RWorksheet_Octavio#4c

2023-11-22

#1. Use the dataset mpg #1a.Show your solutions on how to import a csv file into the environment.

```r
library(readr)
mpg <- read_csv("mpg.csv")
```

```
## New names:
## Rows: 234 Columns: 12
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (6): manufacturer, model, trans, drv, fl, class dbl (6): ...1, displ, year,
## cyl, cty, hwy
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```r
head(mpg)
```

```
## # A tibble: 6 x 12
##    ...1 manufacturer model displ  year   cyl trans drv     cty   hwy fl    class
##   <dbl> <chr>        <chr> <dbl> <dbl> <dbl> <chr> <chr> <dbl> <dbl> <chr> <chr>
## 1     1 audi         a4      1.8  1999     4 auto~ f        18    29 p     comp~
## 2     2 audi         a4      1.8  1999     4 manu~ f        21    29 p     comp~
## 3     3 audi         a4      2    2008     4 manu~ f        20    31 p     comp~
## 4     4 audi         a4      2    2008     4 auto~ f        21    30 p     comp~
## 5     5 audi         a4      2.8  1999     6 auto~ f        16    26 p     comp~
## 6     6 audi         a4      2.8  1999     6 manu~ f        18    26 p     comp~
```

#1b Which variables from mpg dataset are categorical?

```r
str(mpg)
```

```
## spc_tbl_ [234 x 12] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ...1        : num [1:234] 1 2 3 4 5 6 7 8 9 10 ...
##  $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
##  $ model       : chr [1:234] "a4" "a4" "a4" "a4" ...
##  $ displ       : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
##  $ year        : num [1:234] 1999 1999 2008 2008 1999 ...
##  $ cyl         : num [1:234] 4 4 4 4 6 6 6 4 4 4 ...
##  $ trans       : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
##  $ drv         : chr [1:234] "f" "f" "f" "f" ...
##  $ cty         : num [1:234] 18 21 20 21 16 18 18 18 16 20 ...
##  $ hwy         : num [1:234] 29 29 31 30 26 26 27 26 25 28 ...
##  $ fl          : chr [1:234] "p" "p" "p" "p" ...
##  $ class       : chr [1:234] "compact" "compact" "compact" "compact" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ...1 = col_double(),
##   ..   manufacturer = col_character(),
##   ..   model = col_character(),
##   ..   displ = col_double(),
##   ..   year = col_double(),
##   ..   cyl = col_double(),
##   ..   trans = col_character(),
##   ..   drv = col_character(),
##   ..   cty = col_double(),
##   ..   hwy = col_double(),
##   ..   fl = col_character(),
##   ..   class = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
# manufacturer, model, trans, drv, fl, class variables are categorical
```

#1c. Which are continuous variables?

```
summary(mpg)
```

```
##       ...1       manufacturer          model              displ
##  Min.   :  1.00   Length:234         Length:234         Min.   :1.600
##  1st Qu.: 59.25   Class :character   Class :character   1st Qu.:2.400
##  Median :117.50   Mode  :character   Mode  :character   Median :3.300
##  Mean   :117.50                                         Mean   :3.472
##  3rd Qu.:175.75                                         3rd Qu.:4.600
##  Max.   :234.00                                         Max.   :7.000
##       year           cyl           trans               drv
##  Min.   :1999   Min.   :4.000   Length:234         Length:234
##  1st Qu.:1999   1st Qu.:4.000   Class :character   Class :character
##  Median :2004   Median :6.000   Mode  :character   Mode  :character
##  Mean   :2004   Mean   :5.889
##  3rd Qu.:2008   3rd Qu.:8.000
##  Max.   :2008   Max.   :8.000
##       cty             hwy             fl               class
##  Min.   : 9.00   Min.   :12.00   Length:234         Length:234
##  1st Qu.:14.00   1st Qu.:18.00   Class :character   Class :character
##  Median :17.00   Median :24.00   Mode  :character   Mode  :character
##  Mean   :16.86   Mean   :23.44
##  3rd Qu.:19.00   3rd Qu.:27.00
##  Max.   :35.00   Max.   :44.00
```

```
##The categorical variables are the  X, displ, year, cyl, cty, hwy are continuous variables
```

#2.1 Which manufacturer has the most models in this data set? Which model has the most variations? Show your answer.

```
manufacturer_asTable <- table(mpg$manufacturer)
manufacturer_most_models <- names(manufacturer_asTable)[which.max(manufacturer_asTable)]

manufacturer_most_models
```

```
## [1] "dodge"
```

```
# The dodge manufacturer has the most models



model_asTable <- table(mpg$model)
model_most_vars <- names(model_asTable)[which.max(model_asTable)]
```

```
model_most_vars
```

```
## [1] "caravan 2wd"
```

```
# The caravan 2wd has the most variations
```

#2.1a Group the manufacturers and find the unique models. Show your codes and result.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
manufacturers_models <- data.frame(Manufacturer = mpg$manufacturer, Model = mpg$model)
manufacturers_models
```

