J677: Concepts and Tools for Data Analysis and Visualization

Spring 2025

School of Journalism and Mass Communication
UNIVERSITY OF WISCONSIN–MADISON

Course Logistics

Class Hours: Monday & Wednesday 2:30–3:45 PM
Class Room: Vilas 5145
Course Website: Canvas

Instructor and Teaching Assistant Information

Instructor: Ross Dahlke, PhD
Email: ross.dahlke@wisc.edu
Office: 5166 Vilas Hall
Office Hours: Monday 3:45–4:45 PM

Teaching Assistant: Wil M. Dubree, MA
Email: dubree@wisc.edu
Office: 5165 Vilas Hall
Office Hours: Wednesday 1:30–2:30 PM or by appointment

Instructor Bio: Ross is an Assistant Professor at the University of Wisconsin-Madison School of Journalism and Mass Communication. From Westfield, WI, Ross graduated with his Bachelor's from UW-Madison with degrees in Journalism and Political Science. After undergrad, Ross was a Data Scientist at a digital marketing and measurement firm where he developed and deployed Bayesian machine learning models, conducted large-scale A/B testing, and worked with client teams to optimize digital marketing budgets for Fortune 500 companies such as Coach Bags, Calvin Klein, and Facebook Marketplace. He has also worked in politics, consulting for over 80 political campaigns. Before returning to UW, Ross received his PhD from Stanford University, where he was a Knight-Hennessy Scholar and Stanford Data Science Scholar. In his free time, he enjoys attending UW sporting events, attending cultural and artistic events around Madison, and trying to perfect his espresso-making.

Teaching Assistant Bio: Wil is a first-year PhD student at the UW-Madison School of Journalism and Mass Communication, where he employs state-of-the-art computational methods to study the profound influence of local journalism in maintaining an informed and just society. Wil is from Anderson, Indiana and grew up raising llamas.

How Credit Hours Are Met

Students are expected to attend the scheduled class meeting each week and work on course learning activities (reading, writing, problem sets, group project, etc.) outside class for about 6 hours in total. About half of the out-of-class hours should be devoted to hands-on exercises and group projects.

Course Description

Like no other time, our world is recorded in digital formats through social networks, online news platforms, mobile devices, and more. This constant flow of information has given rise to new possibilities for understanding social phenomena, communicating insights, and driving data-informed decisions in fields like journalism, strategic communication, and beyond. As researchers, journalists, strategic communicators, and organizations rapidly innovate ways to create and analyze these digital data, they need professionals who can ethically and effectively harness data to answer important questions, extract key insights, and share those insights in clear, compelling visual formats while remaining aware of what data can and cannot reveal.

This class introduces the processes and tools for working with data in modern digital media contexts. Students will learn to collect, wrangle, and analyze raw data, how to apply ethical and inclusive practices in data handling, and how to design meaningful visualizations that bring data stories to life. By emphasizing the critical connections between exploratory data analysis and effective data visualization, the course shows how to turn statistical findings into narratives that resonate with target audiences.

The course combines conceptual and practical learning designed for upper-level undergraduates and interested graduate students in journalism, strategic communication, social sciences, and related areas. We focus on descriptive and narrative approaches rather than formal hypothesis testing or advanced causal inference, making it suitable for those seeking hands-on data skills applications in media-rich environments. Students who wish to avoid data manipulation or coding completely may find this course unsuitable. At the same time, those eager to explore real-world datasets and digital storytelling techniques will benefit from the curriculum.

The course structure blends lectures, discussions, and lab sessions, ensuring that students not only learn analytical and visualization tools but also understand the broader social context, ethical considerations, and creative aspects of data-driven storytelling. In keeping with the spirit of the Digital Media Analytics Certificate, students will cultivate critical thinking, inclusive communication strategies, and the ability to communicate insights effectively to diverse stakeholders. A final project provides the opportunity to integrate all these skills, producing a portfolio-ready piece demonstrating mastery of data analysis, visualization, and narrative design within digital media environments.

