RWorksheet_Saludo#4c

```
install.packages("readxl")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.5'
## (as 'lib' is unspecified)
```

```
library(ggplot2)
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
```

#1 #a

```
mpg_data <- read.csv("mpg.csv")
str(mpg)
```

```
## tibble [234 x 11] (S3: tbl_df/tbl/data.frame)
##  $ 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        : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
##  $ cyl         : int [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         : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...
##  $ hwy         : int [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" ...
```

#b manufacturer, model, trans, drv, fl, class, year (often treated as categorical) #c displ, cty, hwy, cyl

#2 #a

```
library(dplyr)

manufacturer_models <- mpg %>%
  group_by(manufacturer) %>%
  summarise(unique_models = n_distinct(model)) %>%
  arrange(desc(unique_models))

manufacturer_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
```

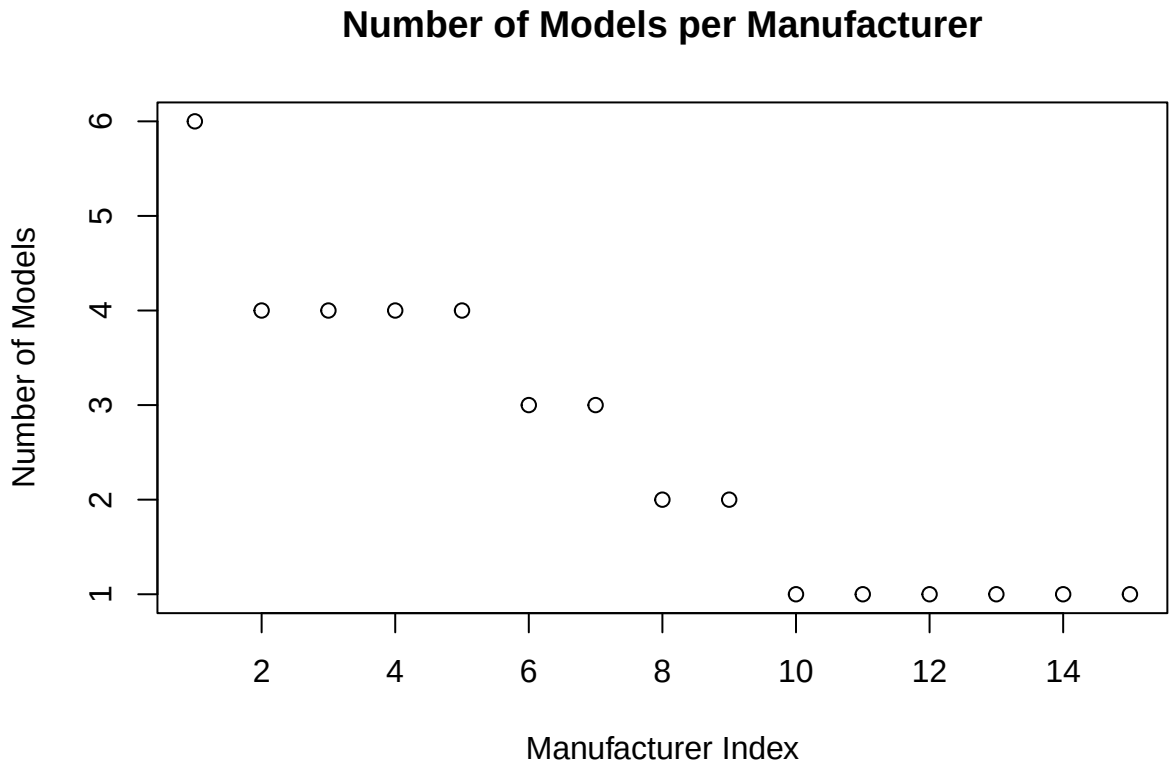
#a

```
model_variations <- mpg %>%
  group_by(model) %>%
  summarise(variations = n()) %>%
  arrange(desc(variations))

model_variations
```

```
## # A tibble: 38 x 2
##    model               variations
##    <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
```

