

Welcome!

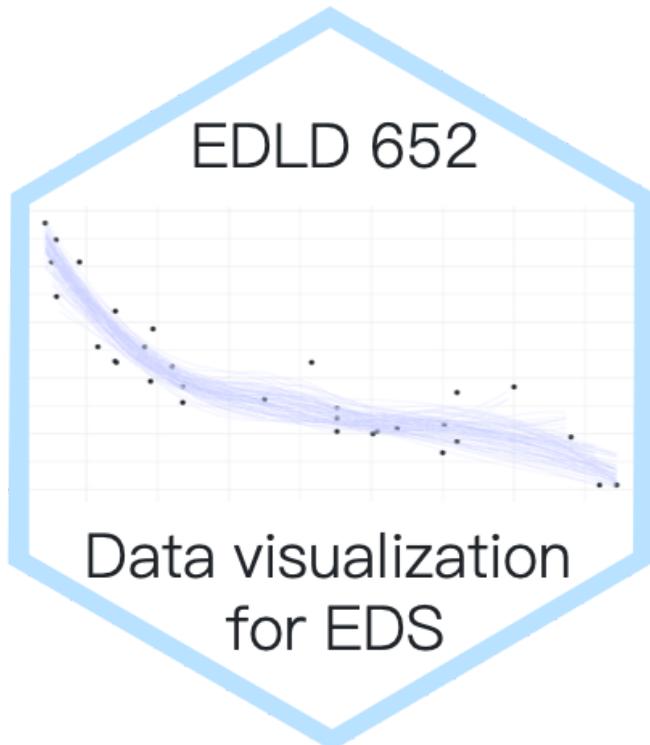
An overview of the course

Maithreyi Gopalan

Week 1

Agenda

- Getting on the same page
- Syllabus
- Accessing and working with the course data
- If time allows - Intro to text data



whoami

- Associate Professor
- Pronouns: she/her/hers
- Primary areas of interest:
educational equity,
policy analysis, social
psychology, systemic
inequities in
opportunities and
achievement, and
variance between
educational institutions
- Big secret: I am not a ❤️
 ❤️ R ❤️ ❤️ ; have been a
SAS and STATA user, so
learning R with y'all!



whoisyou?

- Introduce yourself
- Why are you here?
- What pronouns would you like us to use for you for this class?
- What was one thing you did not related to academic work over winter break?

A few class policies

- Be kind
- Be understanding and have patience, with others **and yourself**
- Help others whenever possible

Truly the most important part of this class. Important not just in terms of decency, but also in your learning, and most importantly, for equity.

A more specific policy

Kiddos in class

- All breastfeeding babies are welcome in class as often as necessary.
- Non-nursing babies and older children are welcome whenever alternate arrangements cannot be made. As a parent of two young children, I understand that babysitters fall through, partners have conflicting schedules, children get sick, and other issues arise that leave parents with few other options.

- In cases where children come to class, I invite parents/caregivers to sit close to the door so as to more easily excuse yourself to attend to your child's needs. Non-parents in the class: please reserve seats near the door for your parenting classmates.
- All students are expected to join with me in creating a welcoming environment that is respectful of your classmates who bring children to class.

In-person class

In-person class

- This class is in-person
- Your class participation grade comes exclusively from your active participation in the class through discussions and hands-on lab sessions
- If you are not feeling well, please do not attend in person
- See syllabus for attendance policy

Last intro thing

- I'm here for you
- We won't have specific office hours, but know I'm always willing to meet
- This course, like all in the sequence, can be difficult. Don't suffer in silence. Don't do this alone.

Syllabus

Course Website(s)

website

repo



[Home](#)



[Schedule](#)



[Assignments](#)



[Syllabus](#)



[Tags](#)

Data Visualization

for educational data science

Welcome to the second course in the [Educational Data Science Specialization](#) taught at the University of Colorado College of Education. This course will be taught through [R](#), a free and open-source statistical software package that provides users with the foundational principles and practice of data visualization, particularly for education data. We will have weekly lectures, covering a wide variety of topics including the history and principles of visual design. We will also cover mediums for communication across diverse different web applications. Weekly hands-on laboratory sessions provide students the opportunity to apply these concepts into practice.

Materials

- Nearly everything will be distributed through the repo and through the website.
- Please clone the repo now, if you haven't already. Pull each week for the most recent changes.
- We'll use Canvas for grading, and that is essentially it.