Course Objectives

By the end of the course, students will:

1. Identify and address the practical, ethical, and inclusive challenges of data collection, management, analysis, and presentation, ensuring responsible use and communication of digital media data.
2. Demonstrate a solid understanding of the grammar and principles of data visualization, applying them to create clear, engaging, and contextually relevant data narratives for diverse audiences.
3. Attain proficiency with industry-relevant tools, including R, tidyverse, and generative AI, to effectively prepare, explore, and visualize data in real-world media and communication settings.
4. Develop the capacity to handle and visualize diverse data types, integrating these skills into compelling, data-driven storytelling projects.

Technical Abilities

We will focus mainly on R (and RStudio). We will be using RStudio Posit, a cloud-based browser interface to R coding, to handle assignments. However, you are encouraged to install R and RStudio on your own computer. Both are open-source and free.

This course is tailored to students with backgrounds in the social sciences, and no familiarity with programming or advanced statistics is expected. The most important thing for you to bring to the course is curiosity and persistence: the willingness to think about interesting questions and how they might be answered with

data, and the patience to endure trial and error in answering them, including a learning curve necessary to pick up programming skills in R.

To be sure, there will be math, statistics, and coding, which are indispensable to data analyses and critical to visual communicators in today's digital age. We will discuss and use concepts in statistics and reasoning skills needed to treat data appropriately and identify misuses of data. However, our approach is fundamentally hands-on: it is not about abstract equations or solving numeric puzzles but about understanding why uses of numbers and data are appropriate in certain instances and not in others. This is intended to give the course an intuitive foundation.

Finally, as you may have done many times, you will learn some new software in the course. The goal here is twofold: first, specific tools are useful and good to have in your toolkit. Second, tools change rapidly, so another goal is to develop the skill of picking up new software skills reasonably quickly. Students with experience in design, image manipulation, or web publishing software will find ways to apply those skills in the course.

Graduate Student Notes

Graduate students are likely to have specific interests or goals to pursue, and the content of the class can be modified to suit these goals. For example, a student may wish to focus on visualizing the results of hypothesis testing or causal inference methods instead of exploratory data analysis. Some graduate students will already be familiar with statistical concepts beyond the level we discuss. If you are comfortable with the concepts or tools we are using, please talk to me in advance of those assignments and deadlines so that we can create a productive plan for you.

For the final project, graduate students are expected to create a product of especially high quality related to academic research. Ideally, graduate students will create publication-ready visualizations for an ongoing research project. As such, it is most useful for graduate students to come to the class with scientific data intended for a future publication that they have or wish to analyze. If you do not have appropriate scientific data, I am happy to discuss appropriate datasets that can be leveraged beyond this course. The final requirements for the course can be tailored to meet the individual goals of graduate students.

All graduate students should meet with me (during office hours or another time) at least once before Week 5 to discuss how the final project fits within their program.

Study Strategies

This is a concept-skill hybrid class, and the end goal is to help you appreciate the power of data visualizations and leverage modern data processing tools such as R towards a project of your choice. Coding is a necessary component of this process. Here are a few tips to help you succeed if you are new to coding:

- **Learn by doing:** Nobody can learn coding by reading books or tutorials. Practice, practice, and practice. We have many in-class coding exercises and assignments to help you establish a routine for practicing coding regularly.
- **Practice more often:** Like learning dancing or swimming, you will need frequent practice to overcome the initial steep learning curve. Avoid waiting until the last minute to complete assignments. Designating a regular block of time (e.g., one hour every other day) to go through lab notes, follow textbook tutorials, and complete assignments is a more efficient strategy.
- **Take advantage of study groups:** You will be part of a study group throughout the semester to help each other and complete group assignments. Group members are a great resource for debugging code, learning new R commands, and asking questions. Of course, when it is your turn, please contribute generously!

Required Textbooks and Resources

Required Textbook:

- Chang, W. (2018). *R Graphics Cookbook: Practical Recipes for Visualizing Data, 2nd Edition*. O'Reilly Media.
- Healy, K. (2018). *Data Visualization: A Practical Introduction*. Princeton University Press.
- Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for Data Science, 2nd Edition*. O'Reilly Media.

