DATS/PSC 2102: Data Visualization

Spring 2025, TR 11:10-12:25, Gelman B108

instructor:	Eric Lawrence
email:	edl@gwu.edu
office:	Monroe 464
office hours:	W 3:30-5:00pm or by appointment

We also have three GTAs working on the class, Ram Mannuru (prudhviram.mannuru@gwu.edu), Hussain Nathan (hussain.nathani@gwu.edu), and Krishna Paila (saikrishna.paila@gwu.edu), with more details below.

Course Description

Do you like exploring data? Are you the type of person who asks, "how do you know that?" and "what does the evidence look like?" If so, this may be the class for you. The class, a required core course for Data Science majors, emphasizes working with data and code to engage scientific questions. Over the course of the semester, you will develop coding, data analysis, and data presentation skills, culminating in a final data driven project on a topic of your choosing. DATS 1001 and STAT 1051/1053/1111/1127 are the prerequisites for both DATS 2102 and PSC 2102. Also note that the class satisfies the "group D" requirement for the Political Science major, so if you are a major and take this class, you do not need to take PSC 2101.

The class has three main elements:

1. Data collection/scraping/munging/wrangling/plumbing/carpentry/cleaning. How can we efficiently collect data in order to tell a story and answer questions?
2. Data visualization. Now that we have data in hand via step 1, how can we describe the data graphically in order to effectively communicate an idea or a story?
3. Statistical modeling. Now that we have the data in hand via step 1, how can we examine the relationships between our outcome(s) of interest and other variables?

The class will not proceed strictly in order 1, 2, 3, above. Rather, we will start with data that is easier to collect and visualize and model it. Then we will repeat the cycle with messier data that is not as simple to collect. You can think of the techniques that we will learn in this class as applied statistics, or if you prefer, you can call it data science as in Figure 1-1 from R4DS 2e below:

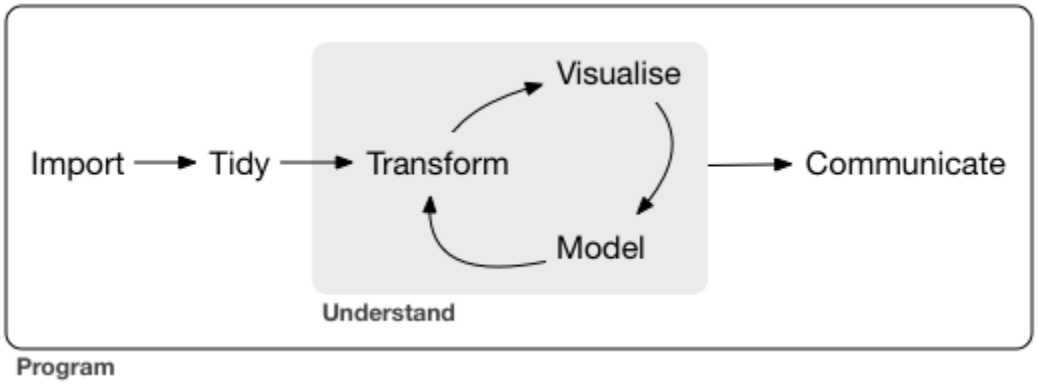

Figure 1: 'Wickham, Çetninkaya-Rundel, and Grolemund, *R for Data Science, 2e*

Learning Objectives

As a result of completing this course, students will:

- Improve their ability to process and analyze data using modern computational tools.
- Have a deeper understanding of the appropriate ways to summarize simple and complex relationships between and among variables.
- Learn better and worse ways to present evidence. You will learn the worse ways to present evidence in order to avoid doing so in the future. You will learn the better ways to present evidence in order to add more beauty to the world.

Time budgeting

We have two and a half hours of in-class meetings a week. You should expect to average about five hours of work outside of class per week. Some weeks will be more than that; some weeks may be less. This class involves a large amount of coding in R, and the amount of time spent on coding projects can often be hard to predict. Starting assignments the day before they are due should be avoided. That said, we will dedicate roughly half of our in person meetings to working on projects in class.

Required Texts

Wickham, Hadley, Çetninkaya-Rundel, and Garrett Grolemund. 2023. *R for Data Science, 2nd edition*. O'Reilly. *You can the order paperback, but the chapters on the syllabus reference the online version of the text, free here:* R4DS2e.

Healy, Kieran. 2018. *Data Visualization: A practical introduction*. Princeton. *Can order in paperback, but the text is also free as a* webpage.

Wilke, Claus. 2019. *Fundamentals of Data Visualization: A Primer on Making Compelling and Accurate Figures*. O'Reilly. *Can order in paperback, but the text also free as a* webpage.

Recommended References

Chang, Winston. 2021. *R Graphics Cookbook*. O'Reilly. Useful reference, full of recipes.

