

## RWorksheet\_Arcenas#4c

Edora Frances Anne V Arcenas

2023-11-22

```
mpg_file <- read.csv("mpg.csv")
```

```
Categor <- str(mpg_file)
```

```
## 'data.frame':    234 obs. of  12 variables:
## $ X              : int  1 2 3 4 5 6 7 8 9 10 ...
## $ manufacturer   : chr  "audi" "audi" "audi" "audi" ...
## $ model          : chr  "a4" "a4" "a4" "a4" ...
## $ displ          : num  1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
## $ year           : int  1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
## $ cyl            : int  4 4 4 4 6 6 6 4 4 4 ...
## $ trans          : chr  "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
## $ drv            : chr  "f" "f" "f" "f" ...
## $ cty            : int  18 21 20 21 16 18 18 18 16 20 ...
## $ hwy            : int  29 29 31 30 26 26 27 26 25 28 ...
## $ fl             : chr  "p" "p" "p" "p" ...
## $ class          : chr  "compact" "compact" "compact" "compact" ...
```

```
print("The variables that is continuous in the MPG data set are the following:
```

Disp, Cyl, Cty, and Hwy")

```
## [1] "The variables that is continuous in the MPG data set are the following:\n\n"
```

```
summary(mpg_file)
```

##	X	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

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

manufacturer_most_models

## [1] "dodge"
print("Dodge manufacturer has the most models")

## [1] "Dodge manufacturer has the most models"
model_asTable <- table(mpg_file$model)
model_most_vars <- names(model_asTable)[which.max(model_asTable)]

model_most_vars

## [1] "caravan 2wd"
print("caravan 2wd has the most variations")

## [1] "caravan 2wd has the most variations"
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_file$manufacturer, Model = mpg_file$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
```

```
unique_mods <- unique(manufacturers_models)
unique_mods
```

```
##      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
```

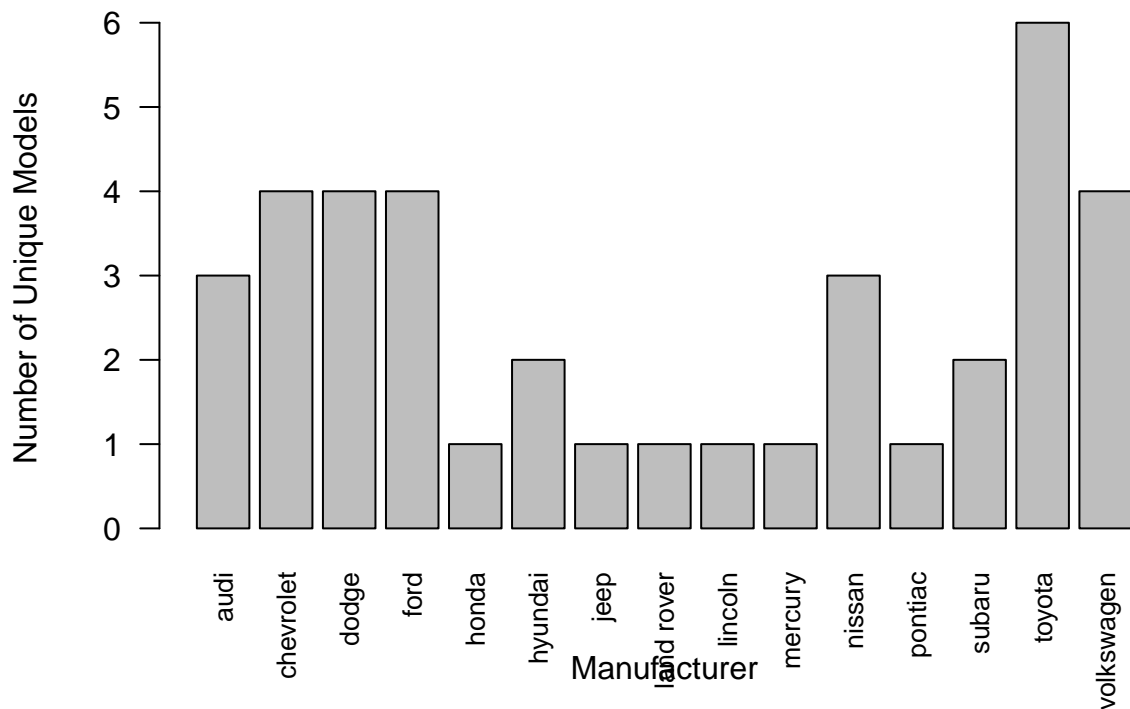
```
unique_mods_factor <- factoredManufacturer <- as.factor(unique_mods$Manufacturer)
```

```
library(ggplot2)
library(dplyr)
```

```
uniqueModMan <- plot(as.factor(factoredManufacturer),
  main = "Unique Models of Manufacturers",
  xlab = "Manufacturer",
  ylab = "Number of Unique Models",
```

```
cex.names = 0.8, las = 2)
```

## Unique Models of Manufacturers

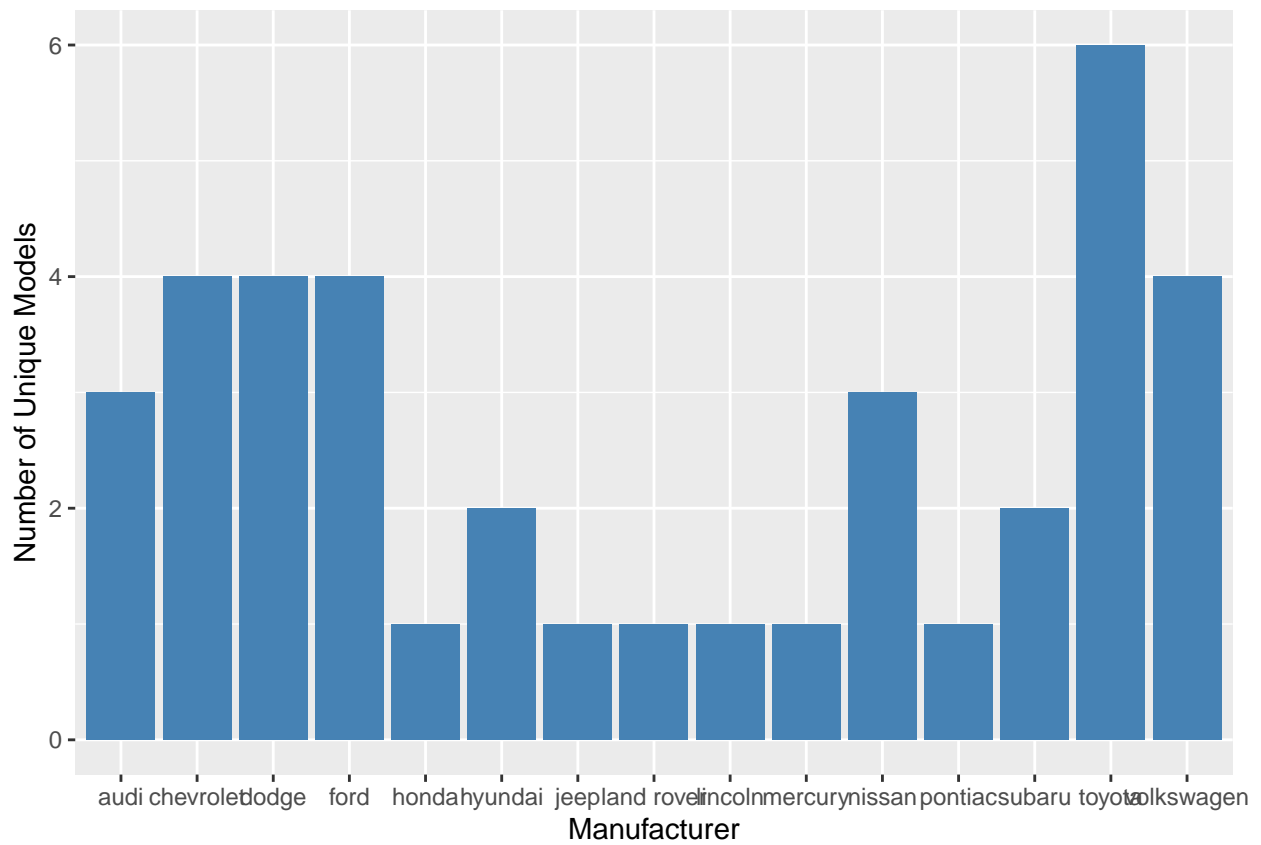


