Data Viz Considerations

Grayson White

Math 241

Week 1 | Spring 2026

Week 1 Goals

Day 1 Lecture

- Course overview

Day 2 Lecture

- Develop language to talk about the components of a graphic

- Discuss considerations for good graphical design

- Graphics consideration activity

- PSET 0 assigned

First: P-set 0

Let's start with the language we will use to describe the components of a graph:

The Grammar of Graphics

Background

Leland Wilkinson wrote a book called "The Grammar of Graphics"

As part of his PhD in Statistics at Iowa State, Hadley Wickham wrote the `R` package `ggplot2`, which we will use to create static graphs.
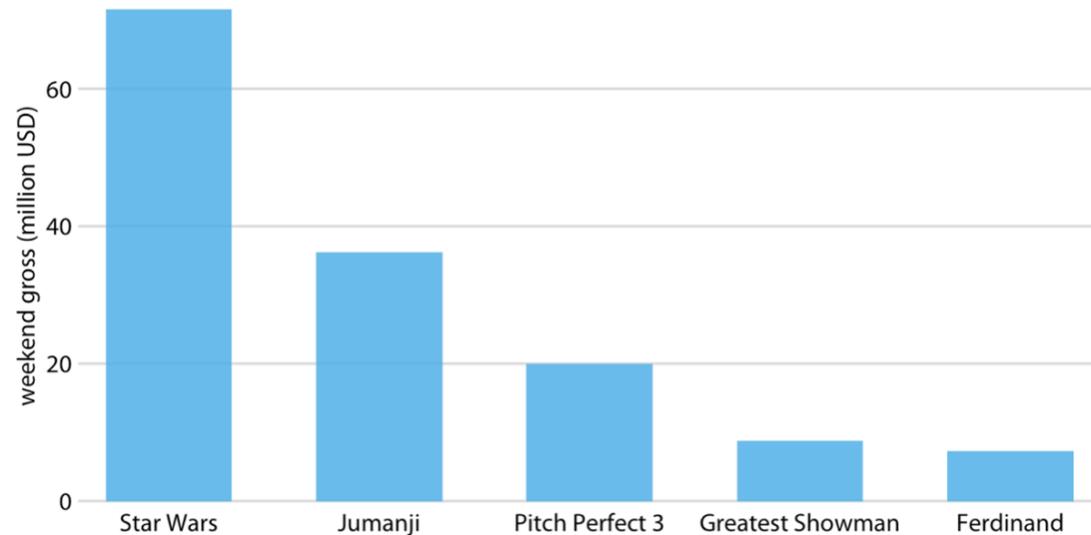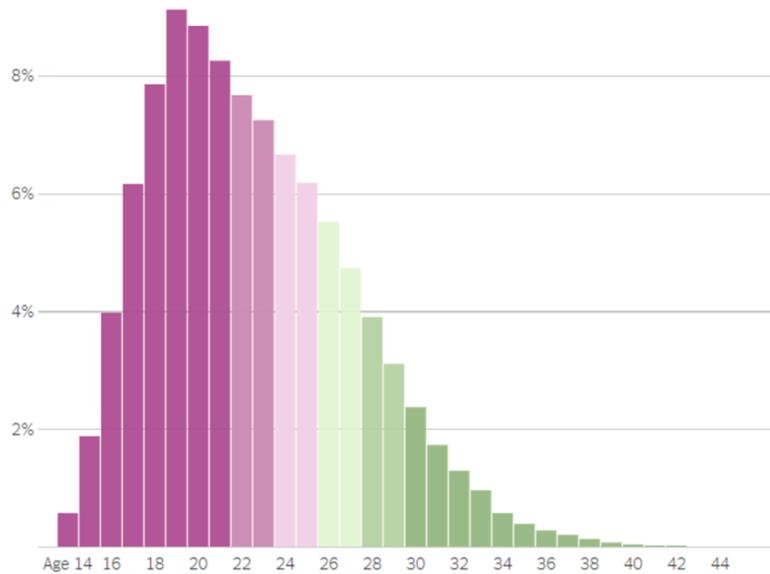
The Grammar of Graphics

- **data**: dataset that contains the data

- **geom**: geometric shape that the data are mapped to

  - point, line, bar, text, …

- **aes**thetic: visual properties of the **geom**

  - x position, y position, color, fill, shape

- **coord**: coordinate system

  - Cartesian, polar, geographic

- **scale**: controls how data are mapped to the visual values of the aesthetic

  - EX: particular colors, linear

- **guide**: legend to help user convert visual display back to the data

Geoms versus Names
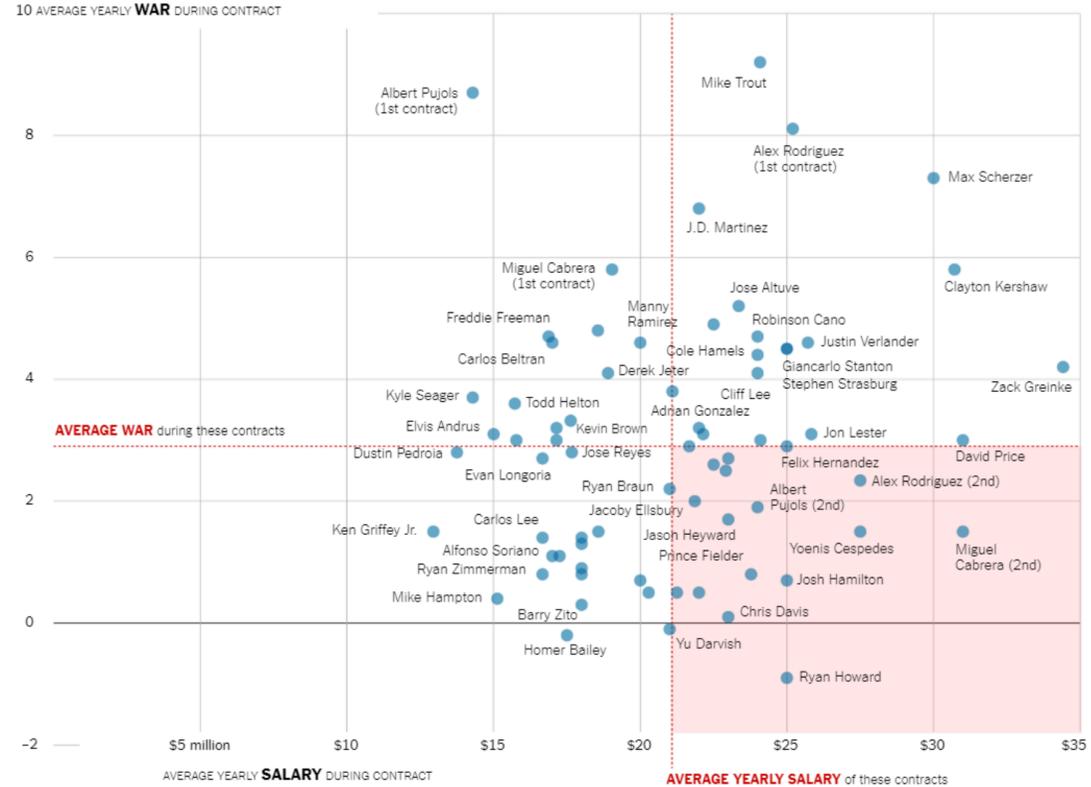
- What are the names of these graphs?

- Focus on the **shapes** (i.e. geoms) and how the variables are **mapped** to those shapes.

Data Viz Example

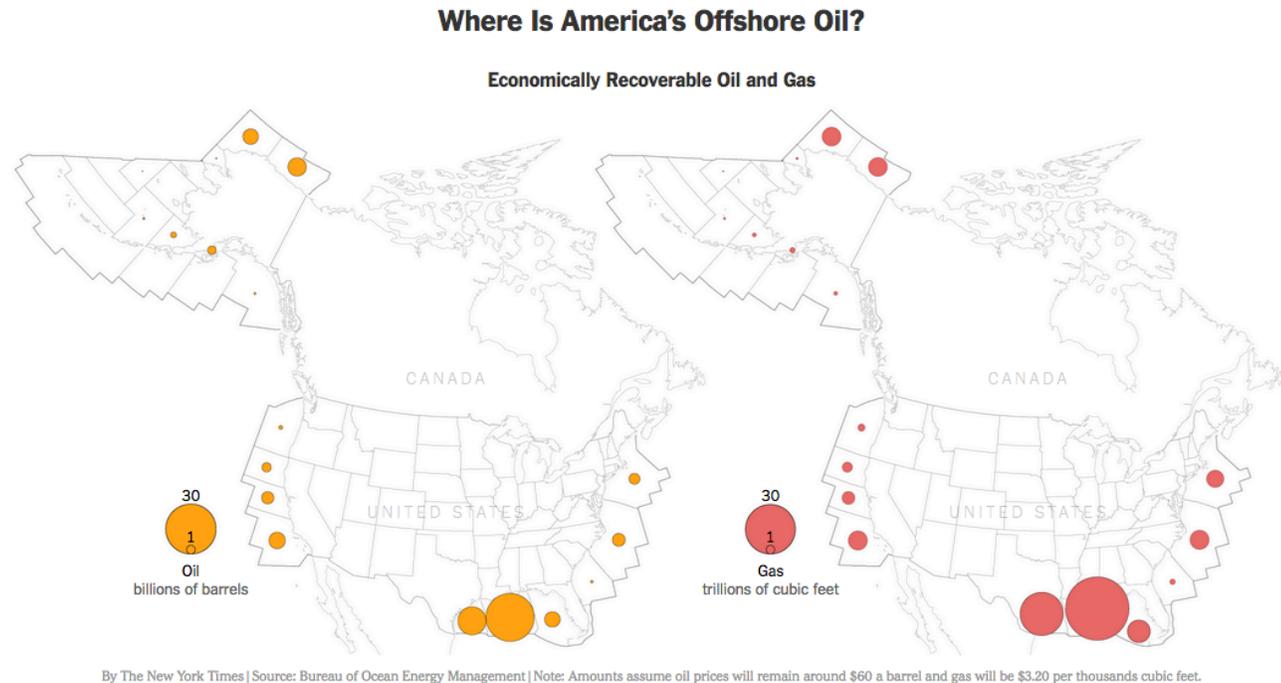Let's practice deconstructing this graph using the grammar of graphics.

- Geom(s)?
- **Aesthetic**s of the **geom**?
  - ▪ Mapping of variables?
- Coord?
- Scales?

Data Viz Example

Let's practice deconstructing this graph using the grammar of graphics.

- Geom(s)?

- **Aesthetic**s of the **geom**?

  - Mapping of variables?

- Coord?

- Scales?

Where Is America's Offshore Oil?

Economically Recoverable Oil and Gas

CANADA

UNITED STATES

30
1
Oil
billions of barrels

CANADA

UNITED STATES

30
1
Gas
trillions of cubic feet

By The New York Times | Source: Bureau of Ocean Energy Management | Note: Amounts assume oil prices will remain around $60 a barrel and gas will be $3.20 per thousands cubic feet.

Choices

For most data, there won't be just **one** way to graph it. Decisions to be made:

- What **geom** to use
  - Point, line, bar, …

- For a given geom, how to map variables to its **aesthetics**
  - Size, location, color, …

- For each **aesthetic**, what scale to use
  - Linear, diverging colors, …

- For the graph, what **coordinate system** to use
  - Cartesian, polar

Let's discuss some considerations that can help guide these decisions. But…

> "Data visualization is part art and part science. The challenge is to get the art right without getting the science wrong and vice versa." – Claus Wilke

Recommendation: Try out different options and make sure to iterate!

Consideration: Consider variable type when picking the aesthetic mapping.