Grades

This course offers a generous late policy. My late policy is as follows: for each week that an assignment is late, I will deduct ten percent of the available points on that assignment. Under my definition, "late" starts one week after the target date in the table below. I think you will be better off meeting the target dates, but I want to maintain as much flexibility as possible. Life happens. If you have unexpected circumstances arise in your life that cause problems for meeting assignment deadlines, let me know as soon as possible (ideally prior to the deadline), and we can likely work out an arrangement. Your course grade will be based on the following components:

Item	**Target date**	**Weight**
Leaflet map	January 23, 2025	5%
Dear Data	February 6, 2025	10%
Graphing single variables	February 20, 2025	10%
Making maps	March 20, 2025	20%
Presentations and tables	April 3, 2025	15%
Graphing relationships	April 17, 2025	15%
Final project	final exam date, TBD	25%

Participation and Reading

I taught this class for the first time in the spring of 2014. For the first few years that I taught it, it was a smaller sized class, but with the growth of the data science major, the class has expanded in size to accomodate demand. Larger classes often deter participation, but I encourage you all to ask questions throughout the semester, even if your question seems narrow or trivial. If you have encountered a dead end or pitfall, chances are you are not alone. We will be figuring out the best way to have online discussions after a field a survey of you all.

I selected the readings for the class carefully, but many of the readings will be most helpful to you as you work through the assignments. If you took driver's ed, you know that there is a difference between reading about how to drive a car and actually driving a car. Working through the assignments in this class requires that you drive the car.

Slides will be posted prior to class sessions except under exceptional situations.

Staff biographies

Eric Lawrence has been at GW for over 20 years and has been teaching data visualization since 2014, first as a political science class, and now jointly as a data science class. Currently chair of the political science department, Eric served as interim director of the Data Science program from July 2108 through June 2020. During that time, Eric worked with a faculty committee and the CCAS dean's office to help create the current Data Science major. Eric grew up in northwest Iowa and went to Stanford thinking he would major in math, but graduated as a political science major after trying out math and two other majors. After working at the Brookings Institution as an RA, Eric went to the University of Minnesota, where he received his PhD. He and his wife live in the Maryland suburbs, with a college graduate son living in New York City and a 2024 college graduate living and working from home.

Ram Mannuru is in his final semester of the Master's in Data Science program at George Washington University, graduating in May 2025. He holds a Bachelor's degree in Computer Science and Engineering. With professional experience as a Software and Machine Learning Engineer, Ram has worked on predictive modeling, backend development, and cloud-based deployments. His technical skills include Python, R, SQL, TensorFlow, PyTorch, AWS, and building RESTful APIs. Ram's interests lie in machine learning, data visualization, and creating impactful real-time applications. As a Graduate Instructional Assistant, he is eager to support students in their learning journey.

Hussain Nathani is currently pursuing his Master's in Data Science at George Washington University. He completed his Bachelor's in Computer Science in 2024 and has developed a strong interest in applying data science to solve real-world problems. With a strong foundation in Python and R programming languages, Hussain's journey into the data science field began with projects focusing on predictive modeling and data visualization, which further

deepened his passion for the subject. He also gained internship experience in the data science field, working on projects that involved data wrangling, analysis, statistical modeling, and Generative AI. Hussain's areas of interest include Natural Language Processing, Data Analytics, and Machine Learning. He is excited to assist students with course content and any questions they may have this semester.

Saikrishna Paila is currently pursuing a Master's in Data Science. He completed his Bachelor of Technology in Computer Science and Engineering, where he built a strong foundation in programming and data analysis. Saikrishna has experience as a Graduate Instructional Assistant for the Data Science Capstone course, guiding students in data preprocessing, feature engineering, model optimization, and effective data visualization techniques. He has worked on projects such as AI-powered legal research tools, stock market prediction models, and interactive dashboards for crime pattern analysis, showcasing his expertise in Python, R, SQL, and visualization tools like Tableau. Saikrishna is passionate about creating insightful visual narratives to drive decision-making.

Course outline

Week	Topic
1	getting rolling
2	the language of graphs
3	description: distributions and variation
4	wrangling data
5	perceptions and principles
6	comparisons
7	labelling is communicating
8	making maps
9	colors: how to choose? what to consider?
10	(modeling) relationships between variables
11	summarizing uncertainty
12	annotations & interactivity
13	final projects I
14	final projects II

Preliminary Course Calendar

Key to readings: The *R for Data Science* readings will be denoted as R4DS, followed by the name of the chapter. The numbering system on the website version sometimes differs from the print version (in case you buy the paperback), so go by the links on the syllabus. The Healy readings will be denoted as socviz, followed by the chapter number. The Wilke readings will be denoted as Wilke, followed by the chapter and section numbers.