```
unique_count <- unique_mods %>%  
  count(unique_mods$Manufacturer)  
unique_count
```

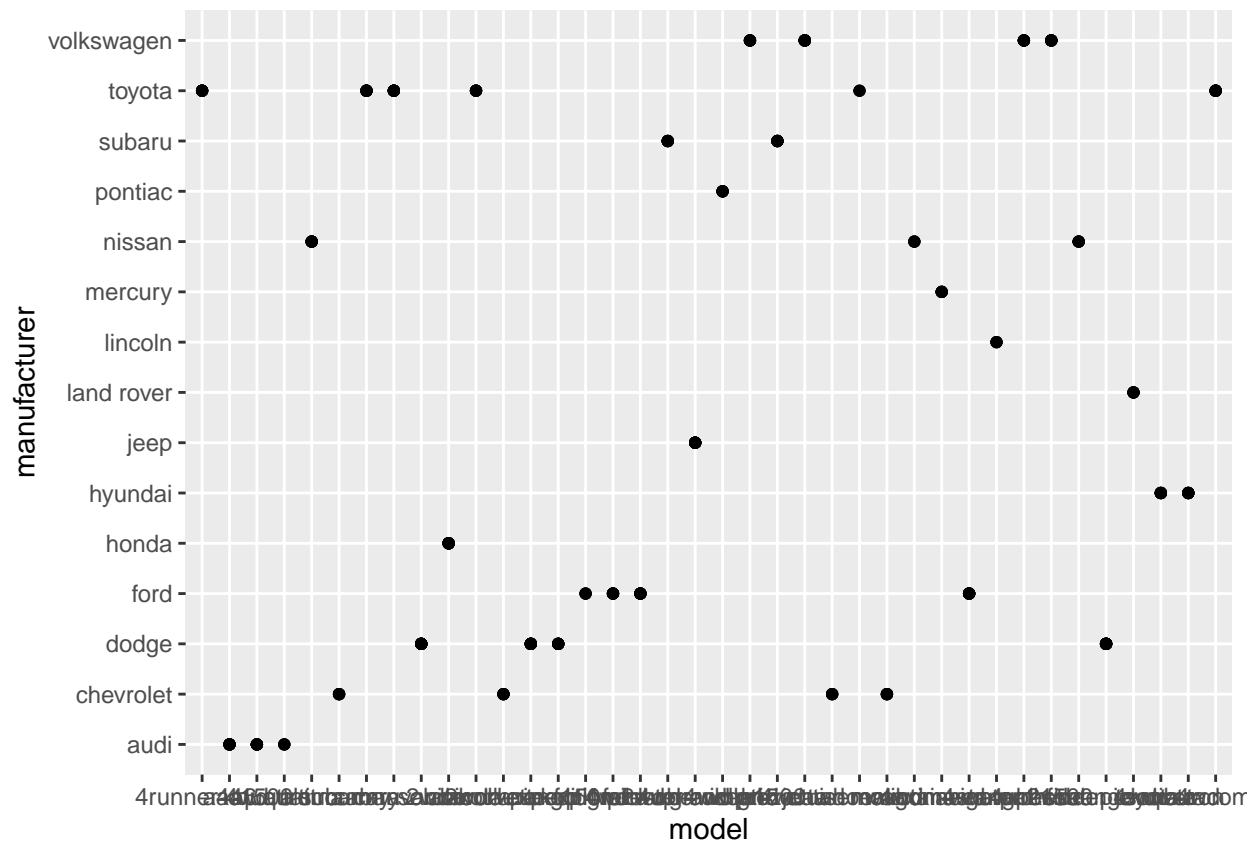
```
##   unique_mods$Manufacturer n  
## 1             audi 3  
## 2        chevrolet 4  
## 3           dodge 4  
## 4            ford 4  
## 5           honda 1  
## 6          hyundai 2  
## 7            jeep 1  
## 8    land rover 1  
## 9          lincoln 1  
## 10         mercury 1  
## 11           nissan 3  
## 12         pontiac 1  
## 13          subaru 2  
## 14           toyota 6  
## 15        volkswagen 4
```

```
ggplot(unique_count, aes(x = `unique_mods$Manufacturer`, y = n)) +  
  geom_bar(stat = "identity", fill = "steelblue") +  
  labs(x = "Manufacturer", y = "Number of Unique Models")
```





