

Lab 2: ggplot2 and tidyverse

0.1 ggplot2

1. Make a scatterplot of `cty` vs. `hwy` mpg using the `mpg` dataset.
2. Describe the relationship that you see.
3. Map color and shape to type of drive the car is (see `?mpg` for details on the variables.). Do you see any patterns?
4. Alter your plot from part 3. to make all the points be larger.

Using the `mpg` dataset,

5. Make a histogram of `hwy`, faceted by `drv`.
6. Make a scatterplot that incorporates color, shape, size, and facets.
7. BONUS - Color your histograms from 1. by `cyl`. Did this do what you thought it would? (Look at `fill` and `group` as options instead).

0.2 readr

1. Read the NFL salaries dataset from https://raw.githubusercontent.com/ada-lovecraft/ProcessingSketches/master/Bits%20and%20Pieces/Football_Stuff/data/nfl-salaries.tsv into R.
2. What is the highest NFL salary in this dataset? Who is the highest paid player?
3. Make a histogram and describe the distribution of NFL salaries.

0.3 dplyr

Using the NFL salaries from https://raw.githubusercontent.com/ada-lovecraft/ProcessingSketches/master/Bits%20and%20Pieces/Football_Stuff/data/nfl-salaries.tsv that you loaded into R in the section, perform the following.

1. What is the team with the highest paid roster?
2. What are the top 5 paid players?

3. What is the highest paid position on average? the lowest? the most variable?

0.4 `tidyr`

1. Is the NFL salaries from https://raw.githubusercontent.com/ada-lovecraft/ProcessingSketches/master/Bits%20and%20Pieces/Football_Stuff/data/nfl-salaries.tsv that you loaded into R in a previous your turn tidy? Why or why not?
2. There is a data set in `tidyr` called `world_bank_pop` that contains information about population from the World Bank (<https://data.worldbank.org/>). Why is this data not tidy? You may want to read more about the data to answer (? `world_bank_pop`).
3. Use functions in `tidyr` to turn this into a tidy form.