```
##     Manufacturer                  Model
## 1           audi                     a4
## 2           audi                     a4
## 3           audi                     a4
## 4           audi                     a4
## 5           audi                     a4
## 6           audi                     a4
## 7           audi                     a4
## 8           audi             a4 quattro
## 9           audi             a4 quattro
## 10          audi             a4 quattro
## 11          audi             a4 quattro
## 12          audi             a4 quattro
## 13          audi             a4 quattro
## 14          audi             a4 quattro
## 15          audi             a4 quattro
## 16          audi             a6 quattro
## 17          audi             a6 quattro
## 18          audi             a6 quattro
## 19     chevrolet     c1500 suburban 2wd
## 20     chevrolet     c1500 suburban 2wd
## 21     chevrolet     c1500 suburban 2wd
## 22     chevrolet     c1500 suburban 2wd
## 23     chevrolet     c1500 suburban 2wd
## 24     chevrolet               corvette
## 25     chevrolet               corvette
## 26     chevrolet               corvette
## 27     chevrolet               corvette
## 28     chevrolet               corvette
## 29     chevrolet        k1500 tahoe 4wd
## 30     chevrolet        k1500 tahoe 4wd
## 31     chevrolet        k1500 tahoe 4wd
## 32     chevrolet        k1500 tahoe 4wd
## 33     chevrolet                 malibu
## 34     chevrolet                 malibu
## 35     chevrolet                 malibu
## 36     chevrolet                 malibu
## 37     chevrolet                 malibu
## 38         dodge            caravan 2wd
## 39         dodge            caravan 2wd
## 40         dodge            caravan 2wd
## 41         dodge            caravan 2wd
## 42         dodge            caravan 2wd
## 43         dodge            caravan 2wd
## 44         dodge            caravan 2wd
## 45         dodge            caravan 2wd
## 46         dodge            caravan 2wd
## 47         dodge            caravan 2wd
## 48         dodge            caravan 2wd
## 49         dodge      dakota pickup 4wd
## 50         dodge      dakota pickup 4wd
## 51         dodge      dakota pickup 4wd
## 52         dodge      dakota pickup 4wd
## 53         dodge      dakota pickup 4wd
## 54         dodge      dakota pickup 4wd
## 55         dodge      dakota pickup 4wd
## 56         dodge      dakota pickup 4wd
## 57         dodge      dakota pickup 4wd
## 58         dodge            durango 4wd
## 59         dodge            durango 4wd
## 60         dodge            durango 4wd
## 61         dodge            durango 4wd
## 62         dodge            durango 4wd
## 63         dodge            durango 4wd
## 64         dodge            durango 4wd
## 65         dodge    ram 1500 pickup 4wd
## 66         dodge    ram 1500 pickup 4wd
## 67         dodge    ram 1500 pickup 4wd
## 68         dodge    ram 1500 pickup 4wd
## 69         dodge    ram 1500 pickup 4wd
## 70         dodge    ram 1500 pickup 4wd
## 71         dodge    ram 1500 pickup 4wd
## 72         dodge    ram 1500 pickup 4wd
## 73         dodge    ram 1500 pickup 4wd
## 74         dodge    ram 1500 pickup 4wd
## 75          ford         expedition 2wd
## 76          ford         expedition 2wd
## 77          ford         expedition 2wd
## 78          ford           explorer 4wd
## 79          ford           explorer 4wd
## 80          ford           explorer 4wd
## 81          ford           explorer 4wd
## 82          ford           explorer 4wd
## 83          ford           explorer 4wd
## 84          ford        f150 pickup 4wd
## 85          ford        f150 pickup 4wd
## 86          ford        f150 pickup 4wd
## 87          ford        f150 pickup 4wd
## 88          ford        f150 pickup 4wd
## 89          ford        f150 pickup 4wd
## 90          ford        f150 pickup 4wd
## 91          ford                mustang
## 92          ford                mustang
## 93          ford                mustang
## 94          ford                mustang
## 95          ford                mustang
## 96          ford                mustang
## 97          ford                mustang
## 98          ford                mustang
## 99          ford                mustang
## 100        honda                  civic
## 101        honda                  civic
## 102        honda                  civic
## 103        honda                  civic
## 104        honda                  civic
## 105        honda                  civic
## 106        honda                  civic
## 107        honda                  civic
## 108        honda                  civic
## 109      hyundai                 sonata
## 110      hyundai                 sonata
## 111      hyundai                 sonata
## 112      hyundai                 sonata
## 113      hyundai                 sonata
## 114      hyundai                 sonata
## 115      hyundai                 sonata
## 116      hyundai                tiburon
## 117      hyundai                tiburon
## 118      hyundai                tiburon
## 119      hyundai                tiburon
## 120      hyundai                tiburon
## 121      hyundai                tiburon
## 122      hyundai                tiburon
## 123         jeep     grand cherokee 4wd
## 124         jeep     grand cherokee 4wd
## 125         jeep     grand cherokee 4wd
## 126         jeep     grand cherokee 4wd
## 127         jeep     grand cherokee 4wd
## 128         jeep     grand cherokee 4wd
## 129         jeep     grand cherokee 4wd
## 130         jeep     grand cherokee 4wd
## 131   land rover            range rover
## 132   land rover            range rover
## 133   land rover            range rover
## 134   land rover            range rover
## 135      lincoln          navigator 2wd
## 136      lincoln          navigator 2wd
## 137      lincoln          navigator 2wd
## 138      mercury        mountaineer 4wd
## 139      mercury        mountaineer 4wd
## 140      mercury        mountaineer 4wd
## 141      mercury        mountaineer 4wd
## 142       nissan                 altima
## 143       nissan                 altima
## 144       nissan                 altima
## 145       nissan                 altima
## 146       nissan                 altima
## 147       nissan                 altima
## 148       nissan                 maxima
## 149       nissan                 maxima
## 150       nissan                 maxima
## 151       nissan         pathfinder 4wd
## 152       nissan         pathfinder 4wd
## 153       nissan         pathfinder 4wd
## 154       nissan         pathfinder 4wd
## 155      pontiac             grand prix
## 156      pontiac             grand prix
## 157      pontiac             grand prix
## 158      pontiac             grand prix
## 159      pontiac             grand prix
## 160       subaru           forester awd
## 161       subaru           forester awd
## 162       subaru           forester awd
## 163       subaru           forester awd
## 164       subaru           forester awd
## 165       subaru           forester awd
## 166       subaru            impreza awd
## 167       subaru            impreza awd
## 168       subaru            impreza awd
## 169       subaru            impreza awd
## 170       subaru            impreza awd
## 171       subaru            impreza awd
## 172       subaru            impreza awd
## 173       subaru            impreza awd
## 174       toyota            4runner 4wd
## 175       toyota            4runner 4wd
## 176       toyota            4runner 4wd
## 177       toyota            4runner 4wd
## 178       toyota            4runner 4wd
## 179       toyota            4runner 4wd
## 180       toyota                  camry
## 181       toyota                  camry
## 182       toyota                  camry
## 183       toyota                  camry
## 184       toyota                  camry
## 185       toyota                  camry
## 186       toyota                  camry
## 187       toyota           camry solara
## 188       toyota           camry solara
## 189       toyota           camry solara
## 190       toyota           camry solara
## 191       toyota           camry solara
## 192       toyota           camry solara
## 193       toyota           camry solara
## 194       toyota                corolla
## 195       toyota                corolla
## 196       toyota                corolla
## 197       toyota                corolla
## 198       toyota                corolla
## 199       toyota land cruiser wagon 4wd
## 200       toyota land cruiser wagon 4wd
## 201       toyota      toyota tacoma 4wd
## 202       toyota      toyota tacoma 4wd
## 203       toyota      toyota tacoma 4wd
## 204       toyota      toyota tacoma 4wd
## 205       toyota      toyota tacoma 4wd
## 206       toyota      toyota tacoma 4wd
## 207       toyota      toyota tacoma 4wd
## 208   volkswagen                    gti
## 209   volkswagen                    gti
## 210   volkswagen                    gti
## 211   volkswagen                    gti
## 212   volkswagen                    gti
## 213   volkswagen                  jetta
## 214   volkswagen                  jetta
## 215   volkswagen                  jetta
## 216   volkswagen                  jetta
## 217   volkswagen                  jetta
## 218   volkswagen                  jetta
## 219   volkswagen                  jetta
## 220   volkswagen                  jetta
## 221   volkswagen                  jetta
## 222   volkswagen             new beetle
## 223   volkswagen             new beetle
## 224   volkswagen             new beetle
## 225   volkswagen             new beetle
## 226   volkswagen             new beetle
## 227   volkswagen             new beetle
## 228   volkswagen                 passat
## 229   volkswagen                 passat
## 230   volkswagen                 passat
## 231   volkswagen                 passat
## 232   volkswagen                 passat
## 233   volkswagen                 passat
## 234   volkswagen                 passat
```