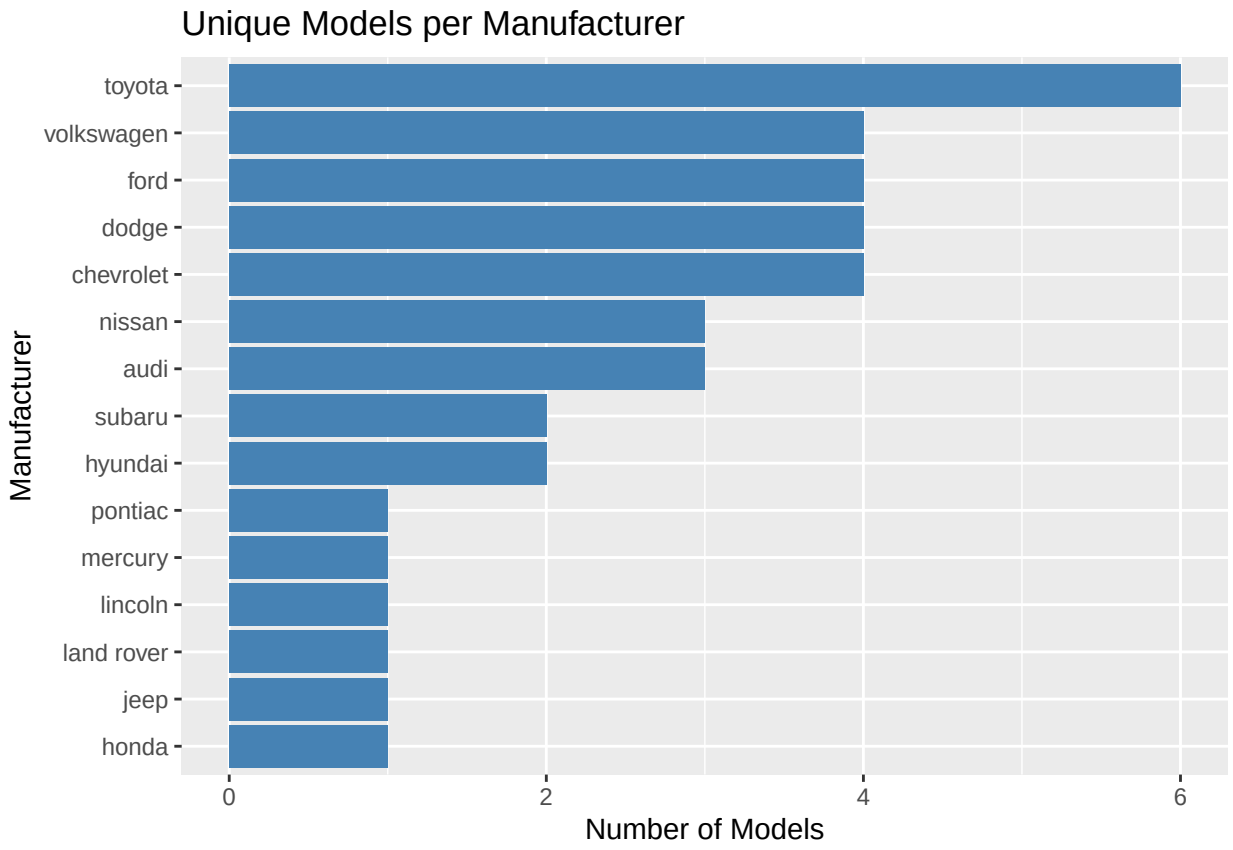
#b

```
plot(manufacturer_models$unique_models,
     main = "Number of Models per Manufacturer",
     xlab = "Manufacturer Index",
     ylab = "Number of Models")
```