```
ggplot(mpg_file, aes(model, manufacturer)) + geom_point()
```

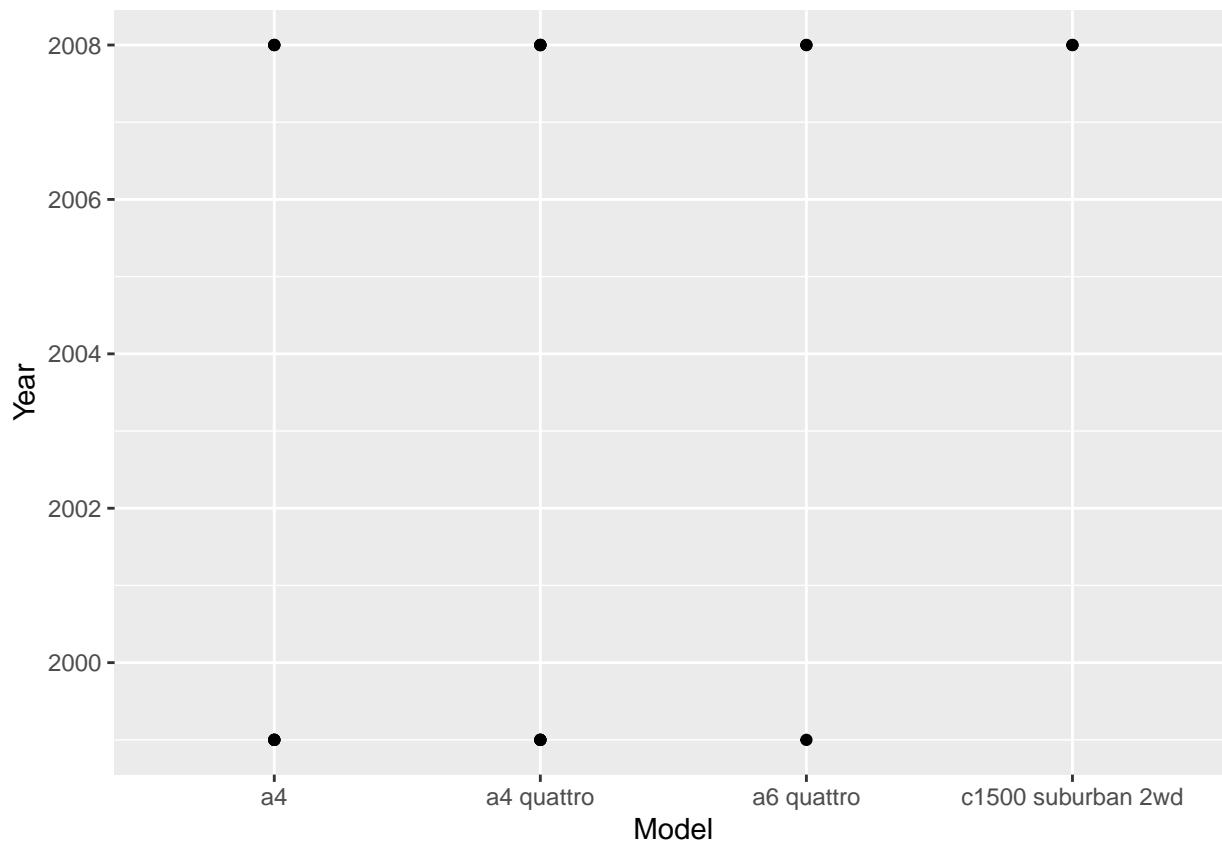


```
print("It plots the mpg dataset with models on the x-axis and manufacturers on the y-axis, where each p

## [1] "It plots the mpg dataset with models on the x-axis and manufacturers on the y-axis, where each p
print("It helps you see the number of models each manufacturer has. You can make it more informative by

## [1] "It helps you see the number of models each manufacturer has. You can make it more informative by
top20 <- head(mpg_file,20)

top20Plot <- ggplot(top20, aes(x = model, y = year)) + geom_point() + labs(x = "Model", y = "Year")
top20Plot
```



```
library(dplyr)
```

```
model_car_count <- mpg_file %>%
  group_by(model) %>%
  summarize(number_of_cars = n())
```

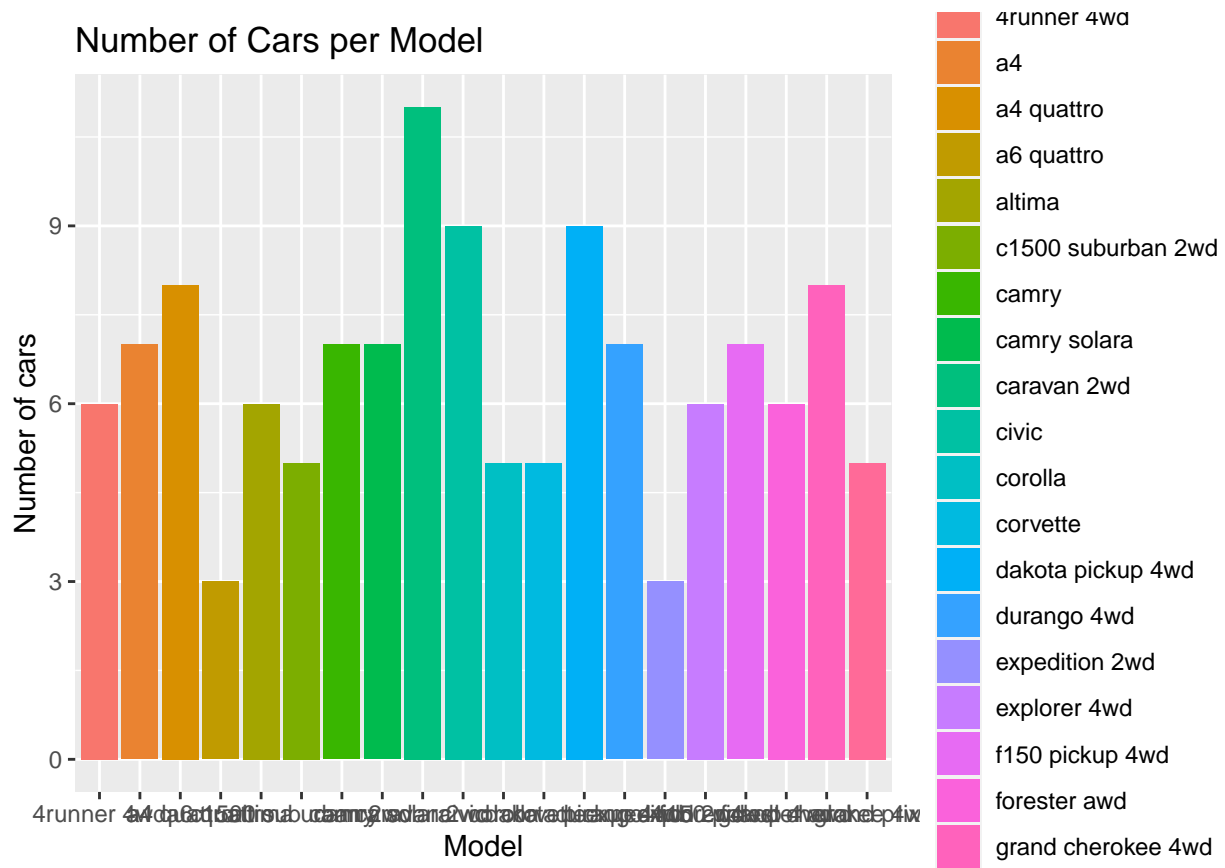
```
model_car_count
```