```r
unique_models <- unique(manufacturers_models)
unique_models
```

```
##     Manufacturer                  Model
## 1           audi                     a4
## 8           audi             a4 quattro
## 16          audi             a6 quattro
## 19     chevrolet     c1500 suburban 2wd
## 24     chevrolet               corvette
## 29     chevrolet        k1500 tahoe 4wd
## 33     chevrolet                 malibu
## 38         dodge            caravan 2wd
## 49         dodge      dakota pickup 4wd
## 58         dodge            durango 4wd
## 65         dodge    ram 1500 pickup 4wd
## 75          ford         expedition 2wd
## 78          ford           explorer 4wd
## 84          ford        f150 pickup 4wd
## 91          ford                mustang
## 100        honda                  civic
## 109      hyundai                 sonata
## 116      hyundai                tiburon
## 123         jeep     grand cherokee 4wd
## 131   land rover            range rover
## 135      lincoln          navigator 2wd
## 138      mercury        mountaineer 4wd
## 142       nissan                 altima
## 148       nissan                 maxima
## 151       nissan         pathfinder 4wd
## 155      pontiac             grand prix
## 160       subaru           forester awd
## 166       subaru            impreza awd
## 174       toyota            4runner 4wd
## 180       toyota                  camry
## 187       toyota           camry solara
## 194       toyota                corolla
## 199       toyota land cruiser wagon 4wd
## 201       toyota      toyota tacoma 4wd
## 208   volkswagen                    gti
## 213   volkswagen                  jetta
## 222   volkswagen             new beetle
## 228   volkswagen                 passat
```

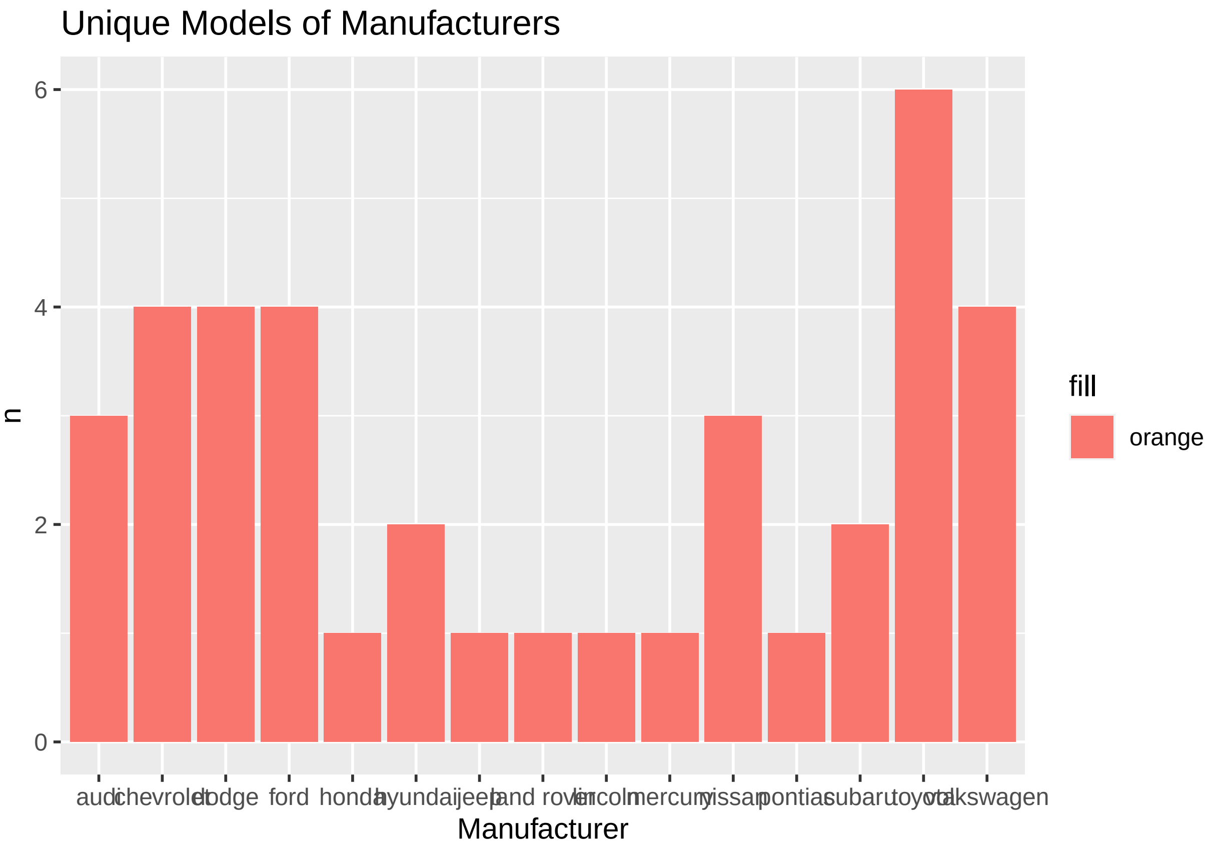
```
unique_models_factor <- factoredManufacturer <- as.factor(unique_models$Manufacturer)
```

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

```
#install.packages("ggplot2")
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'

## The following object is masked _by_ '.GlobalEnv':
##
##     mpg
```

```
#install.packages("dplyr")
library(dplyr)

unique_count <- unique_models %>%
  count(Manufacturer)

ggplot(unique_count, aes(x = Manufacturer, y = n, fill = "orange")) +
  geom_bar(stat = "identity") +
  labs(title = "Unique Models of Manufacturers",
       x = "Manufacturer")
```



#2. Same dataset will be used. You are going to show the relationship of the modeland the manufacturer.

#2a. . Group the manufacturers and find the unique models. Show your codes and result.

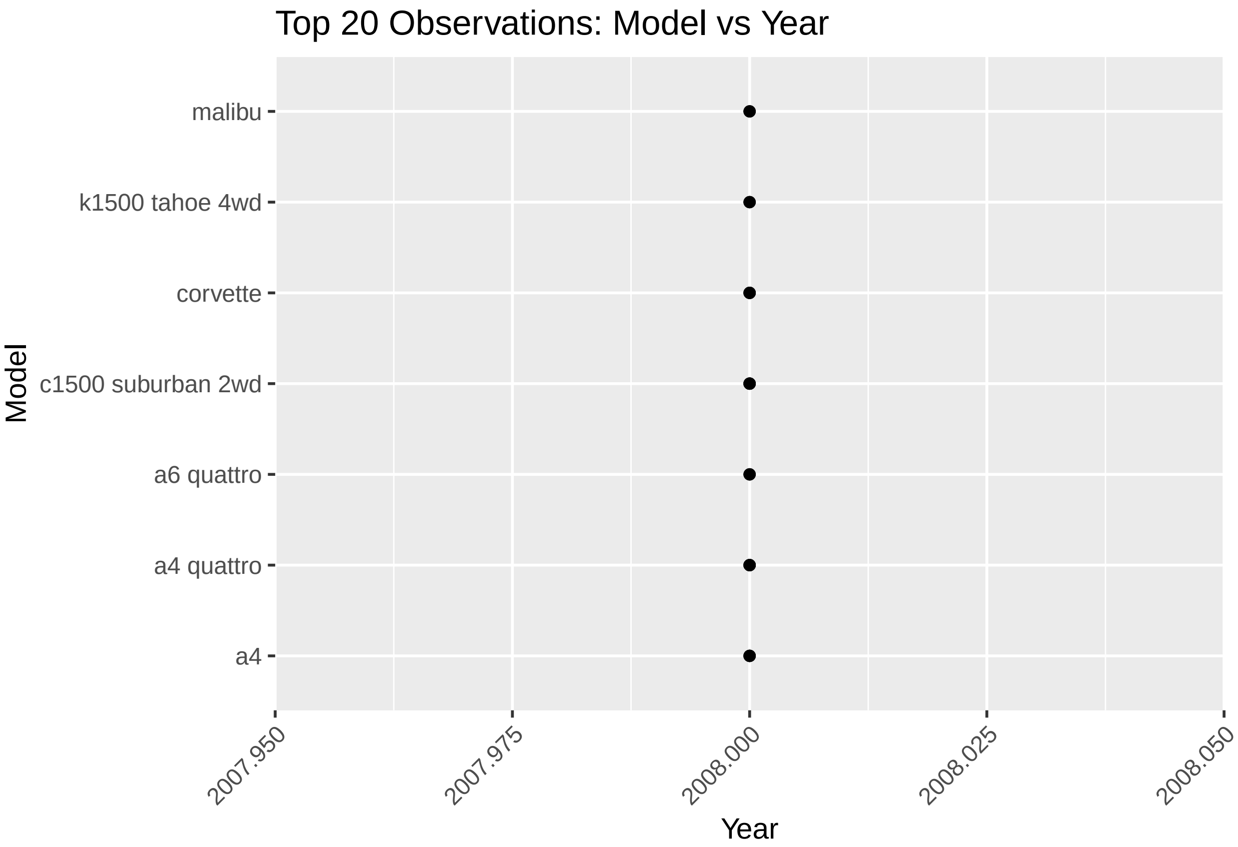
```
# It creates ascatterplot of the mpg dataset with model on the x-axis and manufacturer on the y-axis.
# In this plot,  Each point on the plot represents a specific model and its corresponding manufacturer.
```

2b. For you, is it useful? If not, how could you modify the data to make it more informative?