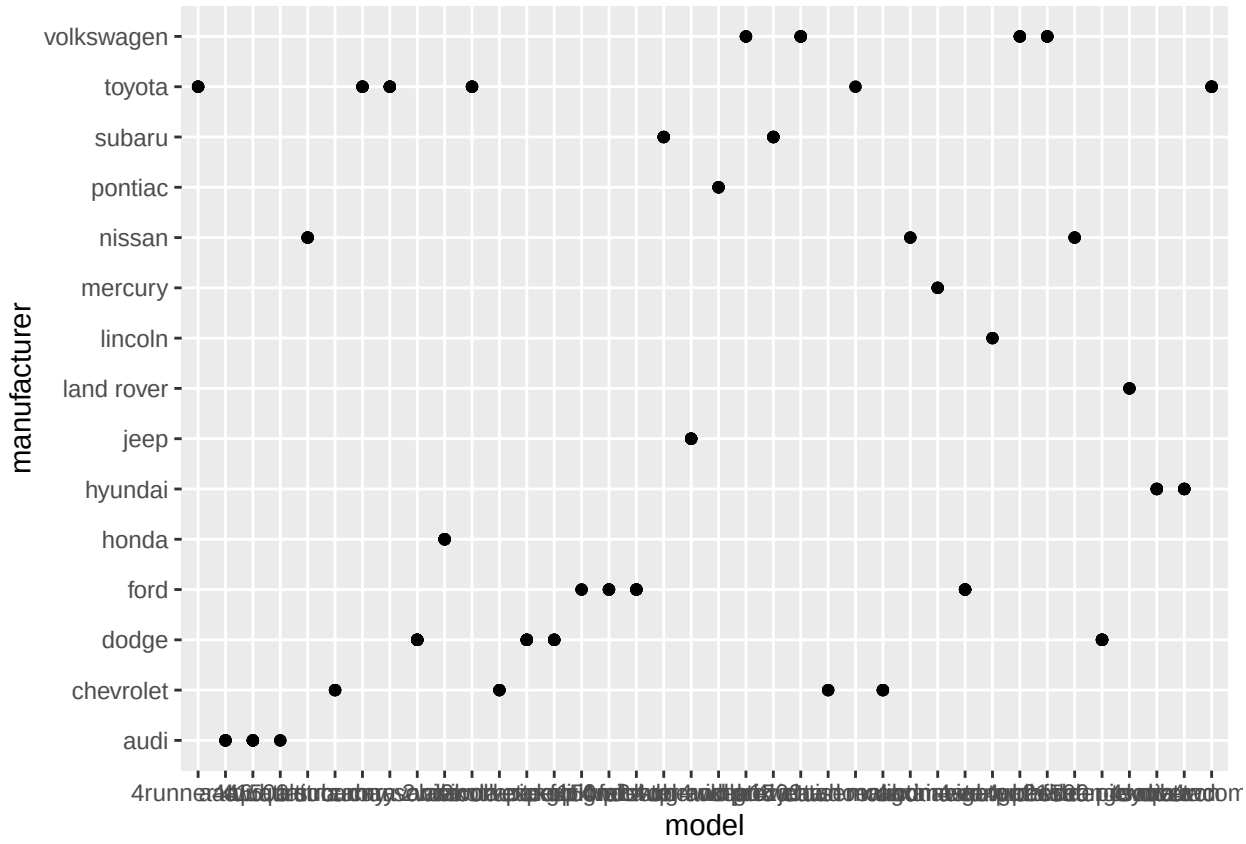
#b

```
library(ggplot2)

ggplot(manufacturer_models,
       aes(x = reorder(manufacturer, unique_models),
           y = unique_models)) +
  geom_col(fill = "steelblue") +
  coord_flip() +
  labs(title = "Unique Models per Manufacturer",
       x = "Manufacturer",
       y = "Number of Models")
```



