

RWorksheet_Sobusa#4c.Rmd

Nexon Sobusa

2024-11-02

```
# 1. Use the dataset mpg
# 1.a Importing a CSV File
mpg_data <- read.csv("C:/Users/kurts/Desktop/R-Code/RWorksheet_4/mpg.csv")
```

```
# 1.b Categorical Variables
# Categorical Variables: manufacturer, model, trans, drv, fl, and class.
```

```
# 1.c Continuous Variables
# Continuous Variables: displ, year, cyl, cty, and hwy.
```

```
# 2. Analysis of Manufacturers and Models
# 2.a Find Manufacturer with Most Models and Model with 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
```

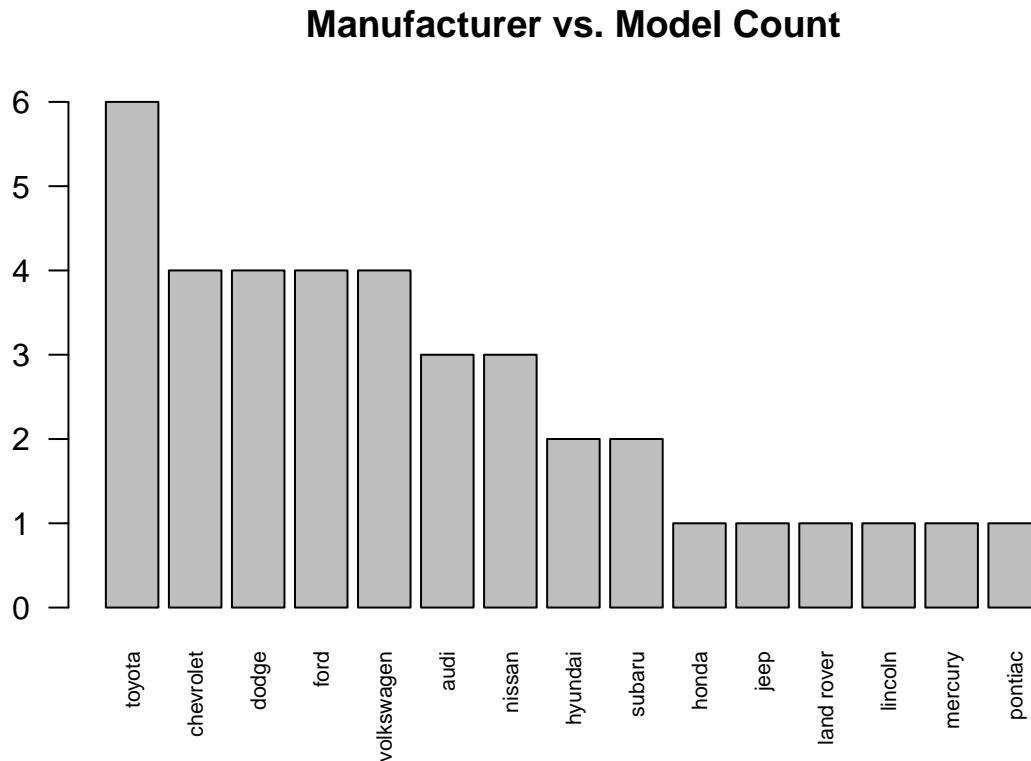
```
most_models <- mpg_data %>%
  group_by(manufacturer) %>%
  summarize(model_count = n_distinct(model)) %>%
  arrange(desc(model_count))
```

```
most_models
```