```
#The plot may be useful for visualizing the distribution of models across manufacturers, but it could b
```

#3. Plot the model and the year using ggplot(). Use only the top 20 observations. Write the codes and its results.

```
top_20_data <- head(mpg[order(mpg$year, decreasing = TRUE), ], 20)
ggplot(top_20_data, aes(x = year, y = model)) +
  geom_point() +
  labs(title = "Top 20 Observations: Model vs Year", x = "Year", y = "Model") + theme(axis.text.x = elem
```



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

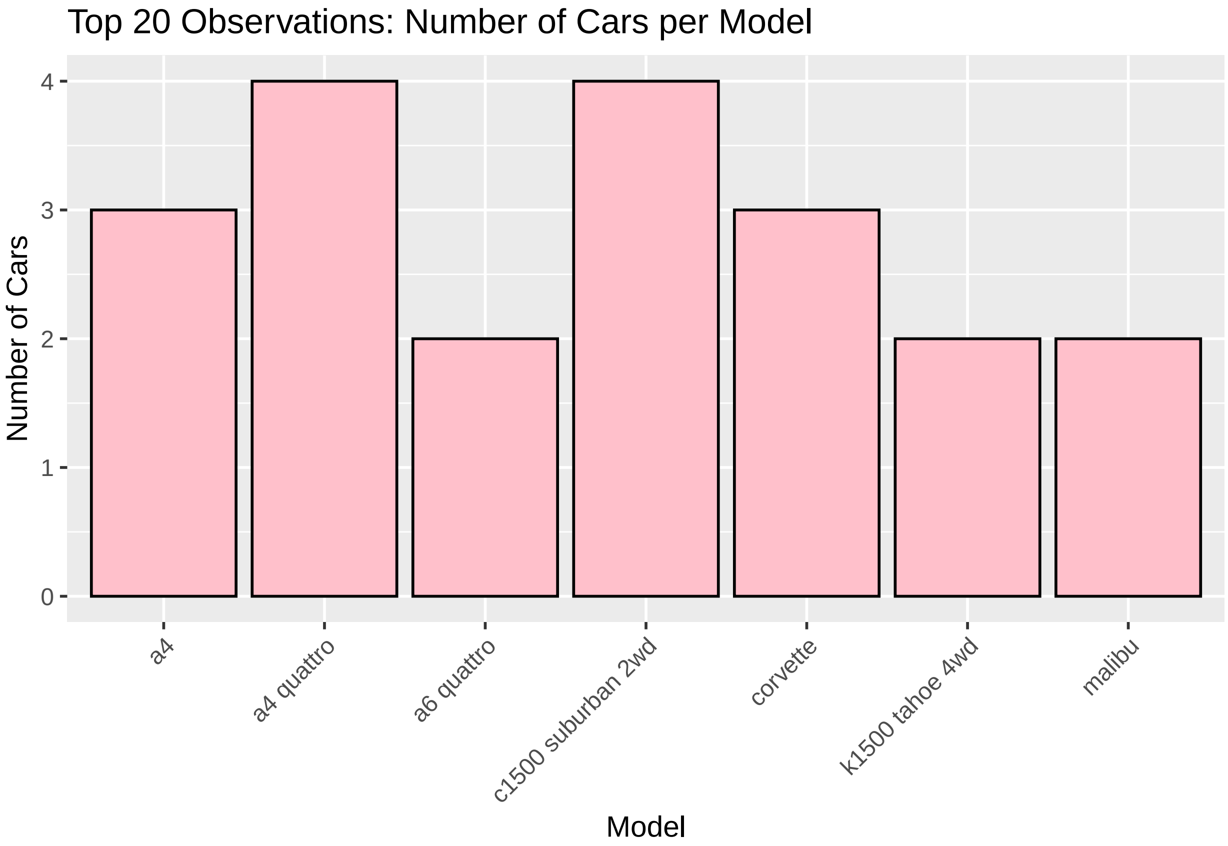
```
cars_per_model <- mpg %>%
  group_by(model) %>%
  summarise(num_cars = n())