What aesthetic options do I have at my disposal?
# Aesthetics: Position/location
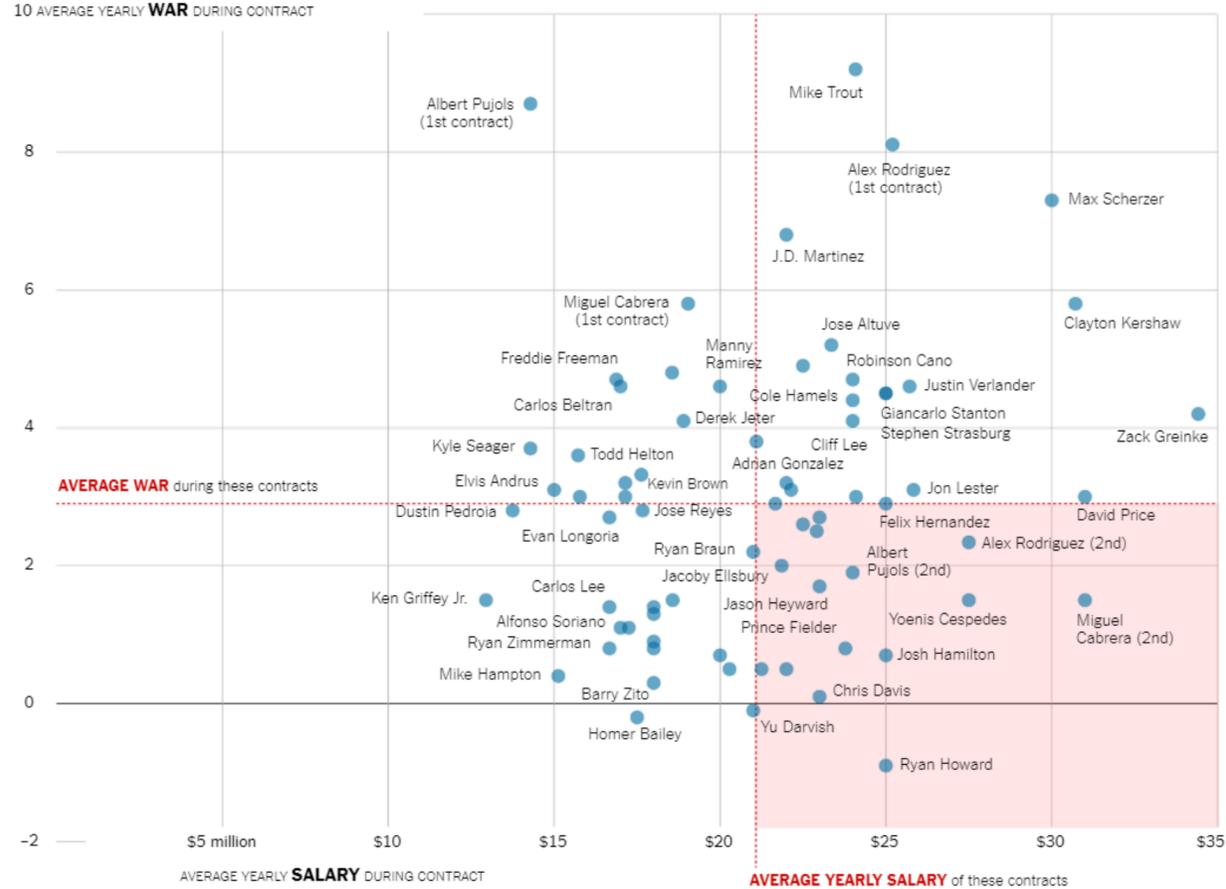
What aesthetic options do I have at my disposal?
Aesthetics: Length

GENDER SPLIT

The percentage of scientific papers with female authors seems to have grown over time. However, the trend is not clear because the algorithm used to assign genders was only able to classify fewer than one-third of authors' names in earlier years — many items were signed using initials — and only three-quarters of names in the current era.

Percentage of female authors
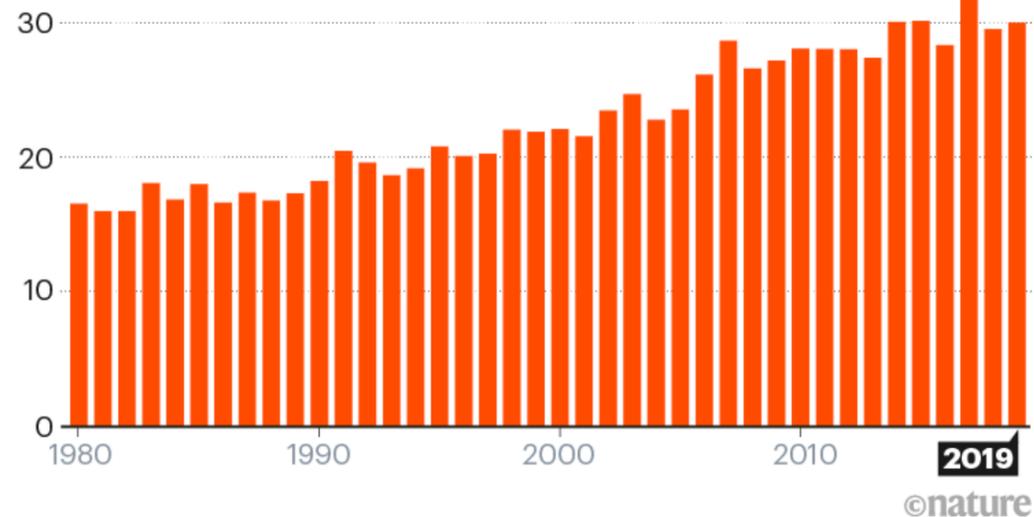
Design: Jasiek Krzysztofiak; Data analysis: Digital Science Consultancy

What aesthetic options do I have at my disposal?
Aesthetics: Area

What aesthetic options do I have at my disposal?
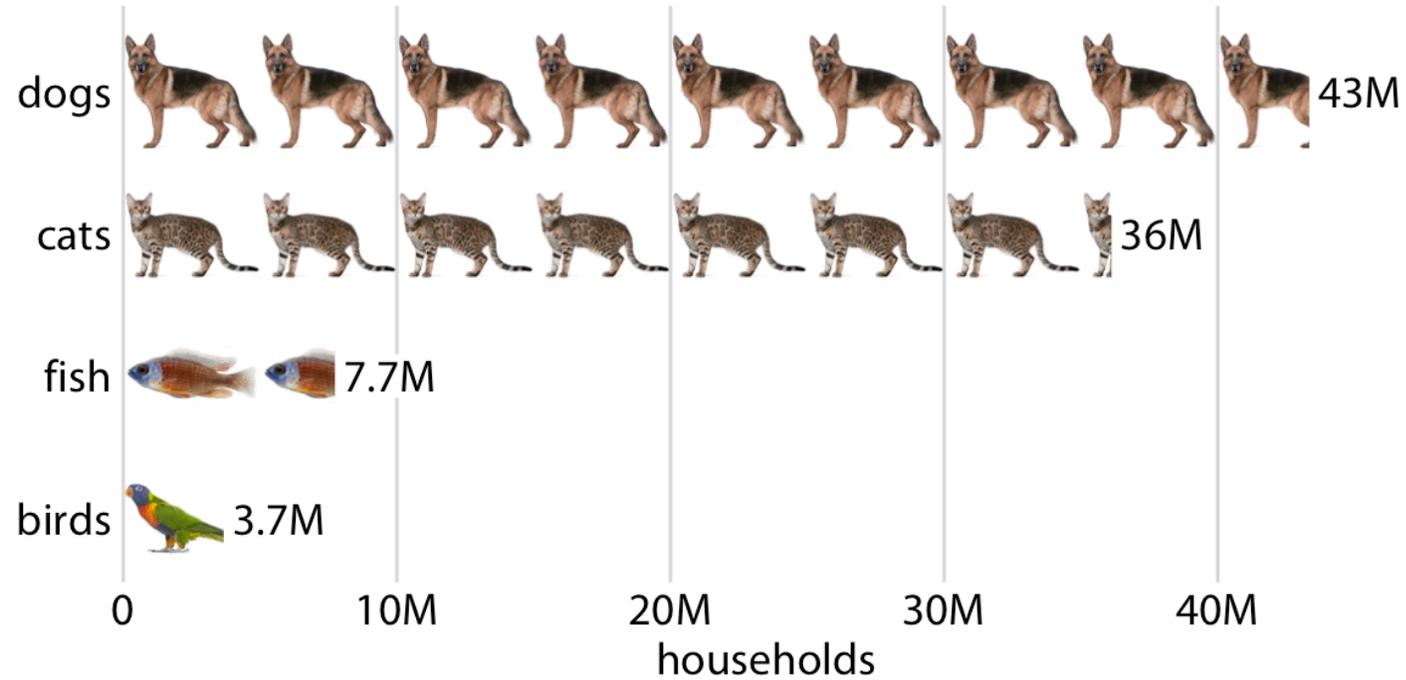## Aesthetics: Angle

Basset Hound
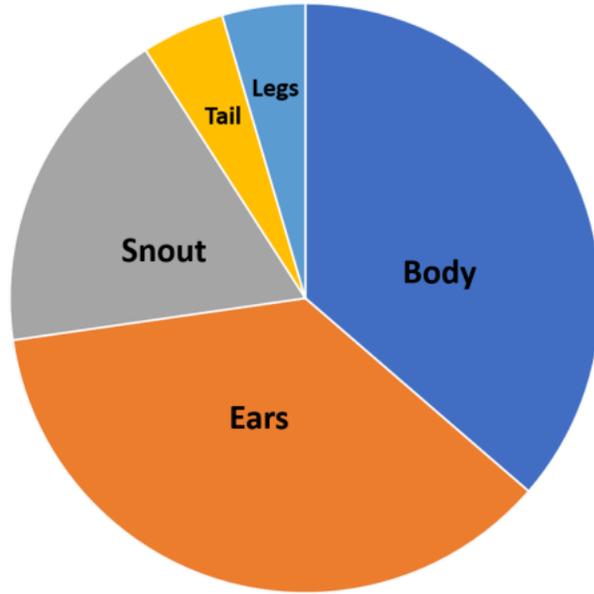
Daschund
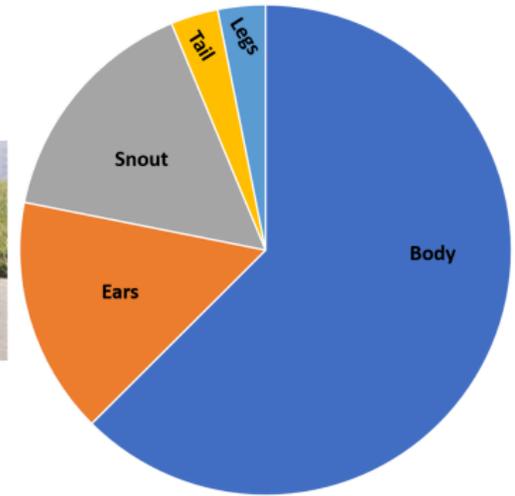
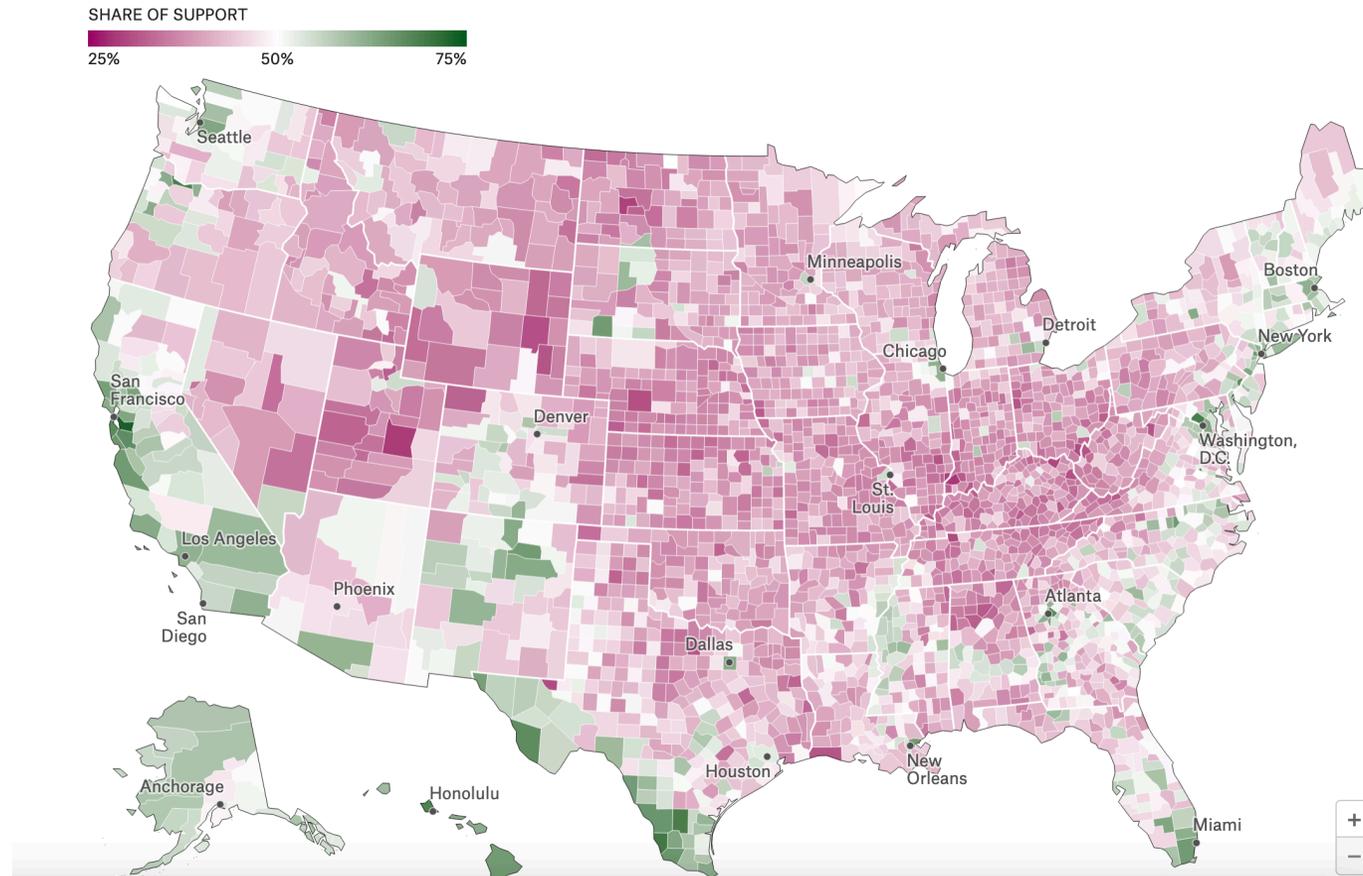# What aesthetic options do I have at my disposal?
Aesthetics: Shapes

