

RWorksheet_CAoyonan#4c

```
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
library(ggplot2)
library(readr)
library(readxl)

# 1. Import mpg CSV

mpg <- read_csv("mpg.csv")

## Rows: 234 Columns: 11
## -- Column specification -----
## Delimiter: ","
## chr (6): manufacturer, model, trans, drv, fl, class
## dbl (5): 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.

# 1b. Categorical variables
categorical_vars <- c("manufacturer", "model", "trans", "drv", "fl", "class")
categorical_vars

## [1] "manufacturer" "model"          "trans"          "drv"           "fl"
## [6] "class"

# 1c. Continuous variables
continuous_vars <- c("displ", "year", "cyl", "cty", "hwy")
continuous_vars

## [1] "displ" "year"  "cyl"   "cty"   "hwy"

# 2. Manufacturer with most models

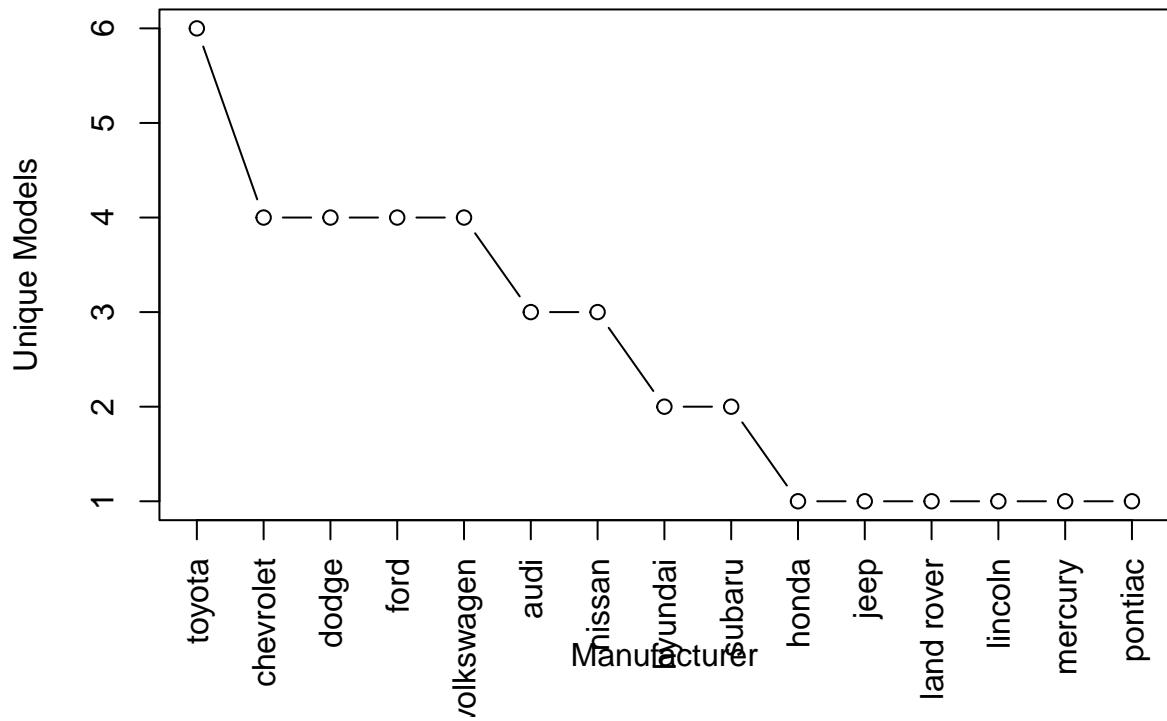
# 2a. Group and count models
manu_models <- mpg %>%
  group_by(manufacturer) %>%
  summarise(unique_models = n_distinct(model)) %>%
  arrange(desc(unique_models))
```

```
manu_models
```