print(cars_per_model)
```

```
## # A tibble: 38 x 2
##    model              num_cars
##    <chr>                 <int>
##  1 4runner 4wd               6
##  2 a4                        7
##  3 a4 quattro                8
##  4 a6 quattro                3
##  5 altima                    6
##  6 c1500 suburban 2wd        5
##  7 camry                     7
##  8 camry solara              7
##  9 caravan 2wd              11
## 10 civic                     9
## # i 28 more rows
```

4a. Plot using geom_bar() using the top 20 observations only. The graphs shoudl have a title, labels and colors. Show code and results.

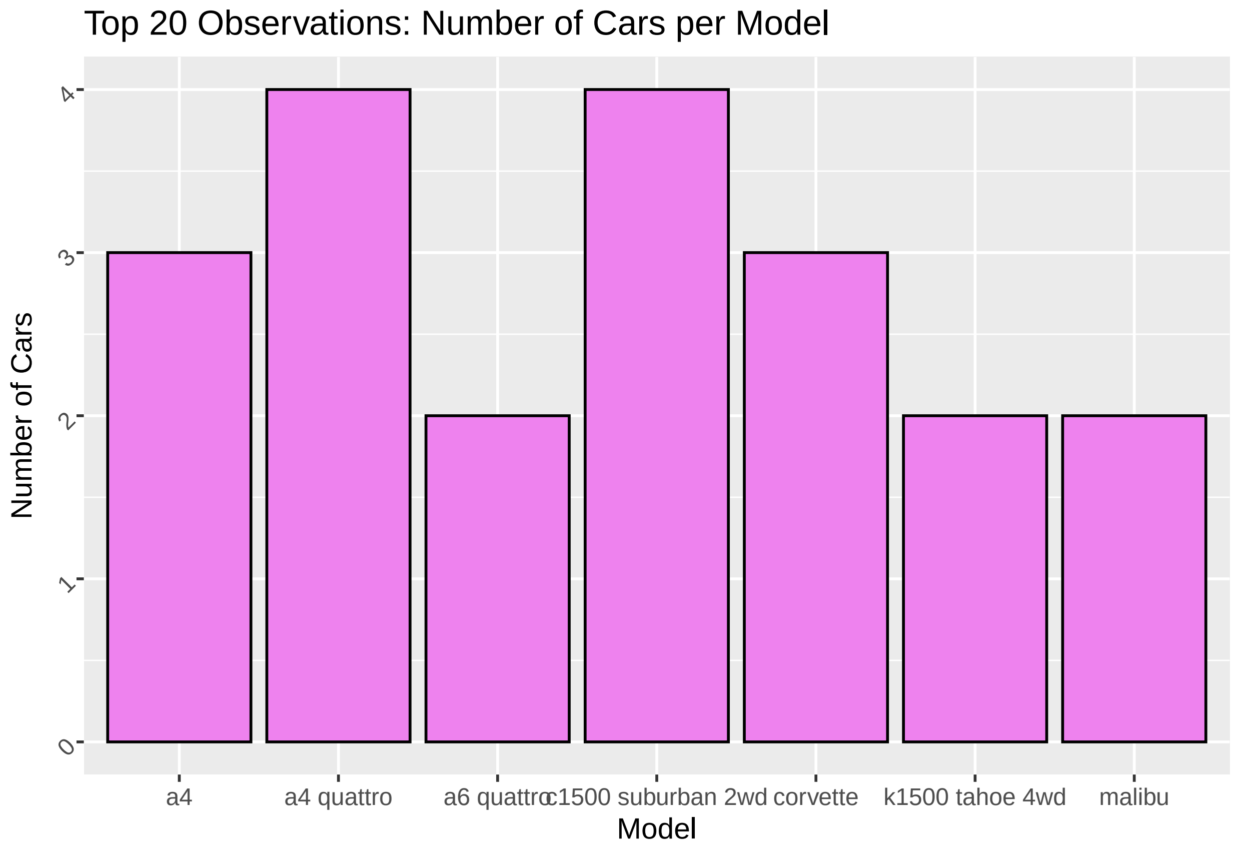
```
top_20_data <- head(mpg[order(mpg$year, decreasing = TRUE), ], 20)
```

```r
ggplot(top_20_data, aes(x = model)) +
  geom_bar(fill = "pink", color = "black") +
  labs(title = "Top 20 Observations: Number of Cars per Model",
       x = "Model", y = "Number of Cars") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



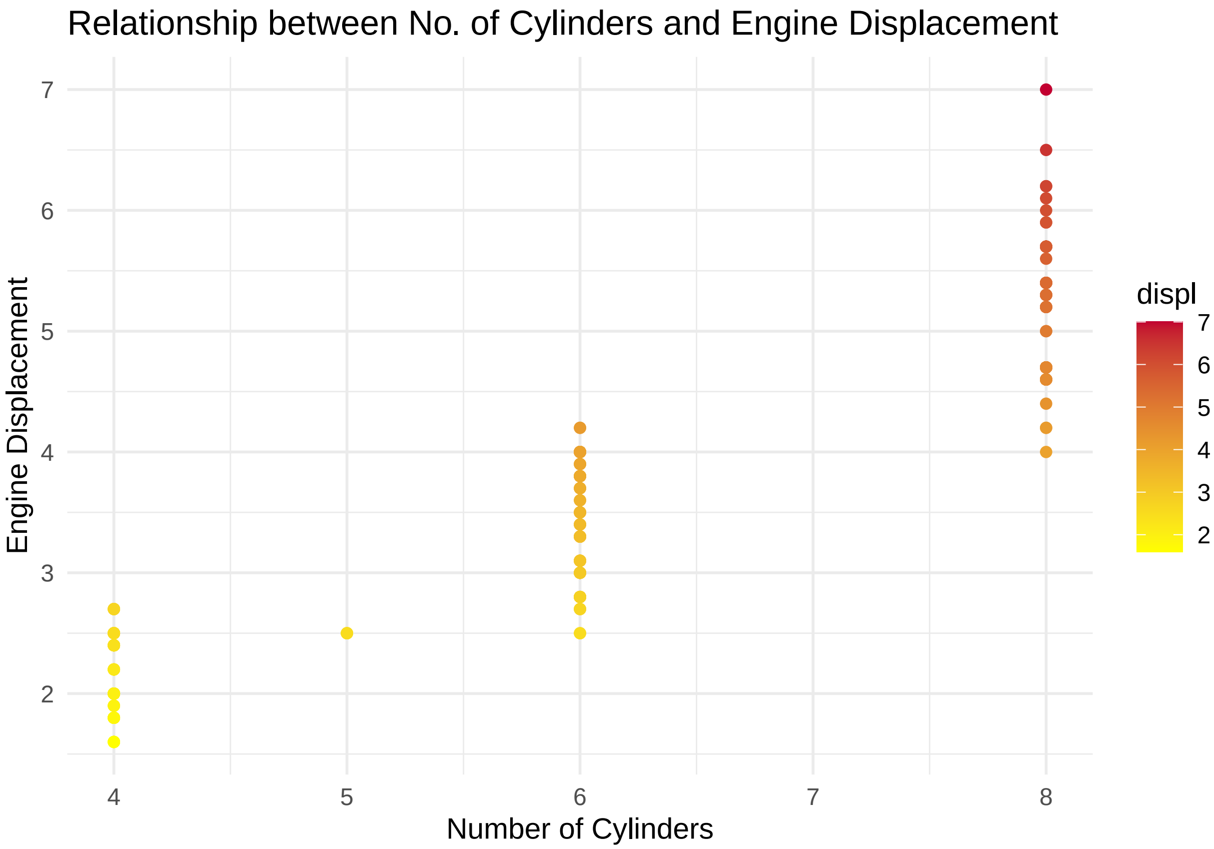
#4b. Plot using the geom_bar() + coord_flip() just like what is shown below. Show codes and its result.

```r
ggplot(top_20_data, aes(y = model)) +
  geom_bar(fill = "violet", color = "black") +
  labs(title = "Top 20 Observations: Number of Cars per Model",
       x = "Number of Cars", y = "Model") +
  coord_flip() +
  theme(axis.text.y = element_text(angle = 45, hjust = 1))
```



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

```
ggplot(mpg, aes(x = cyl, y = displ, color = displ)) +
  geom_point() +
  labs(title = "Relationship between No. of Cylinders and Engine Displacement",
       x = "Number of Cylinders", y = "Engine Displacement") +
  scale_color_gradient(low = "yellow", high = "#C20030") +
  theme_minimal()
```