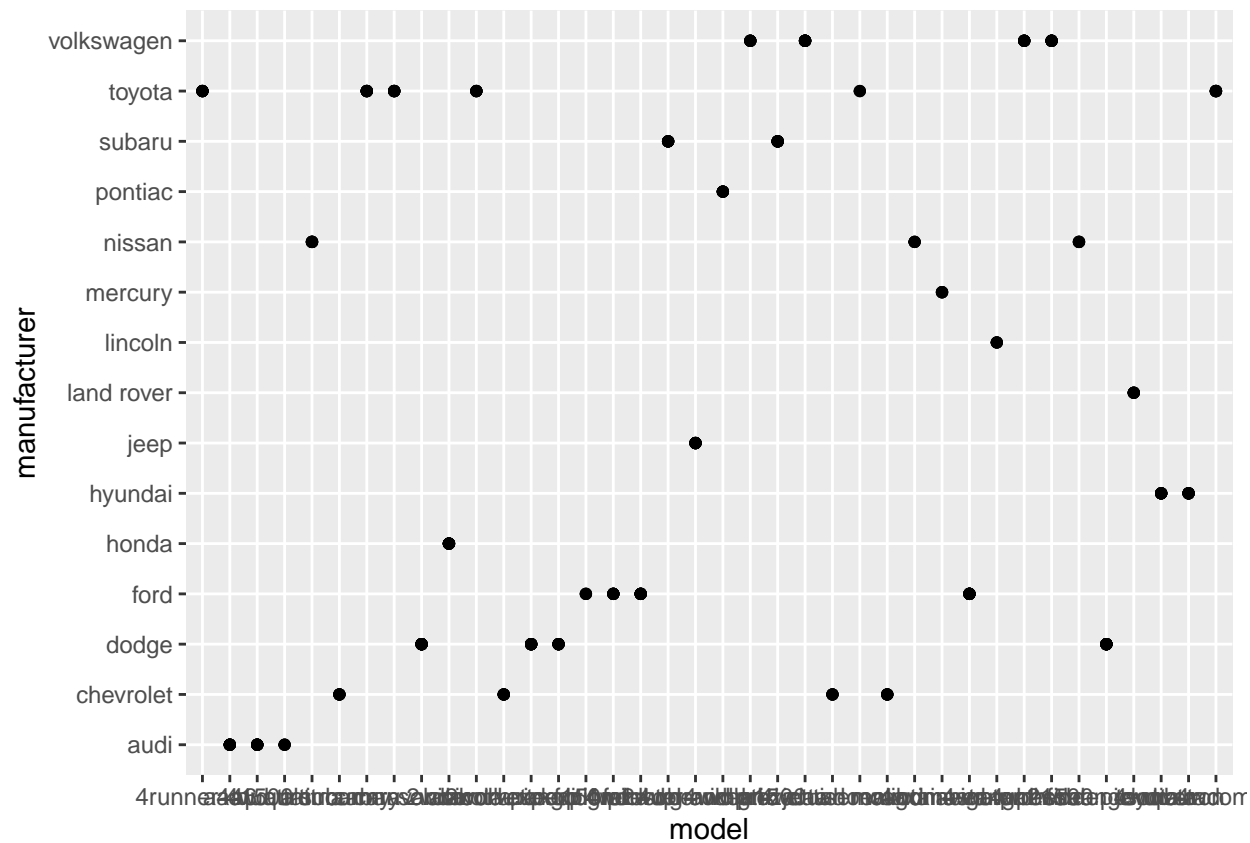
```
## # A tibble: 15 x 2
##   manufacturer model_count
##   <chr>           <int>
## 1 toyota             6
## 2 chevrolet          4
## 3 dodge             4
## 4 ford              4
## 5 volkswagen        4
## 6 audi              3
## 7 nissan             3
## 8 hyundai           2
## 9 subaru            2
## 10 honda            1
```

```
## 11 jeep                1
## 12 land rover          1
## 13 lincoln             1
## 14 mercury             1
## 15 pontiac             1
```

```
# 2.b Plotting Manufacturer and Model Counts with barplot()
barplot(most_models$model_count, names.arg = most_models$manufacturer,
        main = "Manufacturer vs. Model Count", las = 2, cex.names = 0.7)
```



```
# 2. Exploring Model and Manufacturer Relationship
# 2.a Plotting the Relationship with ggplot
library(ggplot2)
ggplot(mpg_data, aes(x = model, y = manufacturer)) + geom_point()
```