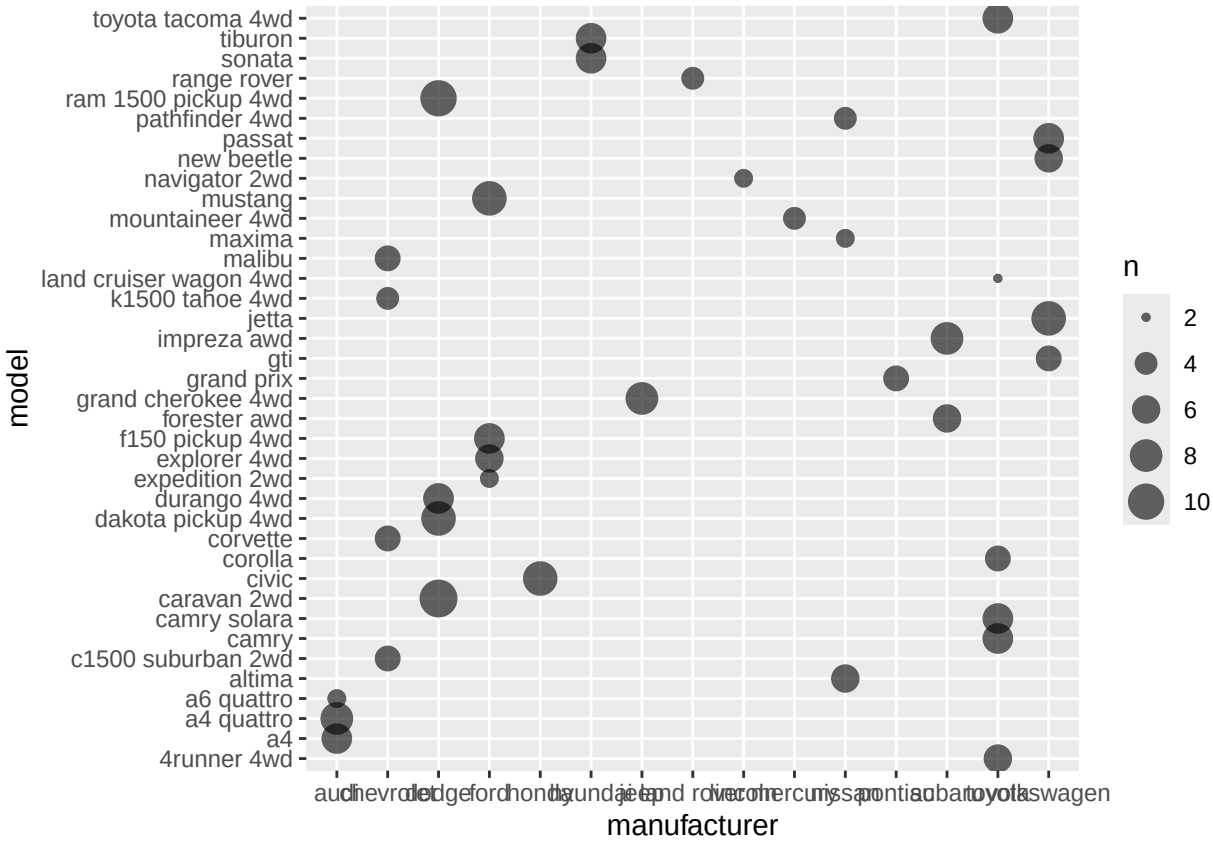
#2b #a

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



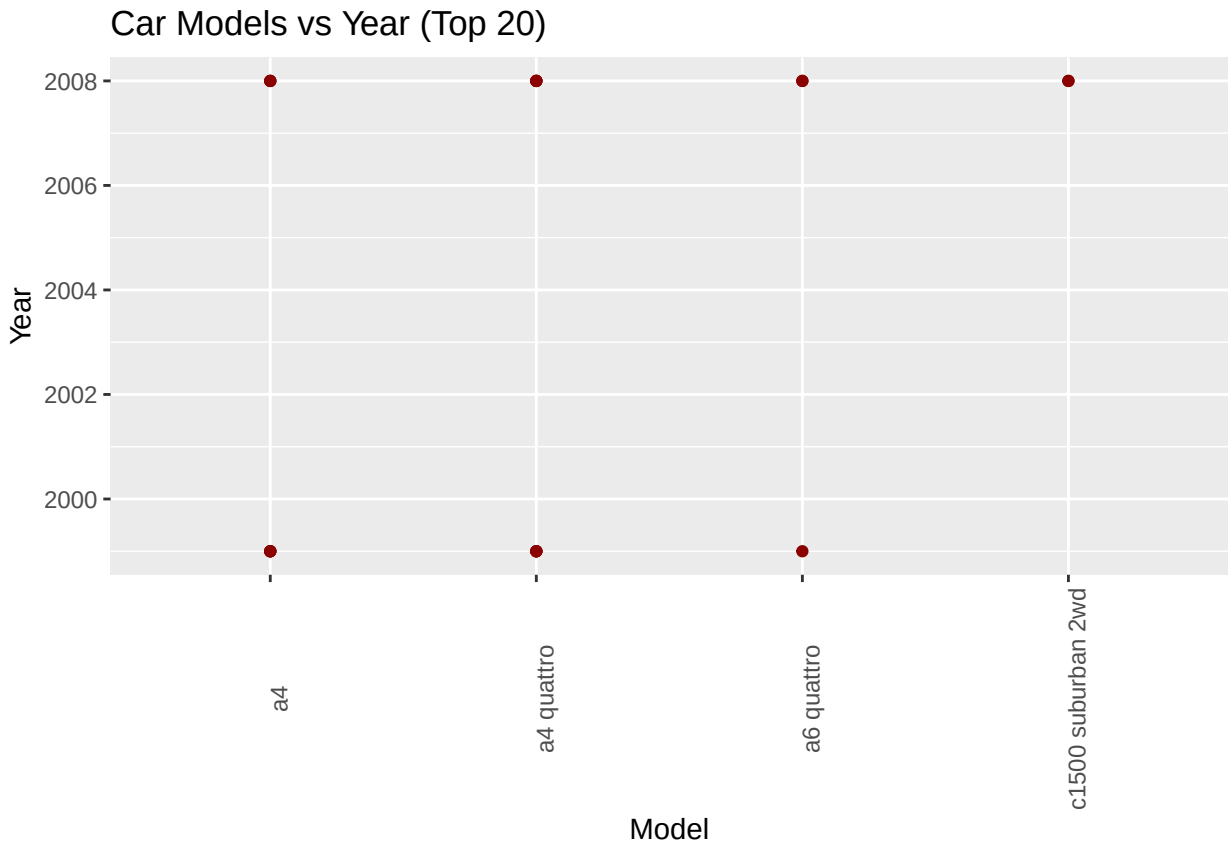
#b #not useful

```
mpg %>%
  count(manufacturer, model) %>%
  ggplot(aes(x = manufacturer, y = model, size = n)) +
  geom_point(alpha = 0.6)
```