Week 1 [1.16-1.18] getting rolling, warming up, hitting the ground running

- 1.14: Course overview and tech issues, R and RStudio + Quarto
  - Install R first and then install RStudio. Also install the Quarto package inside RStudio if needed.
  - You may already have R and RStudio installed, but as of 1.13.25, the latest version of R is 4.4.2, and the latest version of RStudio is 2024.12.0+467. I recommend installing the latest versions, as conflicts from old versions can often cause errors.
- 1.16: R4DS: Workflow: scripts and projects R4DS: chapters 28-29 on Quarto, socviz: preface, socviz: ch. 2

Week 2 [1.21-1.23] the language of graphs and ggplot

- Wilke: ch. 1-5
- R4DS: Visualization, Communication
- socviz: ch. 1-1.2, ch. 3

Week 3 [1.28-1.30] description: distributions and variation

- Wilke:
  - chs. 6-11
- R4DS:
  - Exploratory Data Analysis
  - Data Transformation (with dplyr)
  - Pipes
- socviz:
  - ch. 3 - ch. 4
- Aran Lunzer and Amelia McNamara, "Exploring Histograms"

Recommended:

- Cédric Scherer: Visualizing distributions with raincloud plots with ggplot2

Week 4 [2.4-2.6] wrangling data

- R4DS: chs. 6, also see "transform" section of R4DS2e here

Week 5 [2.11-2.13] perceptions and principles

- Tufte, Edward. 1997. *Visual Explanations*, chapter 2. Graphics Press. [On BB]
- Cleveland, W.S. and McGill, R., 1984. "Graphical perception: Theory, experimentation, and application to the development of graphical methods." *Journal of the American Statistical Association*, 79(387): 531-554
- Franconeri, S.L., Padilla, L.M., Shah, P., Zacks, J.M. and Hullman, J. 2021 "The science of visual data communication: What works." *Psychological Science in the Public Interest*, 22(3): 110-161
- socviz, ch. 1.3-1.8

Recommended:

- John Rauser, "How humans see data," Youtube video
- Elliot, Kennedy. 2017. "39 Studies about Human Perception in 30 Minutes." *Medium*

Week 6 [2.18-2.20] comparisons: the heart of science and statistics

- Wilke: ch. 9, ch. 21
- socviz: ch. 4

Recommended:

- Lisa Charlotte Muth: How to read a log scale (also see parts 2 and 3)
- Mike Bostock: Methods of comparisons compared

Week 7 [2.25-2.27] labeling is communicating

- Wilke: chs. 22-24
- socviz: ch. 5
- Lisa Charlotte Muth, "What to consider when using text in data visualizations"

Week 8 [3.4-3.6] making maps

- Wilke: ch. 15
- socviz: ch. 7
- Lisa Charlotte Muth: What to consider when creating choropleth maps
- Kyle Walker, Analyzing US Census Data, chapter 6

Recommended:

- Lovelace et al.: Geocomputation with R, chs. 8-9

Week 9 [3.18-3.20] colors: how to choose? what to consider?

- Wilke: ch. 19
- Lisa Charlotte Muth: (1) How to pick more beautiful colors for your data visualizations, (2) When to use quantitative and when to use qualitative color scales, (3) When to use sequential and when to use diverging color scales
- All of Muth's color posts are collected at datawrapper.

Week 10 [3.25-3.27] (modeling) relationships between variables

- Wilke: chs. 12-14
- socviz: ch. 6.1-6.4

Week 11 [4.1-4.3] summarizing uncertainty

- Wilke: ch. 15
- socviz: ch. 6.5-6.9

Week 12 [4.8-4.10] annotations and interactivity

- Wilke: chs. 17, 18, & 20
- sociviz: ch. 8

Weeks 13-14 [4.15-4.24] final project workshops

- Wilke: ch. 29

Course AI policy

AI Policy (Original version by Robert Betz, lightly modified by Eric Lawrence)

We are entering a new technological era with the rise of generative Artificial Intelligence (AI), such as GPT, LLaMA, laMDA and other large language models, that are driving an ongoing conversation about their academic uses. Writing aid products like Grammarly, QuillBot, Caktus.ai, etc. advertise their AI features (GrammarlyGO, etc.). We are also learning about the potential benefits and misuse of AI and how it can be applied in the classroom. Learning to use generative AI is an emerging skill, but we must use generative AI tools effectively and responsibly.

Generative AI has been discussed at length within academia, but other sectors are also grappling with its use due to its rapid rise and increased access to the tools. In the private sector, e.g., many workplaces are considering banning its use, partly due to security risks (see "Most businesses to ban ChatGPT, generative AI apps on work devices," August 8, 2023, CSOonline). We are all trying to figure out the right way to use GAI in the long and medium run, but below I have outlined the expectations in our class of its permitted and prohibited use.