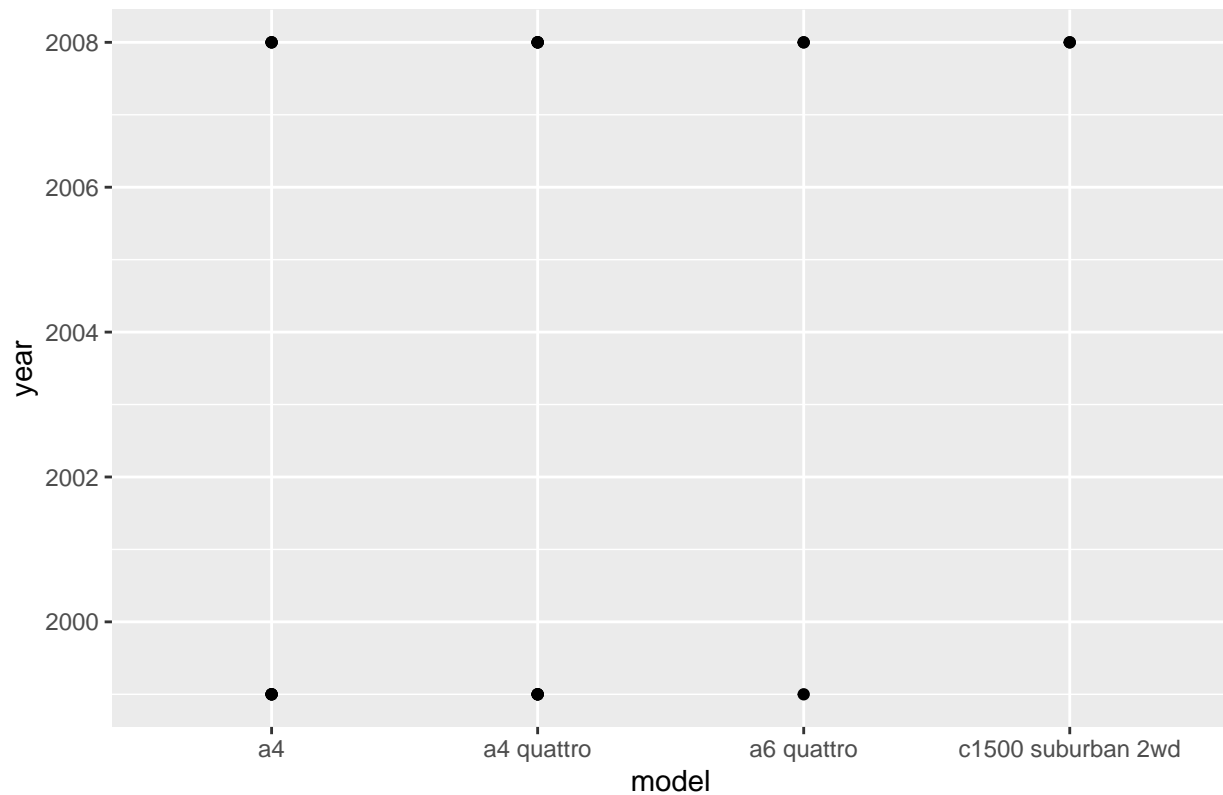
2.b Improving Data Presentation

The scatterplot might look messy because it shows a lot of different models. To make it easier to read

3. Plotting Model and Year (Top 20 Observations)

```
top20_data <- head(mpg_data, 20)
ggplot(top20_data, aes(x = model, y = year)) + geom_point() +
  labs(title = "Model vs Year (Top 20 Observations)")
```

Model vs Year (Top 20 Observations)