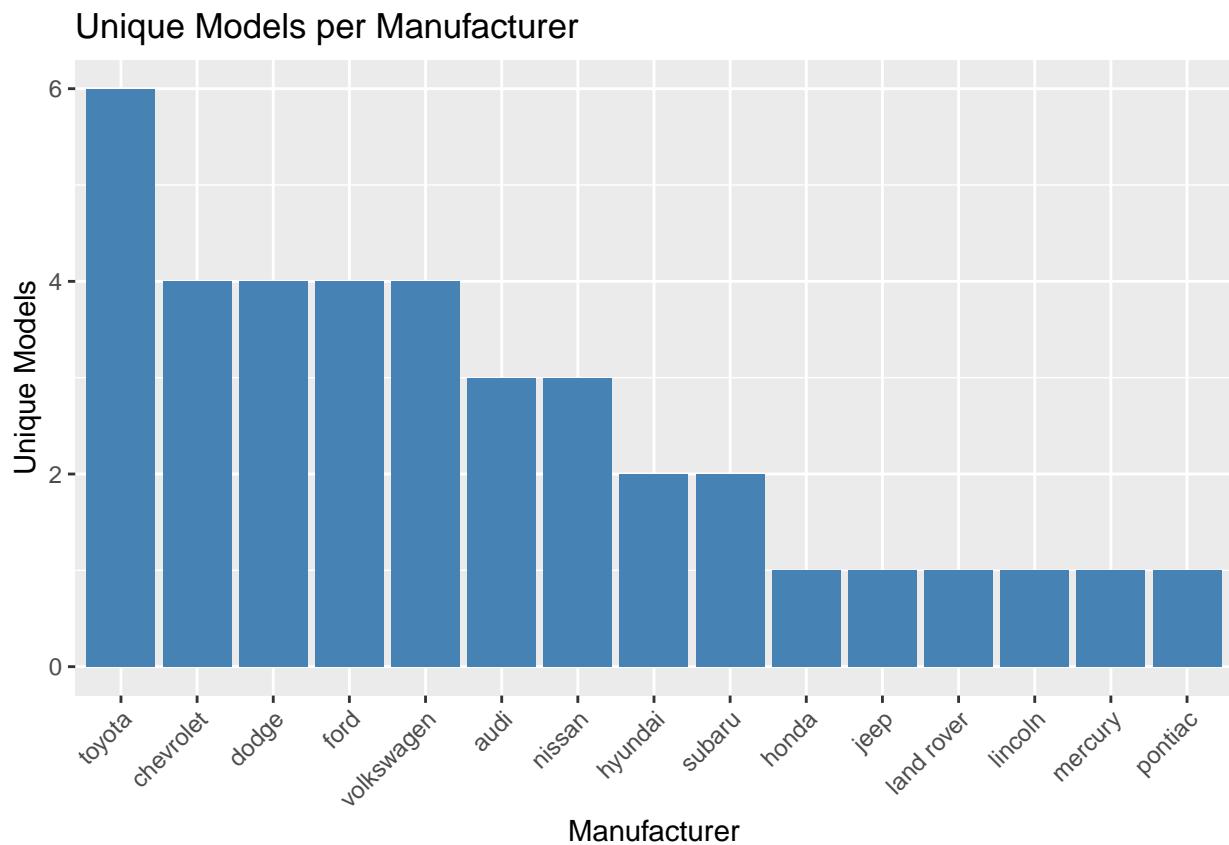
```
## # A tibble: 15 x 2
##   manufacturer unique_models
##   <chr>           <int>
## 1 toyota            6
## 2 chevrolet         4
## 3 dodge             4
## 4 ford              4
## 5 volkswagen        4
## 6 audi              3
## 7 nissan            3
## 8 hyundai           2
## 9 subaru            2
## 10 honda             1
## 11 jeep              1
## 12 land rover        1
## 13 lincoln           1
## 14 mercury           1
## 15 pontiac           1

# 2b. Base R plot
plot(manu_models$unique_models, type="b",
      xaxt="n", xlab="Manufacturer", ylab="Unique Models",
      main="Unique Models per Manufacturer")
axis(1, at=1:nrow(manu_models), labels=manu_models$manufacturer, las=2)
```

