

Welcome!

An overview of the course

*Daniel Anderson
Week 1, Class 1*



Agenda

- Getting on the same page
- Syllabus
- A little bit of git/GitHub



A whale is breaching the ocean surface, its body arched out of the water. Its mouth is wide open, likely performing a 'bubble net' or 'gaping' maneuver. The whale's dark grey skin is visible against the bright blue sky and the darker blue of the ocean. The water is slightly choppy around the whale.

WHALE HELLO THERE

We're a bit of an odd mix





So let's help each other out!

What's so odd?

Before Fall '18

- 4 credits
- Only R

During Fall '18

- 3 credits
- R + git/GitHub

What's so odd?

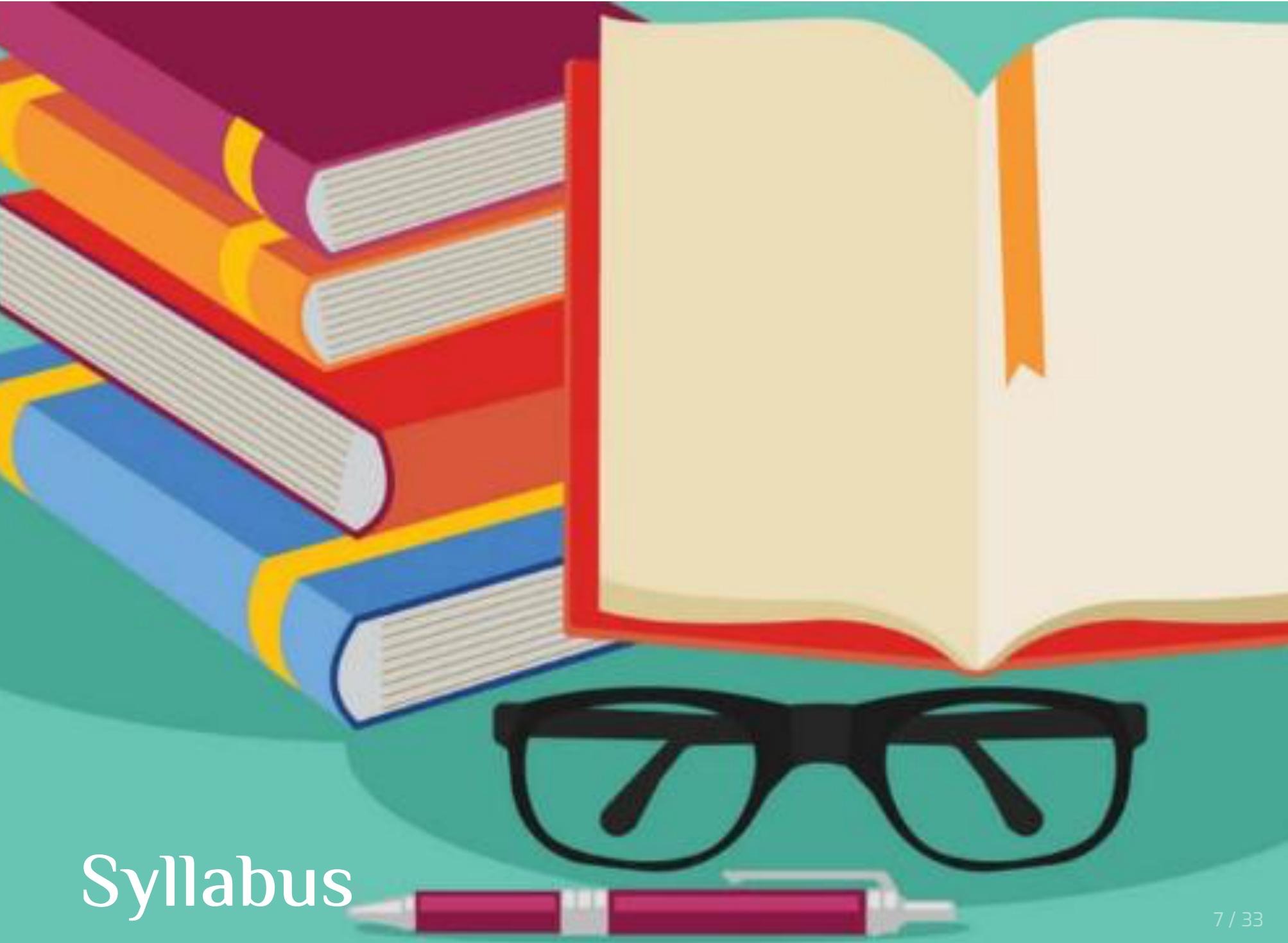
Before Fall '18

- 4 credits
- Only R

During Fall '18

- 3 credits
- R + git/GitHub



The background features a vibrant illustration of several books stacked together. One book is open, showing blank white pages. In front of the books, a pair of black-rimmed glasses rests on a teal surface. A pink and white striped pen lies next to the glasses. The overall theme is education or learning.

Syllabus

Course Website(s)

- [website](#)
- [repo](#)

Communicating and Transforming Data

This is the second course in a sequence of courses that will eventually lead to a *data science in educational research* specialization. This course will be taught through R (<https://cran.r-project.org>), a free and open-source statistical computing environment. This course, in particular, will rely heavily on building on the first course using RStudio. This course will give students a foundation in the principles and practice of data visualization, particularly as applied to scientific and technical data. We will have weekly lectures, covering a wide variety of topics including human perception, color theory, principles of visual design, etc. We will also have weekly hands-on laboratory sessions in which students will have the opportunity to put the lecture material into practice. Often when creating effective data visualizations, it is necessary to combine and transform different data sources, and this will be a considerable focus of the course as well.



(<http://creativecommons.org/licenses/by-nc/4.0/>)

GitHub

If you haven't already:

- Get *git* installed on your computer
- Register for a GitHub account

We'll be using it all term long.

I assume you:

1. Understand the R package ecosystem (how to find, install, load, and learn about them)

I assume you:

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

I assume you:

1. Understand the R package ecosystem (how to find, install, load, and learn about them)
2. Can read "flat" (i.e., rectangular) datasets into R
 - I don't care what you use, but you should be using RStudio Projects & I highly recommend [{here}](#)
 - See [Jenny Bryan's blog post](#) for why.
3. 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

I assume you:

1. Understand the R package ecosystem (how to find, install, load, and learn about them)
2. Can read "flat" (i.e., rectangular) datasets into R
 - I don't care what you use, but you should be using RStudio Projects & I highly recommend [{here}](#)
 - See [Jenny Bryan's blog post](#) for why.
3. 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
4. Use version control with R via git and GitHub
 - I am, of course, there will need to be some catching up here

I assume you:

1. Understand the R package ecosystem (how to find, install, load, and learn about them)
2. Can read "flat" (i.e., rectangular) datasets into R
 - I don't care what you use, but you should be using RStudio Projects & I highly recommend [{here}](#)
 - See [Jenny Bryan's blog post](#) for why.
3. 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
4. Use version control with R via git and GitHub
 - I am, of course, there will need to be some catching up here
5. Use R Markdown to create reproducible dynamic reports

Learning objects for this class

- Transform data in a variety of ways to create effective data visualizations

Learning objects for this class

- Transform data in a variety of ways to create effective data visualizations
- Understand and fluently apply different types of data joins

Learning objects for this class

- Transform data in a variety of ways to create effective data visualizations
- Understand and fluently apply different types of data joins
- Customize ggplot2 graphics by reordering factors, creating themes, and applying ggthemes

Learning objects for this class

- Transform data in a variety of ways to create effective data visualizations
- Understand and fluently apply different types of data joins
- Customize ggplot2 graphics by reordering factors, creating themes, and applying ggthemes
- Create reproducible slides using ioslides/xaringan