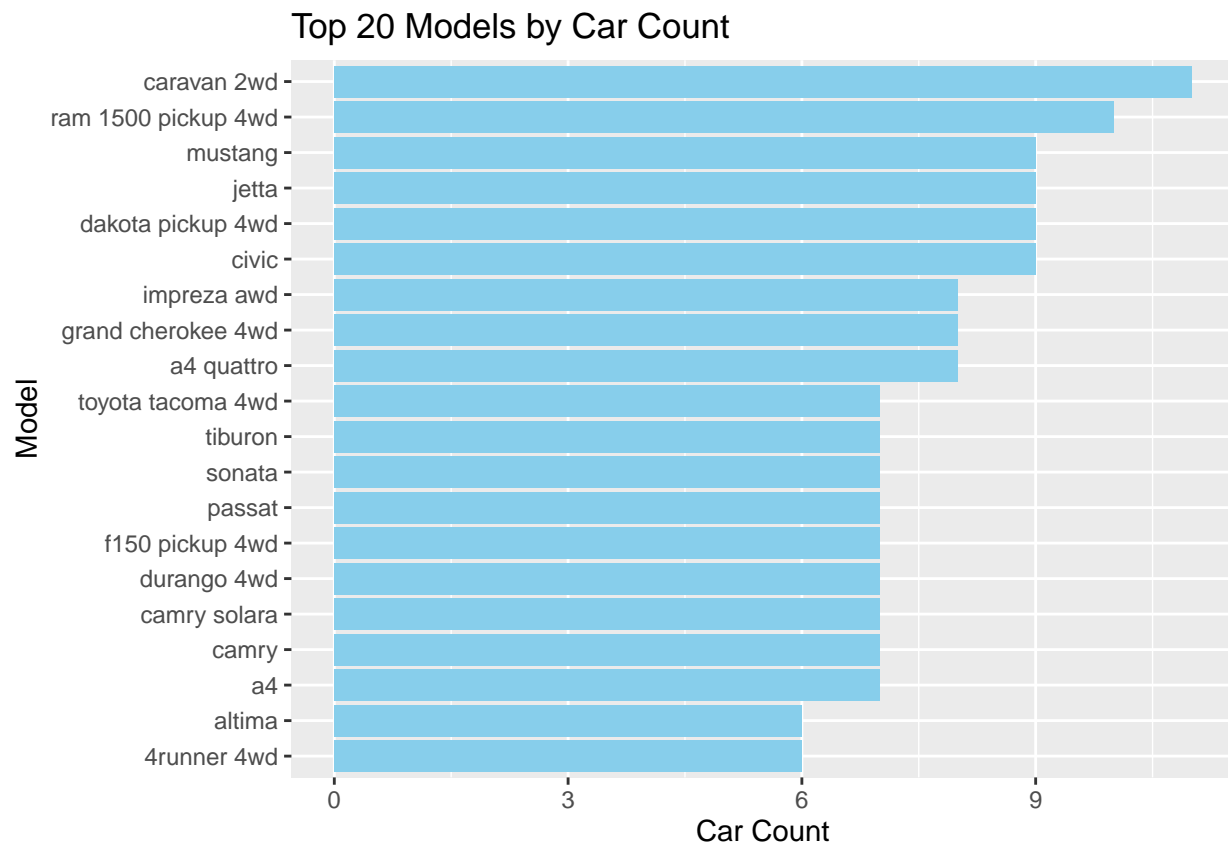


```
# 4. Count Cars per Model with Pipe and Plot
# 4.a Grouping and Counting Models
model_count <- mpg_data %>%
  group_by(model) %>%
  summarize(car_count = n()) %>%
  arrange(desc(car_count))
model_count
```