Unique Models per Manufacturer

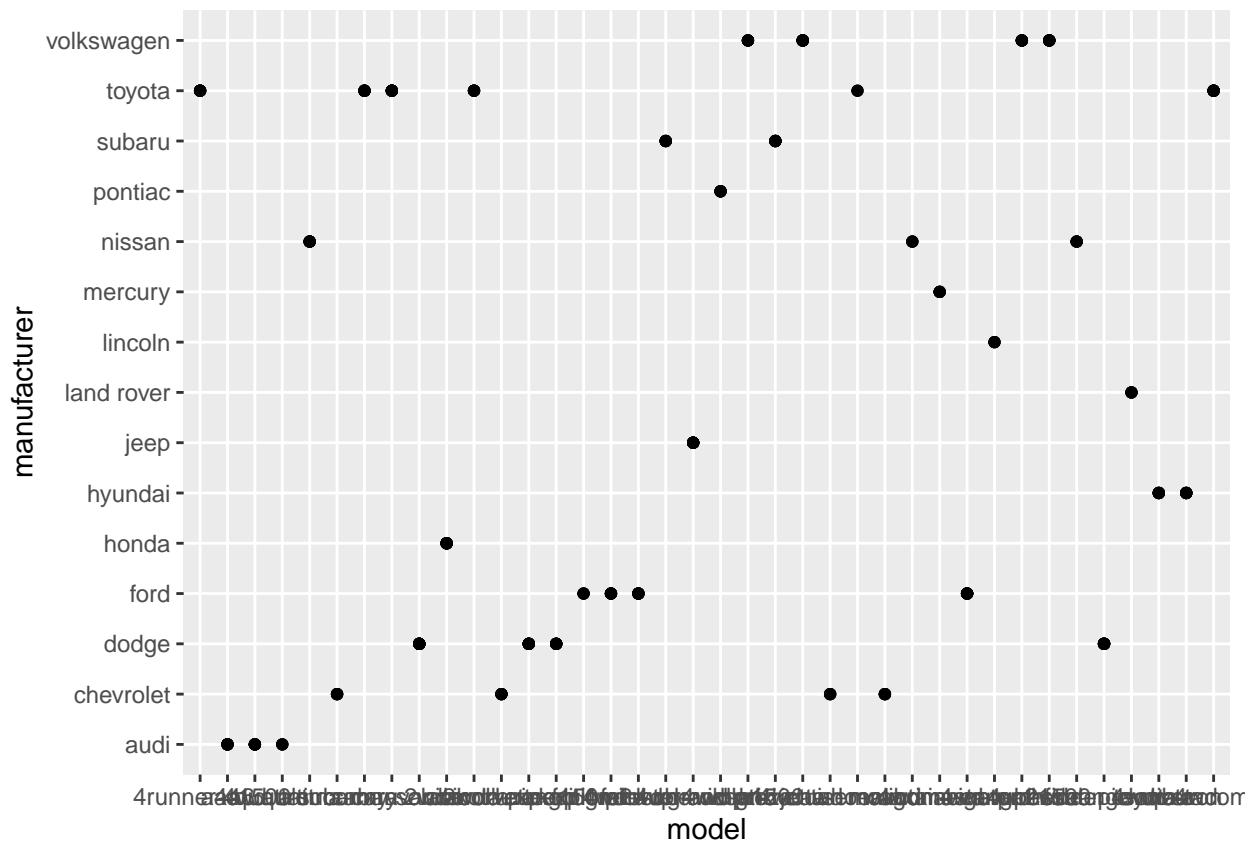


```
# 2b. ggplot version
ggplot(manu_models, aes(x=reorder(manufacturer, -unique_models), y=unique_models)) +
  geom_col(fill="steelblue") +
  theme(axis.text.x = element_text(angle=45, hjust=1)) +
  labs(title="Unique Models per Manufacturer",
       x="Manufacturer", y="Unique Models")
```



```
# 3. Relationship between model and manufacturer

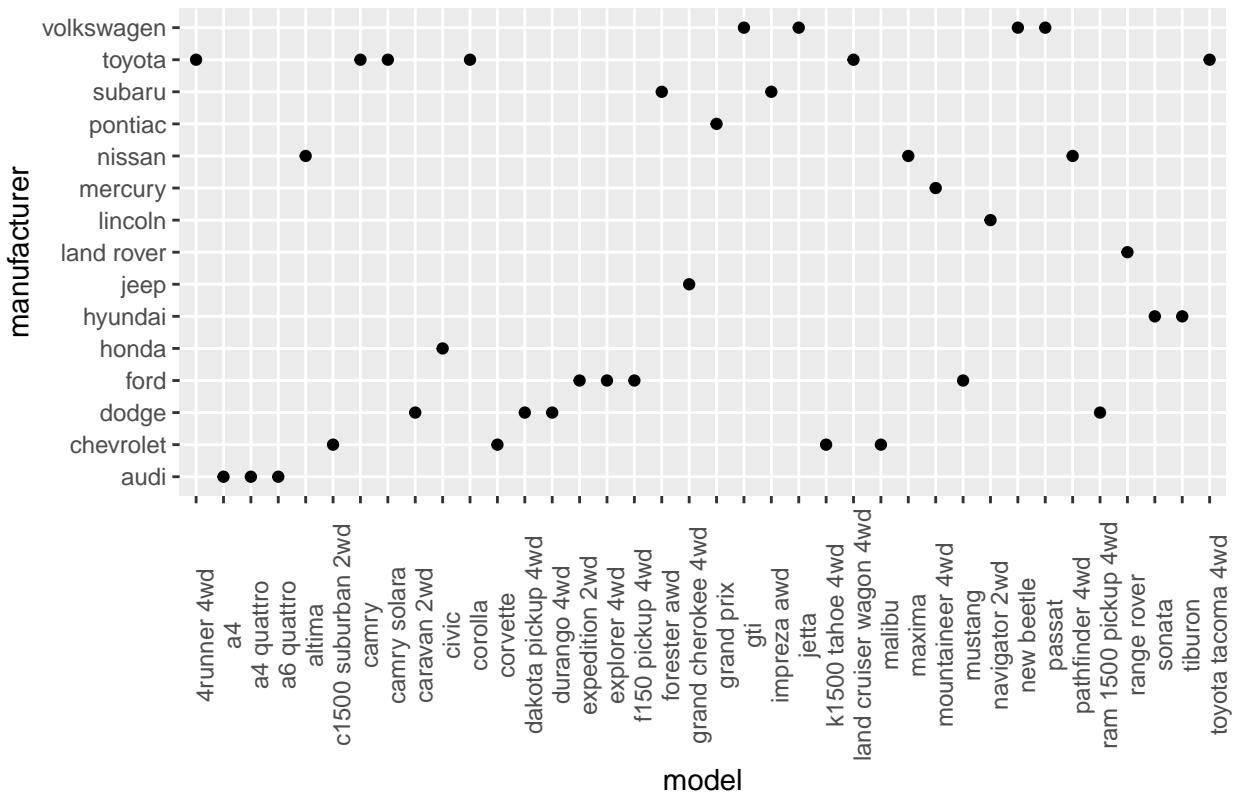
# 3a. Initial plot
ggplot(mpg, aes(model, manufacturer)) + geom_point()
```



```
# 3b. Improved version (remove duplicates)
mpg_unique <- mpg %>% distinct(model, manufacturer)

ggplot(mpg_unique, aes(model, manufacturer)) +
  geom_point() +
  theme(axis.text.x = element_text(angle=90)) +
  labs(title="Unique Model-Manufacturer Combinations")
```