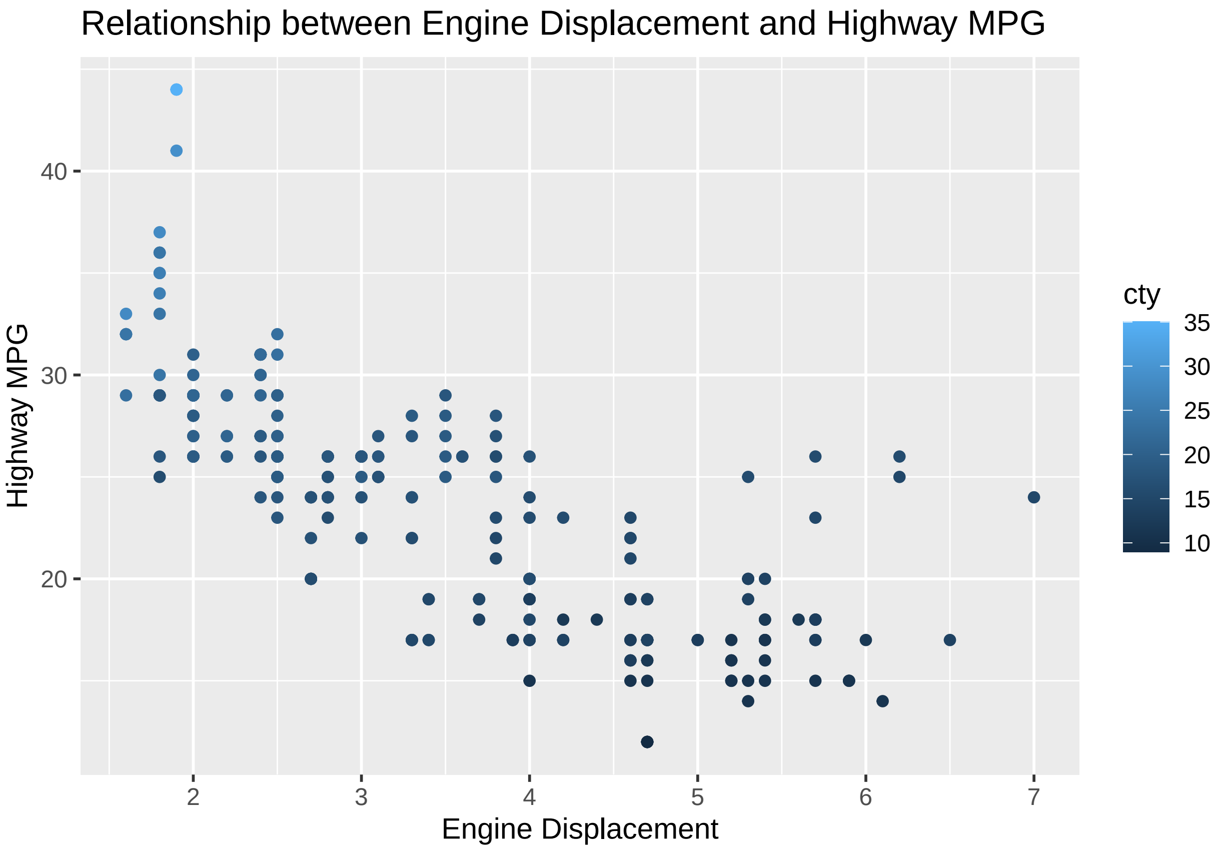


5a. How would you describe its relationship? Show the codes and its result.

```
# It will generate a scatter plot showing the relationship between the number of cylinders and engine d

# As the number of cylinders increases, the engine displacement tends to increase as well. This suggest
```

6. Plot the relationship between displ (engine displacement) and hwy(highway miles per gallon). Mapped it with a continuous variable you have identified in #1-c. What is its result? Why it produced such output?

```
displ_hwy_plot <- ggplot(mpg, aes(x = displ, y = hwy, color = cty)) +
  geom_point() +
  labs(title = "Relationship between Engine Displacement and Highway MPG",
       x = "Engine Displacement",
       y = "Highway MPG")

displ_hwy_plot
```



```
# This is a scatterplot with engine displacement on the x-axis and highway miles per gallon on the y-ax

# Using this plot, we can understand the relationship between the displ, hwy, and cty. By mapping the c

# This can provide understanding of the fuel efficiency of vehicle with different engine sizes
```

6. Import the traffic.csv onto your R environment.

```
library(readr)
traffic <- read_csv("traffic.csv")
```

```
## Rows: 48120 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (3): Junction, Vehicles, ID
## dttm (1): DateTime
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
traffic
```

```
## # A tibble: 48,120 x 4
##    DateTime            Junction Vehicles          ID
##    <dttm>                 <dbl>    <dbl>       <dbl>
##  1 2015-11-01 00:00:00        1       15 20151101001
##  2 2015-11-01 01:00:00        1       13 20151101011
##  3 2015-11-01 02:00:00        1       10 20151101021
##  4 2015-11-01 03:00:00        1        7 20151101031
##  5 2015-11-01 04:00:00        1        9 20151101041
##  6 2015-11-01 05:00:00        1        6 20151101051
##  7 2015-11-01 06:00:00        1        9 20151101061
##  8 2015-11-01 07:00:00        1        8 20151101071
##  9 2015-11-01 08:00:00        1       11 20151101081
## 10 2015-11-01 09:00:00        1       12 20151101091
## # i 48,110 more rows
```

```r
num_obs <- nrow(traffic)
num_obs
```

```
## [1] 48120
```

```r
num_vars <- ncol(traffic)
num_vars
```

```
## [1] 4
```

```r
vars <- colnames(traffic)
vars
```

```
## [1] "DateTime" "Junction" "Vehicles" "ID"
```

6a. How many numbers of observation does it have? What are the variables of the traffic dataset the Show your answer.

```r
num_obs <- nrow(traffic)
num_obs
```

```
## [1] 48120
```

```r
num_vars <- ncol(traffic)
num_vars
```

```
## [1] 4
```

```r
vars <- colnames(traffic)
vars
```

```
## [1] "DateTime" "Junction" "Vehicles" "ID"
```

6b. subset the traffic dataset into junctions. What is the R codes and its output?

```r
junctions_subset_1 <- subset(traffic, Junction == 1)

junctions_subset_2 <- subset(traffic, Junction == 2)

junctions_subset_3 <- subset(traffic, Junction == 3)

junctions_subset_4 <- subset(traffic, Junction == 4)
```

6c. Plot each junction in a using geom_line(). Show your solution and output.

#7. From alexa_file.xlsx, import it to your environment

```r
library(readxl)
alexa_file <- read_excel("alexa_file.xlsx")
alexa_file
```

```
## # A tibble: 3,150 x 5
##    rating date                variation           verified_reviews      feedback
##     <dbl> <dttm>              <chr>               <chr>                    <dbl>
##  1      5 2018-07-31 00:00:00 Charcoal Fabric     Love my Echo!               1
##  2      5 2018-07-31 00:00:00 Charcoal Fabric     Loved it!                   1
##  3      4 2018-07-31 00:00:00 Walnut Finish       Sometimes while play~       1
##  4      5 2018-07-31 00:00:00 Charcoal Fabric     I have had a lot of ~       1
##  5      5 2018-07-31 00:00:00 Charcoal Fabric     Music                       1
##  6      5 2018-07-31 00:00:00 Heather Gray Fabric I received the echo ~       1
##  7      3 2018-07-31 00:00:00 Sandstone Fabric    Without having a cel~       1
##  8      5 2018-07-31 00:00:00 Charcoal Fabric     I think this is the ~       1
##  9      5 2018-07-30 00:00:00 Heather Gray Fabric looks great                 1
## 10      5 2018-07-30 00:00:00 Heather Gray Fabric Love it! I've listen~       1
## # i 3,140 more rows
```

7a. How many observations does alexa_file has? What about the number of columns? Show your solution and answer.

```
number_obs <- nrow(alexa_file)
number_obs
```

```
## [1] 3150
```

```
number_cols <- ncol(alexa_file)
number_cols
```

```
## [1] 5
```

7b. group the variations and get the total of each variations. Use dplyr package. Show solution and answer.