What aesthetic options do I have at my disposal?
Aesthetics: Color Shade

Americans in the hardest-hit regions want climate change to be a priority
Estimated share of adults who say global warming should be a high priority for the next president and Congress, by county

SHARE OF SUPPORT

25%   50%   75%

What aesthetic options do I have at my disposal?
Aesthetics: Color Hue

What Young People Think is the Most Important Factor in Determining Success
How 15- to 24-year-olds in each country responded.
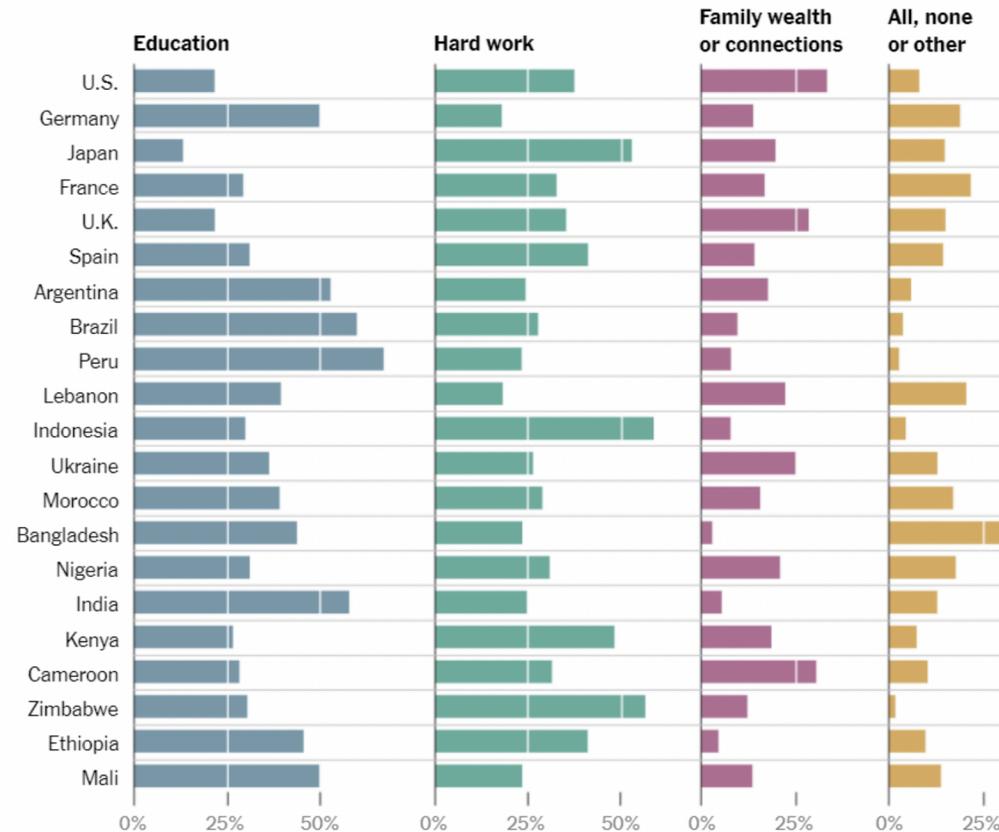
Note: Countries are sorted from richest to poorest according to a measure of **gross national income per capita** used by the World Bank. • Source: UNICEF-Gallup survey of more than 21,000 people conducted by telephone in 21 countries between February and June 2021.

What aesthetic options do I have at my disposal?
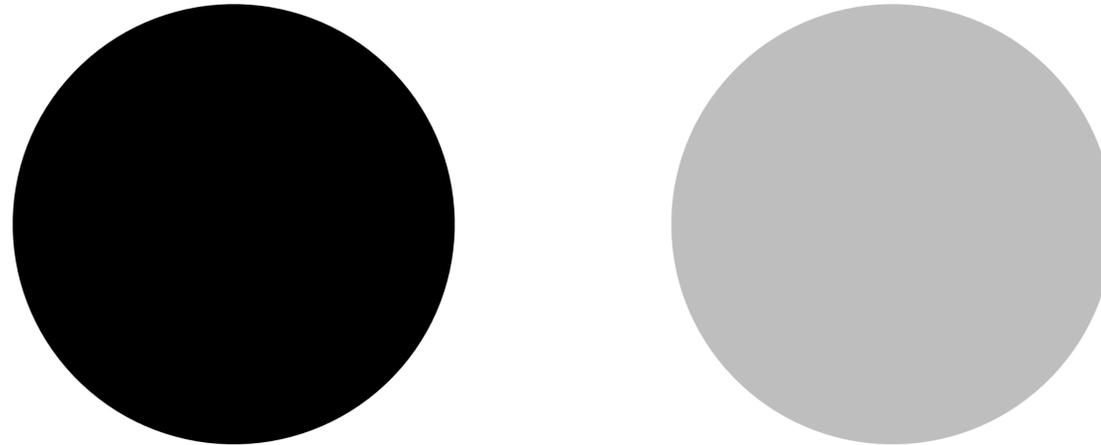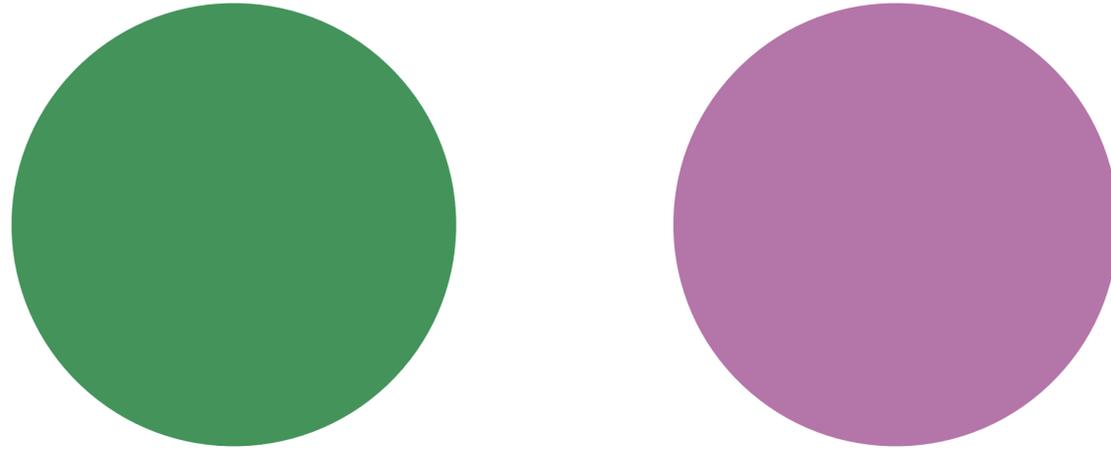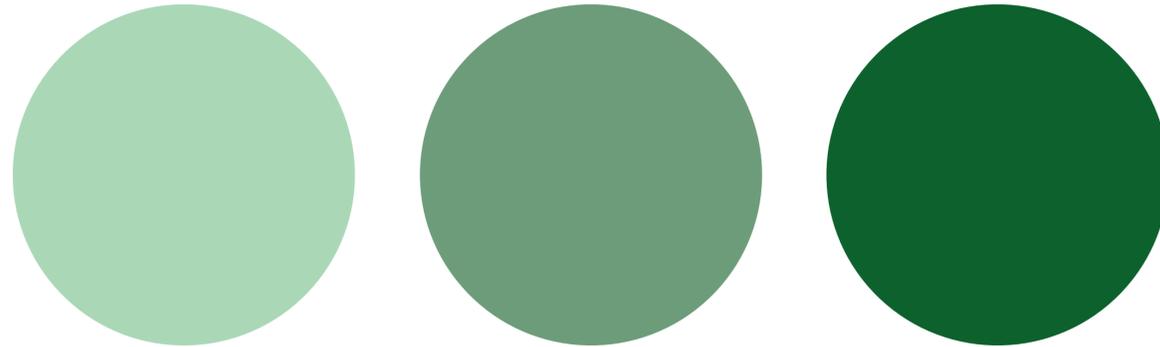# Which represents the larger value?
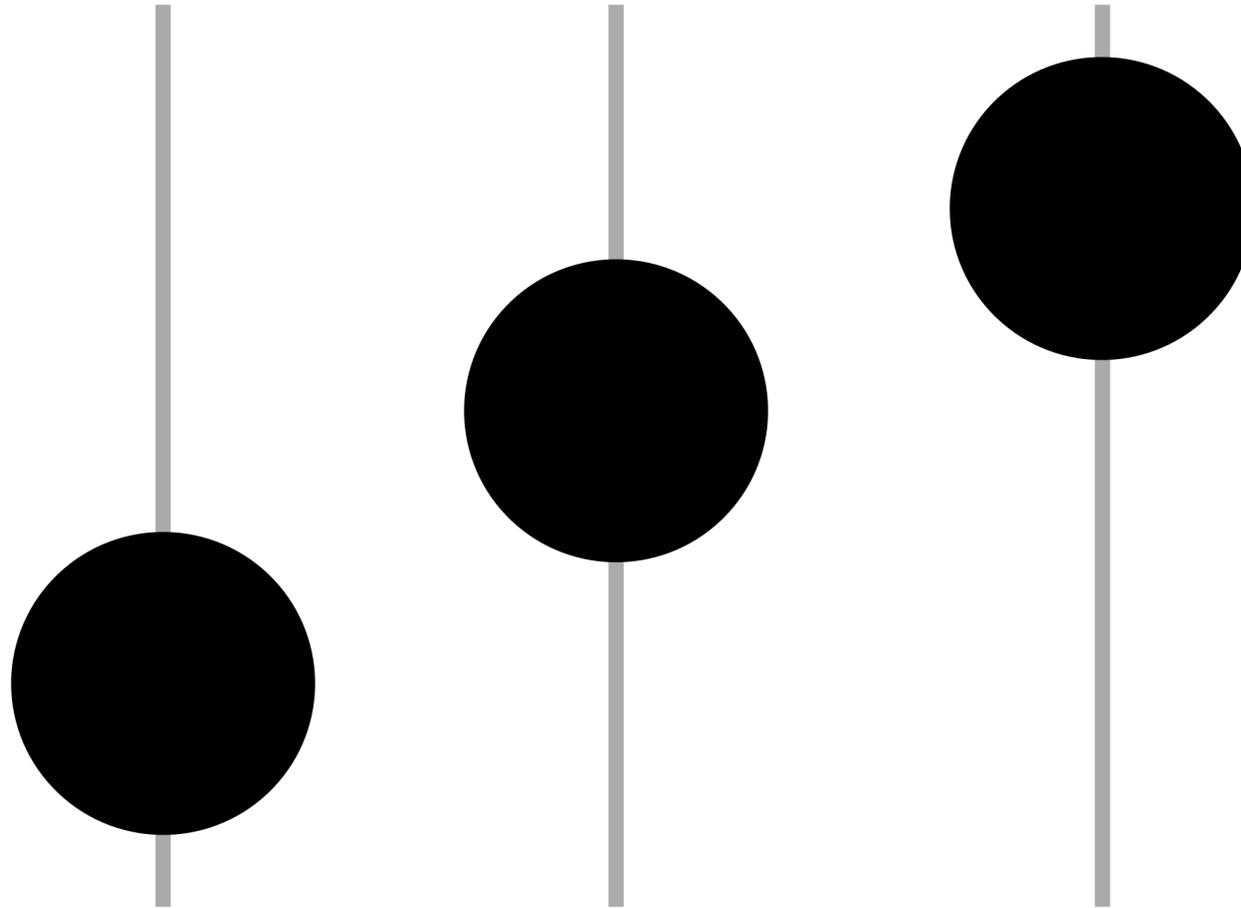
What aesthetic options do I have at my disposal?
Which represents the larger value?
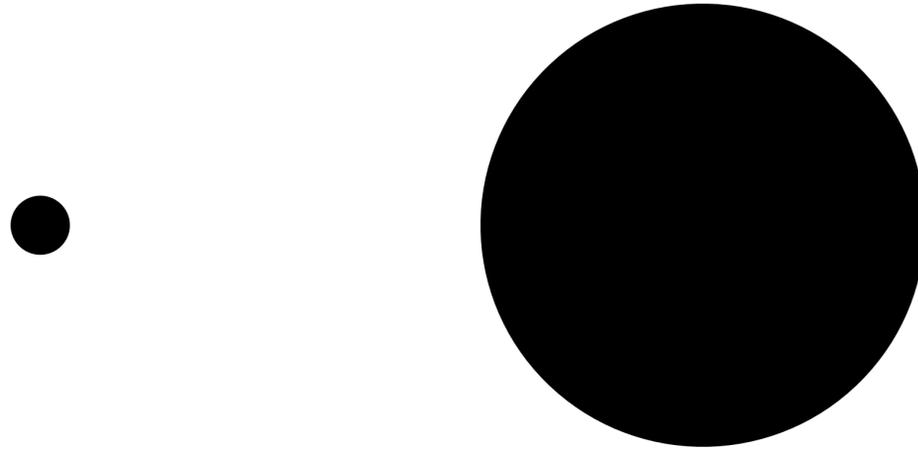
From Wickham (2012)

What aesthetic options do I have at my disposal?
## Which represents the larger value?