Learning objects for this class

- Transform data in a variety of ways to create effective data visualizations
- Understand and fluently apply different types of data joins
- Customize ggplot2 graphics by reordering factors, creating themes, and applying ggthemes
- Create reproducible slides using ioslides/xaringan
- Create an online data visualization portfolio using radix and/or flexdashboards to demonstrate key learning

Weekly learning objectives

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

Weekly learning objectives

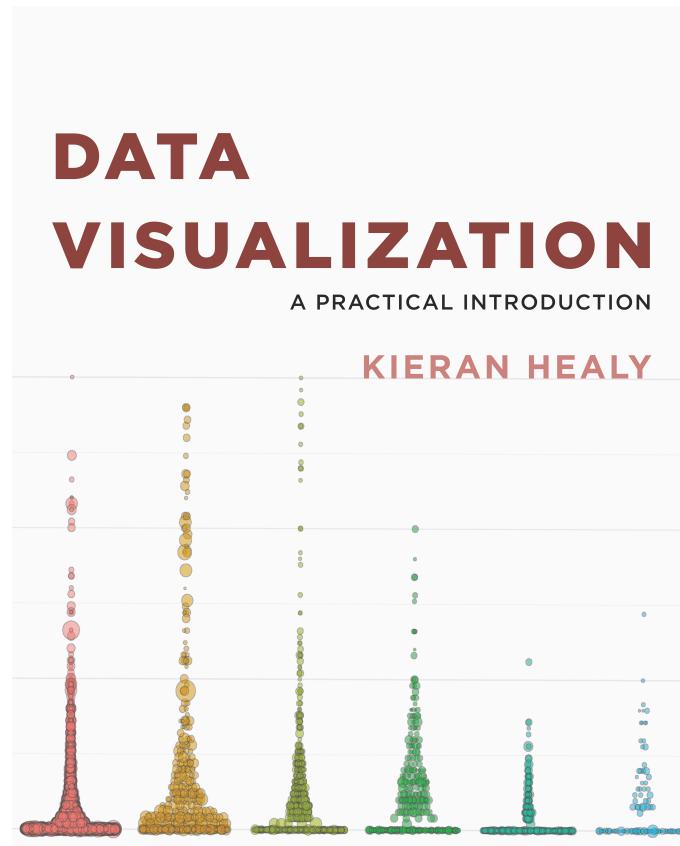
Provide you a frame for what you should be working to learn for that specific week.

This week's objectives

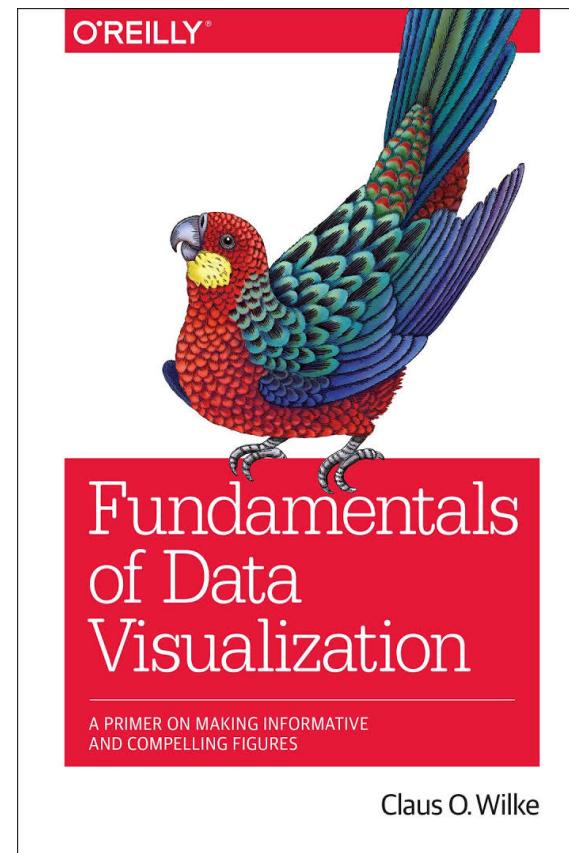
- Understand the requirements of the course
- Understand the requirements of the final project
- Be ready to go with *git* and GitHub