Unique Model–Manufacturer Combinations

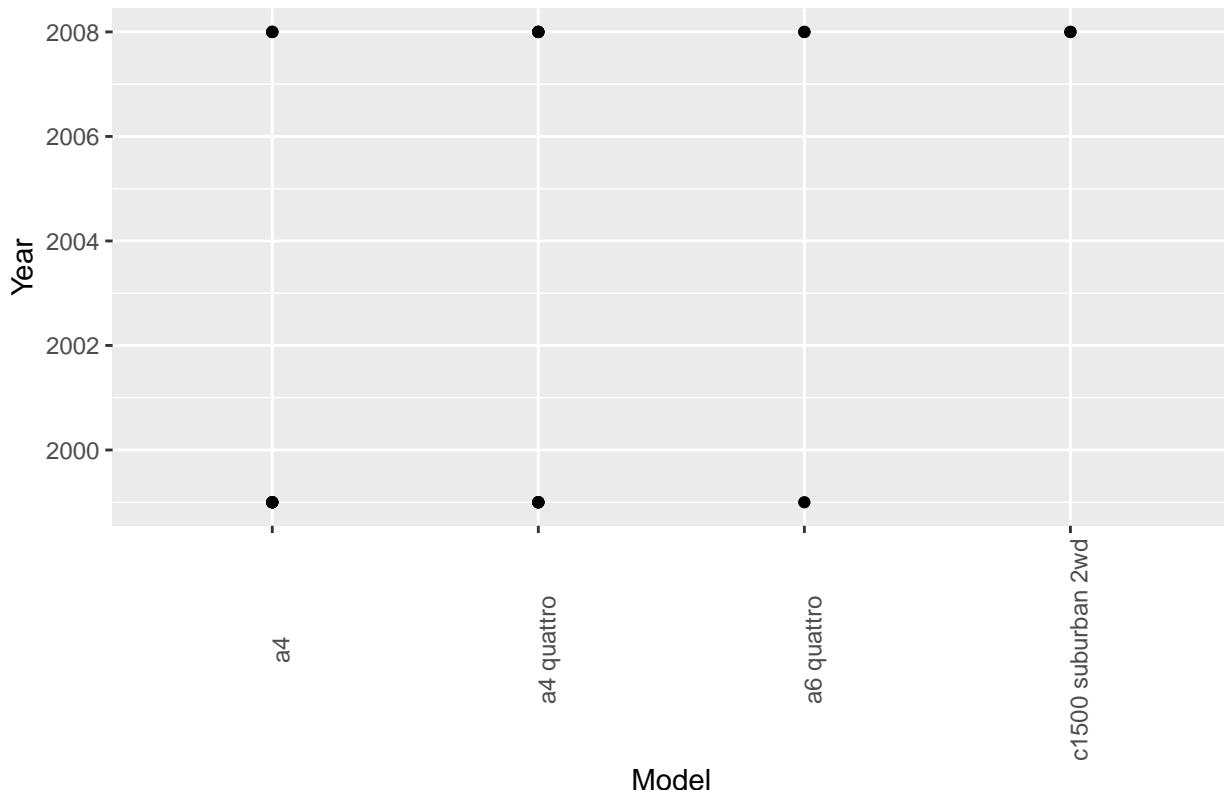


```
# 4. Plot model vs year (top 20)
```

```
top20 <- mpg %>% slice(1:20)

ggplot(top20, aes(x=model, y=year)) +
  geom_point() +
  theme(axis.text.x = element_text(angle=90)) +
  labs(title="Model vs Year (Top 20 Rows)",
       x="Model", y="Year")
```