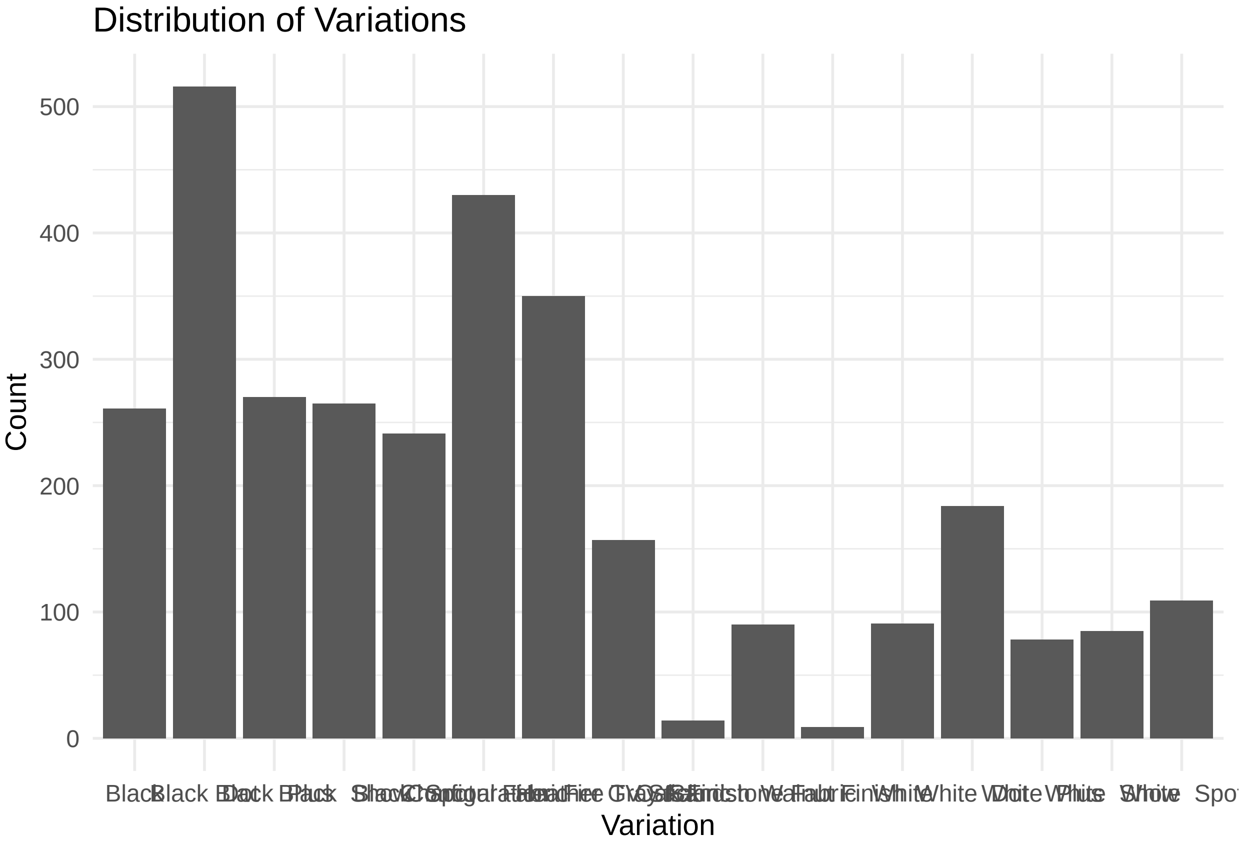
```
groupA <- alexa_file %>%
  group_by(variation) %>%
  summarise(totalcount_=n())
groupA
```

```
## # A tibble: 16 x 2
##    variation                    totalcount_
##    <chr>                              <int>
##  1 Black                                261
##  2 Black  Dot                           516
##  3 Black  Plus                          270
##  4 Black  Show                          265
##  5 Black  Spot                          241
##  6 Charcoal Fabric                      430
##  7 Configuration: Fire TV Stick         350
##  8 Heather Gray Fabric                  157
##  9 Oak Finish                            14
## 10 Sandstone Fabric                      90
## 11 Walnut Finish                          9
## 12 White                                 91
## 13 White  Dot                           184
## 14 White  Plus                           78
## 15 White  Show                           85
## 16 White  Spot                          109
```

7c. Plot the variations using the ggplot() function. What did you observe? Complete the details of the graph. Show solution and answer.

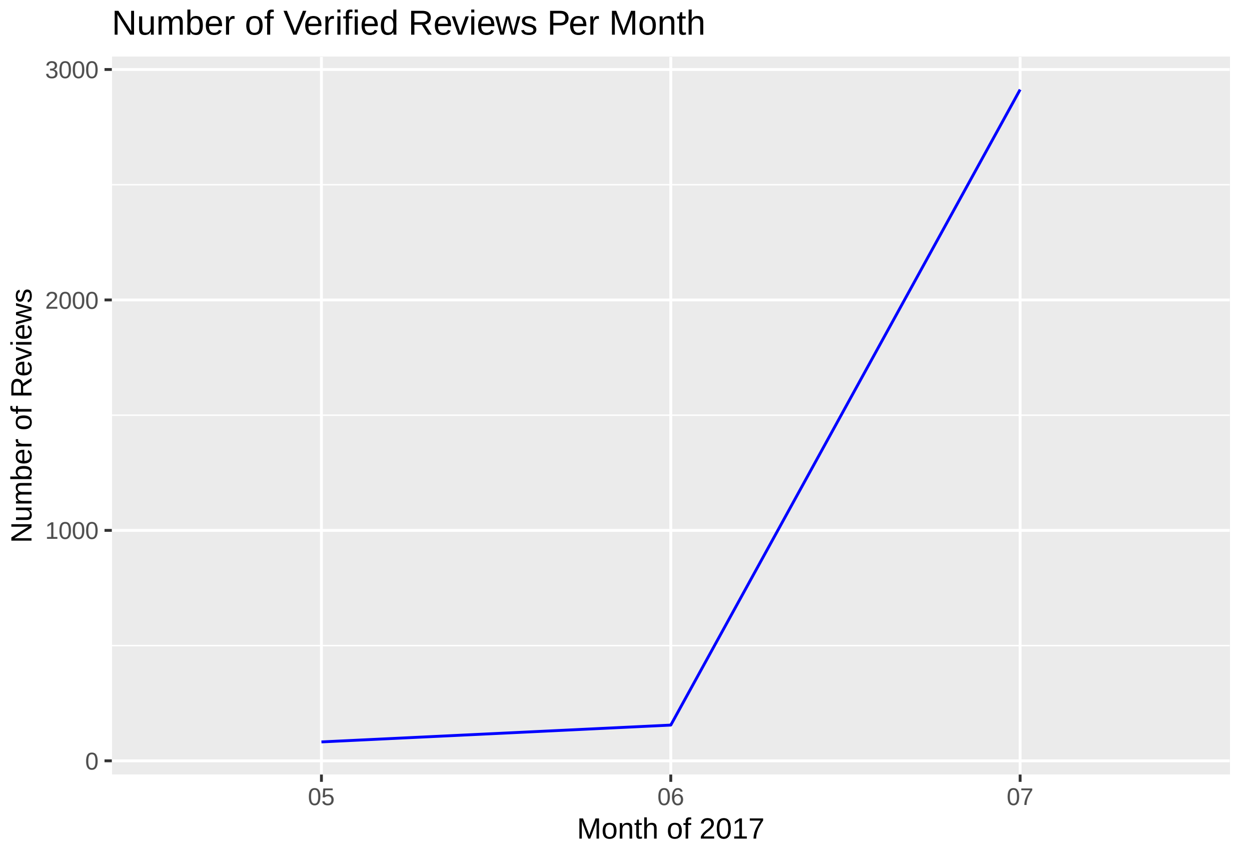
```
library(ggplot2)

ggplot(alexa_file, aes(x = variation)) +
```

```r
  geom_bar() +
  labs(title = "Distribution of Variations",
       x = "Variation",
       y = "Count") +
  theme_minimal()
```