Required Textbooks (free)

Healy (<http://socviz.co>)



Wilke
(<https://serialmentor.com/dataviz/>)

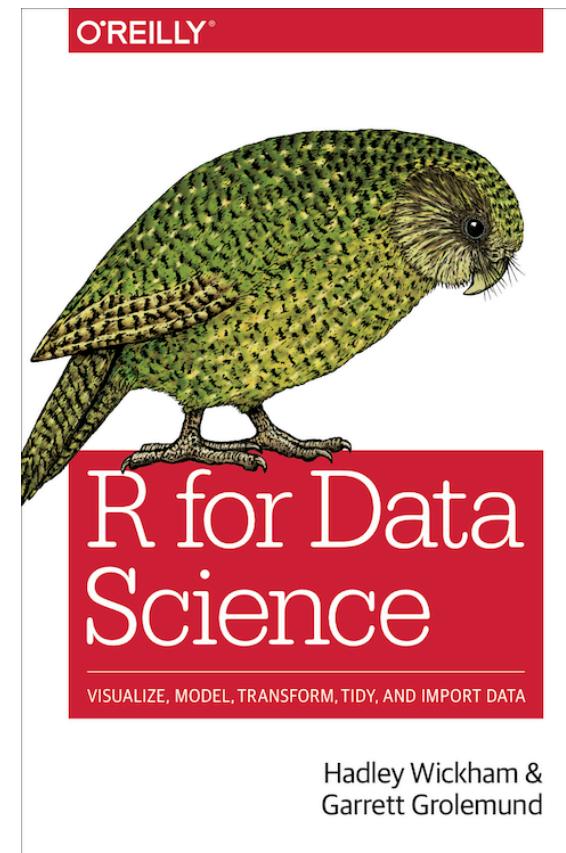


Other books (also free)

Bryan (<http://happygitwithr.com>)

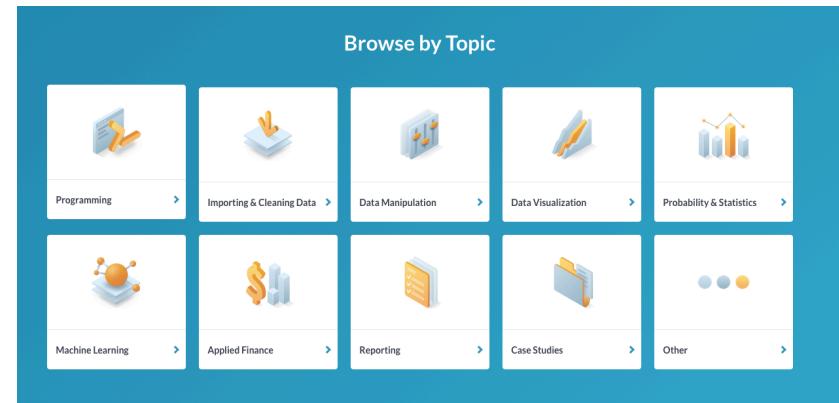
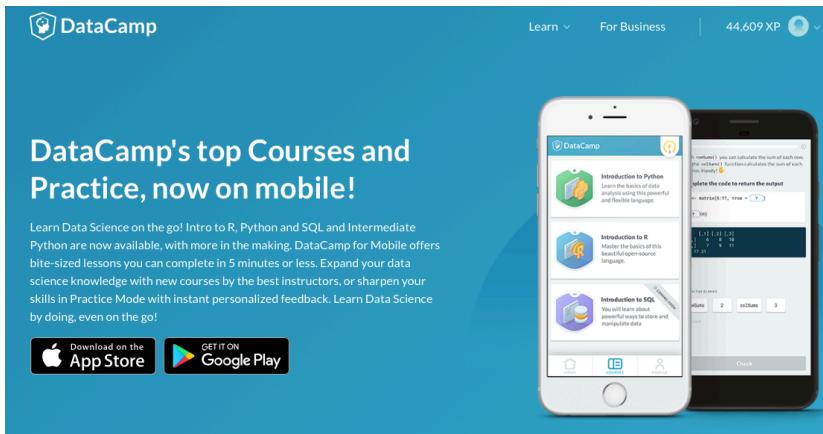