```
## # A tibble: 38 x 2
##   model          number_of_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
```

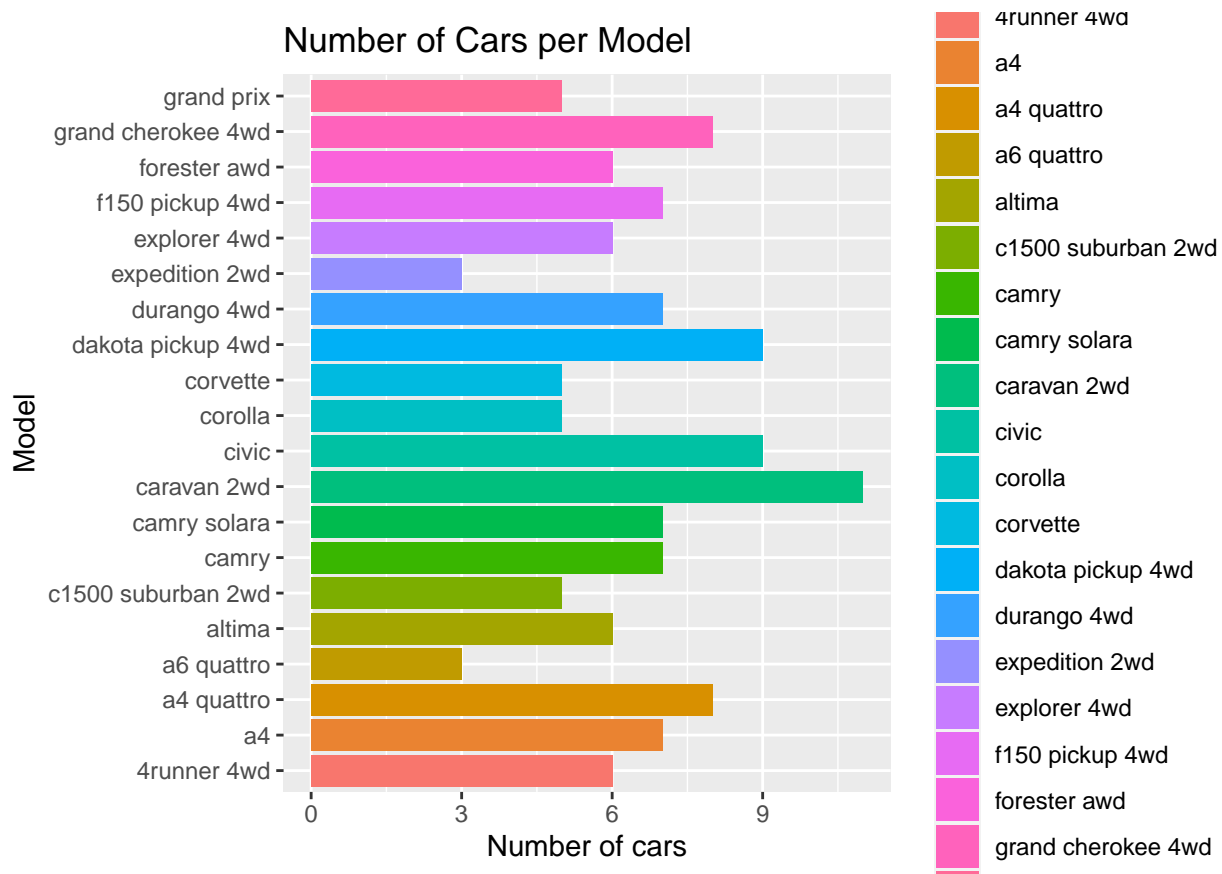
```
obs_20 <- head(model_car_count, 20)
```

```
top_20 <- ggplot(obs_20, aes(x = model, y = number_of_cars, fill = model)) + geom_bar(stat = "identity")
```

```
top_20
```

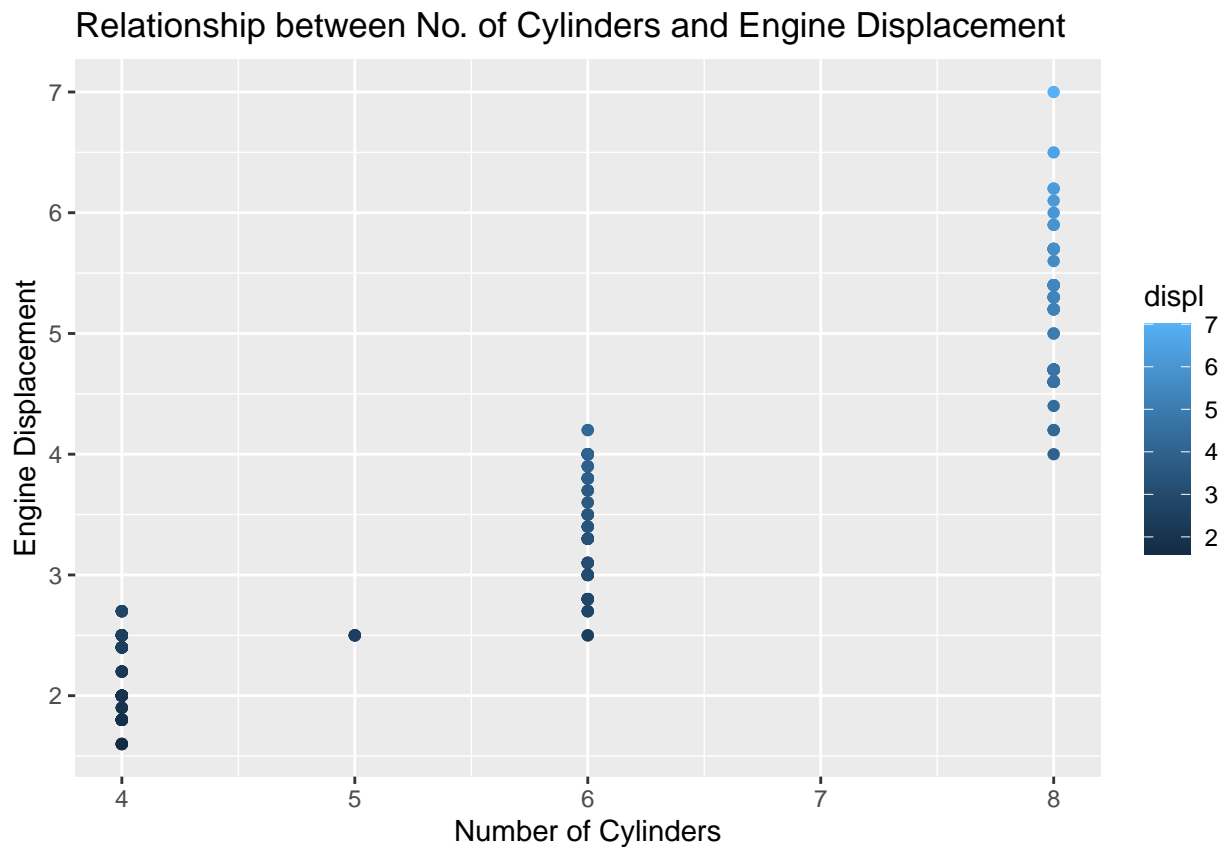


```
flipped_top_20 <- ggplot(obs_20, aes(x = model, y = number_of_cars, fill = model)) + geom_bar(stat = "identity")
flipped_top_20
```