Recommended Books and Resources:

- Berinato, S. (2016). *Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations*. Harvard Business Press (or the updated 2023 edition).
- Cairo, A. (2012). *The Functional Art*. New Riders.
- Cairo, A. (2019). *How Charts Lie: Getting Smarter About Visual Information*. W.W. Norton & Company.

Grading

Lecture/ Lab Activities

#	Topic	Points
1	Intro to R, RStudio, & Tidyverse	1
2	Intro to Data & Data Structures	1
3	More R & Tidyverse	1
4	Intro to ggplot & Univariate Visualization	1
5	Bivariate Analysis	1
6	Bivariate Visualization	1
7	Data Sources	1
8	Data Cleaning	1
9	Themes, Facets, & Combining Graphs	1
10	Plot Axes	1
11	Color, Color Theory, & Accessibility	1
12	Visualizing Uncertainty	1
13	Visual Focus	1
14	Annotations, Labels, Legends, & Guides	1
15	Resume (Writing Center)	1
Total		**15**

Group Assignments

#	Assignment	Points
1	The Best and Worst of Data Visualization	10
2	Tabular Data Visualization	10
3	AI Client Simulation	10
Total		**30**

Final Project

#	Component	Points
1	Idea Pitch	7
2	Cleaned Dataset	7
3	Instagram Post	7
4	Infographic	7
5	AI Role Playing	7
6	Final Poster + Presentation	20
Total		**55**

Total: 100

Grading Scale

Range	Grade
93–100	A
88–92	AB
83–87	B
78–82	BC
70–77	C
60–69	D
Below 60	F

Course Schedule

**Indicates materials are due by the end of the week (Friday at 11:59 pm).*

***This schedule is tentative and subject to change by the instructor.*

Week 1: January 22, 2025

- Wednesday, January 22: Lecture Syllabus and Intro to Data Visualization
  - **Readings:**
    - R4DS: Welcome & Chapter 1 (Welcome - Introduction)
    - Healy: Chapter 1 (Look at Data)
  - *Why: Establish an overview of how data visualization underpins effective storytelling in media professions, setting the stage for ethical and impactful data-driven communication.*

Week 2: January 27–29, 2025

- Monday, January 27: Lecture Intro to R, RStudio, & Tidyverse*
  - **Readings:**
    - R4DS: Chapter 2 (Explore - Introduction)
    - R4DS: Chapter 4 (Basics)
    - Healy: Chapter 2 (Get Started)

  - *Why: Lay the groundwork for using core R workflows so you can handle large, real-world datasets with efficiency and clarity in professional settings.*

- Wednesday, January 29: Lecture Intro to Data & Data Structures*
  - **Readings:**
    * R4DS: Chapter 6 (Scripts)
    * RGC: Chapter 1 (R Basics)
  - *Why: See how R organizes and handles different data types so you can confidently work with varied file formats and structures that you may encounter in professional settings.*

Week 3: February 3–5, 2025

- Monday, February 3: Lecture More R & Tidyverse*
  - **Readings:**
    * R4DS: Chapter 5 (Data Transformation)
  - *Why: Strengthen your data cleaning and wrangling skills to create reliable datasets, a foundational step in producing trustworthy content and analysis.*
- Wednesday, February 5: Lecture Intro to ggplot & Univariate Visualization*
  - **Readings:**
    * R4DS: Chapter 3 (Data Visualization)
    * Healy: Chapter 3 (Make a Plot)
    * RGC: Chapter 2 (Quickly Exploring Data)
  - *Why: Learn the fundamentals of ggplots (data, aesthetics, layers) and how to create essential single-variable plots with ggplot, building a foundation for transforming raw data into clear, compelling visual stories for a range of stakeholders.*