Wickham & Grolemund
(<https://r4ds.had.co.nz>)



DataCamp

- Everyone should have received a datacamp invite this morning. Contact me if not.
- Great platform for supplemental learning.
- We'll use it as part of your homework



Two required modules

5 points each

Due before class next week

- Data Visualization with ggplot2 (Part 1)

Due before class, January 21st

- Visualization Best Practices in R

Two elective modules

5 points each

Pick two

Data Visualization with ggplot2 (2)	Working with Geospatial Data in R
Data Visualization with ggplot2 (3)	Interactive Maps with leaflet in R
Data Visualization in R	Visualizing Time Series Data in R
Data Visualization in R with ggvis	Communicating with Data in the Tidyverse
Interactive Data Visualization with plotly in R	Network Science in R - A Tidy Approach
Interactive Data Visualization with rbokeh	Visualizing Big Data with Trelliscope

Extra credit opportunities

- 5 points: Do one extra DataCamp module (one of the electives, or get approval for a different one)
- 10 points: Deep dive into a topic not covered by the course

Extra credit opportunities

- 5 points: Do one extra DataCamp module (one of the electives, or get approval for a different one)
- 10 points: Deep dive into a topic not covered by the course

Some options for 2

We'll talk about this more in a minute

- Geographic data
- Network data
- Flow data (alluvial diagrams)
- Relational data (SQL & friends)
- Interactive plots
- Animated plots
- Text data, text nets

Lab Schedule

5 points each (30 points total; 15%)

1. GitHub collabo (**Assigned: 1/9/19; Due: 1/14/19**)
2. Mapping data to aesthetics (**Assigned: 1/16/19; Due: 1/21/19**)

Homework 1

3. Visual perception (**Assigned: 1/30/19; Due: 2/4/19**)
4. Color (**Assigned: 2/6/19; Due: 2/11/19**)

Homework 2

5. Tables/Uncertainty (**Assigned: 2/20/19; Due: 2/25/19**)
6. Sharing & Web deployment (**Assigned: 2/27/19; Due: 3/4/19**)

Homework

10 points each (20 points; 10%)

- Basically the same as the labs, but scored correct/incorrect, and no in-class time devoted to them.
- Okay to work on collaboratively - I actively encourage you to do so as long as you're using a shared repo

Homework

10 points each (20 points; 10%)

- Basically the same as the labs, but scored correct/incorrect, and no in-class time devoted to them.
- Okay to work on collaboratively - I actively encourage you to do so as long as you're using a shared repo

First will be focused on joins and basically relational data (with visualizations built in)

Second will ask you to reproduce some data visualizations "in the wild" with publicly available data.

Final Project

120 points total (60%)

6 parts:

- Proposal (10 points): Due 2/6/19
- Draft (15 points): Due 2/27/19
- Peer review (15 points): Assigned, 2/27/19; Due 3/6/19
- Presentation (20 points): 3/11/19 and 3/13/19 (Week 10)
- Product (60 points): Due 11:59:59 PM, 3/20/19

Product

Four components:

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

Proposal

Four components:

- Description of the data source (**must** be publicly available)
- Preliminary ideas (even hand sketches) of different visualizations
- Identification of the intended audience for each visualization
 - Note, you might consider displaying the same data/relations more than once, with each plot displayed for a different audience.
- The intended message to be communicated for each plot.