```
cyl_displ_plot <- ggplot(mpg_file, 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")
```

```
cyl_displ_plot
```

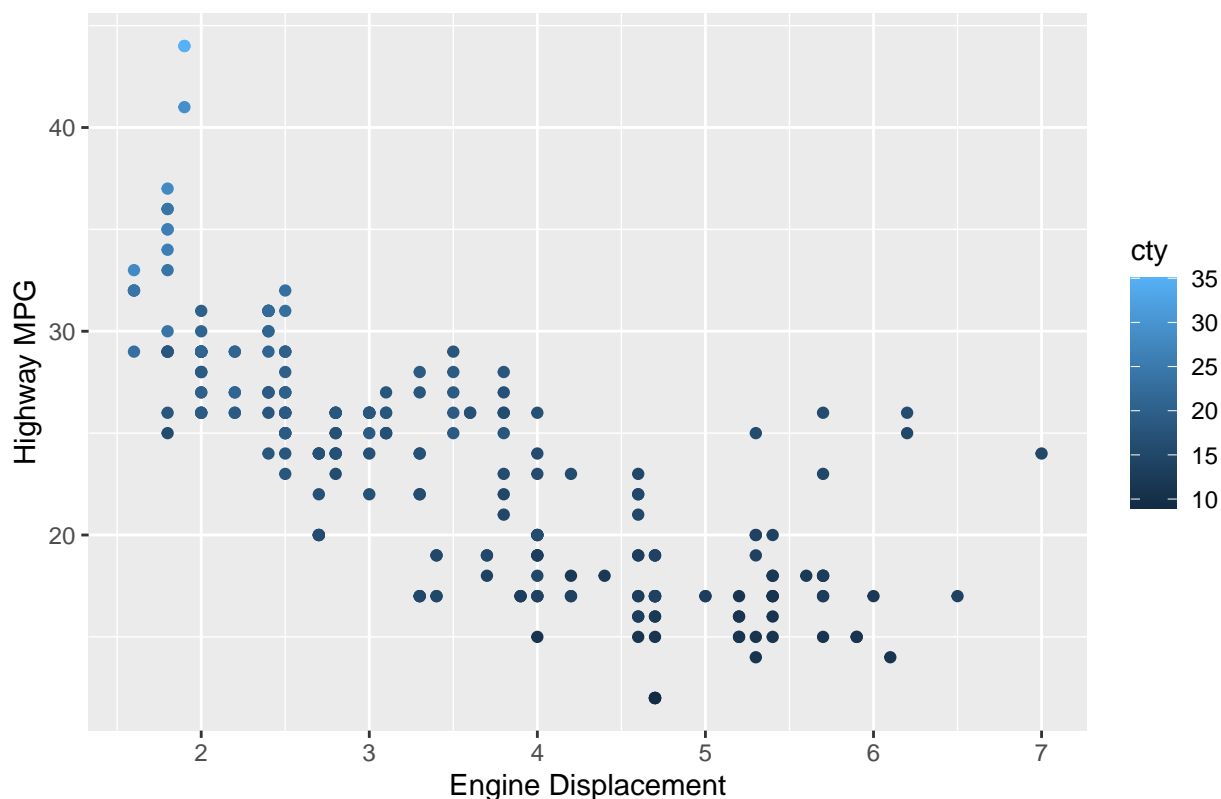


```
print("It makes a scatter plot to show how the number of cylinders and engine size are connected. Each p

## [1] "It makes a scatter plot to show how the number of cylinders and engine size are connected. Each p
displ_hwy_plot <- ggplot(mpg_file, 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
```

# Relationship between Engine Displacement and Highway MPG



```
print("It's a graph with engine size on one side and highway miles per gallon on the other. Each dot's cty")

## [1] "It's a graph with engine size on one side and highway miles per gallon on the other. Each dot's cty"

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

num_obs <- nrow(traffic)
num_obs

## [1] 48120

num_vars <- ncol(traffic)
num_vars

## [1] 4

vars <- colnames(traffic)
vars

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

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

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

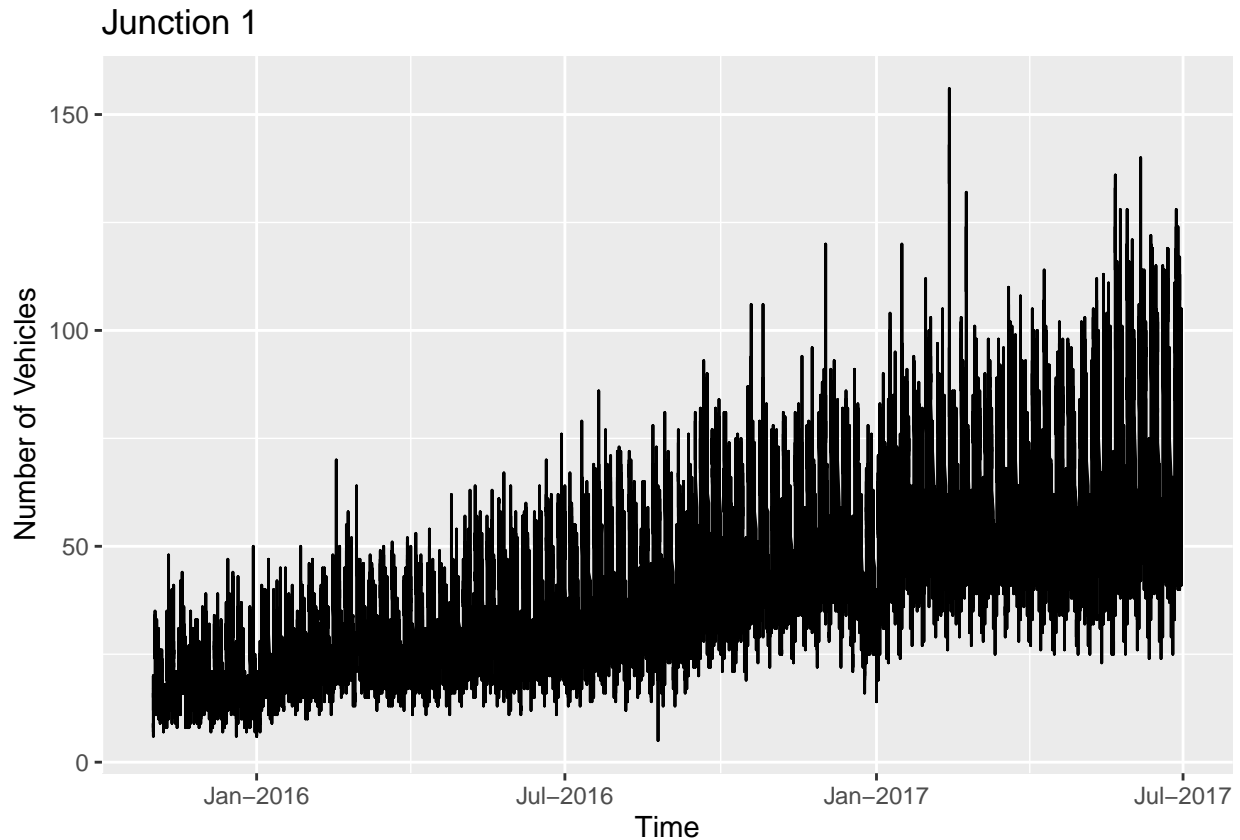
junctions_subset_3 <- subset(traffic, Junction == 3)