From Wickham (2012)

What aesthetic options do I have at my disposal?
Which represents the larger value?

From Wickham (2012)

What aesthetic options do I have at my disposal?
Which represents the larger value?

From Wickham (2012)

What aesthetic options do I have at my disposal?
Which represents the larger value?

From Wickham (2012)

What aesthetic options do I have at my disposal?
Consideration: Consider variable type when picking the aesthetic mapping.

Some aesthetics are ordinal. Some are not. Some can be both!

- Color palettes, for example, can be:

  - **Sequential**: Ordered data with one direction

  - **Diverging**: Ordered data with two directions

  - **Qualitative**: No order to the data
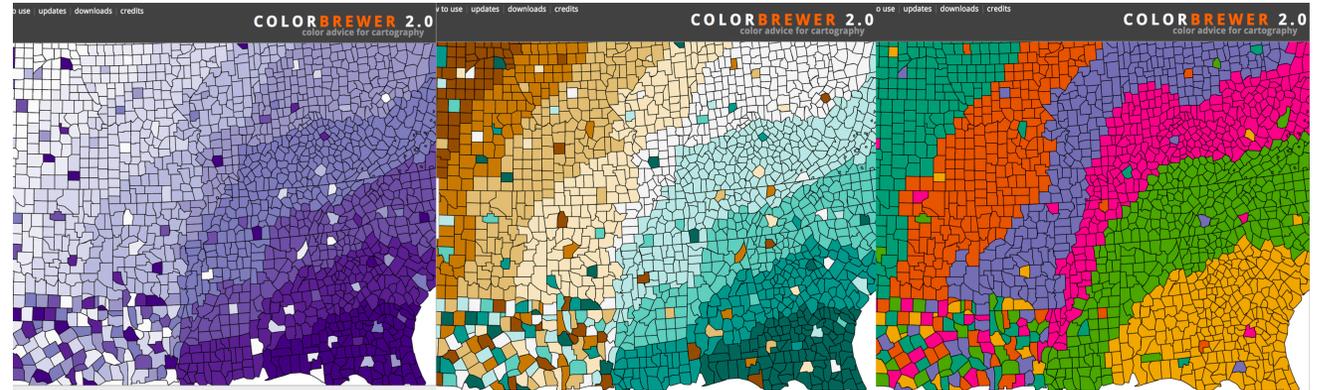
What aesthetic options do I have at my disposal?
Consideration: Consider variable type when picking the aesthetic mapping.

Our ability to perceive differences varies by aesthetic!
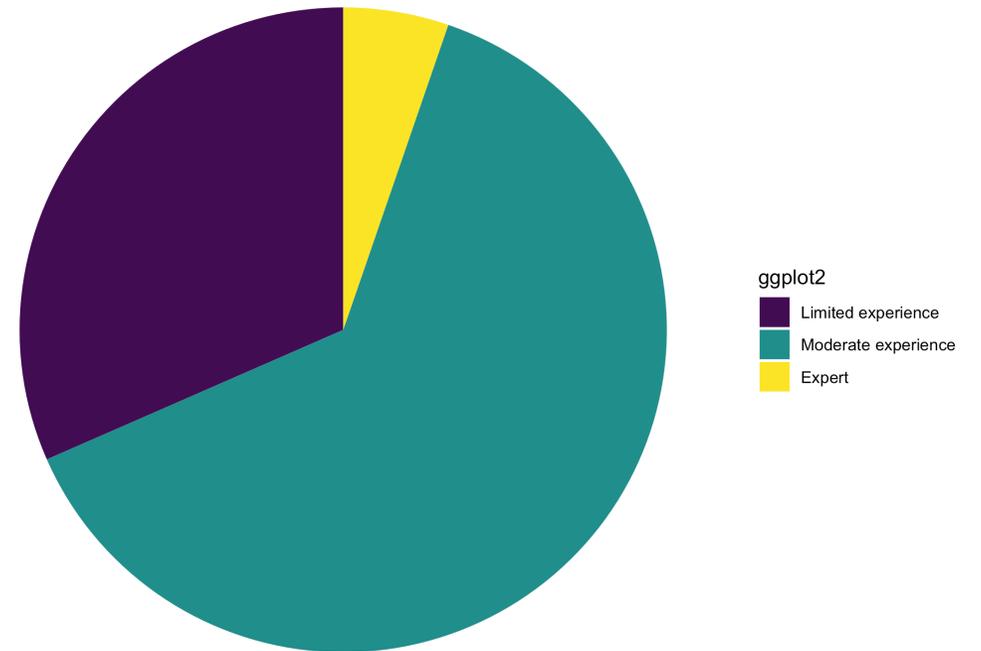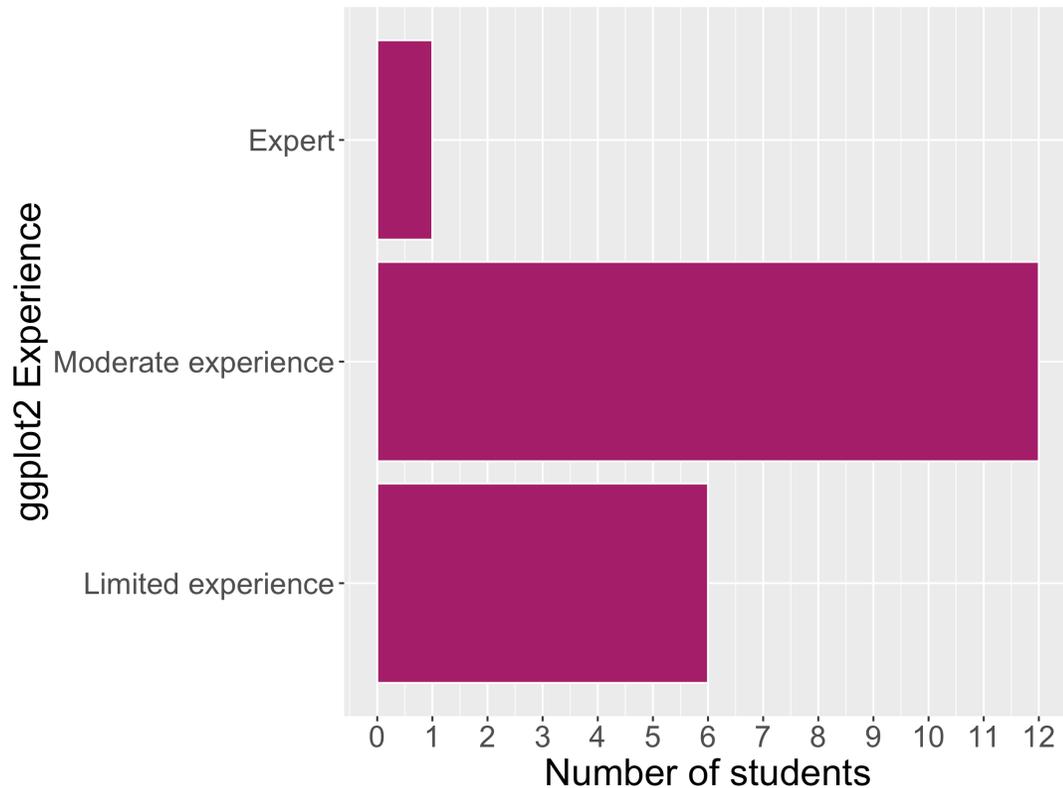
What aesthetic options do I have at my disposal?
Consideration: Pick the aesthetic that makes it easy to assess value.

Generally, humans are good at making comparisons involving:

- Position

- Length

- Angle (but less so than position and length!)

Generally, humans are NOT good at making comparisons involving:

- Area (especially with irregular shapes)

- Volume

- Color shade

- Color hue

What aesthetic options do I have at my disposal?
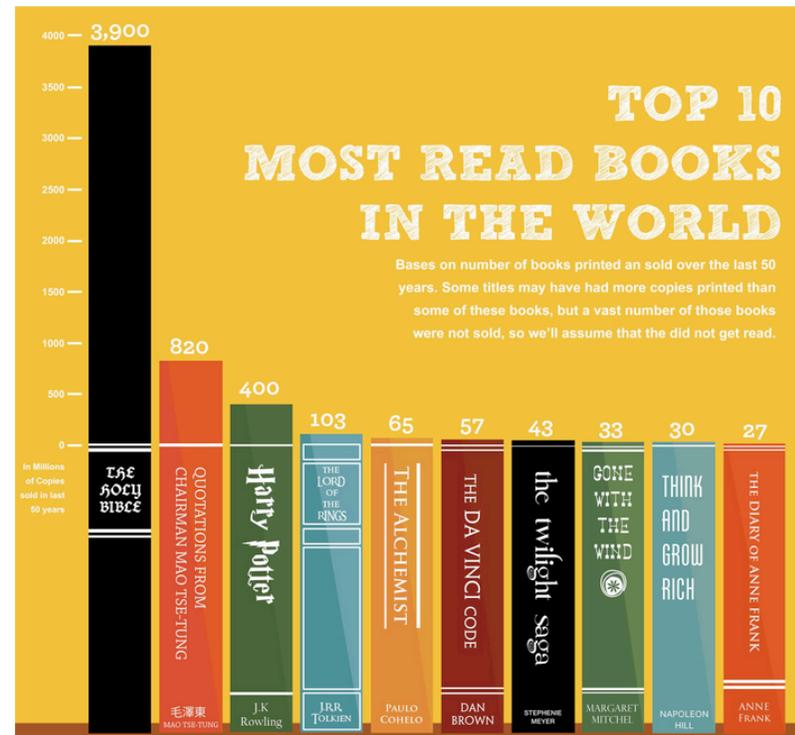# Consideration: Respect the principle of proportional ink.

Principle of proportional ink: The size of shaded areas need to be proportional to the data values they represent.
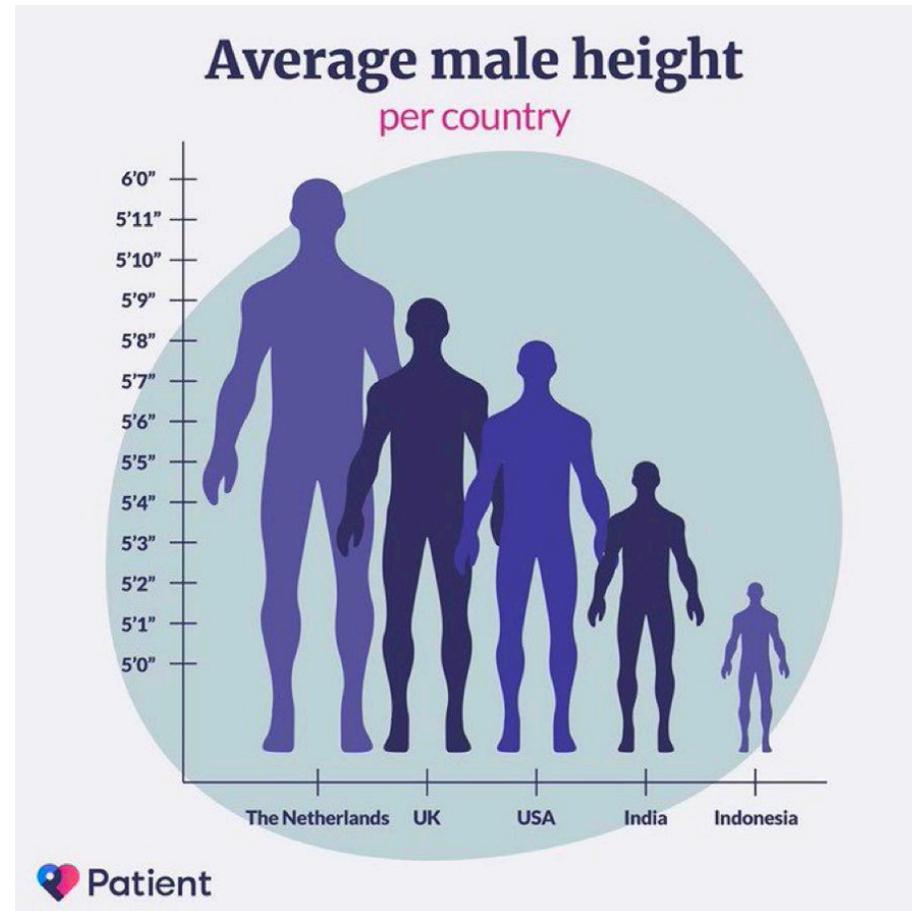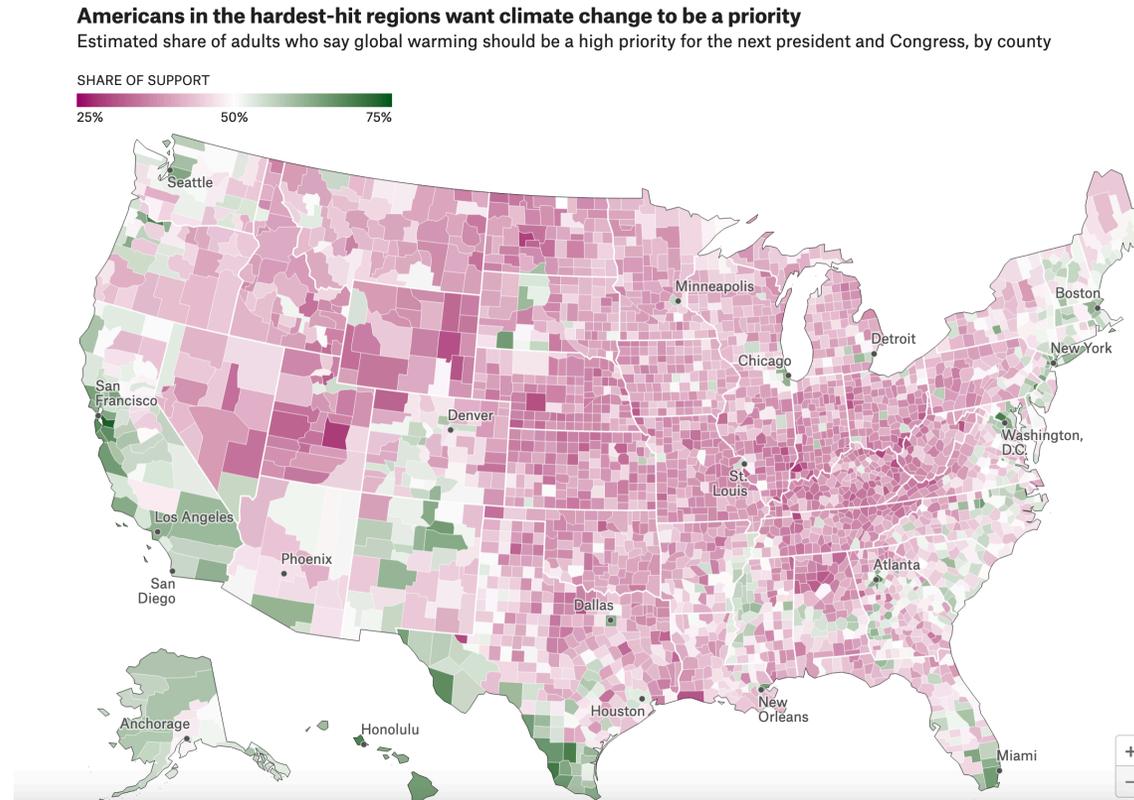
What aesthetic options do I have at my disposal?
Consideration: Respect the principle of proportional ink.

