Introduction to Tidyverse Worksheet

Bios 338/538

All answers to questions will be listed at the end but please try to figure out the code first! The data set you will be working with today is called mtcars. This is an introductory dataset on 32 vehicles and some of their characteristics:

* Mpg: Miles per gallon – a metric of fuel efficiency
* Cyl: Cylinders – the more cylinders the faster the car
* Disp: Displacement – a metric for size of the engine
* Hp: Horsepower – engine power
* Drat: Driveshaft Ratio: ¯\\_(ツ)\_/¯ (don’t worry about this one)
* Wt: Weight (in tons)
* Qsec: ¼ mile time – measurement of acceleration
* Vs: engine shape – Binary (1 is V shape, 0 is other)
* Am: transmission – automatic or manual
* Gear: number of gears
* Carb: number of carborateurs

Whenever you work with new data it is always important as a first step to familiarize yourself with what each variable means!

1. **Load the data**

This first one is a freebie because loading data in R can be challenging!

*Data(“mtcars”)*

\*You may get the dataset under the Values section in your environment saying <promise>. Just click on the word “mtcars” if this happens!

1. **Filter data for cars that have > 20 miles per gallon**
2. **Filter data for cars that have both < 110 horsepower and have 6 cylinders**
3. **Make a new column called kpg (kilometers per gallon) and multiply mpg by the conversion factor of kilometers per mile (1.6) save this new dataframe as mtcars\_kpg**
4. **Make a new column called power\_ratio that displays the horsepower per weight of each car and arrange it by least to greatest. Save dataframe as mtcars\_power\_ratio (\*bonus figure out how to arrange it greatest to least!)**
5. **Using the mtcars\_kpg dataframe, rename the column wt to “weight” and save the dataframe**
6. **Make a histogram of the distribution of mpg**
7. **Make a scatterplot (geom\_point()) of the relationship between miles per gallon and power ratio.**

Look at what we see looks like there may be a linear relationship! But we will get there in future lectures!

1. **To further investigate the relationship in the previous plot let’s visualize how the number of cylinders plays a role in this relationship. Add color = factor(cyl) into the aesthetic! Do you notice any relationships?**
2. **As a teaser for next week, let’s make the size of the dots bigger, and change the axis titles to “MPG (Miles Per Gallon)” and “Power Ratio” with the labs function. Look at the labs function documentation using the ?labs function and read the documentation to figure out how it works! To make the dots bigger add size = 2 inside geom\_point()**

**Answers! Don’t Go past unless you gave the question a try!**

*2. mtcars %>% filter(mpg > 20)*

*3. mtcars %>% filter(hp < 110, cyl == 6)* **Or** *mtcars %>% filter(hp < 110 & cyl == 6)*

*4. mtcars\_kpg <- mtcars %>% mutate(kpg = mpg \* 1.6)*

*5. mtcars %>% mutate(power\_ratio = hp / wt) %>% arrange(power\_ratio)*

*Bonus: mtcars\_power\_ratio<- mtcars %>% mutate(power\_ratio = hp / wt) %>% arrange(desc(power\_ratio))*

*6. mtcars\_kpg <- mtcars\_kpg %>% rename(weight = wt)*

*7. mtcars\_kpg %>% ggplot(aes(x = mpg)) + geom\_histogram()*

*8. mtcars\_power\_ratio %>% ggplot(aes(x = mpg, y = power\_ratio)) + geom\_point()*

*9. mtcars\_power\_ratio %>% ggplot(aes(x = mpg, y = power\_ratio, color = factor(cyl))) + geom\_point()*

*10. mtcars\_power\_ratio %>% ggplot(aes(x = mpg, y = power\_ratio, color = factor(cyl))) + geom\_point(size = 2) + labs(x = “MPG (Miles Per Gallon)”, y = “Power Ratio”)*