R Markdown notes

- These slides were produced with `{xaringan}`, an R Markdown variant. I encourage you to try it out and use it for your final project presentation.
- The website was also produced with R Markdown (sort of)
 - It's a `{blogdown}` website with some custom CSS and Hugo shortcodes
- This course is not just about data viz, but also mediums for communication. This includes websites and `data dashboards` among other possibilities.

My
assumptions
about you

I assume you

- Understand the R package ecosystem (how to find, install, load, and learn about them)
- Can read "flat" (i.e., rectangular) datasets into R
 - I don't care what you use, but you should be using RStudio Projects & the `{here}` package
 - See Jenny Bryan's blog post for why.

- Can perform basic data wrangling and transformations in R, using the tidyverse
 - Leverage appropriate functions for introductory data science tasks (pipeline)
 - "clean up" the dataset using scripts and reproducible workflows
- Use R Markdown to create reproducible dynamic reports
- Use version control with R via git and GitHub
- Indeed, all of today's lab is going to be about that!

Learning objectives

- Transform data in a variety of ways to create effective data visualizations
- Understand best practices in data visualization
- Create and customize graphics in a variety of ways using best practices (e.g., visual perception, color choices, text annotations, categorical axis ordering, uncertainty)
- Build web-based platforms for sharing data visualizations

Examples

Below are some links to final projects from students who have taken this class previously.

Dashboards

- Songyi Ahn
- Sophia Angleton
- Maiko Hata

Blog post

- Teresa Chen
- Ouafaa Hmaddi
- Murat Kezer

Weekly learning objectives

Provide you a frame for what you should be working to learn for that specific week.

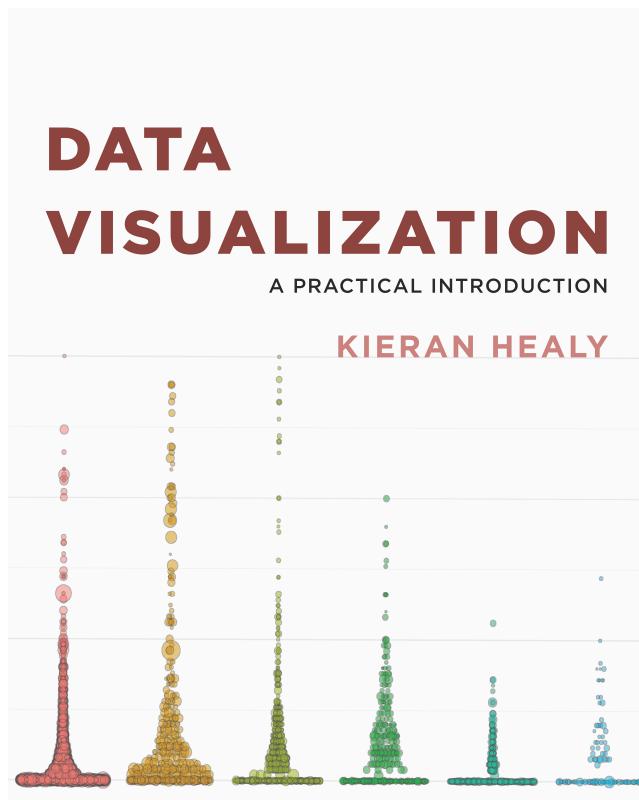
This week's objectives

- Understand the requirements of the course
- Understand the requirements of the final project
- Be ready to go with *git* and GitHub

Required Textbooks

(free)