Principle of proportional ink: The size of shaded areas need to be proportional to the data values they represent.

- Bars on a linear scale should start at 0.

What aesthetic options do I have at my disposal?
Consideration: Respect the principle of proportional ink.

Difficult to respect with spatial data. Why?

Americans in the hardest-hit regions want climate change to be a priority
Estimated share of adults who say global warming should be a high priority for the next president and Congress, by county

SHARE OF SUPPORT
25%　50%　75%

Land doesn't vote, people do.

What aesthetic options do I have at my disposal?
Consideration: Respect the principle of proportional ink.

Instead of using geographic boundaries, pick a standardized shape and place "near" geographic location.

Which districts could flip parties?
Each party's chances to flip every House seat
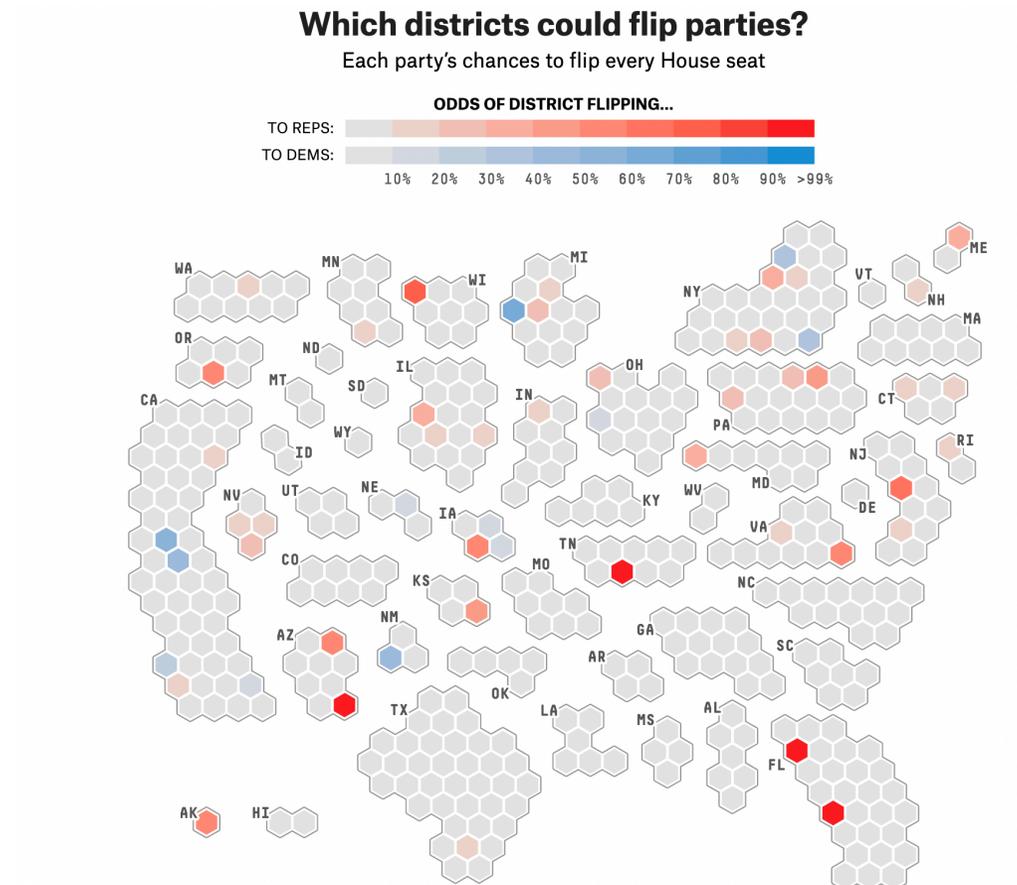
What aesthetic options do I have at my disposal?
Consideration: Try to have a high data-ink ratio.

Data-ink ratio: "proportion of a graphic's ink devoted to the non-redundant display of data-information." – Edward Tufte
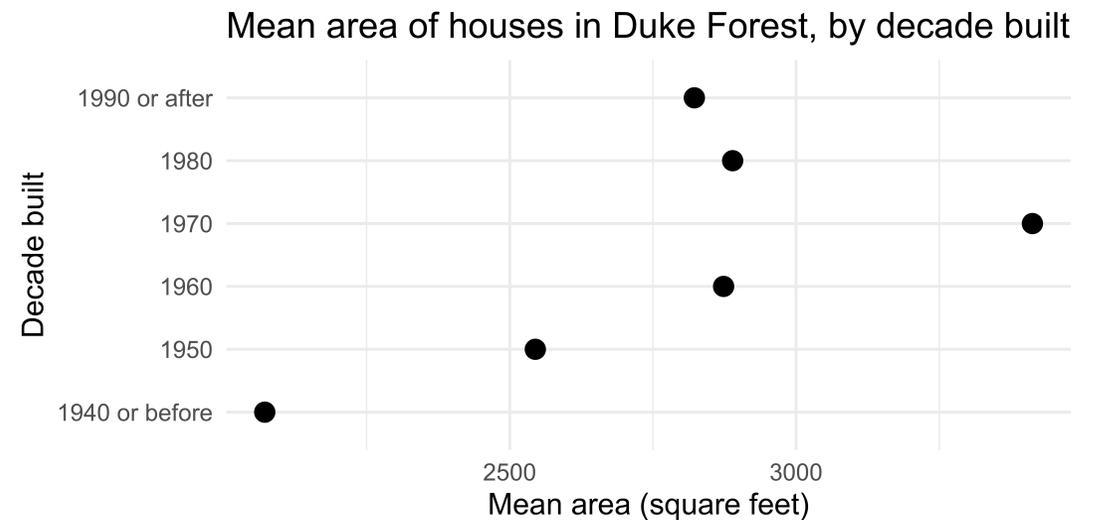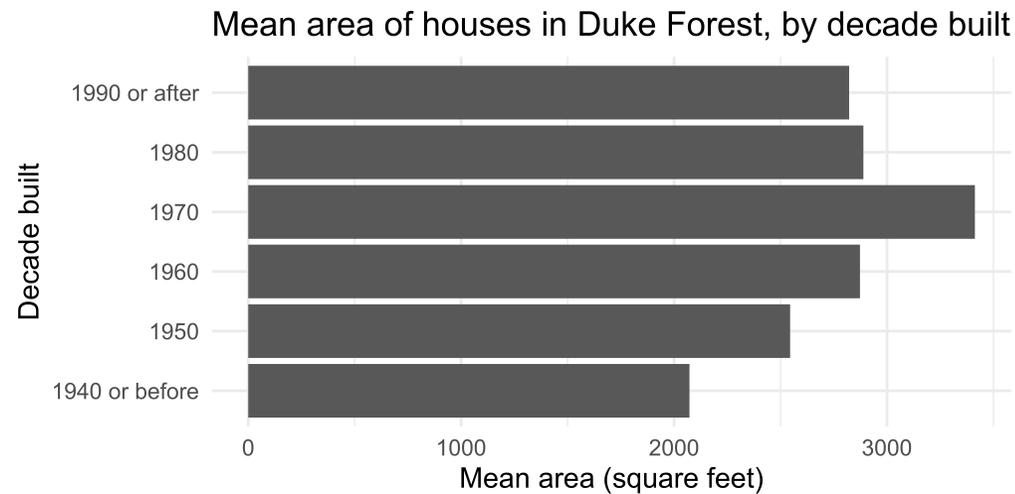
Credit: Mine Çetinkaya-Rundel
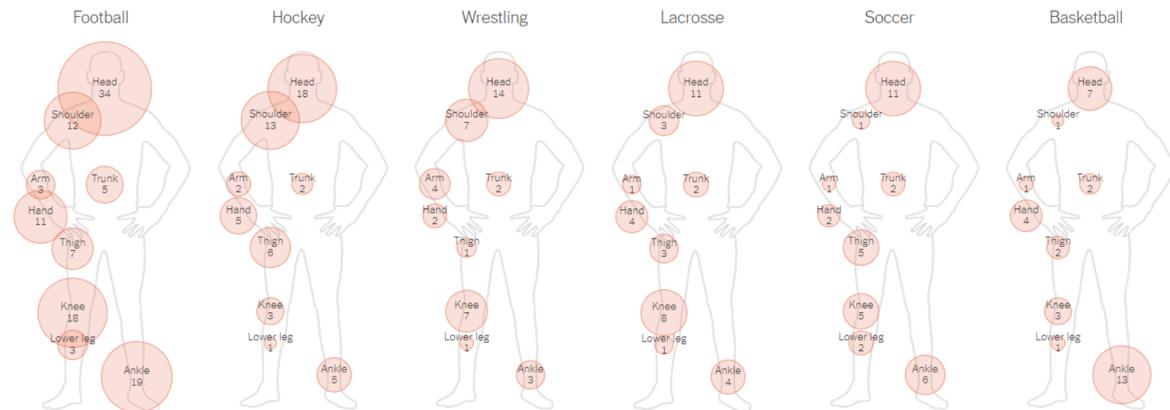
What aesthetic options do I have at my disposal? Consideration: Think carefully about context.

Consider including:

- Title (or Figure Caption)

- Subtitle with maker and data source

- Caption with key points

- Legends/helpers (with units)

- Axis labels (with units)

- Other annotations or reference points

What to add depends **greatly** on the research question or the story you are telling!

Ruth Bader Ginsburg was a stalwart liberal on the court
Estimated ideologies of Supreme Court justices since Roberts joined the court in 2005
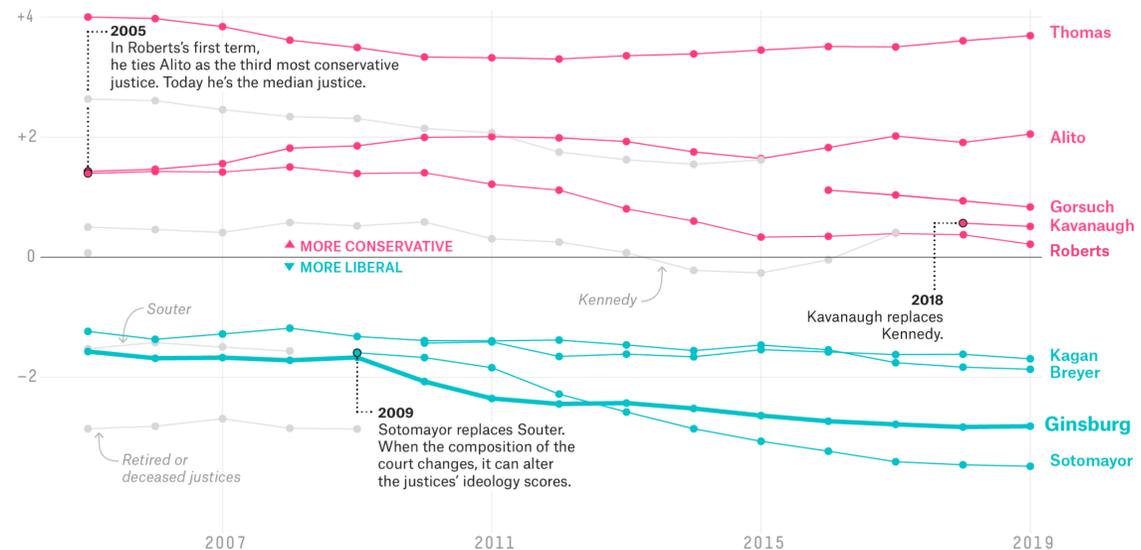
What aesthetic options do I have at my disposal?
Consideration: Think carefully about context.