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

```
junction_1_plot <- ggplot(junctions_subset_1, aes(x = as.Date(junctions_subset_1$DateTime), y = Vehicles)) +
  geom_line() +
  scale_x_date(date_labels = "%b-%Y") + theme(legend.position = "none") +
  labs(title = "Junction 1", x = "Time", y = "Number of Vehicles")
```

```
junction_1_plot
```

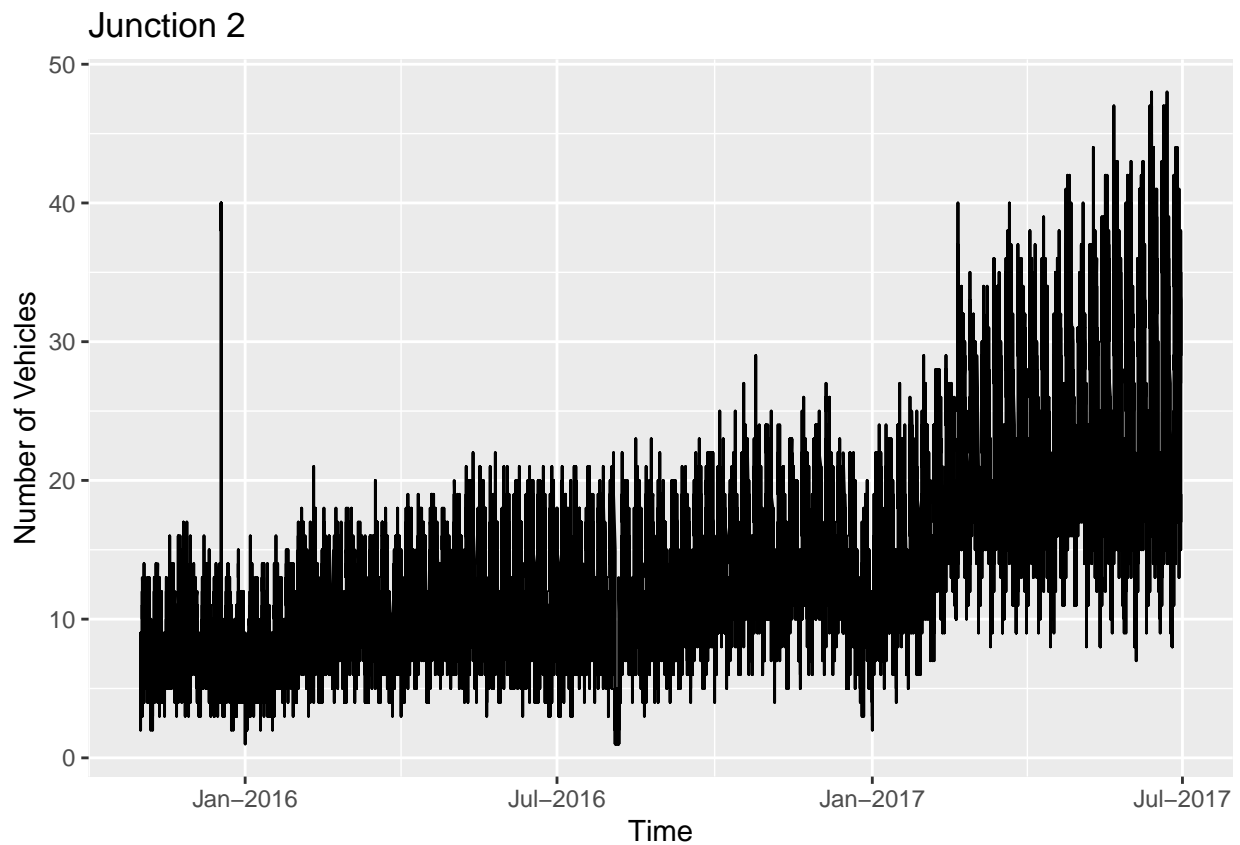
```
## Warning: Use of `junctions_subset_1$DateTime` is discouraged.
## i Use `DateTime` instead.
```



```
junction_2_plot <- ggplot(junctions_subset_2, aes(x = as.Date(junctions_subset_2$DateTime), y = Vehicles)) +
  geom_line() +
  scale_x_date(date_labels = "%b-%Y") + theme(legend.position = "none") +
  labs(title = "Junction 2", x = "Time", y = "Number of Vehicles")
```

```
junction_2_plot
```