Permitted:

- A student types a prompt into an AI tool and reviews the generated content to help them study for a quiz or exam (i.e., a study guide).
- A student types a prompt into an AI tool and uses the generated content to help them brainstorm ideas for a paper or research project.
- A student types a prompt into an AI tool and uses the generated content to help them create a citation for a source and/or reference list.
- A student types a prompt into an AI tool and uses the generated content to help them with small group discussion.

Citing GAI and Verifying its Accuracy:

- By submitting work for evaluation in this course, you represent it as your own intellectual product. If you include content (e.g., ideas, text, code, images) that was generated, in whole or in part, by generative AI tools (including, but not limited to, ChatGPT and other large language models) in work submitted for evaluation in this course, you must document and credit your source. Material generated using other tools should be cited accordingly.
- If you include material generated by a generative AI tool and it is substantively incorrect you will lose points as appropriate. You should verify the accuracy of all content you include in your work.

Sample citation:

- "ChatGPT-4. (YYYY, Month DD of query). 'Text of your query.' Generated using OpenAI. https://chat.openai.com/.
- "ChatGPT-4 (2023, August 9) 'What is a pressing policy issue in the District of Columbia?" Generated using OpenAI. https://chat.openai.com/.

Prohibited:

- While taking an out-of-class ("take-home") or an in-class quiz, a student types a prompt into a generative AI tool and incorporates some or all of the generated content into their submitted answer.

Be aware of the limits of GAI:

- Generative AI is a tool, but you need to cite it when you use it. Always. No exception. And you are prohibited from using it as stated above.
- It may stifle your own independent thinking, creativity, and understanding of class concepts. Minimum effort into both generative AI prompts and your assignments will produce low quality results. Effectively and correctly using AI in academic work takes time and effort.
- Don't trust anything or everything AI says. If it gives you a number or fact, assume it is wrong unless you either know the answer or can check it with another non-AI source. This is an opportunity for you to practice your critical analysis skills. As noted above, you will be responsible for any errors of omissions provided by the tool.
- AI tools are based on data that can include biases and reflect historical or social inequities and thus the AI tool can replicate those biases and inequities. Be aware that it can also produce problematic and potentially offensive answers.

University policies

Academic Integrity Code

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 of record 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.

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.

Use of Electronic Course Materials and Class Recordings

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. Please contact Disability Support Services at disability-support.gwu.edu if you have questions or need assistance in accessing electronic course materials.

Academic support

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.

Academic Commons

Academic Commons provides tutoring and other academic support resources to students in many courses. Students can schedule virtual one-on-one appointments or attend virtual drop-in sessions. Students may schedule an appointment, review the

tutoring schedule, access other academic support resources, or obtain assistance at academiccommons.gwu.edu.

Support for students outside the classroom

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 disabilitysupport.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-and-psychological-services.

GW Campus Emergency Information

GW Emergency Services: 202-994-6111

For situation-specific instructions, refer to GW's Emergency Procedures guide.

GW Alert

GW Alert is an emergency notification system that sends alerts to the GW community. GW requests students, faculty, and staff maintain current contact information by logging on to alert.gwu.edu. Alerts are sent via email, text, social media, and other means, including the Guardian app. The Guardian app is a safety app that allows you to communicate quickly with GW Emergency Services, 911, and other resources. Learn more at safety.gwu.edu.

Protective Actions

GW prescribes four protective actions that can be issued by university officials depending on the type of emergency. All GW community members are expected to follow directions according to the specified protective action. The protective actions are Shelter, Evacuate, Secure, and Lockdown (details below). Learn more at safety.gwu.edu/gw-standard-emergency-statuses.

Shelter

- Protection from a specific hazard

- The hazard could be a tornado, earthquake, hazardous material spill, or other environmental emergency.
- Specific safety guidance will be shared on a case-by-case basis.

Action:

- Follow safety guidance for the hazard.

Evacuate

- Need to move people from one location to another.
- Students and staff should be prepared to follow specific instructions given by first responders and University officials.

Action:

- Evacuate to a designated location.
- Leave belongings behind.
- Follow additional instructions from first responders.

Secure

- Threat or hazard outside of buildings or around campus.
- Increased security, secured building perimeter, increased situational awareness, and restricted access to entry doors.

Action:

- Go inside and stay inside.
- Activities inside may continue.

Lockdown

- Threat or hazard with the potential to impact individuals inside buildings.
- Room-based protocol that requires locking interior doors, turning off lights, and staying out of sight of corridor window.

Action:

- Locks, lights, out of sight
- Consider Run, Hide, Fight

Classroom emergency lockdown buttons

Some classrooms have been equipped with classroom emergency lockdown buttons. If the button is pushed, GWorld Card access to the room will be disabled, and GW Dispatch will be alerted.The door must be manually closed if it is not closed when the button is pushed.Anyone in the classroom will be able to exit, but no one will be able to get in.