Week 4: February 10–12, 2025

- Monday, February 10: Lecture Bivariate Analysis*
  - **Readings:**
    * R4DS: Chapter 7 (Exploratory Data Visualization)
  - *Why: Expand your ability to uncover meaningful relationships in data, an essential skill for developing evidence-based insights in media and strategic communication.*
- Wednesday, February 12: Lecture Bivariate Visualization*
  - **Readings:**
    * Healy: Chapter 4 (Show the Right Numbers)
    * RGC: Chapter 5 (Scatter Plots)
  - *Why: Translate those relationships into effective charts, helping you communicate nuanced trends and comparisons ethically and clearly.*

Week 5: February 17–19, 2025

- Monday, February 17: Lecture Data Sources*
  - **Readings:**

        - R4DS: Chapter 9 (Introduction to Wrangle)
        - R4DS: Chapter 10 (Tibbles)
        - R4DS: Chapter 11 (Data Import)
    - *Why: Practice importing data from various sources to ensure you can locate, load, and manage real-world datasets.*
- Wednesday, February 19: Final Project Idea Pitch (**Due before class**)

Week 6: February 24–26, 2025

- Monday, February 24: Lecture Data Cleaning*
    - **Readings:**
        - R4DS: Chapter 12 (Tidy Data)
        - R4DS: Chapter 13 (Relational Data)
        - R4DS: Chapter 14 (Strings)
        - R4DS: Chapter 15 (Factors)
        - R4DS: Chapter 16 (Dates and Times)
    - *Why: Transform messy data into a tidy, analyzable format by handling missing values, joining tables, and fixing structural issues.*
- Wednesday, February 26: Group Assignment The Best and Worst of Data Visualization*
    - *Why: Sharpen critical thinking by identifying different visual examples' strengths and weaknesses, refining your best practices to uphold ethical standards and avoid misleading audiences.*

Week 7: March 3–5, 2025

- Monday, March 3: Lecture Themes, Facets, & Combining Graphs*
    - **Readings:**
        - RGC: Chapter 9 (Controlling the Overall Appearance of Graphs)
        - RGC: Chapter 11 (Facets)
    - *Why: Polish the look of your plots with customizable themes and compare data across facets, to present complex data insights in a user-friendly way.*
- Wednesday, March 5: Final Project Cleaned Dataset & Dictionary*

Week 8: March 10–12, 2025

- Monday, March 10: Lecture Plot Axes*
    - **Readings:**
        - RGC: Chapter 8 (Axes)
    - *Why: Control and fine-tune plot axes for accurate scaling, labeling, and readability in your visualizations.*
- Wednesday, March 12: Group Assignment Tabular Data Visualization*

Week 9: March 17–19, 2025

- Monday, March 17: Lecture Color, Color Theory, & Accessibility*
    - **Readings:**

    - Healy: Chapter 8 (Refine Your Plots)
    - RGC: Chapter 12 (Using Colors in Plots)
  - *Why: Apply color thoughtfully to enhance audience comprehension and ensure your designs remain inclusive for all audiences.*

- Wednesday, March 19: Final Project Instagram Post*

Week 10: March 31–April 2, 2025

- Monday, March 31: Lecture Visualizing Uncertainty*
  - **Readings:**
    - RGC: Chapter 6 (Summarized Data Distributions)
  - *Why: Represent uncertainty in your data openly, fostering transparency and building trust in your data-driven storytelling.*
- Wednesday, April 2: Group Assignment AI Client Simulation*

Week 11: April 7–9, 2025

- Monday, April 7: Lecture Visual Focus*
  - **Readings:**
    - R4DS: Chapter 28 (Graphics for Communication)
  - *Why: Identify and highlight the most important elements in your plots, refining the overall impact of your data storytelling.*
- Wednesday, April 9: Final Project Infographic*

Week 12: April 14–16, 2025

- Monday, April 14: Lecture Annotations, Labels, Legends, & Guides*
  - **Readings:**
    - Healy: Chapter 5 (Graph Tables, Add Labels, Make Notes)
    - RGC: Chapter 7 (Annotations)
    - RGC: Chapter 10 (Legends)
  - *Why: Add clear, informative elements to your charts—such as text, legends, and highlights—to guide viewers seamlessly through the data.*
- Wednesday, April 16: Final Project Final Project AI Role Playing*