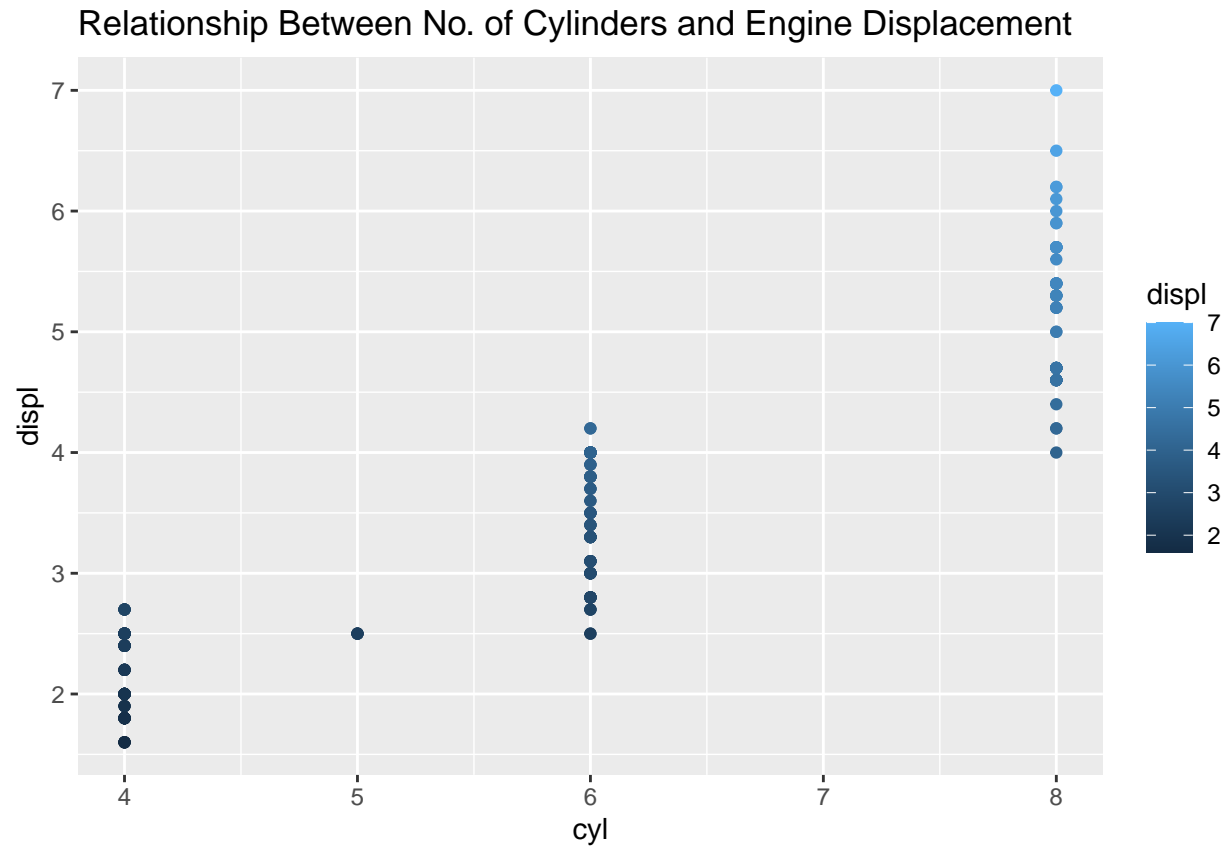
```
## # A tibble: 38 x 2
##   model                car_count
##   <chr>                 <int>
## 1 caravan 2wd           11
## 2 ram 1500 pickup 4wd    10
## 3 civic                 9
## 4 dakota pickup 4wd      9
## 5 jetta                 9
## 6 mustang               9
## 7 a4 quattro            8
## 8 grand cherokee 4wd     8
## 9 impreza awd           8
## 10 a4                   7
## # i 28 more rows
```

```
# 4.b Plotting with geom_bar() and coord_flip()
ggplot(model_count[1:20, ], aes(x = reorder(model, car_count), y = car_count)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  coord_flip() +
```

```
labs(title = "Top 20 Models by Car Count", x = "Model", y = "Car Count")
```



```
# 5. Plotting Cylinders vs. Displacement with Color
ggplot(mpg_data, aes(x = cyl, y = displ, color = displ)) +
  geom_point() +
  labs(title = "Relationship Between No. of Cylinders and Engine Displacement")
```