#3

```
top20 <- mpg[1:20, ]

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



#4

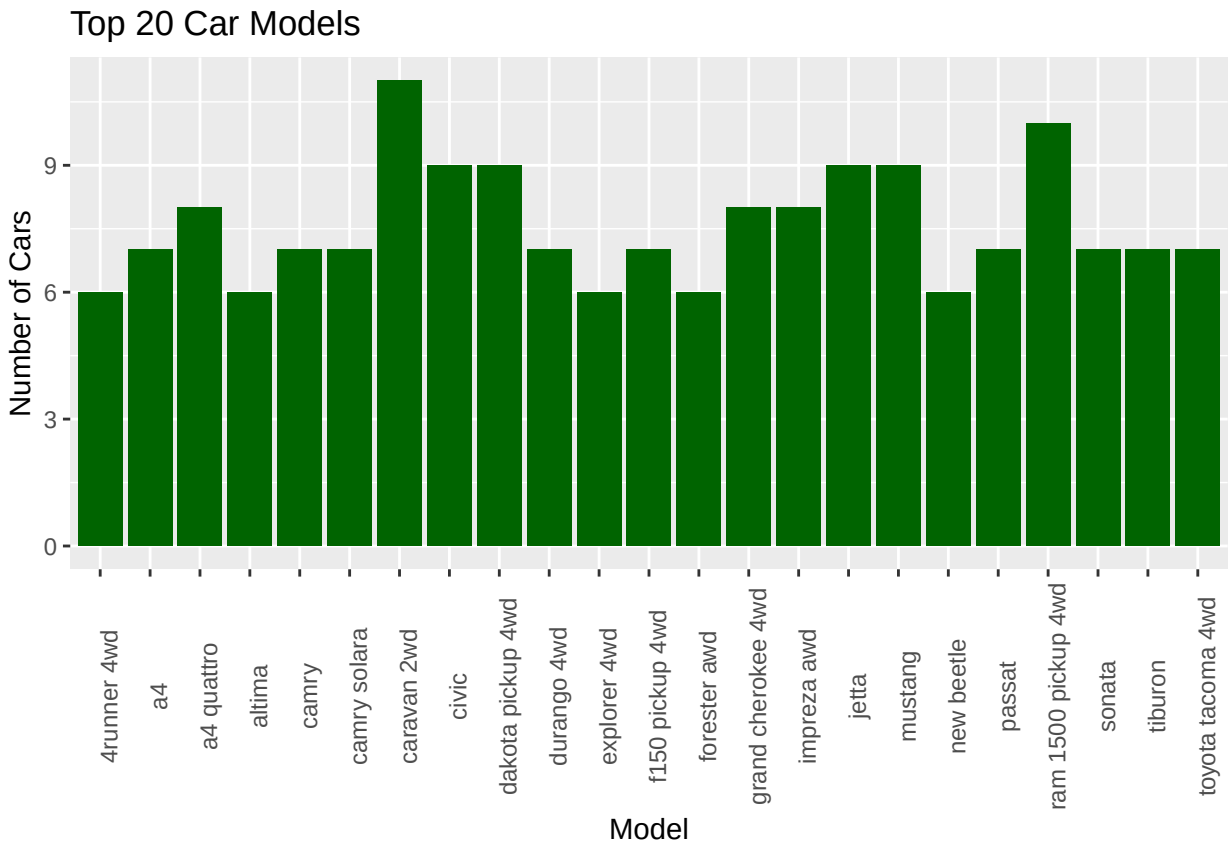
```
model_count <- mpg %>%
  group_by(model) %>%
  summarise(count = n()) %>%
  arrange(desc(count))