Week 13: April 21–23, 2025

- Monday, April 21: Final Project Final Project Peer Feedback
- Wednesday, April 23: Lecture Writing Center (Resume)

Week 14: April 28–30, 2025

- Monday, April 28: Final Project Final Project Instructor Feedback
- Wednesday, April 30: Final Project Poster Presentations (**Due before class**)

Course Policies

Readings

The assigned readings for each class period must be completed before class. While materials from the readings and lectures may sometimes feel redundant, this repetition is intentional. Since this course is skills-based, repeatedly practicing new skills is the most effective way of fully comprehending the materials. The course moves fast—completing the readings before lecture is the best way to maintain progress.

Assignments

All assignments are due on the designated dates. Keep a copy of your assignments for your records, and remember that it is your responsibility to ensure the instructor receives your submissions. On assignment due dates, come to class prepared to discuss the material covered in the assignment. If you struggled with an assignment, be ready to ask questions. Assignments turned in after the due date will be docked 10% of the points per day it is late.

Attendance and Participation

Regular attendance is essential for this course. You will develop an original visualization project throughout the semester, requiring consistent class attendance. Individual lecture/ lab assignments are due the day they are covered in class. Group assignments, which make up a significant portion of your grade, also necessitate regular interactions with group members. However, given the ongoing COVID-19 pandemic and related challenges, occasional absences may be allowed on a case-by-case basis.

Usage of Audio Recorded Lectures

Lecture materials and recordings for J677 are protected intellectual property of UW-Madison. Students may use these materials and recordings for personal use related to participation in this class. Students are not authorized to record lectures without explicit permission unless they qualify for accommodations under university policies. Unauthorized use of lecture materials, including sharing or selling them, constitutes copyright infringement and may result in penalties as outlined in university policies (UWS Chapters 14 and 17).

Students with Disabilities

Students with a McBurney VISA requiring special accommodations should notify the instructor by the end of Week 2 to facilitate arrangements. If other accommodations are needed, please inform the instructor as soon as possible.

Academic Dishonesty

All work in this course must be original and completed individually unless otherwise stated. This class adheres to UW-Madison guidelines on scholastic misconduct. If academic misconduct occurs, the minimum penalty will be a zero for the assignment, with a report filed to the Office of Student Conduct and Community Standards. Additional penalties may apply depending on the severity of the offense.

Familiarize yourself with what constitutes academic misconduct, as outlined in Wisconsin State Legislature UWS 14.03: https://docs.legis.wisconsin.gov/code/admin_code/uws/14/03.

Turnitin software will be used to evaluate assignments for originality and potential plagiarism. Assignments will be compared to various sources, including publications and databases. The instructor will review the similarity scores to assess potential misconduct.

Creating an Inclusive Learning Environment

This class may occasionally discuss challenging or sensitive topics. All participants are expected to remain respectful of individual differences, including but not limited to gender, age, race, religion, sexual orientation, national origin, and disability. Recognize your own biases and privilege, and strive to create a "safe" space for diverse perspectives. Statements of hate or intolerance will not be tolerated.

Course Policy on Generative Artificial Intelligence (AI)

Students can use generative AI tools (e.g., ChatGPT, DALL-E) to support course objectives, such as exploring code solutions or testing visualizations. However, students are responsible for ensuring the accuracy and ethicality of AI-generated content. Use of AI tools must be documented transparently and cited properly.

You may not submit AI-generated text, images, graphs, or code as your own work. Instead, use these tools to support your learning and articulate your own knowledge of the material.

For more information, refer to UW-Madison policies on generative AI: https://ctlm.wisc.edu/self-serve-resources/generativeai/.

How to Document AI Usage:

- Clearly describe how AI was used (e.g., for translation, grammar editing, research methods, data analysis).
- Cite the AI tool and version (e.g., ChatGPT-3.5, GPT-4) in the text and reference list.