Consider including:

- Title (or Figure Caption)

- Subtitle with maker and data source

- Caption with key points

- Legends/helpers (with units)

- Axis labels (with units)

- Other annotations or reference points

Context should add both **memorability** and **clarity**.

Common injuries for boys among popular high school sports
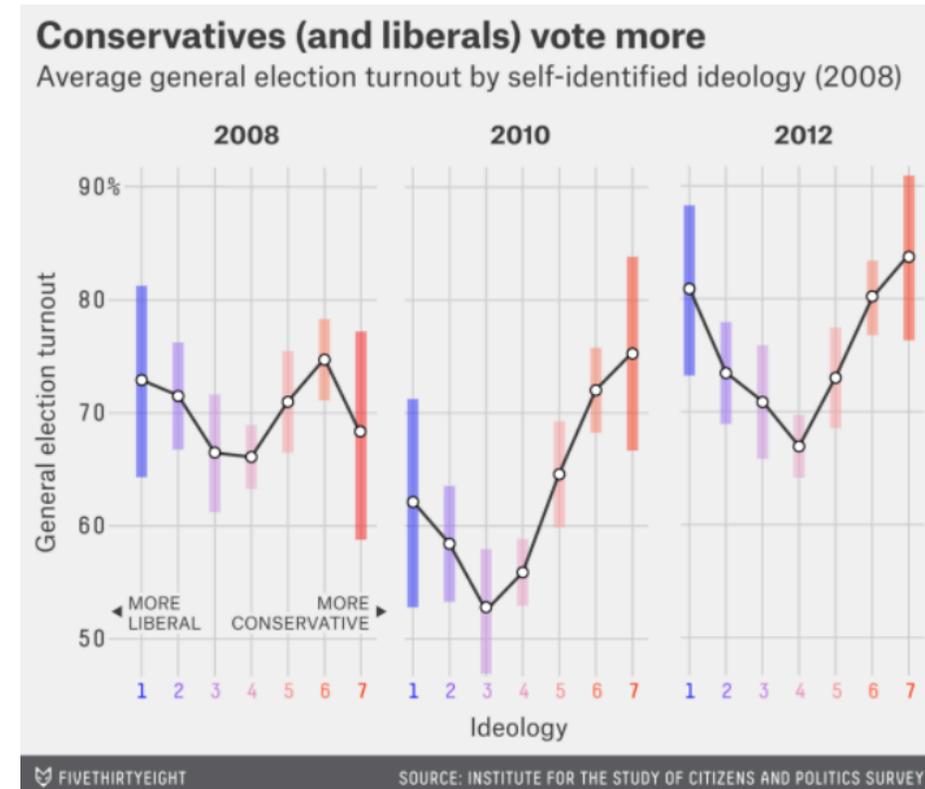Injuries per 10,000 competition plays

Source: National High School Sports-Related Injury Surveillance Study

What aesthetic options do I have at my disposal? Consideration: Simplify as much as you can.

- Faceting is a great way to add another variable without over-complicating your graphic.

- But only add additional variables that are **useful** to the story!
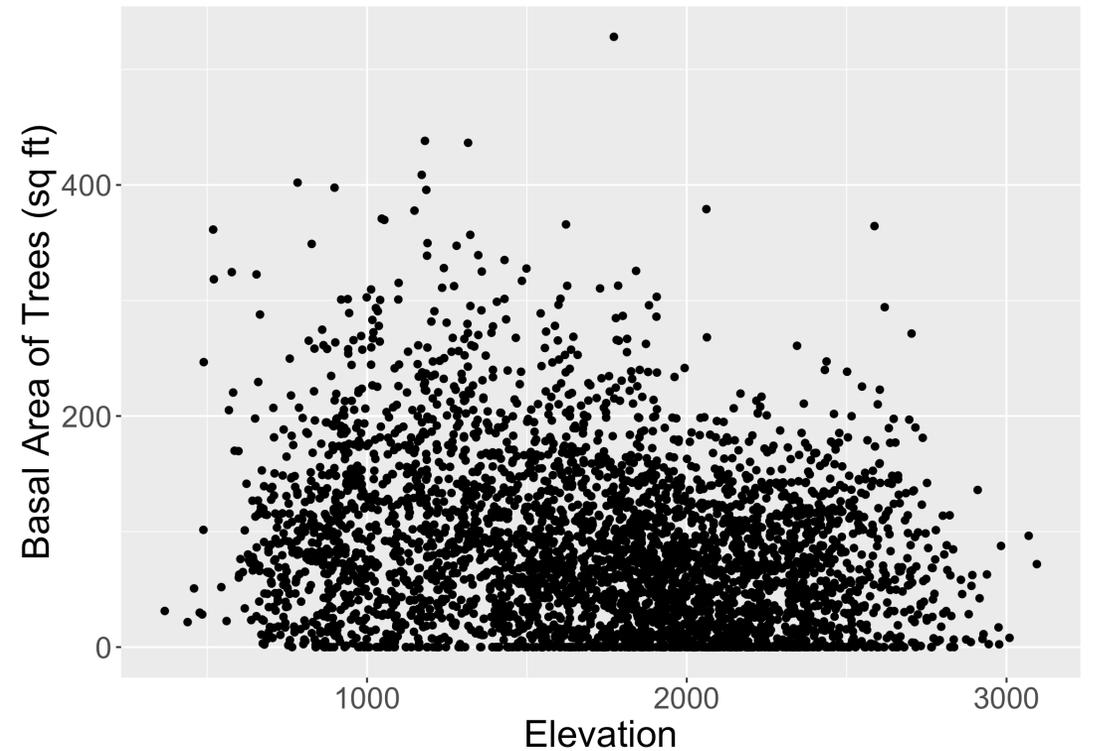
What aesthetic options do I have at my disposal?
Consideration: Simplify as much as you can.

- Over-plotting is very common in the Age of Big Data!

- Example from my own work with the US Forest Inventory and Analysis Program
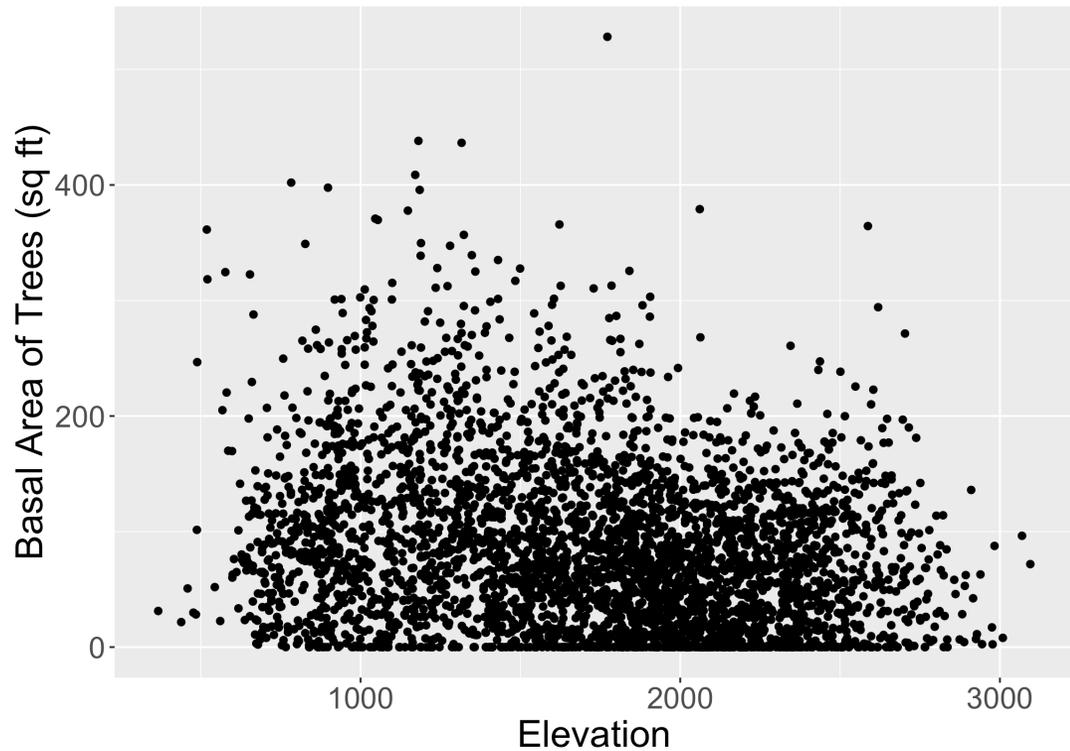
Sampled Forest Inventory plots in Idaho 2011 - 2020

Sampled Forest Inventory plots in Idaho 2011 - 2020

- Jitter the points.

What aesthetic options do I have at my disposal?
Consideration: Simplify as much as you can.
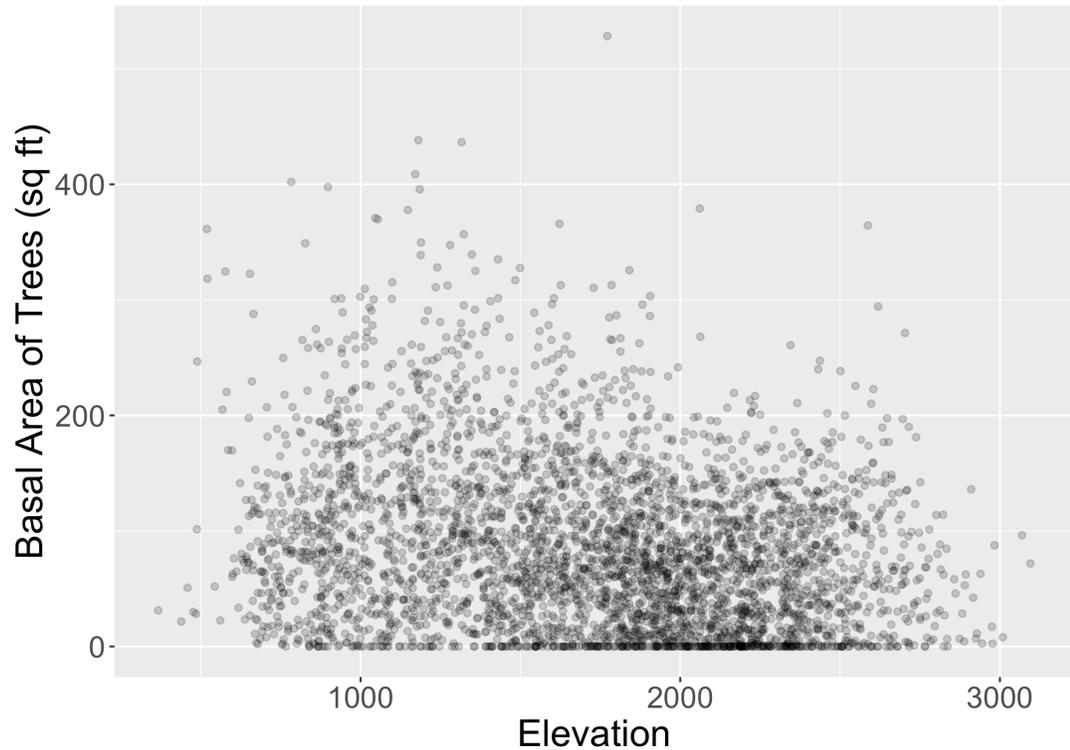
Sampled Forest Inventory plots in Idaho 2011 - 2020
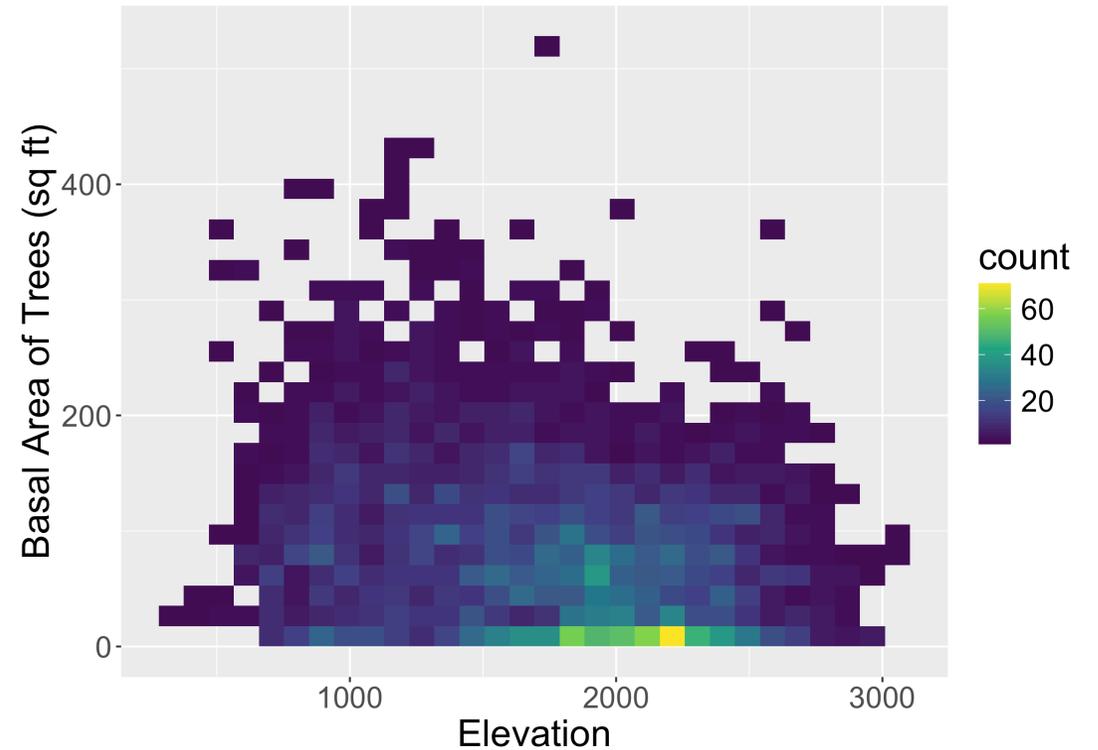
Sampled Forest Inventory plots in Idaho 2011 - 2020

