Data Visualization - part 1

Willa

9/27/2022

Check-in

Start Recording

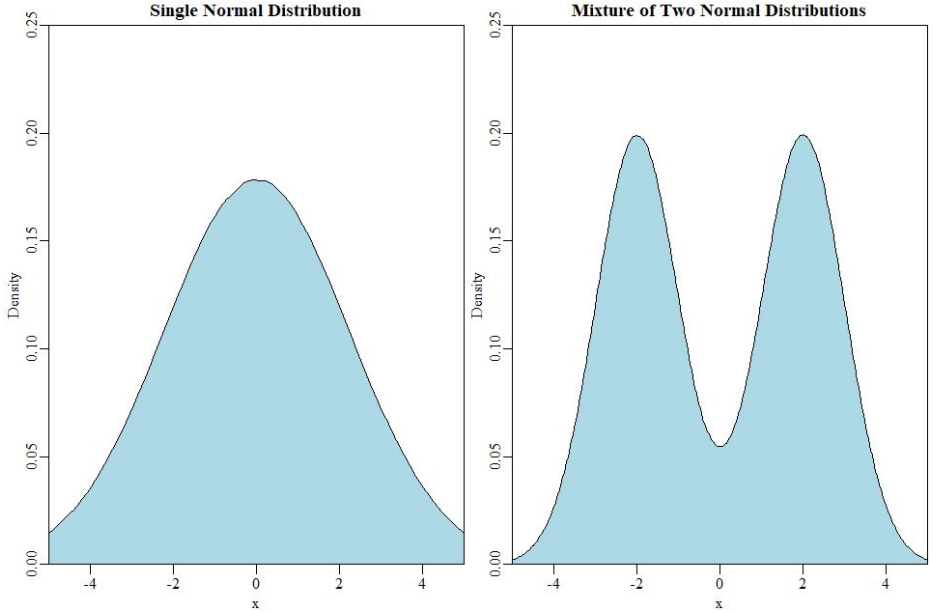
So you have clean data… now what?

Are you ready to run analyses?

How do you decide what analyses to run?

Why do we visualize our data ?

Why do we visualize our data ?



Same Mean

Same Variance

Very different data

Exploring your data lets you…

Know what your data looks like.

Figure out what questions you can ask.

Formalize hypotheses and questions

Figure out what statistics are appropriate.

Identify outliers

etc.


tidyverse

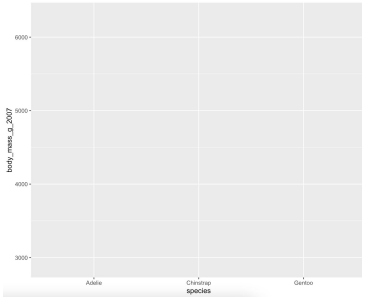

dplyr : go wrangling
dplyr
data
Horst '18


ggplot2:
Build a data
MASTERPIECE
THEMES
GEOMS
Horst '18

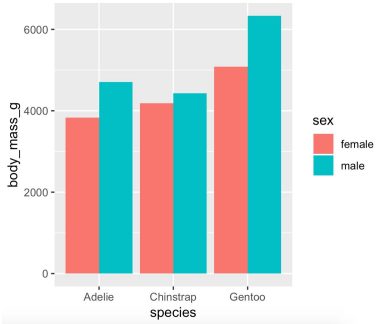
Build any plot with ggplot

You can think of each element of your plot as a layer that you can build up. This is what makes ggplot so flexible!

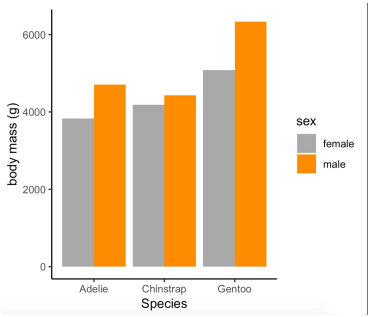
Step 1: create a plot and designate variables



Step 2: Decide how to represent the data



Step 3: customize

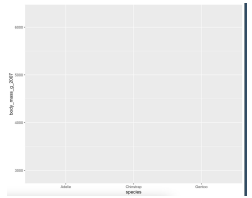


In ggplot terminology:

Step 1: Create a ggplot object

aesthetic mapping: Tell ggplot how to map the variables in your dataframe to visual properties . (ie how do you want to arrange the data).

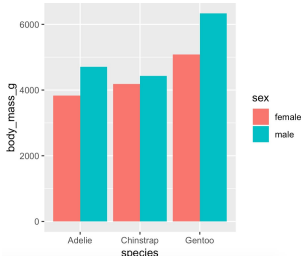
```
ggplot(data = myDataFrame, aes(x = xaxis_values, y = yaxis_values, fill = grouping_var))
```



Step 2: Add geometric objects to your plot. These tell ggplot how to represent your data.

```
ggplot(data = myDataFrame, aes(x = xaxis_values, y = yaxis_values, fill = grouping_var)) +
  geom_bar()
```

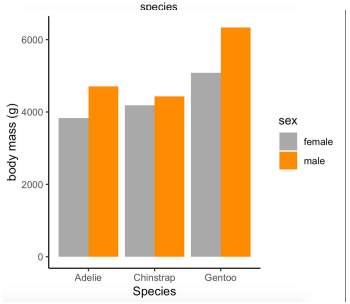
Notice "+" connects ggplot objects together



Step 3: Add other layout and design features

```
ggplot(data = myDataFrame, aes(x = xaxis_values, y = yaxis_values, fill = grouping_var)) +
    scale_fill_manual(values = c("dark gray", "dark orange")) +

    theme_classic()
```



Colors in ggplot

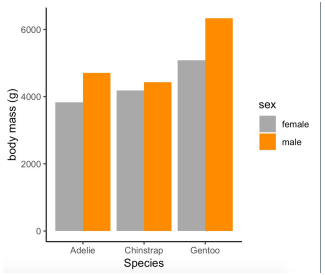
1. Color can be used as a grouping variable when you want to compare groups.

```
ggplot(data, aes( species, body_mass_g, fill = sex)) +

geom_col() +

scale_fill_manual(values = c("dark gray", "dark orange"))
```

Add another object to customize your color mapping



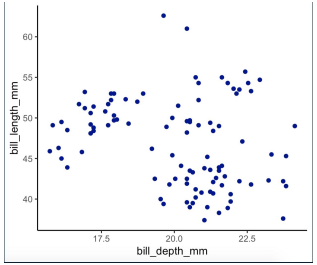
2. Color doesn't have to correspond to a variable. You can just set the color of a geom_object() for aesthetic reasons.

```
ggplot(data, aes( species, body_mass_g)) +

geom_point(color = "dark blue")
```

Color = outline color

Fill = filled in color



Now lets practice exploring a dataset through data viz