Model vs Year (Top 20 Rows)



```
# 5. Group by model and count

model_count <- mpg %>%
  group_by(model) %>%
  summarise(n = n()) %>%
  arrange(desc(n))

model_count

## # A tibble: 38 x 2
##   model           n
##   <chr>        <int>
## 1 caravan      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

top20_models <- model_count %>% slice(1:20)

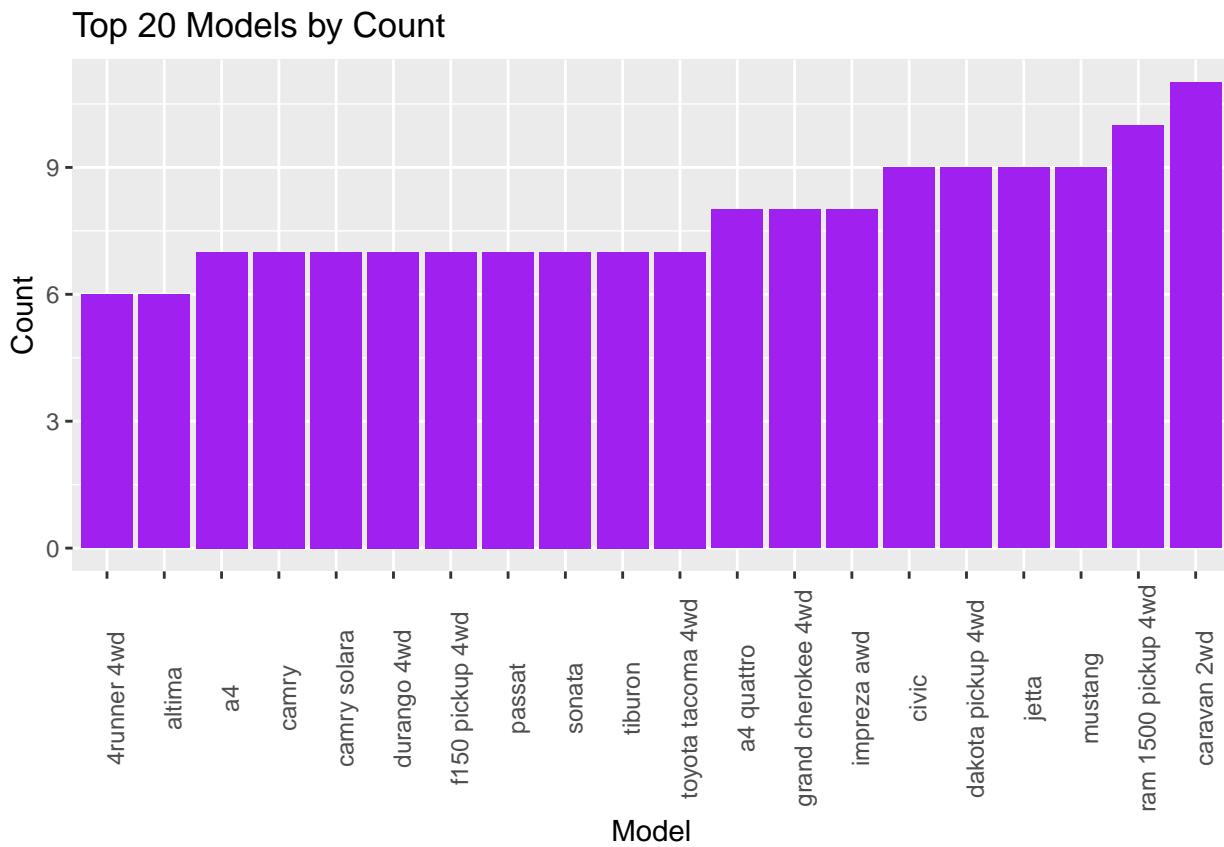
# 5a. Bar plot
```

```
ggplot(top20_models, aes(x=reorder(model, n), y=n)) +
```

```

geom_bar(stat="identity", fill="purple") +
labs(title="Top 20 Models by Count",
x="Model", y="Count") +
theme(axis.text.x = element_text(angle=90))