model_count
```

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

#a

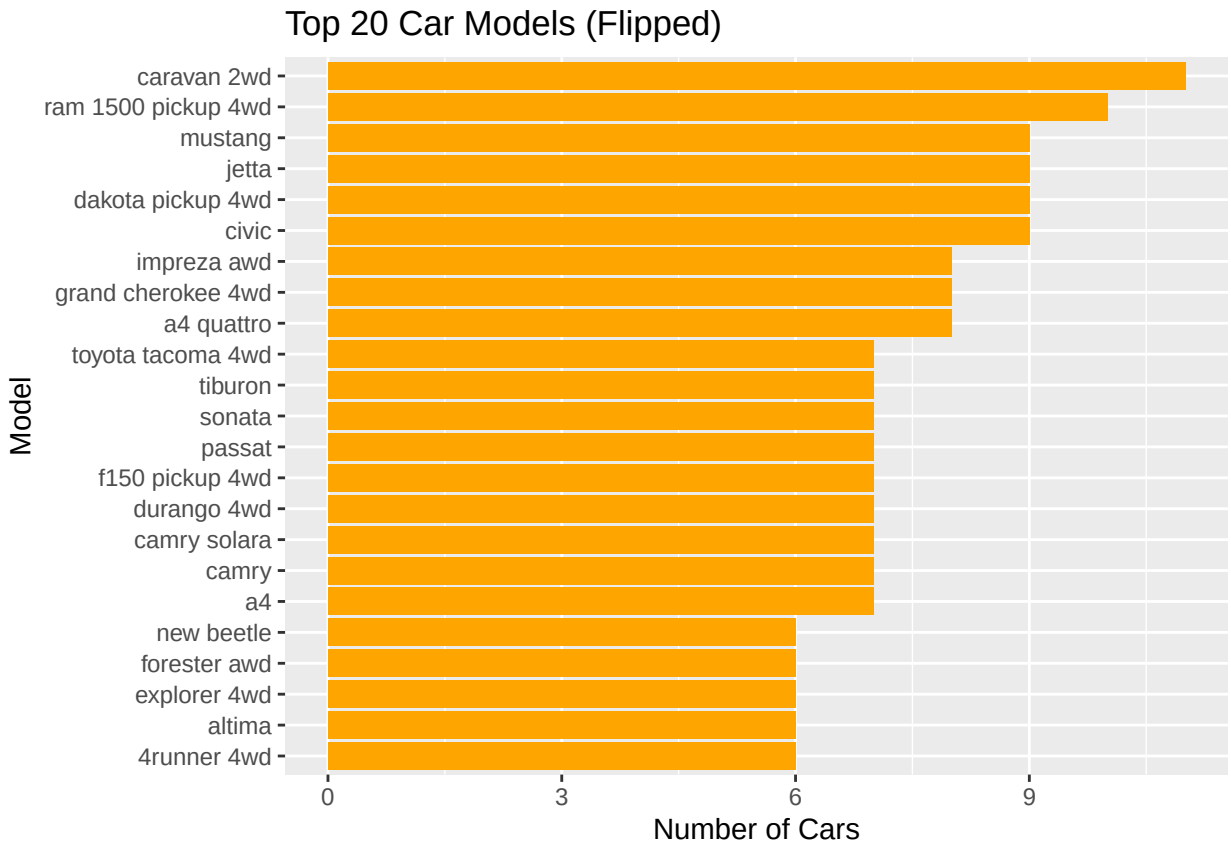
```
top20_models <- model_count %>% slice_max(count, n = 20)