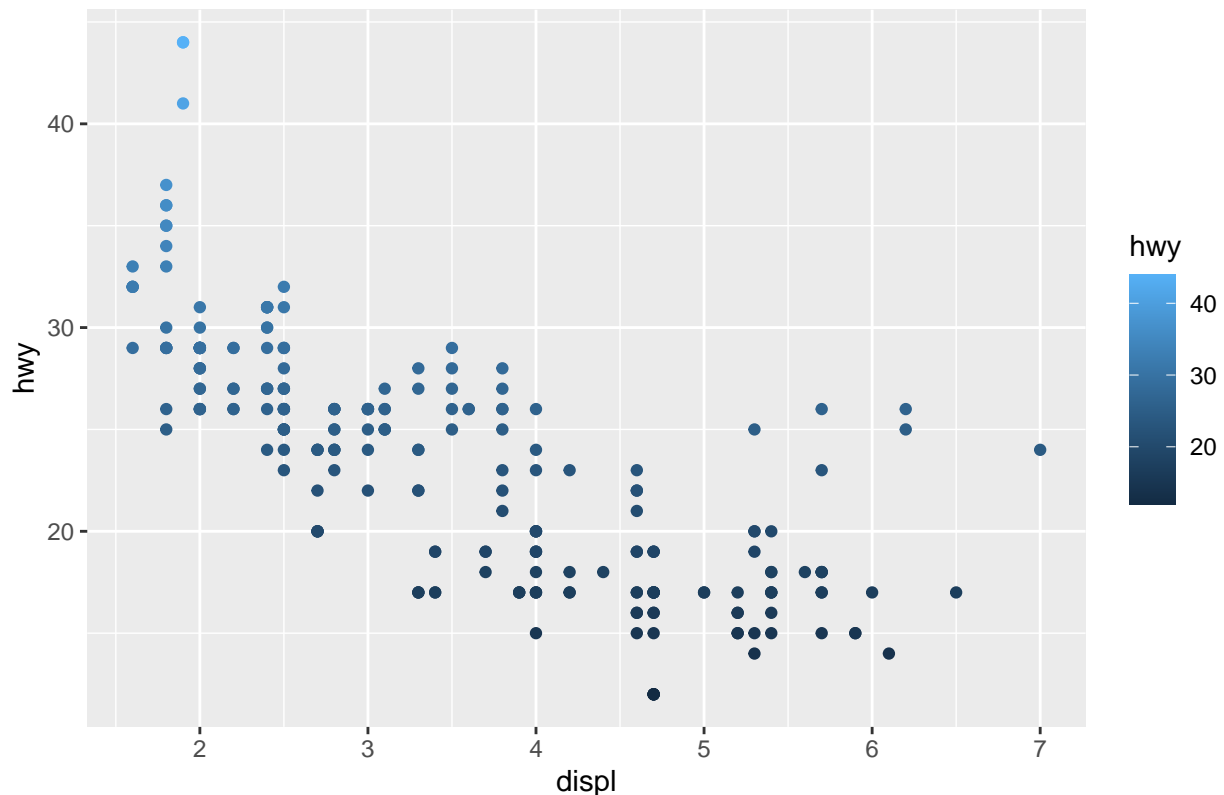


This plot shows the connection between the number of cylinders (cyl) and engine size (displ), with co

6. Plotting Displacement vs. Highway MPG

```
ggplot(mpg_data, aes(x = displ, y = hwy, color = hwy)) + geom_point() +  
  labs(title = "Relationship Between Displacement and Highway MPG")
```

Relationship Between Displacement and Highway MPG



This plot illustrates how highway miles per gallon (hwy) changes with engine displacement (displ), us

6. Importing traffic.csv into R

6.a Number of Observations and Variables

```
traffic_data <- read.csv("C:/Users/kurts/Desktop/R-Code/RWorksheet_4/traffic.csv")
dim(traffic_data)
```

```
## [1] 48120      4
```

```
names(traffic_data)
```

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

6.b Subsetting Traffic Dataset by Junctions

```
traffic_junctions <- traffic_data %>%
  group_by(Junction) %>%
  summarize(count = n())
traffic_junctions
```

```
## # A tibble: 4 x 2
```

```
##   Junction count
```

```
##   <int> <int>
```

```
## 1     1 14592
```

```
## 2     2 14592
```

```
## 3     3 14592
```