```

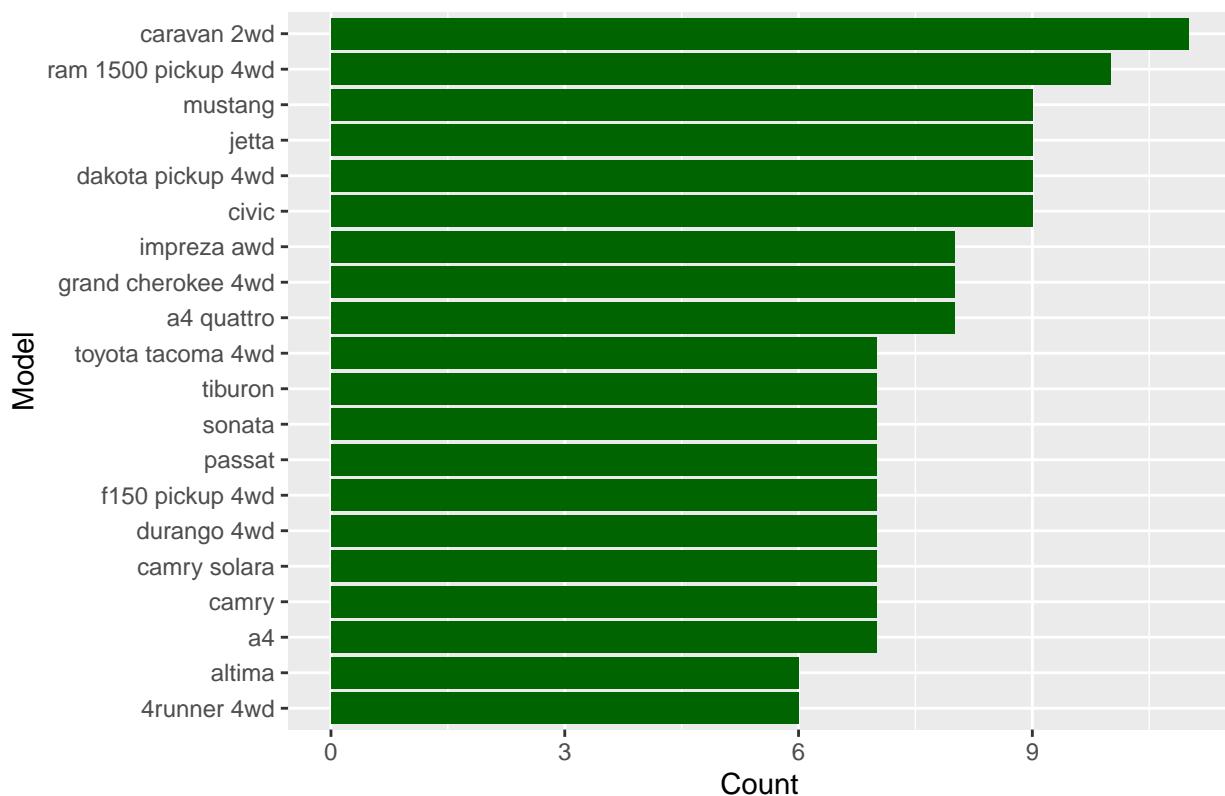


```

# 5b. Bar plot flipped
ggplot(top20_models, aes(x=reorder(model, n), y=n)) +
  geom_bar(stat="identity", fill="darkgreen") +
  coord_flip() +
  labs(title="Top 20 Models by Count (Flipped)",
x="Model", y="Count")

```

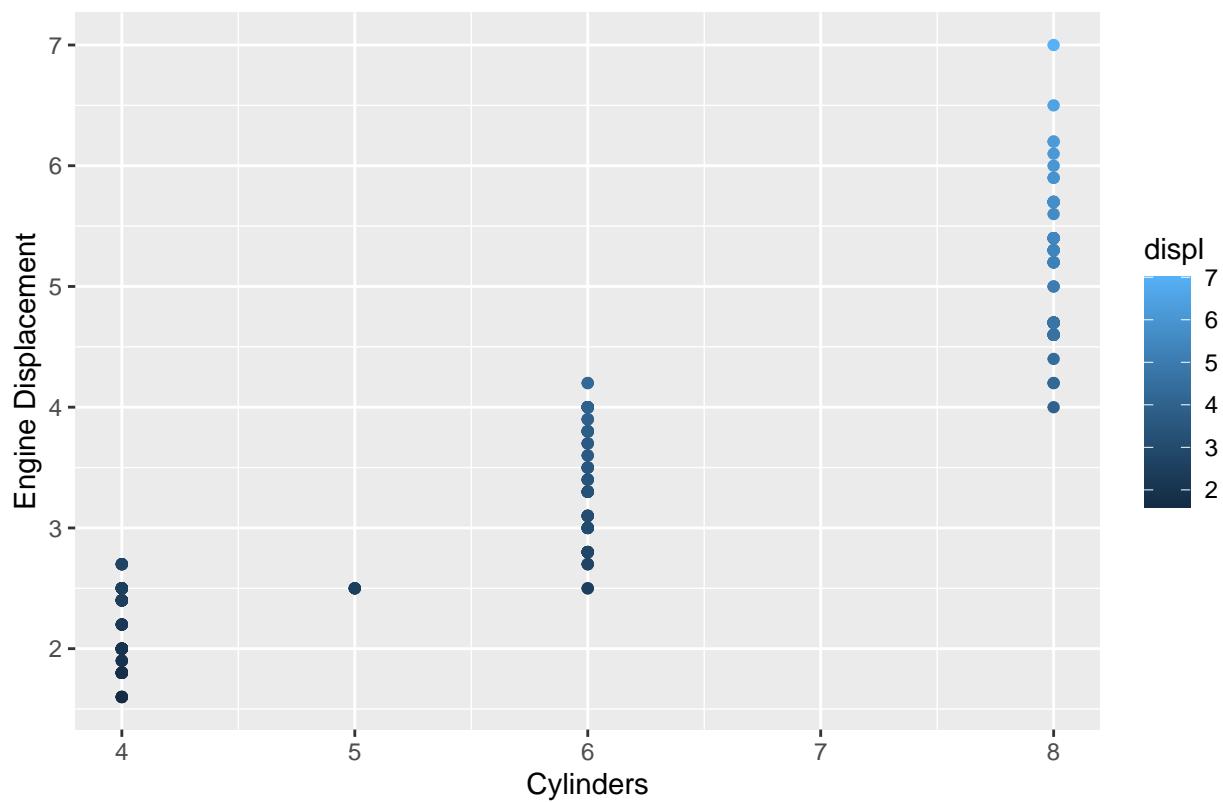
Top 20 Models by Count (Flipped)



```
# 6. cyl vs displ
```

```
ggplot(mpg, aes(x=cyl, y=displ, color=displ)) +  
  geom_point() +  
  labs(title="Relationship between No. of Cylinders and Engine Displacement",  
       x="Cylinders", y="Engine Displacement")
```