Healy

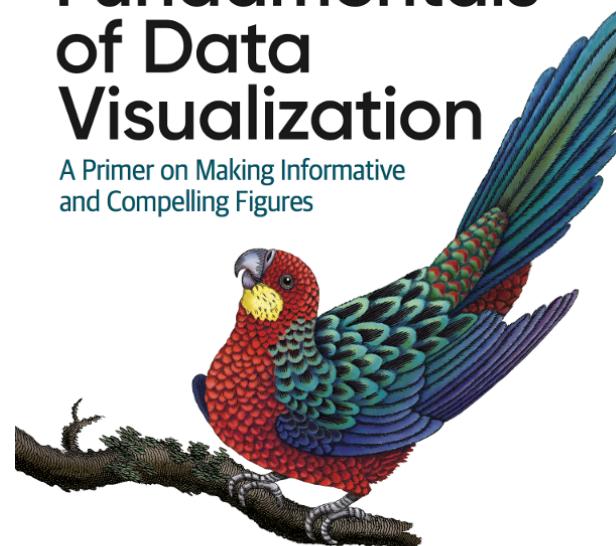


Wilke

O'REILLY®

Fundamentals of Data Visualization

A Primer on Making Informative
and Compelling Figures



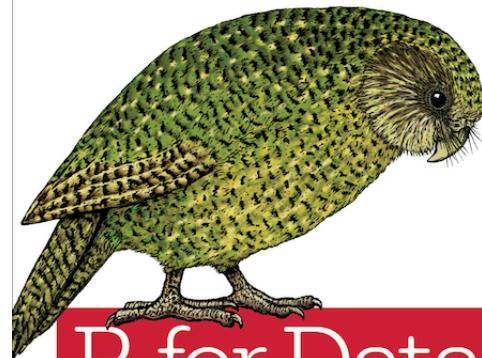
Claus O. Wilke

Other books (also free)



Bryan

O'REILLY®



R for Data
Science

VISUALIZE, MODEL, TRANSFORM, TIDY, AND IMPORT DATA

Hadley Wickham &
Garrett Grolemund

Wickham & Grolemund

Extra credit opportunity

5 points: Deep dive into a topic not covered by the course

Some options

- Geographic data (we will discuss this, but it's relatively late and there's a ton we won't be able to get to)
- Network data
- DAGs
- Flow data (e.g., alluvial diagrams)
- Interactive plots
- Animated plots

