Welcome to QuACK!

Week 3

7/16/2025

Nina, Sophie, and Victoria

Structure of the bootcamp

Session 1 (6/11)	Session 2 (7/2)	Session 3 (7/16)	Session 4 (7/23)
Intro to R and RStudio	Data cleaning and wrangling	Data visualization	Intro to stats with R

We will work with some of the same data throughout the bootcamp to get a feel for the process of exploring data and getting it ready for analysis

Sessions build on each other and we will pick up where we left off in the last session – you will get the most out of the bootcamp if you attend all four weeks!

UC Berkeley

Today's Agenda: Plotting

- Warm-up activity
- Intro to ggplot2
- Demo: data visualization

UC Berkeley

Week 3 Warm-up – 10 minutes

1. Download this week's materials from this link: https://tinyurl.com/quack-summer
2. Open the practice PDF from **last week** (session 2)
3. If you did the practice exercises during last week's session, refresh your memory on what you did and what the answers were
4. If you didn't get a chance last week, try the practice exercises now!

UC Berkeley

When should you visualize your data?

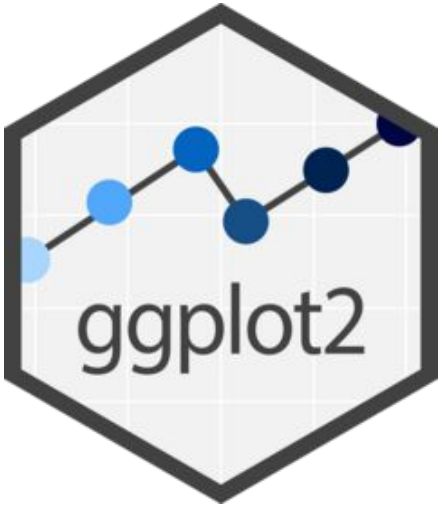
UC Berkeley

Ok…but why?

- Helps to easily identify trends in your data – maybe some you expected, and others you didn't expect
- Helps find outliers in your data – these may be throwing off your stats without you even knowing!
- In some cases, it can help you decide which statistical tests are appropriate for the kind of data you have
- Can help you easily communicate your findings to collaborators and the broader scientific community

UC Berkeley

Introducing ggplot2



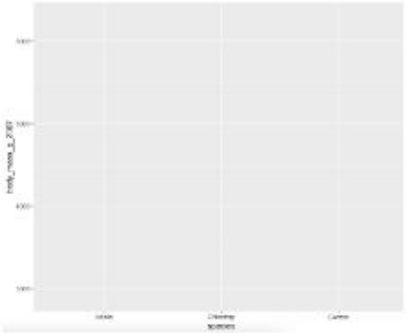
- Last week we learned about Tidyverse. We used some Tidyverse packages like `readr` and `dplyr` to clean our data
- Today we are learning about another package in Tidyverse called `ggplot2` that is used for data visualization
- Why ggplot?
  - It's easy to use
  - It works well with the rest of Tidyverse (and R in general)
  - It creates very flexible and easily customizable plots

UC Berkeley

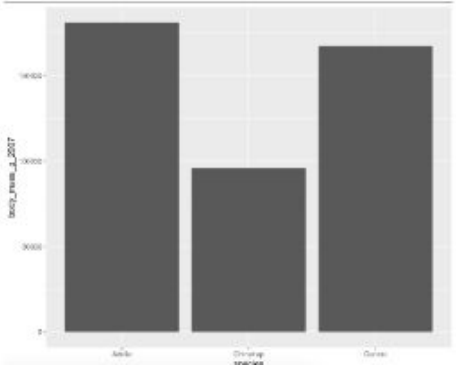
Build any plot with ggplot

- You can combine layers in ggplot2 to create a plot
- The way ggplot2 is so flexible is by offering many different layers, each with many different customization options

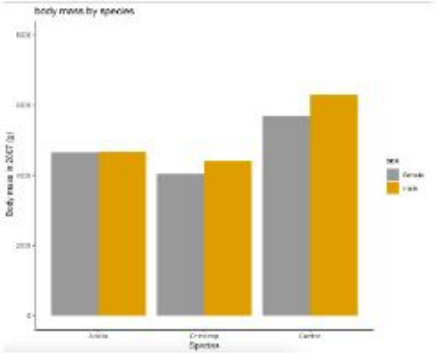
Step 1: create a plot and designate variables



Step 2: Decide how to represent the data



Step 3: customize



UC Berkeley

ggplot2 terminology

- **Creating a ggplot object (this is your "canvas")**

```
ggplot(data = myDataFrame, aes(x = x_series, y = y_series))
```

- **Adding geometric objects to your plot. These tell ggplot how to represent your data.**

```
ggplot(data = myDataFrame, aes(x = x_series, y = y_series))+
    geom_col()
```

Note that ggplot layers are connected by "+" not "%>%"

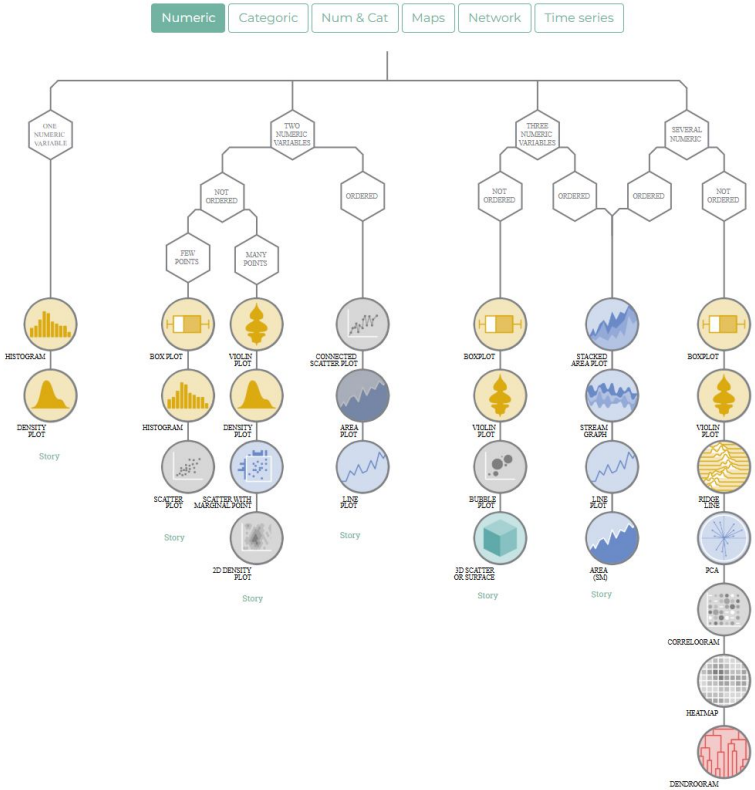
- **Add other layout and design features**

```
ggplot(data = myDataFrame, aes(x = x_series, y = y_series))+
    geom_col() +
    theme_classic()
```

UC Berkeley

And lastly, a helpful dataviz resource:

https://www.data-to-viz.com/



UC Berkeley

Time to get started in R!

- Go to https://tinyurl.com/quack-summer
- Download and unzip the course materials
- You will see many files:
  - Two R Markdown (.rmd) files
  - A .csv file
  - Two PDF files – slides and a practice doc
  - Some other folders – ignore these for now
- Open the R Markdown file marked “starter_code” – this is where we will code together for the first part of the session
- The .csv file contains the data we will be working with
- The practice PDF contains practice exercises for you to do during the second part of the session

UC Berkeley