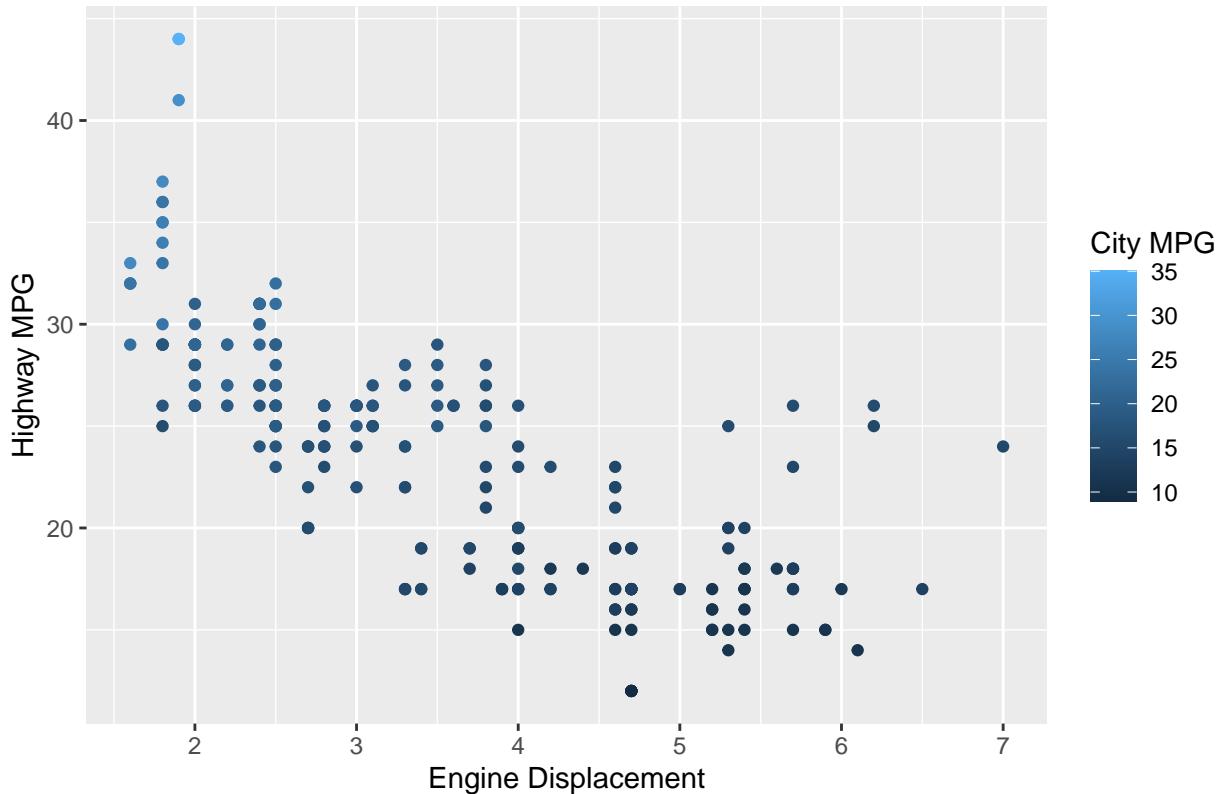
Relationship between No. of Cylinders and Engine Displacement



```
# 7. displ vs hwy mapped to continuous variable

ggplot(mpg, aes(x=displ, y=hwy, color=cty)) +
  geom_point() +
  labs(title="Engine Displacement vs Highway MPG",
       x="Engine Displacement", y="Highway MPG",
       color="City MPG")
```

