ITA04 STATISTICS WITH R PROGRAMMING

**Assignment–Day 4**

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?

SOURCE CODE:

library(ggplot2)

ggplot(mtcars, aes(x = disp, y = mpg)) +

geom\_point() +

labs(x = "Engine size (cubic inches)", y = "Miles per gallon")

cor(mtcars$disp, mtcars$mpg)

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

SOURCE CODE:

dim(mpg)

There are 234 rows and 11 columns in the mpg data frame.

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

SOURCE CODE:

str(mpg)

This code will output the structure of the "mpg" dataset, including the data type of each variable. Based on this output, we can identify the categorical variables as those with a data type of "factor" or "character," while the continuous variables will have a data type of "numeric." Here is the list of variables and their data types:

Categorical variables:

manufacturer (character)

model (character)

displ (numeric, but can be treated as categorical if we group by engine size)

year (numeric)

cyl (numeric)

trans (character)

drv (character)

class (character)

Continuous variables:

cty (numeric)

hwy (numeric)

Based on this information, we can conclude that there are eight categorical variables and two continuous variables in the "mpg" dataset.

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?

Faceting and coloring are both useful techniques for visualizing relationships in data. The advantages of using faceting instead of the color aesthetic are:

Faceting allows us to compare relationships between variables within subsets of the data more easily. In the given example, we can see the relationship between engine displacement and highway miles per gallon for each class of car, allowing us to identify any patterns or differences between classes more easily.

Faceting allows us to use different scales for each subset of the data, which can be important if the variables have different ranges or units.

Faceting is more accessible to colorblind individuals who may have difficulty distinguishing between different colors.

The disadvantages of using faceting are:

Faceting can result in a large number of plots, which may be overwhelming or difficult to interpret if there are many subsets of the data.

Faceting can result in small plots if there are many subsets, which may make it difficult to see patterns or differences between subsets.

The balance between advantages and disadvantages may change if we had a larger dataset. With a larger dataset, faceting may become more challenging due to the potentially large number of subsets, making it difficult to interpret the results. However, faceting can still be useful in identifying patterns and relationships within subsets of the data, and can help to identify any differences or similarities between subsets. In contrast, using the color aesthetic may be more feasible with a larger dataset as it may be more difficult to visualize many subsets using faceting. However, using color may result in a loss of information due to colorblindness or the inability to distinguish between different colors.

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

1.Line chart: To draw a line chart, we use the geom\_line() function.

2.Boxplot: To draw a boxplot, we use the geom\_boxplot() function.

3.Histogram: To draw a histogram, we use the geom\_histogram() function.

4.Area chart: To draw an area chart, we use the geom\_area() function.