
	January 20	• Wickham and Grolemund Ch 13
		The Use and Abuse of Data: Besearcher Choices
		• Reinhart (2015), Statistics Done Wrong, Ch. 9: "Researcher
		Freedom: Good Vibrations?"
	ΙΟΓ	• Silver (2020), "How FiveThirtyEight's 2020 Presidential
2		Forecast Works" (FiveThirtyEight)
3		• Aschwanden (2015), "Science Isn't Broken" (FiveThir-
		tyEight)
	January 27	Data Wrangling: Modifying Data
		Wickham and Grolemund, Chs. 5.1 through 5.5
		The Use and Abuse of Data: Bias
		• Retro Report (2021), "What's in a Number? (video)
	February 1	• Ordway (2021), "Covering Scientific Consensus" (Journal-
4		ISTS RESOURCE)
		• Glass (2010), KHOCK KHOCK. WHO'S THELE? THE HUHL. (THIS American Life [nedeast]: listen only to Act One)
		Data Wrangling: Working with Factors
	February 3	• Wickham and Grolemund Chs 15
		The Use and Abuse of Data: Causation vs. Correlation
		• Spiegelhalter (2019) The Art of Statistics Ch 4 ⁻ "What
	February 8	Causes What?"
		• Bergstrom & West (2020), Calling Bullshit, Ch. 4: "Causality"
5		• Willis (2014), "Professors' Research Project Stirs Political
		Outrage in Montana" (New York Times)
	February 10	Data Wrangling: Summarizing Data
		• Wickham and Grolemund, Chs. 5.6 and 5.7
The Good	l, the Bad, and the U	Jgly Graphs
	February 15	Class cancelled
6	February 17	Catch-up/Review Day
		• No assigned readings
	February 22	Data Visualization: Getting to the Feeling Benind Our Eyes
		• Excerpts of Bernhald (2010), <u>Good Charts</u>
		Misinformation Can Be Beautiful Too"
7		• Fry (2021) "When Granhs Are a Matter of Life and Death"
1		(The New Yorker)
	February 24	Data Visualization in R. Part I
		• Freeman and Ross, Chs. 16.1 and 16.2
		• See also Wickham and Grolemund, Ch. 3

		Data Visualization: Presenting Data to the Public
		• Chalabi (2018), "Sequence, Sequence Surprise! Designing
	March 1	Data for Maximum Impact" (video)
		• Patino (2021), "The Rise of the Pandemic Dashboard"
8		(Bloomberg CityLab)
		• Twitter thread by Crystal Lee (2021)
		Data Visualization in R, Part II
	March 3	• Freeman and Ross, Ch. 16.3
		• See also Chang (2022), <u>R Graphics Cookbook</u> , 2nd ed.
		Designing and Conducting a Data Analysis Project
		• Peng and Matsui (2018), <u>The Art of Data Science</u> , Chs. 2
		and 3 ("Epicycles of Analysis" and "Stating and Refining the
	March 8	Question")
Q		• Michel and Rebma (2021), "Dating Data: How We Used Mul-
		tiple Dating Schemes to Provide the Most Complete Picture
		of the Pandemic" (COVID Tracking Project/The Atlantic)
		Final Exam Instructions Distributed
	March 10	Data Visualization in R, Part III
		• No assigned readings
Spring Bi	reak	
Polling is	s dead. Long live po	lling!
		Political Polling, Part I
	March 22	• Wheelan (2014), <u>Naked Statistics</u> , Ch. 10: "Polling"
		• Edwards-Levy (2018), "5 Tips for Reading the Polls Like a
		Pro" (Huffington Post)
		• Silver (2014), "How FiveThirtyEight Calculates Pollster
		Ratings" (FiveThirtyEight)
10		Political Polling, Part II
10		• Keeter, Kennedy, and Deane (2020), "Understanding how
	March 24	2020 election polls performed and what it might mean for
		other kinds of survey work" (Pew Research Center)
		• Schulson (2020), "In Fallout Over Polls, 'Margin of Error'
		Gets New Scrutiny" (Undark)
		• Agiesta, (2021), "CNN launches new polling methodology"
		(CNN)
	March 29	Guest Speaker: Walter Hickey, Senior Data Editor at Insider
11		and Founder, Numlock News
	March 31	No Class
	April 5	Working with Polling Data in R, Part I
12		Readings TBD
	April 7	Working with Polling Data in R, Part II
		• Keadings IBD

13	April 12	Ethics in Data Analysis • O'Neil (2017) "The era of blind faith in big data must end" (video)
		• Bouie (2021), "Quantifying the Pain of Slavery" (New York
		• Diakopolous (2016), "BuzzFeed's pro tennis investigation
		displays ethical dilemmas of data journalism" (Columbia Journalism Review)
	April 14	Additional Topics in R, Part I
		• Readings TBD
14	April 19	Managing and Sharing Data
		• Rogers (2012), "Open data journalism" (The Guardian)
		• Eubank (2016), 'Embrace your Fallibility' Inoughts on Code Integrity" The Delitical Methodologist 12(2): 10, 16
		• Alexander (2013) "The Student Who Caught the Profe"
		(BBC)
	April 21	Additional Topics in R, Part II
		• Readings TBD
15	April 26	No Class (meetings with the professor to discuss final projects)

Last Updated: February 16, 2022 Subject to change.