Engine Displacement vs Highway MPG



```
# 9. Import alexa_file.xlsx

alexa <- read_excel("alexa_file.xlsx")

# 9a. Observations and columns
nrow(alexa)

## [1] 3150
ncol(alexa)

## [1] 5

# 9b. Group variations
var_total <- alexa %>%
  group_by(variation) %>%
  summarise(total = n())

var_total

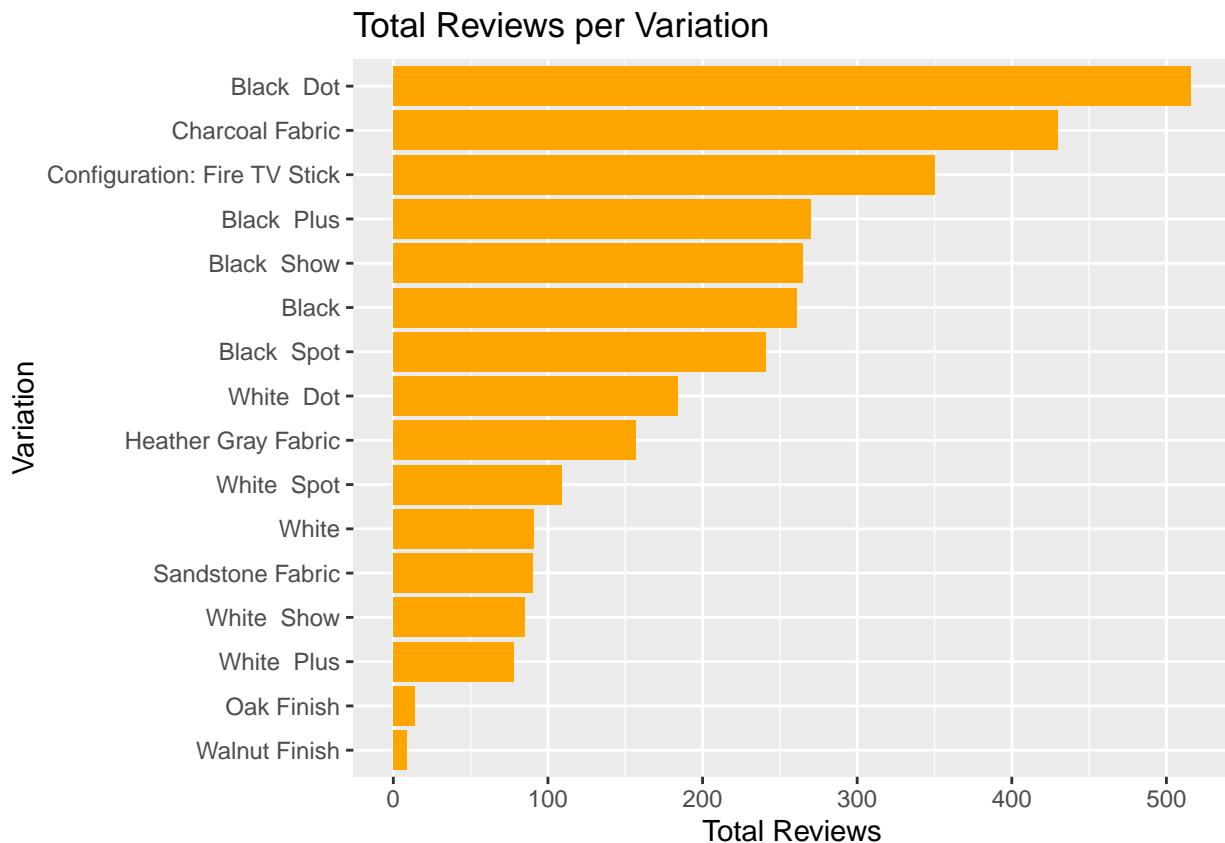
## # A tibble: 16 x 2
##   variation           total
##   <chr>              <int>
## 1 Black                261
## 2 Black Dot             516
## 3 Black Plus            270
## 4 Black Show             265
## 5 Black Spot              241
## 6 Charcoal Fabric        430
```

```

## 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

# 9c. Plot variations
ggplot(var_total, aes(x=reorder(variation, total), y=total)) +
  geom_col(fill="orange") +
  coord_flip() +
  labs(title="Total Reviews per Variation",
       x="Variation", y="Total Reviews")

```



```

# 9d. Line plot for verified reviews over time
ggplot(alexa, 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 inexpensive 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 "Group 2", but lately I've been getting terrible support. The guy that took my call just rambled off a (completely unhelpful) script at me.

Noting to add this bulk to my Alexa Echo Plus. Everything I tried ended in a "Discovery Failed" message. I tried to set it up multiple pages. The one thing that did not work is when screen cards do not really relate what I tried to set them up for.

9e. Variation vs Ratings

```
var_rating <- alexa %>%
  group_by(variation) %>%
  summarise(avg_rating = mean(rating, na.rm=TRUE))

ggplot(var_rating, aes(x=reorder(variation, avg_rating), y=avg_rating)) +
  geom_col(fill="skyblue") +
  coord_flip() +
  labs(title="Average Rating per Variation",
       x="Variation", y="Average Rating")
```