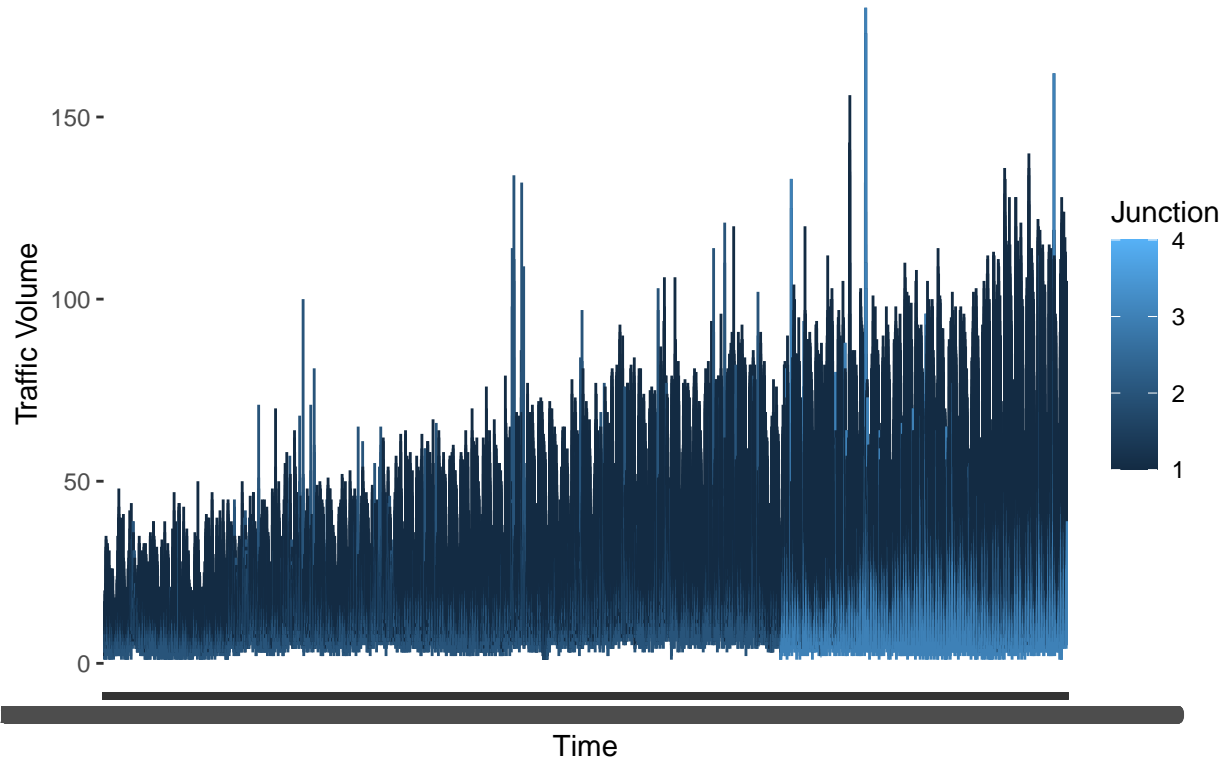
```
## 4     4  4344
```

If the column names are correct, create the plot

```
ggplot(traffic_data, aes(x = DateTime, y = Vehicles, color = Junction)) +
```

```
geom_line() +
labs(title = "Traffic Volume by Junction Over Time", x = "Time", y = "Traffic Volume")
```

Traffic Volume by Junction Over Time



```
# 7. Importing alexa_file.xlsx
# 7.a Number of Observations and Columns
library(readxl)
alexa_data <- read_excel("C:/Users/kurts/Desktop/R-Code/RWorksheet_4/alexa_file.xlsx")
dim(alexa_data)
```

```
## [1] 3150    5
```

```
# 7.b Grouping and Summing Variations
variation_counts <- alexa_data %>%
  group_by(variation) %>%
  summarize(total_count = n())
variation_counts
```