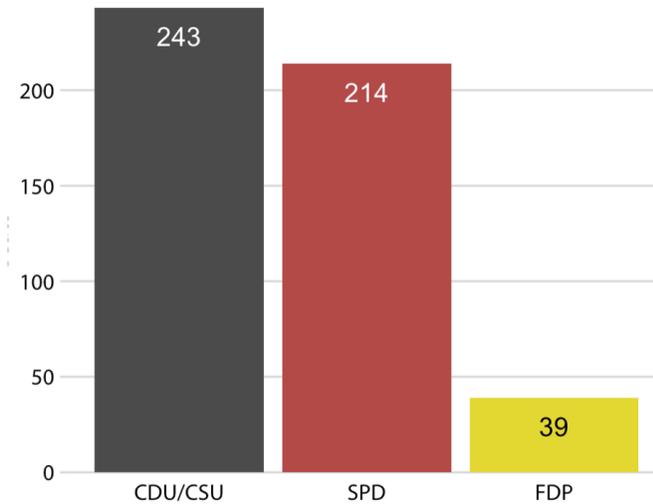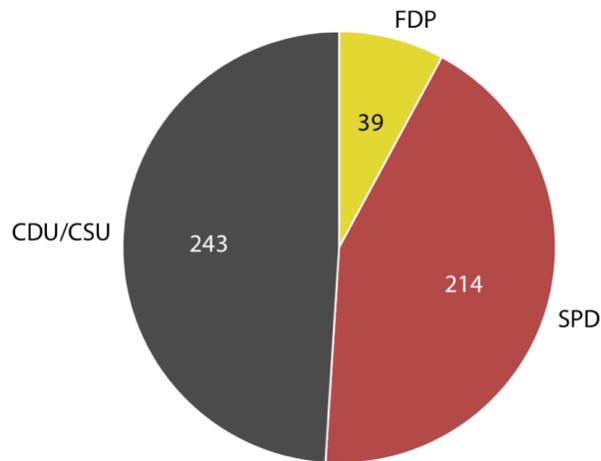- Add transparency.

- Bin the data and then try a different geom.

Pros and cons of different approaches?

What aesthetic options do I have at my disposal?
Consideration: Make important comparisons easy.

Which graph makes it easy to conclude that the ruling coalition (FDP + SPD) have a majority?
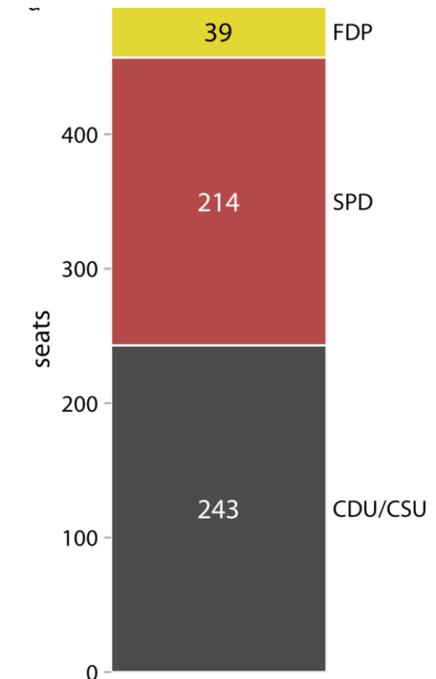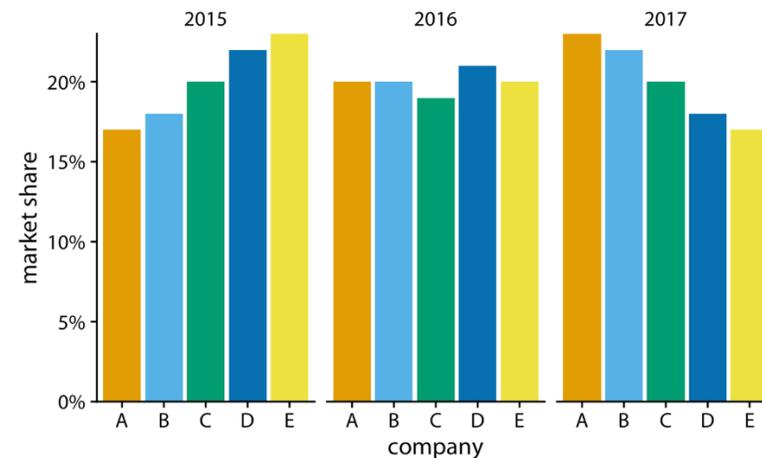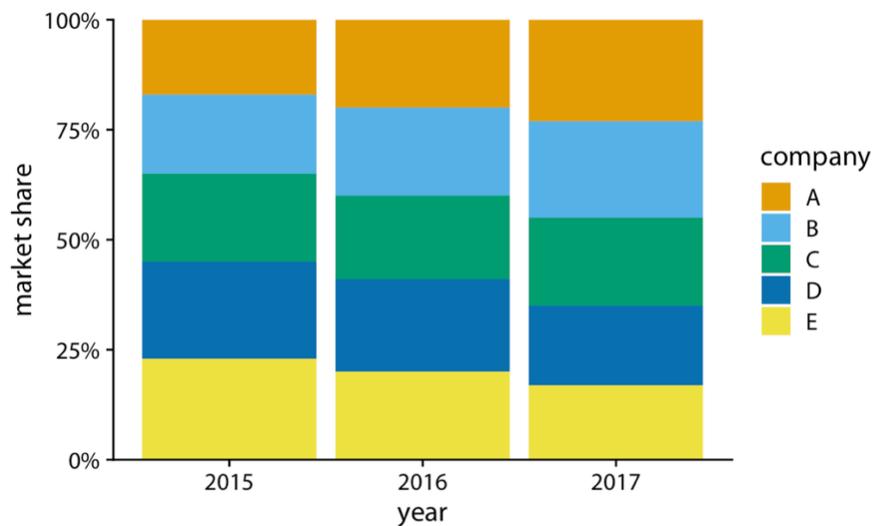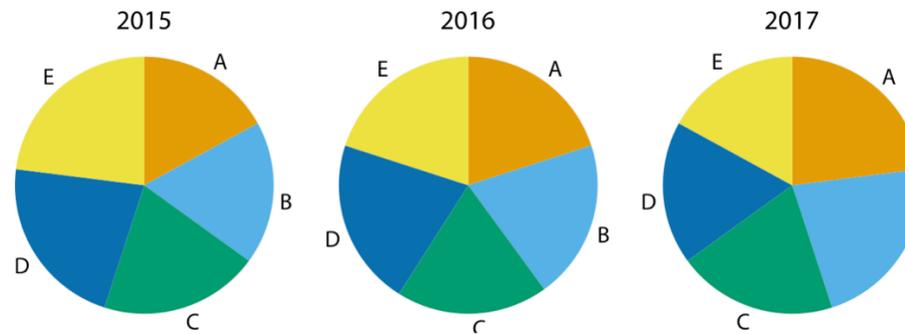

Wilke (2019)

What aesthetic options do I have at my disposal?
Consideration: Make important comparisons easy.

Which graph makes it easy to see how a company's market share changes over time? (Warning: Fake data.)

What aesthetic options do I have at my disposal?
Consideration: Make your graphs accessible!

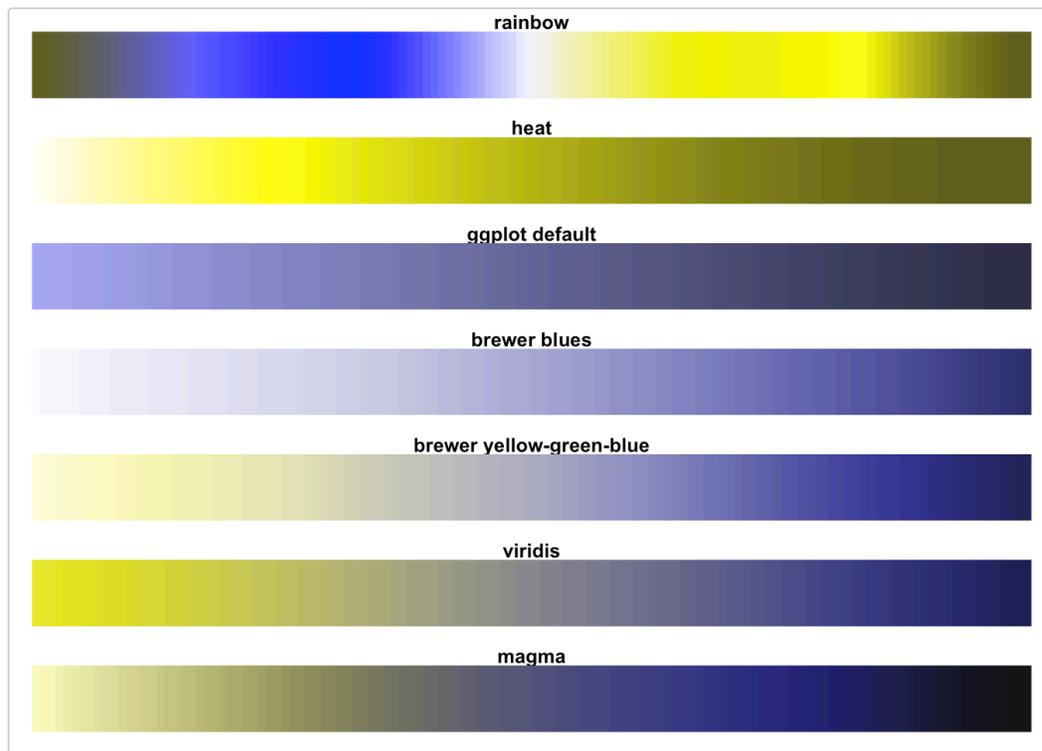Not all R color palettes have been vetted for color blindness.

