NASH JULIAN CABALFIN BSIT 2-A

CS 101

#1. How many columns are in mpg dataset? How about the number of rows? Show the codes and its result.

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

```
datampg <- glimpse(mpg)

nrow(mpg)

#> [1] 234

ncol(mpg)

#> [1] 11
```

#2. Which manufacturer has the most models in this data set? Which model has the most variations? Ans:

#dodge has 37 modelss

```
totalno <- mpg %>%
  group_by(manufacturer) %>%
  tally(sort = TRUE)
```

```
#a. Group the manufacturers and find the unique models. Copy the codes and result.
datampg <- mpg
uniqMods <- datampg %>% group_by(manufacturer, model) %>%
  distinct() %>% count()
uniqMods

colnames(uniqMods) <- c("Manufacturer", "Model","Counts")
uniqMods

#b. Graph the result by using plot() and ggplot(). Write the codes and its result.

#plot
qplot(model, data = mpg,geom = "bar", fill=manufacturer)

#ggplot
ggplot(mpg, aes(model, manufacturer)) + geom_point()
```

#3. Same dataset will be used. You are going to show the relationship of the model and the manufacturer.

```
datampg <- mpg

modfact <- datampg %>% group_by(manufacturer, model) %>%

  distinct() %>% count()

modfact


colnames(modfact) <- c("Manufacturer", "Model")

modfact


#a. What does ggplot(mpg, aes(model, manufacturer)) + geom_point() show?


ggplot(mpg, aes(model, manufacturer)) + geom_point()

#geometric point graph of mpg(model and manufacturer)


#b. For you, is it useful? If not, how could you modify the data to make it more informative?  : Yes, It is useful because you could trackdown the data of each model of the manufacturer

#to modify the data:

ggplot(mpg, aes(model, manufacturer)) +

  geom_point() +

  geom_jitter()
```

#4. Using the pipe (%>%), group the model and get the number of cars per model. Show codes and its result.

```
datampg <- uniqMods %>% group_by(Model) %>% count()

datampg


colnames(datampg) <- c("Model","Counts")
```

#a. Plot using the geom_bar() + coord_flip() just like what is shown below. Show codes and its result

```
qplot(model,data = mpg,main = "Number of Cars per Model", xlab = "Model",ylab = "Number of Cars", geom = "bar", fill = manufacturer) + coord_flip()
```

#b. Use only the top 20 observations. Show code and results.

```
barplot(datampg$count[1:20],names.arg=datampg$manufacturer[1:20])
```

#5. Plot the relationship between cyl - number of cylinders and displ - engine displacement using geom_point with aesthetic colour = engine displacement. Title should be “Relationship between No. of Cylinders and Engine Displacement”.

#a. Show the codes and its result.

```
ggplot(data = mpg , mapping = aes(x = displ, y = cyl, main = "Relationship between No of Cylinders and Engine Displacement")) +
  geom_point(mapping=aes(colour = "engine displacement")) + geom_jitter()
```

#b. How would you describe its relationship?

#I would say according to my data of making cyl the y, the graph is jittered. the pink color indicates the engine displacement and you can see that it is in a straight horizontal position.

#6.Get the total number of observations for drv - type of drive train (f = front-wheel drive, r = rear wheel drive, 4 = 4wd) and class - type of class (Example: suv, 2seater, etc.) Plot using the geom_tile() where the number of observations for class be used as a fill for aesthetics.

#a. Show the codes and its result for the narrative in #6.

```
ggplot(data = mpg, mapping = aes(x = drv, y = class)) +
geom_point(mapping=aes(color=class)) +
  geom_tile()
```

#b. Interpret the result: Areas covered with black are "mapped" using the mapping geometric point graph. y as class and x as drv.

#7. Discuss the difference between these codes. Its outputs for each are shown below.

#Code #1

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, colour = "blue"))
```

#+ Code #2

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy), colour = "blue")
```

#8. Try to run the command ?mpg. What is the result of this command?

```
?mpg
```

#a. Which variables from mpg dataset are categorical?

#Categorical variables in mpg include: manufacturer, model, trans (type of transmission), drv (front-wheel drive, rear-wheel, 4wd), fl (fuel type), and class (type of car).

#b. Which are continuous variables?

#Continuous varibles in R are called doubles or integers.

#c. Plot the relationship between displ (engine displacement) and hwy(highway miles per gallon). Mapped it with a continuous variable you have identified in #5-b.

```
ggplot(mpg, aes(x = displ, y = hwy, colour = cty)) + geom_point()
```

#What is its result? Why it produced such output? : data tracks the cty by placing cty(city miles per gallon) at color having a variation or hues of blue.

#9.Plot the relationship between displ (engine displacement) and hwy(highway miles per gallon) using geom_point(). Add a trend line over the existing plot using geom_smooth() with se = FALSE. Default method is “loess”.

```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
  geom_point(mapping=aes(color=class)) +
  geom_smooth(se = FALSE)
```

#10.Using the relationship of displ and hwy, add a trend line over existing plot. Set the se = FALSE to remove the confidence interval and method = lm to check for linear modeling

```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy, color = class)) +
  geom_point() +
  geom_smooth(se = FALSE)
```