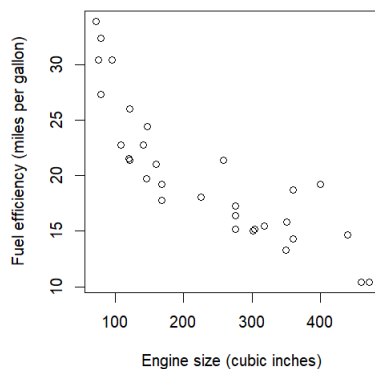


DAY 4 ASSIGNMENT

1. Use the built-in dataset mtcars for this assignment. Do cars with big engines use more fuel than cars with small engines? You probably already have an answer, but try to make your answer precise. What does the relationship between engine size and fuel efficiency look like? Is it positive?

Negative?

```
> # Load the mtcars dataset
> data(mtcars)
> # Print the structure of the dataset
> str(mtcars)
'data.frame':   32 obs. of  11 variables:
 $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
 $ cyl : num   6  6  4  6  8  6  8  4  4  6 ...
 $ disp: num  160 160 108 258 360 ...
 $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
 $ drat: num   3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
 $ wt  : num   2.62 2.88 2.32 3.21 3.44 ...
 $ qsec: num   16.5 17 18.6 19.4 17 ...
 $ vs  : num   0  0  1  1  0  1  0  1  1  1 ...
 $ am  : num   1  1  1  0  0  0  0  0  0  0 ...
 $ gear: num   4  4  4  3  3  3  3  4  4  4 ...
 $ carb: num   4  4  1  1  2  1  4  2  2  4 ...
> # Create a scatterplot of engine size versus fuel efficiency
> plot(mtcars$disp, mtcars$mpg, xlab = "Engine size (cubic inches)", ylab =
"Fuel efficiency (miles per gallon)")
> # Calculate the correlation coefficient between engine size and fuel efficiency
> cor(mtcars$disp, mtcars$mpg)
[1] -0.8475514
```



2. How many rows are in mpg? How many columns?

```
> # Load the mpg dataset
> library(ggplot2)
> data(mpg)
> # Count the number of rows and columns
> nrow(mpg) # Output: 234
[1] 234
> ncol(mpg) # Output: 11
[1] 11
```

3. Which variables in mpg are categorical? Which variables are continuous?

ANS:

In the mpg dataset, the categorical variables are:

manufacturer

model

year

cyl

trans

drv

fl

class

The continuous variables are:

displ

hwy

cty

4. Take the first faceted plot in this section:

```
ggplot(data = mpg) +
```

```
  geom_point(mapping = aes(x = displ, y = hwy)) +
```

```
  facet_wrap(~ class, nrow = 2)
```

What are the advantages to using faceting instead of the colour aesthetic?

What are the disadvantages? How might the balance change if you had a larger dataset?

ANS:

4) The advantages of using faceting instead of the color aesthetic are:

Clarity: Faceting provides a clear and organized way to present multiple plots in a single visualization. It allows us to compare and contrast multiple groups or variables without the need for color legends or additional labeling.

Flexibility: Faceting can be used with a variety of plot types, making it a versatile option for exploring relationships between variables.

Efficient use of space: Faceting allows us to use space more efficiently than simply using a single plot with color, as we can fit multiple smaller plots into the same space.

The disadvantages of faceting are:

Limited color options: Faceting typically relies on a single color scheme, which may not be ideal for certain types of data. For example, if we need to show the relationship between two continuous variables, it might be difficult to find a suitable color palette to differentiate between the two.

Complexity: Faceting can make a visualization more complex, especially if there are many variables to be plotted. This can make it difficult to interpret or communicate findings.

If we had a larger dataset, the balance might change as the size and number of plots would increase, which could make the visualization more complex and difficult to interpret. Additionally, using a color aesthetic might become more useful for highlighting patterns and trends across larger datasets.

5. What geom would you use to draw a line chart? A boxplot? A histogram? An area chart?

ANS:

Line chart: `geom_line` or `geom_path`

Boxplot: `geom_boxplot`

Histogram: `geom_histogram`

Area chart: `geom_area`