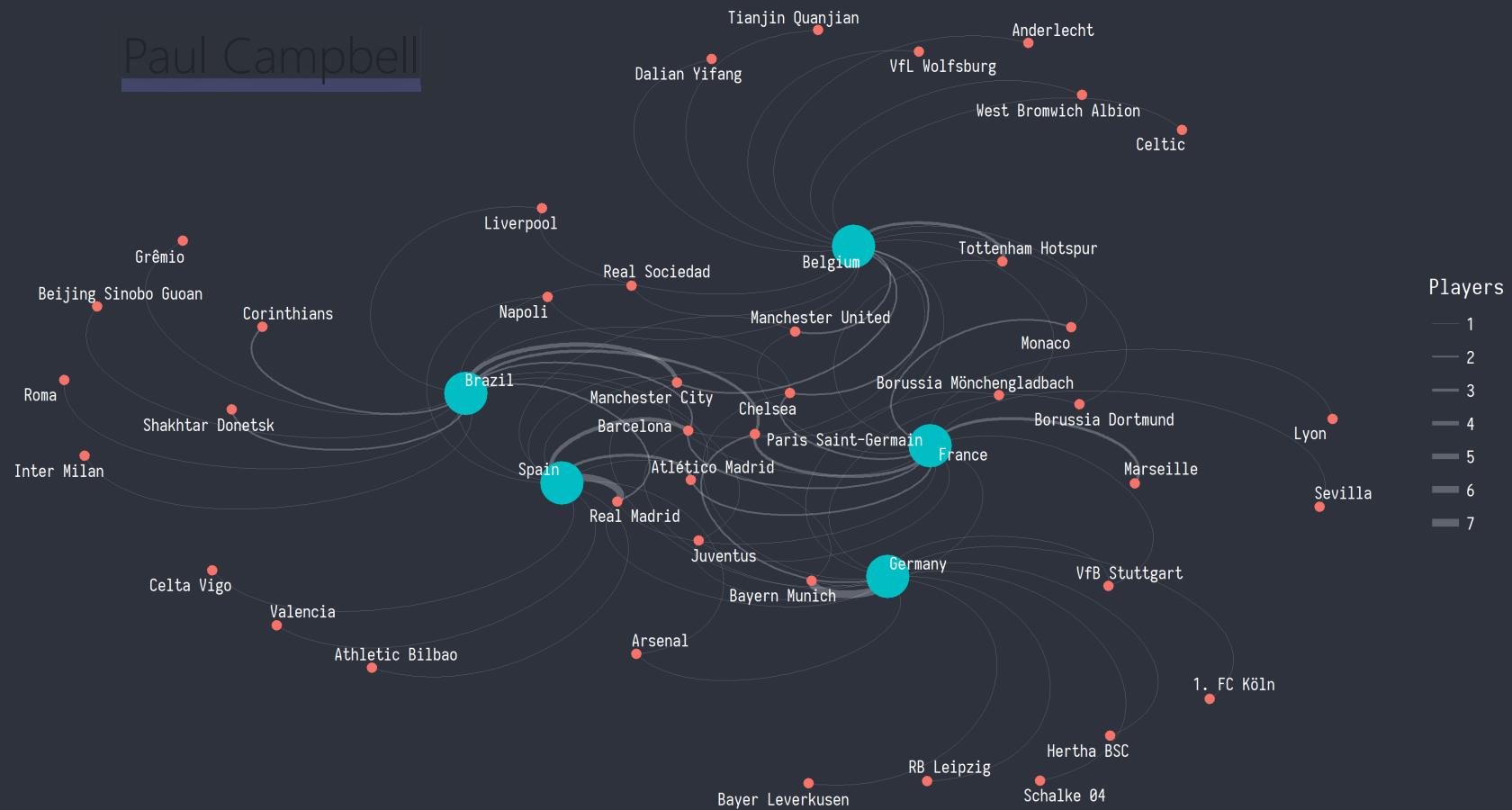
Some examples

Timo Grossenbacher

World Cup 2018 | Club Country Network

Belgium, Brazil, France, Germany, Spain

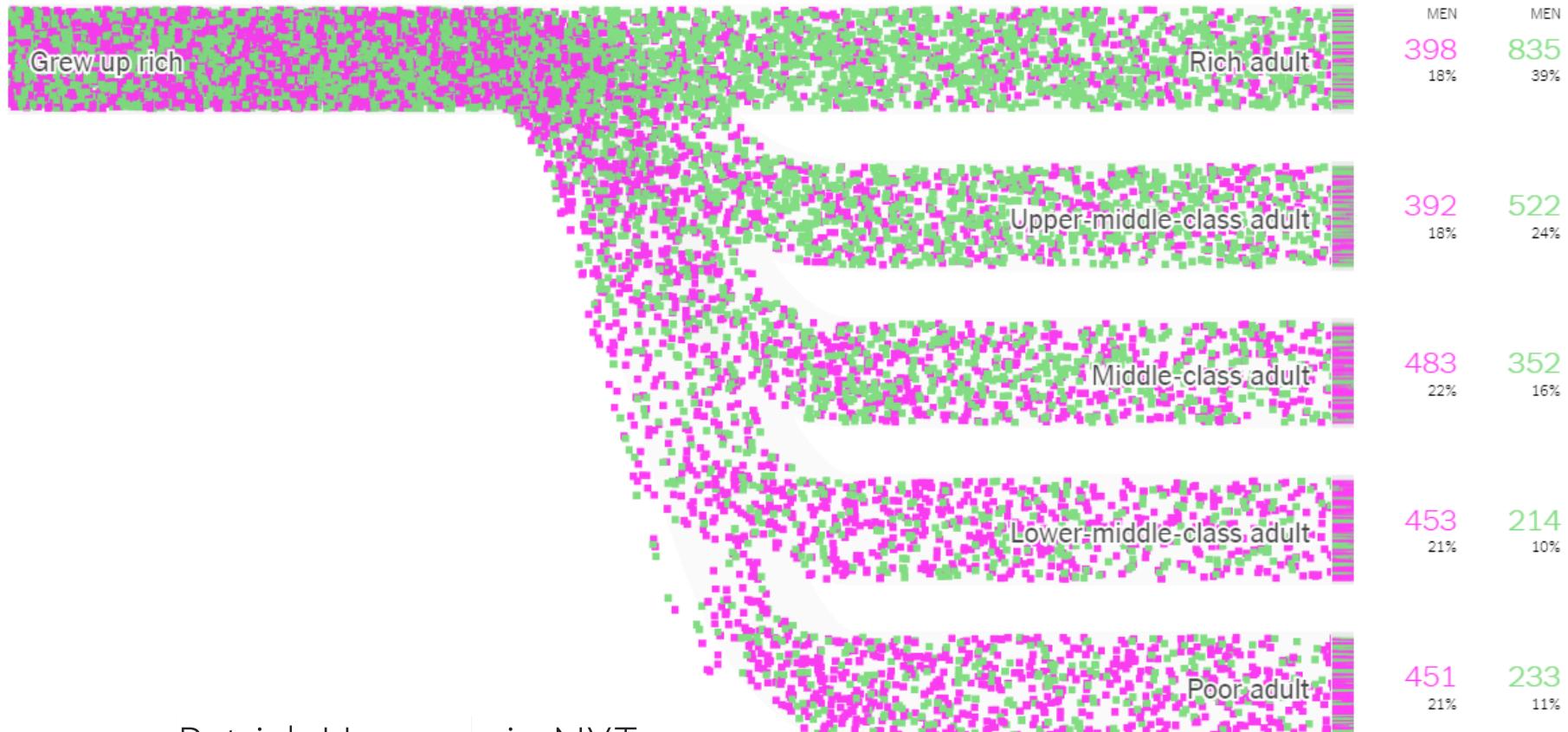
Paul Campbell



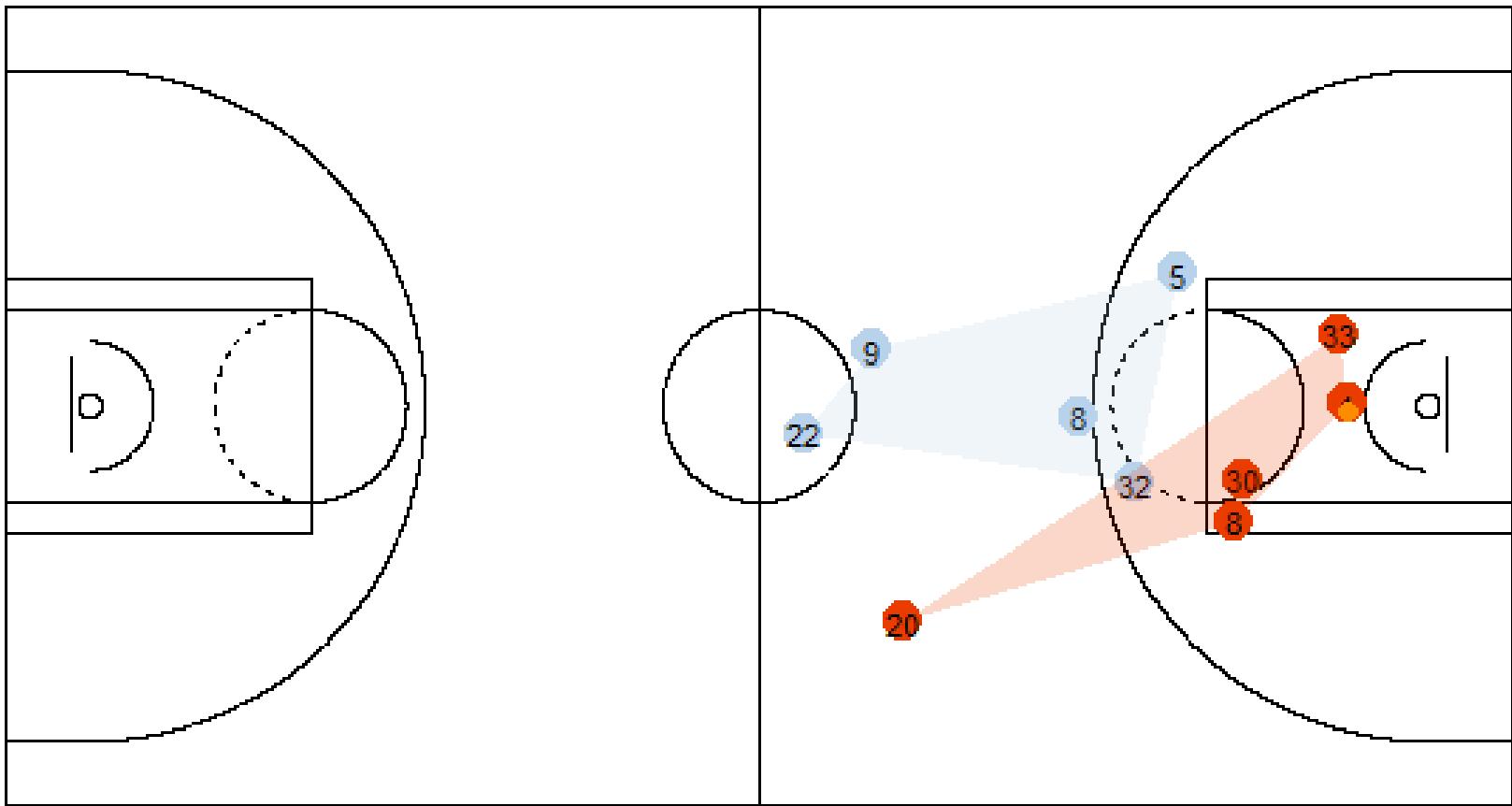
@paulcampbell91 | Source: Wikipedia

Black and white boys raised in wealthy families

Follow the lives of these **17,195** Americans and see where they end up as adults:



Patrick Honner via NYT



James Curley

Labs

See the assignments page of the website.