7d. Plot a geom_line() with the date and the number of verified reviews. Complete the details of the graphs. Show your answer and solution.

```r
alexa_file$date <- as.Date(alexa_file$date)

alexa_file$month <- format(alexa_file$date, "%m")

monthscount <- alexa_file %>%
  group_by(month) %>%
  summarise(num_reviews = n())


monthlyreview <- table(monthscount)

ggplot(monthscount, aes(x = month, y = num_reviews, group = 1)) +
  geom_line(color = "blue") +
  labs(title = "Number of Verified Reviews Per Month",
       x = "Month of 2017", y = "Number of Reviews")
```



7e. Get the relationship of variations and ratings. Which variations got the most highest in rating? Plot a graph to show its relationship. Show your solution and answer.

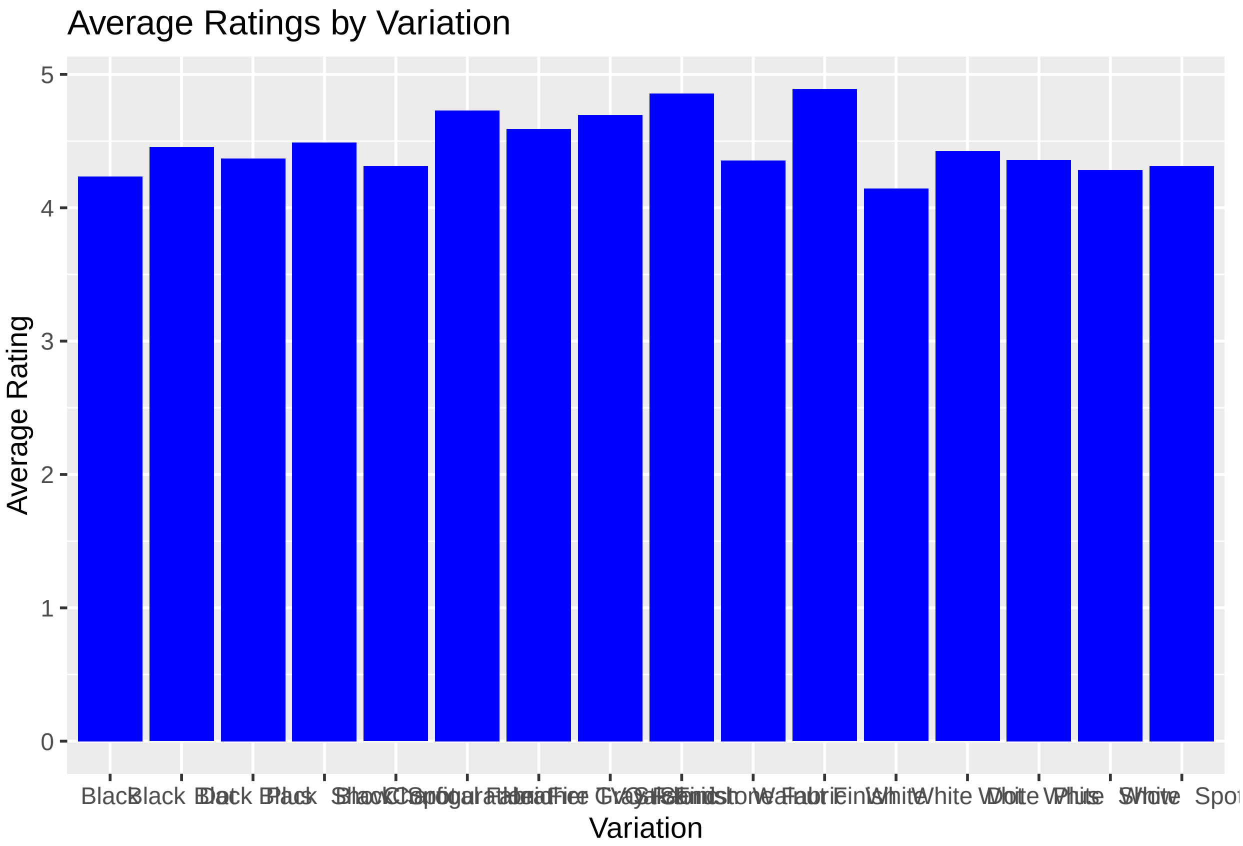
```
library(dplyr)
variation_ratings <- alexa_file %>%
  group_by(variation) %>%
  summarise(avg_rating = mean(rating))

variation_ratings
```

```
## # A tibble: 16 x 2
##    variation                    avg_rating
##    <chr>                             <dbl>
##  1 Black                              4.23
##  2 Black  Dot                         4.45
##  3 Black  Plus                        4.37
##  4 Black  Show                        4.49
##  5 Black  Spot                        4.31
##  6 Charcoal Fabric                    4.73
##  7 Configuration: Fire TV Stick       4.59
##  8 Heather Gray Fabric                4.69
##  9 Oak Finish                         4.86
## 10 Sandstone Fabric                   4.36
## 11 Walnut Finish                      4.89
## 12 White                              4.14
## 13 White  Dot                         4.42
## 14 White  Plus                        4.36
## 15 White  Show                        4.28
## 16 White  Spot                           4.31
```

```
highest_ratings <- variation_ratings %>%
  filter(avg_rating == max(avg_rating))

highest_ratings
```

```
## # A tibble: 1 x 2
##   variation     avg_rating
##   <chr>              <dbl>
## 1 Walnut Finish       4.89
```

```
# The walnut finish variation has the highest rating

ggplot(variation_ratings, aes(x = variation, y = avg_rating)) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(title = "Average Ratings by Variation",
       x = "Variation",
       y = "Average Rating")
```


Average Ratings by Variation
Average Rating
Variation
Black
Spot