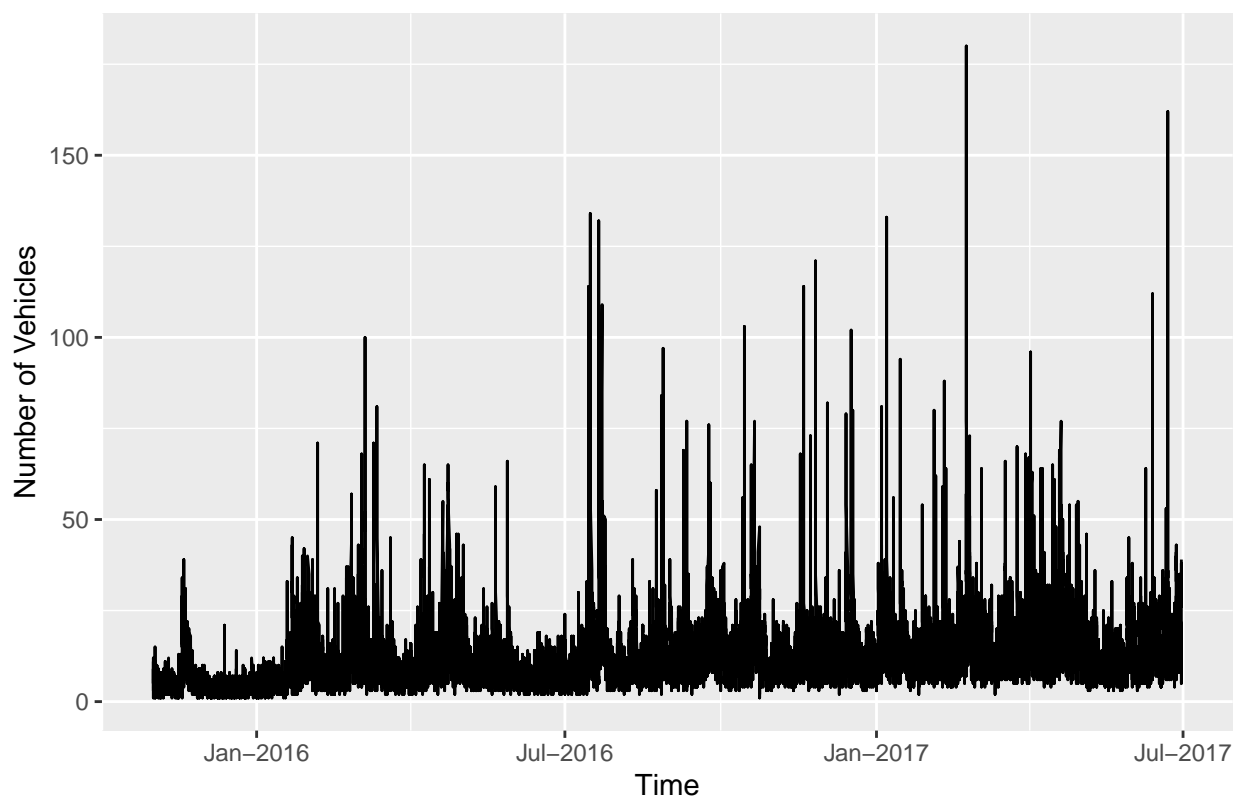




```
junction_3_plot <- ggplot(junctions_subset_3, aes(x = as.Date(junctions_subset_3$DateTime), y = Vehicle
  geom_line() +
  scale_x_date(date_labels = "%b-%Y") + theme(legend.position = "none") +
  labs(title = "Junction 3", x = "Time", y = "Number of Vehicles")

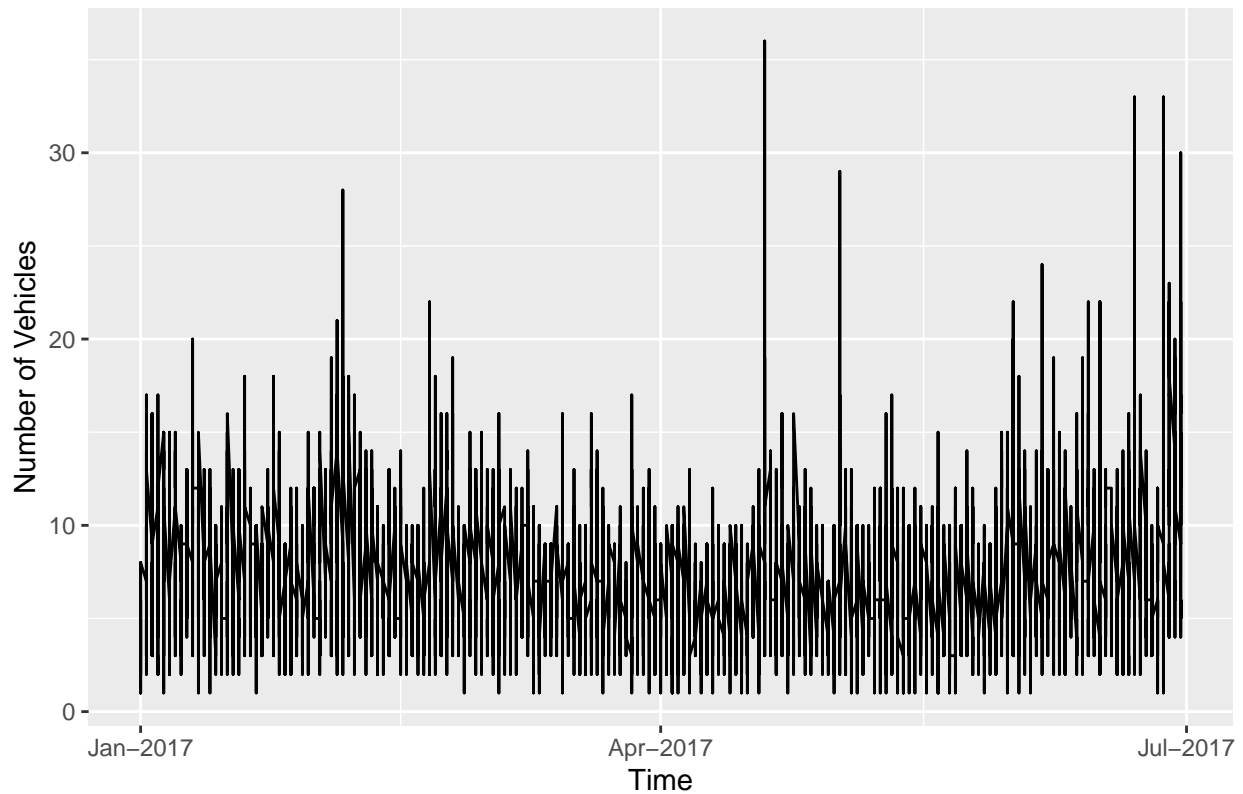
junction_3_plot
```

### Junction 3



```
junction_4_plot <- ggplot(junctions_subset_4, aes(x = as.Date(junctions_subset_4$DateTime), y = Vehicles)) +  
  geom_line() +  
  scale_x_date(date_labels = "%b-%Y") + theme(legend.position = "none") +  
  labs(title = "Junction 4", x = "Time", y = "Number of Vehicles")  
  
junction_4_plot
```