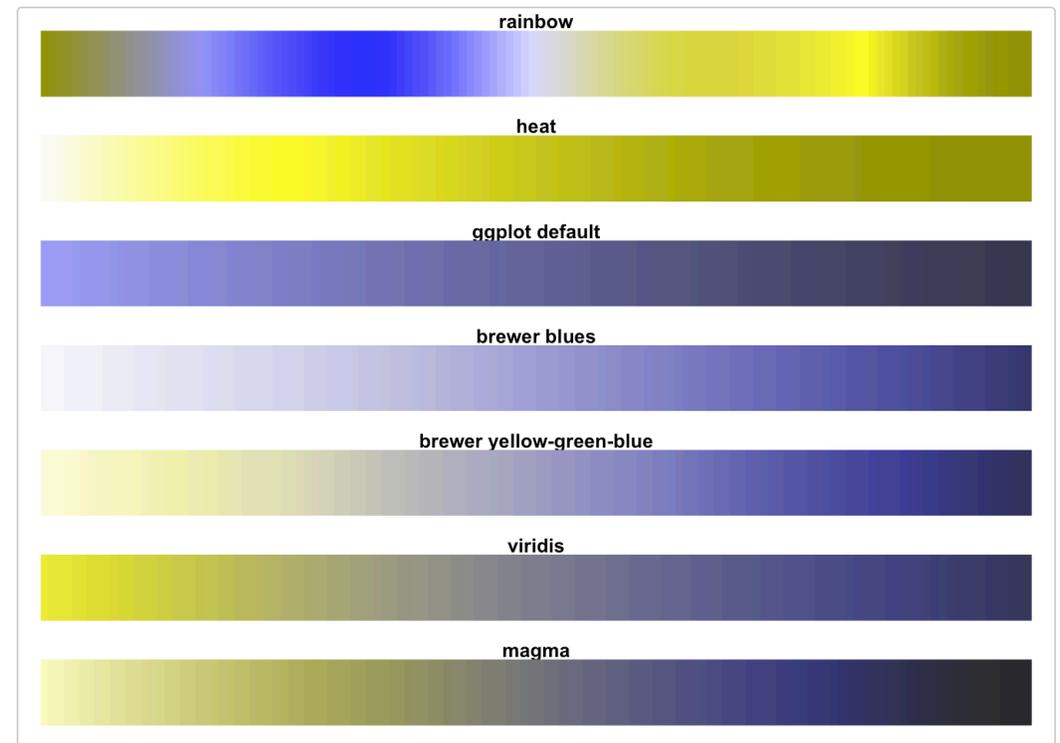# What aesthetic options do I have at my disposal?
Consideration: Make your graphs accessible!
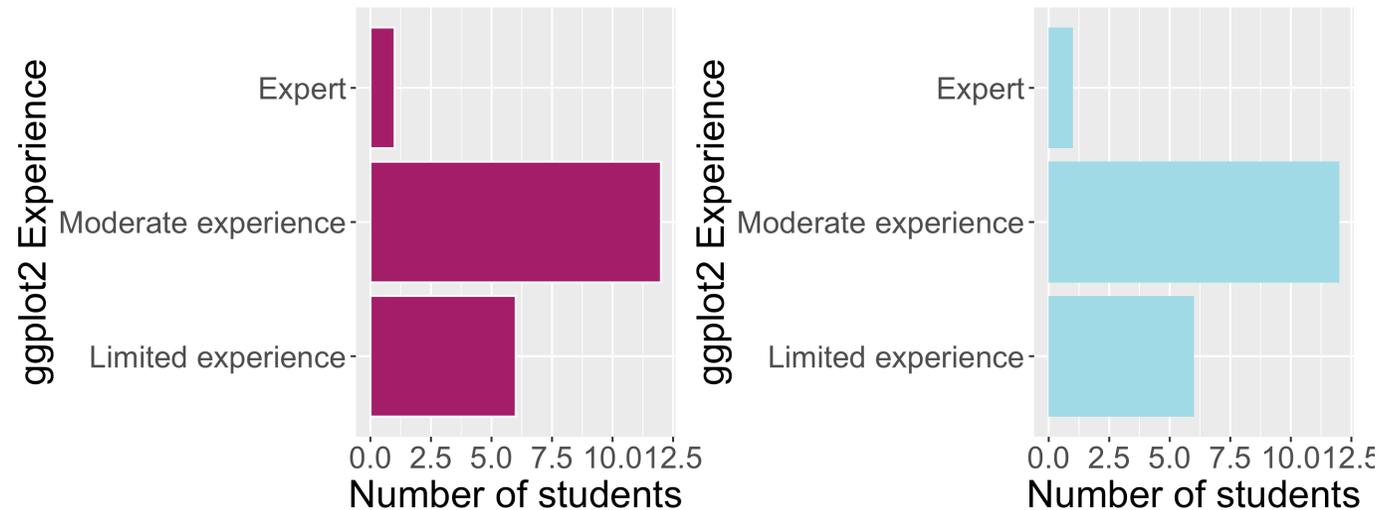
Not all R color palettes have been vetted for color blindness.

What aesthetic options do I have at my disposal? Consideration: Make your graphs accessible!

- Color contrast also matters.
    - Shoot for a ratio of 4.5 or higher between overlapping colors.


```
1  library(coloratio)
2  cr_get_ratio("#A71F69", "#EBEBEB")
```
```
[1] 5.78868
```

```
1  cr_get_ratio("#A4DBE8", "#EBEBEB")
```
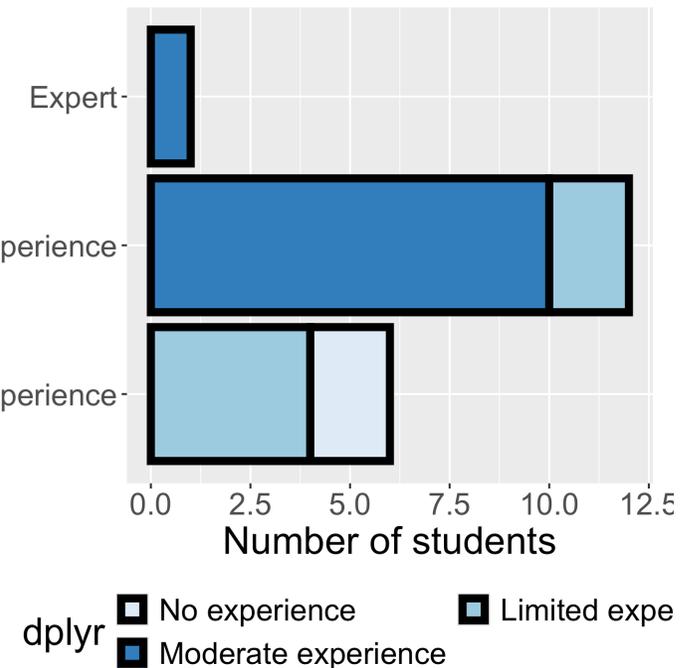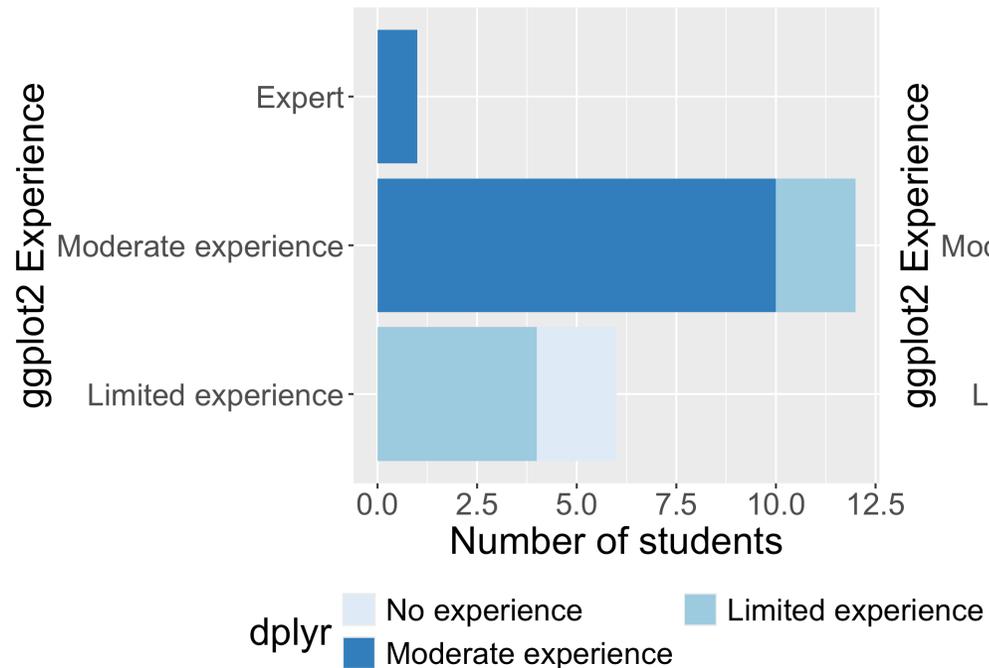```
[1] 1.269455
```

# What aesthetic options do I have at my disposal?
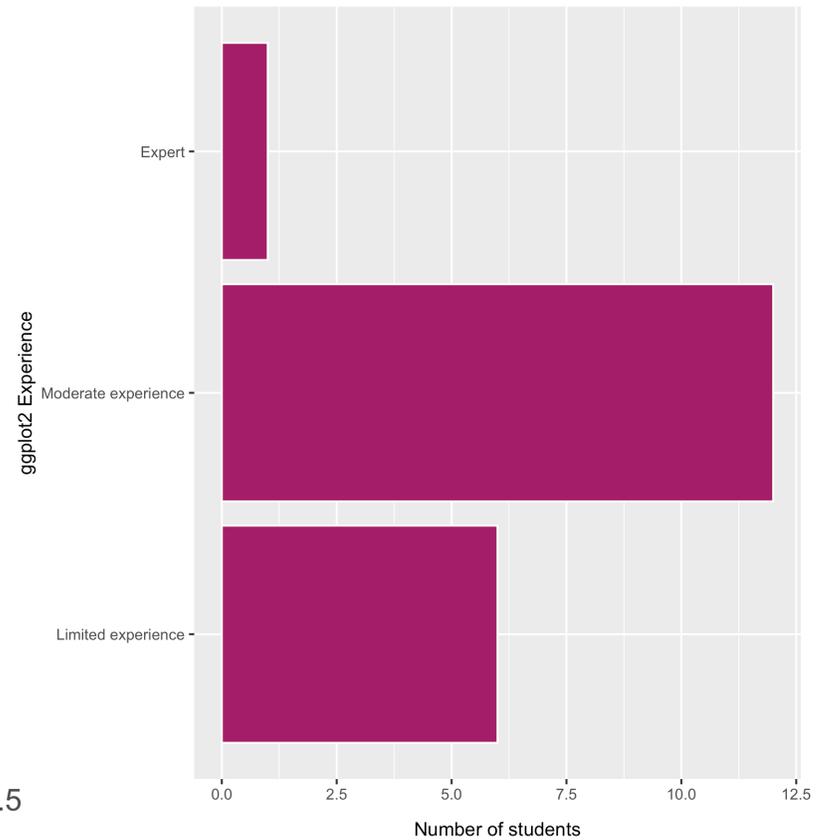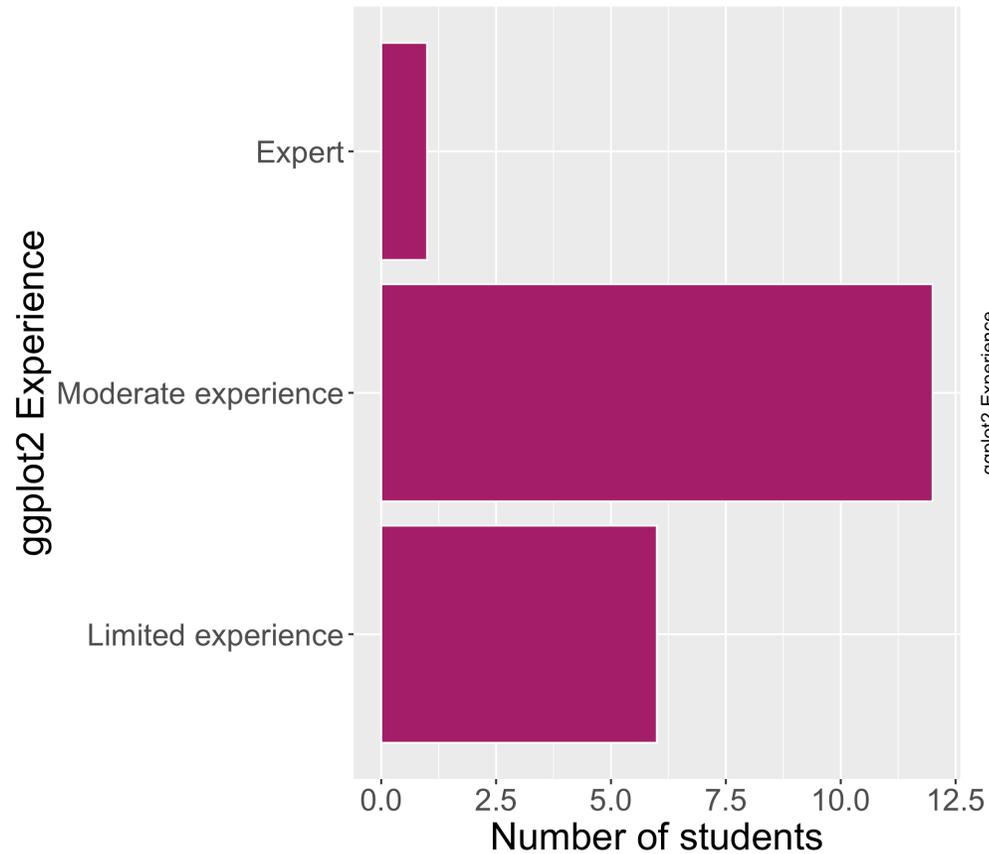# Consideration: Make your graphs accessible!

Use white space to help separate elements.

# What aesthetic options do I have at my disposal?
# Consideration: Make your graphs accessible!

Use a large enough font size!

# What aesthetic options do I have at my disposal?
# Data Viz Considerations

- We could spend all semester on data viz principles.

- Be thoughtful and iterate.

- Let's spend some time considering the strengths and weaknesses of some graphs via an activity.
  - Pick out the:
    - Clearest story
    - Most memorable
    - Best overall

- Be ready to present one of your graphs to the class, if we have time.

15:00