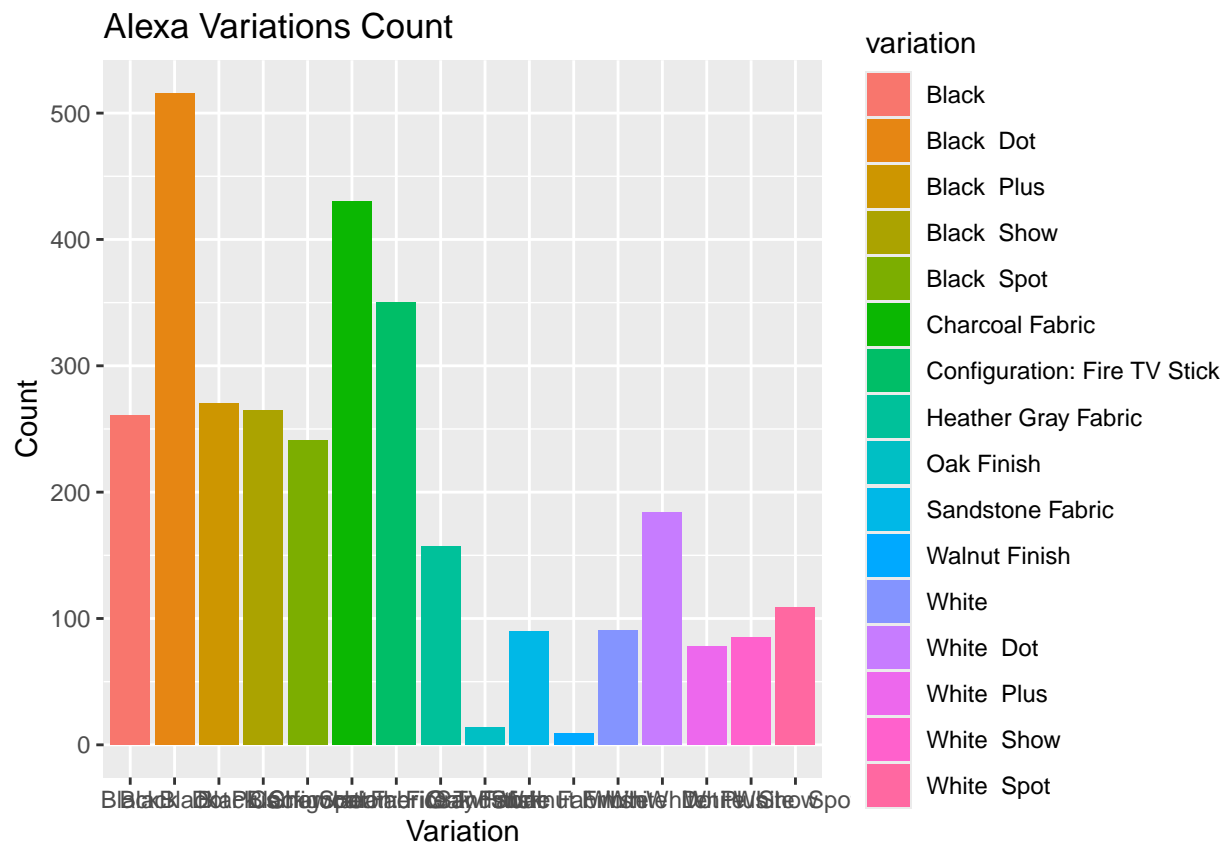
```
## # A tibble: 16 x 2
##   variation          total_count
##   <chr>              <int>
## 1 Black              261
## 2 Black Dot          516
## 3 Black Plus         270
## 4 Black Show         265
## 5 Black Spot         241
## 6 Charcoal Fabric     430
## 7 Configuration: Fire TV Stick 350
```



```
## 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
```

```
# 7.c Plotting Variations with ggplot
```

```
ggplot(variation_counts, aes(x = variation, y = total_count, fill = variation)) +
  geom_bar(stat = "identity") +
  labs(title = "Alexa Variations Count", x = "Variation", y = "Count")
```



```
# 7.d Plotting Date vs. Verified Reviews with geom_line()
```

```
ggplot(alexa_data, aes(x = date, y = verified_reviews)) +
  geom_line() +
  labs(title = "Verified Reviews Over Time", x = "Date", y = "Verified Reviews")
```

are some serious flaws, particularly if you are the last one to bed or the first to wake. It doesn't seem like the engineer

expensive alternative option to fill the gap. Ordered the Amazon Fire Stick from Best Buy. Instructions were short and

one of the lights by saying "Alexa, turn off the second light". In the Alexa app, I created a 'Group' with "but lately I've been getting terrible support. The guy that took my call just rambled off a (completely unhelpful) script an

ng to add this bulb to my Alexa Echo Plus. Everything I tried ended in a Discover Failed message. I tried to set it up in multiple pages. The one thing that I am not a fan of is the home screen cards do not really read that much. They

```
# 7.e Plotting Relationship of Variations and Ratings
variation_ratings <- alexa_data %>%
  group_by(variation) %>%
  summarize(avg_rating = mean(rating, na.rm = TRUE))

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