Data viz "in the wild" presentations

Everyone will be randomly assigned a date to share two data visualizations you have found in publications, websites, or anywhere else IRL.

- Not a formal presentation
- Share the links with me before class - we'll look at it as a group and discuss
- You note where you found it and what you like/dislike about it

Presentation order

Date	Presenter
2026-01-14	Kat
2026-01-14	Ziyuan
2026-01-14	James
2026-01-21	Nishat
2026-01-21	Steven
2026-01-28	Febe
2026-01-28	Cheyna
2026-02-04	Chen
2026-02-04	Tongle
2026-02-11	Everett
2026-02-11	Dodjivi

Date	Presenter
2026-02-18	America
2026-02-18	Rachel
2026-02-25	Erika
2026-02-25	Ramtin
2026-03-04	Kyla
2026-03-04	Isha

I will email this out as well.

Final Project

70 points total (35%)

Group project

- Please try to finalize your group by the end of today. You will have time when exploring the course data to work together.
- No fewer than 2, no more than 3.
- Although the final is the only mandated group project, I encourage you to work with your group for all labs and the lab assignments as well.

Three parts

- Proposal (20 points)
- Presentation (20 points)
- Product (30 points)

Product

Four components:

- A web-deployed portfolio showcasing your #dataviz skills.
 - `distill` (what I'll lecture on), `R Markdown`, or `blogdown` website, or `Quarto`
 - Technical document with `pagedown` or `bookdown`
 - Scientific poster with `pagedown`
 - `flexdashboard`

- At least four finalized data displays, with each accompanied by a strong narrative/story, as well as the history of how the visualization changed over time.
- Housed on GitHub
 - Fully reproducible
- Deployed through GitHub pages (or netlify or similar)

Proposal

Four components:

- Show me some evidence that you've at least played around with some data and that you have some ideas of what you want to do
- *Very* preliminary visualizations, and/or hand-sketches of visuals you'd like to make, noting the data sources/columns to be used
- Identification of the intended audience for each viz
- The intended message to be communicated for each viz

Main point - feedback!

Draft

- Expected to still be a work in progress
 - Data visualizations should be largely complete
- Deployment not expected
- Can be provided to your peers so they can learn from you as much as you can learn from their feedback

Peer Review

- We are all professionals here. It is imperative we act like it.
- Understand the purpose of the exercise.
- Zero tolerance policy for inappropriate comments
- Should be vigorously encouraging

Utilizing GitHub

- Fork their repo, embed comments & suggest changes to their code, submit a PR

Presentation

Order randomly assigned. Basically a chance to share what you created!

- Presentation length will be determined later, but likely to be in the 10-15 minute range (note - you will present as a group)
- Share the final products
- Share the prior iterations
- Discuss the progression along the way and why specific changes were made
- What challenges did you face along the way? What victories did you have that you are particularly proud of?

Grading

Lower percent	Lower point range	Grade	Upper point range	Upper percent
0.970+	(146 pts or more)	A+		
0.930	(140 pts)	A	(145 pts)	0.969
0.900	(135 pts)	A-	(139 pts)	0.929
0.870	(131 pts)	B+	(134 pts)	0.899
0.830	(125 pts)	B	(130 pts)	0.869
0.800	(120 pts)	B-	(124 pts)	0.829
0.770	(116 pts)	C+	(119 pts)	0.799
0.730	(110 pts)	C	(115 pts)	0.769
0.700	(105 pts)	C-	(109 pts)	0.739
		F	(104 pts or less)	0.699



COMMUNICATION

Questions?

Let's jump into a quick Github Workflow refresher

Next time

- Quick refresher of R Markdown/R and visualizations of continuous variables
- If time permits, we will also discuss text data, string manipulations, distributions/binning