Draft

- Expected to still be a work in progress
 - Data visualizations should be largely complete
- Deployment not expected
- 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

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

- You'll be assigned three proposals to review (5 points each)
- Fork their repo
- Embed comments, suggest changes to their code
- Submit a PR
 - Summarize your overall review in the PR

Presentation

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

- Discuss what is trying to be communicated
- Share the final products
- Discuss the progression along the way and why specific changes were made

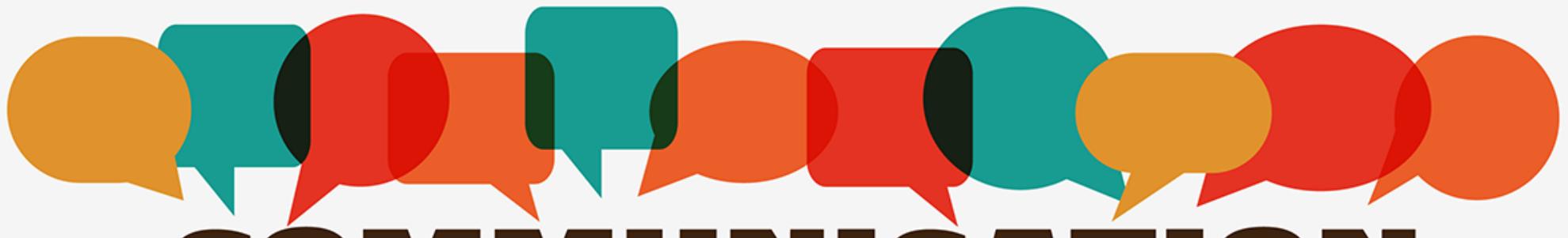
Grading

200 points total

- 6 labs at 5 points each (30 points; 15%)
- 2 homework assignments at 10 points each (20 points; 10%)
- 4 DataCamp modules at 5 points each (20 points; 10%)
- five-minute data visualization "in the wild" presentation (10 points; 5%)
- Final Project (120 points; 60%)
 - Proposal (10 points; 5%)
 - Draft (15 points; 7.5%)
 - Peer review (15 points; 7.5%)
 - Presentation (20 points; 10%)
 - Product (60 points; 30%)

Grading

Lower percent	Lower point range	Grade	Upper point range	Upper percent
0.97	(194 pts)	A+		
0.93	(186 pts)	A	(194 pts)	0.97
0.90	(180 pts)	A-	(186 pts)	0.93
0.87	(174 pts)	B+	(180 pts)	0.90
0.83	(166 pts)	B	(174 pts)	0.87
0.80	(160 pts)	B-	(166 pts)	0.83
0.77	(154 pts)	C+	(160 pts)	0.80
0.73	(146 pts)	C	(154 pts)	0.77
0.70	(140 pts)	C-	(146 pts)	0.73
		F	(140 pts)	0.70



COMMUNICATION

Missing classes



I will be out **all** of next week!

- Trying desperately to get a guest speaker - most people I can think of will be at the conference with me.
- May end up just having to have you do the readings and work on a take-home lab

Missing classes

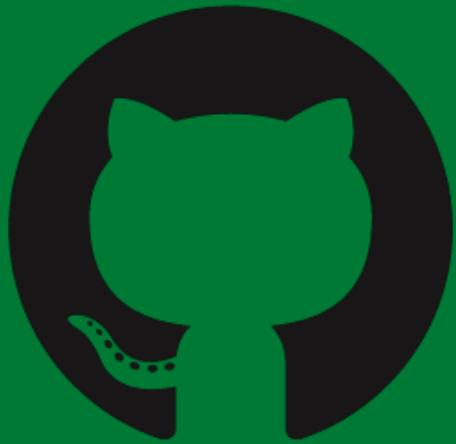


I will be out **all** of next week!

- Trying desperately to get a guest speaker - most people I can think of will be at the conference with me.
- May end up just having to have you do the readings and work on a take-home lab

February 6

- I will also be out February 6 (but should be able to be here on the 4th)
- I *think* I have a guest lecture lined up for that day, who will guide you through the lab.



GitHub

Demo

- Creating a GitHub repo
- Sharing access (or creating an organization)
- Cloning the repo
- stage, commit, push
- pull
- forking and issues
- The [gitkraken GUI](#)

Next time

Collaborating with GitHub lab