## Junction 4



```
library(readxl)

alexa_data <- read_excel("/cloud/project/worksheet#4/alexa_file.xlsx")

num_obs <- nrow(alexa_data)
num_obs

## [1] 3150

num_cols <- ncol(alexa_data)
num_cols

## [1] 5

Var_counts <- alexa_data %>%
  count(variation)

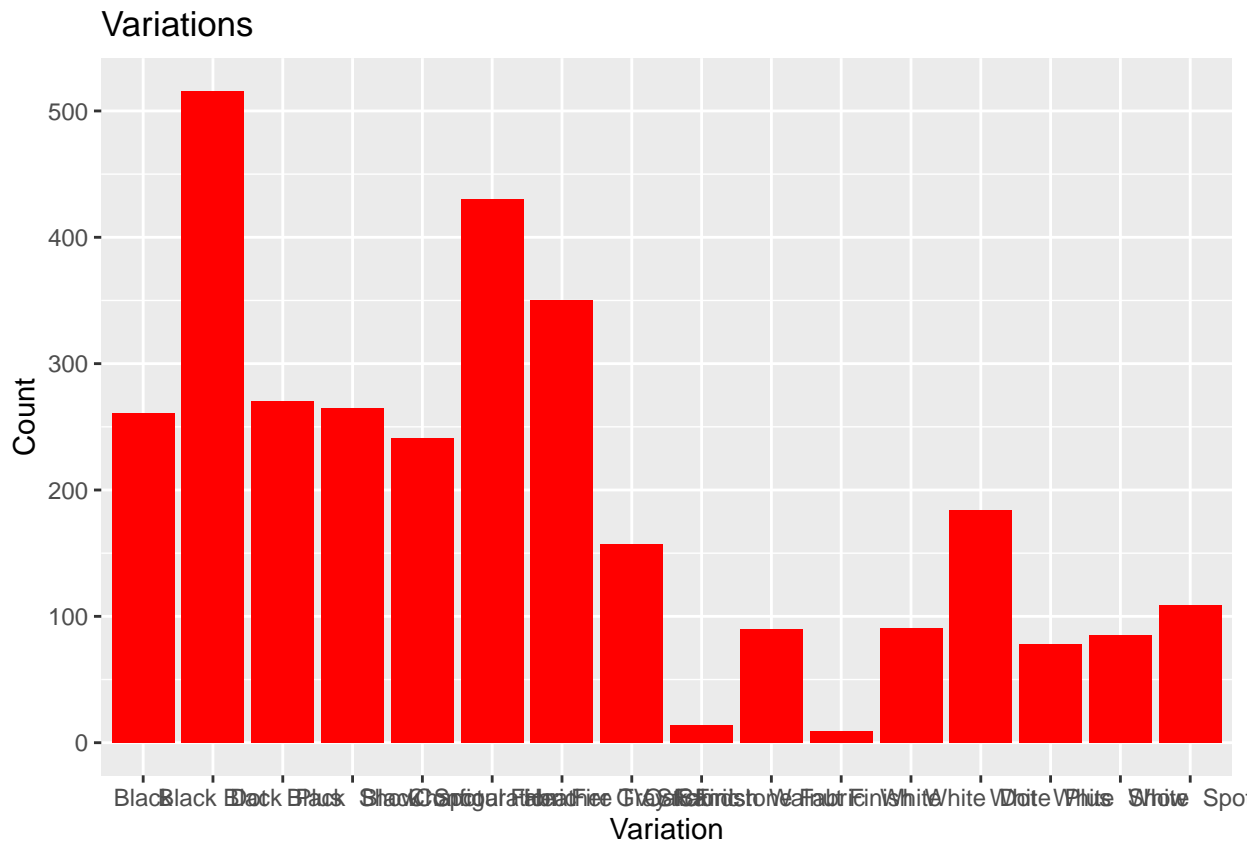
Var_counts

## # A tibble: 16 x 2
##   variation      n
##   <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
```

```
Alexa_plot <- ggplot(alexa_data, aes(x = variation)) +
  geom_bar(fill = "red") +
  labs(title = "Variations",
       x = "Variation",
       y = "Count")
```

Alexa\_plot



```
print("The chart displays different variations and how many times each one occurs. Each bar stands for a variation.")
```

```
## [1] "The chart displays different variations and how many times each one occurs. Each bar stands for a variation."
```

```
library(dplyr)
```

```
alexa_data$date <- as.Date(alexa_data$date)
```

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

```
countMonth <- alexa_data %>%
```

```
count(month)
countMonth
```

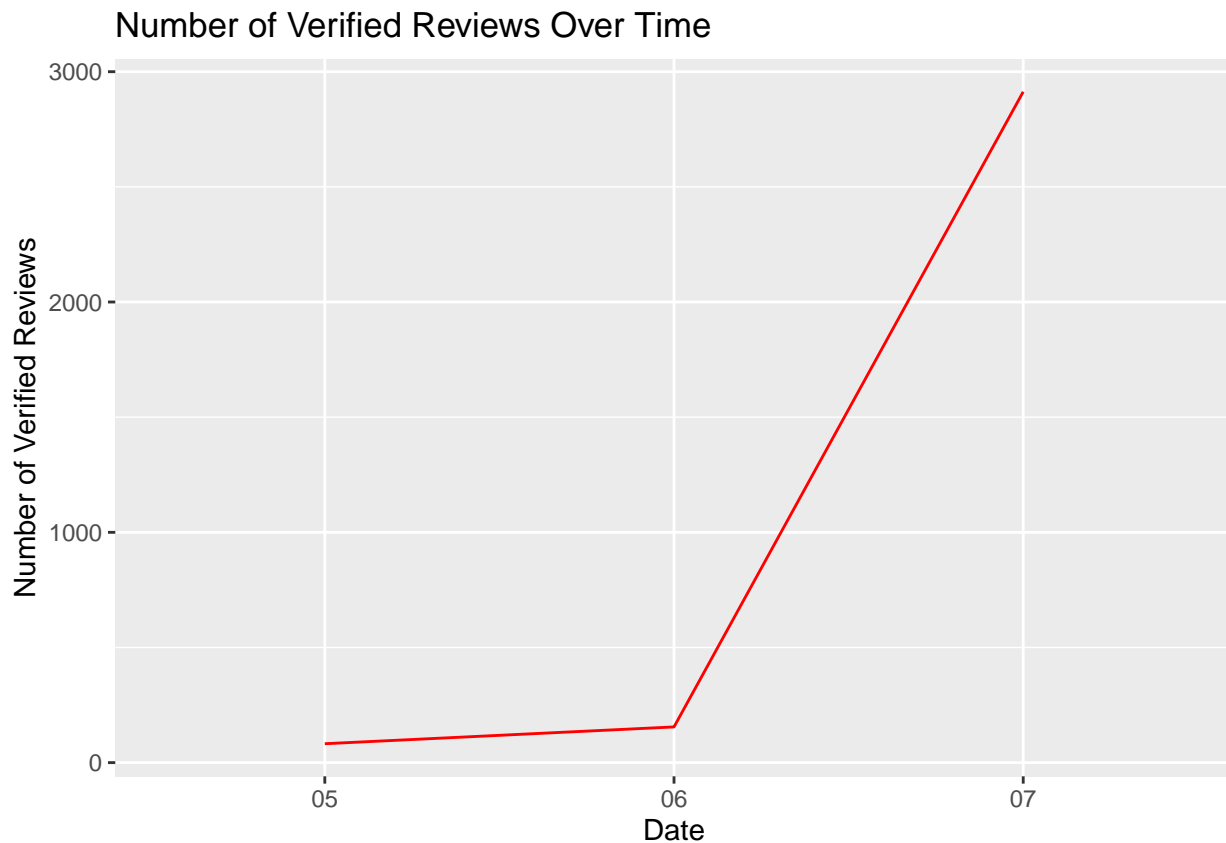
```
## # A tibble: 3 x 2
##   month     n
##   <chr> <int>
## 1 05      82
## 2 06     155
## 3 07    2913
```

```
mon_revCount <- table(countMonth)
mon_revCount
```

```
##           n
## month 82 155 2913
##   05  1   0   0
##   06  0   1   0
##   07  0   0   1
```

```
Alexa_Line <- ggplot(countMonth, aes(x = month, y = n, group = 1)) +
  geom_line(color = "red") +
  labs(title = "Number of Verified Reviews Over Time",
       x = "Date",
       y = "Number of Verified Reviews")
```

Alexa\_Line



```
variation_ratings <- alexa_data %>%
  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

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

```