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

```
  geom_bar(stat = "identity", fill = "darkgreen") +
  labs(title = "Top 20 Car Models",
       x = "Model",
       y = "Number of Cars") +
  theme(axis.text.x = element_text(angle = 90))
```



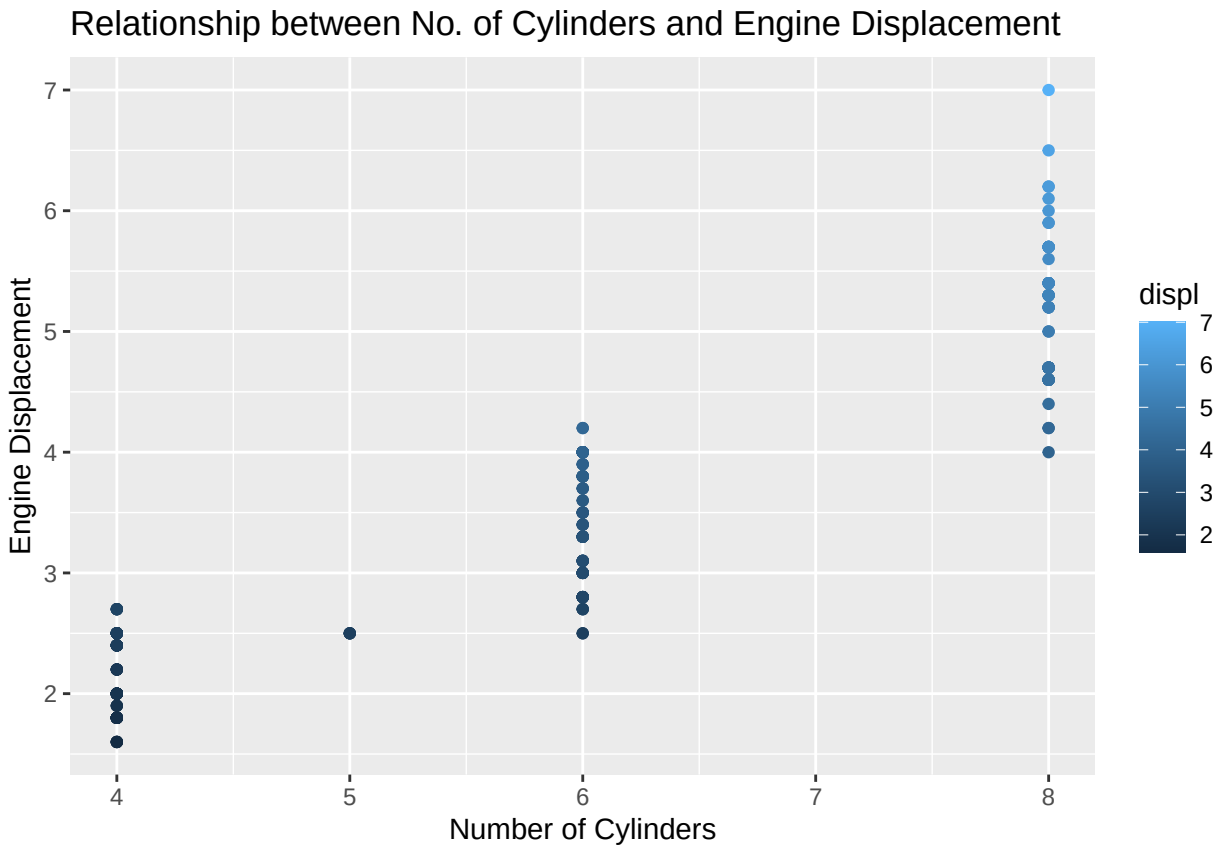
#b

```
ggplot(top20_models, aes(x = reorder(model, count), y = count)) +
  geom_bar(stat = "identity", fill = "orange") +
  coord_flip() +
  labs(title = "Top 20 Car Models (Flipped)",
       x = "Model",
       y = "Number of Cars")
```



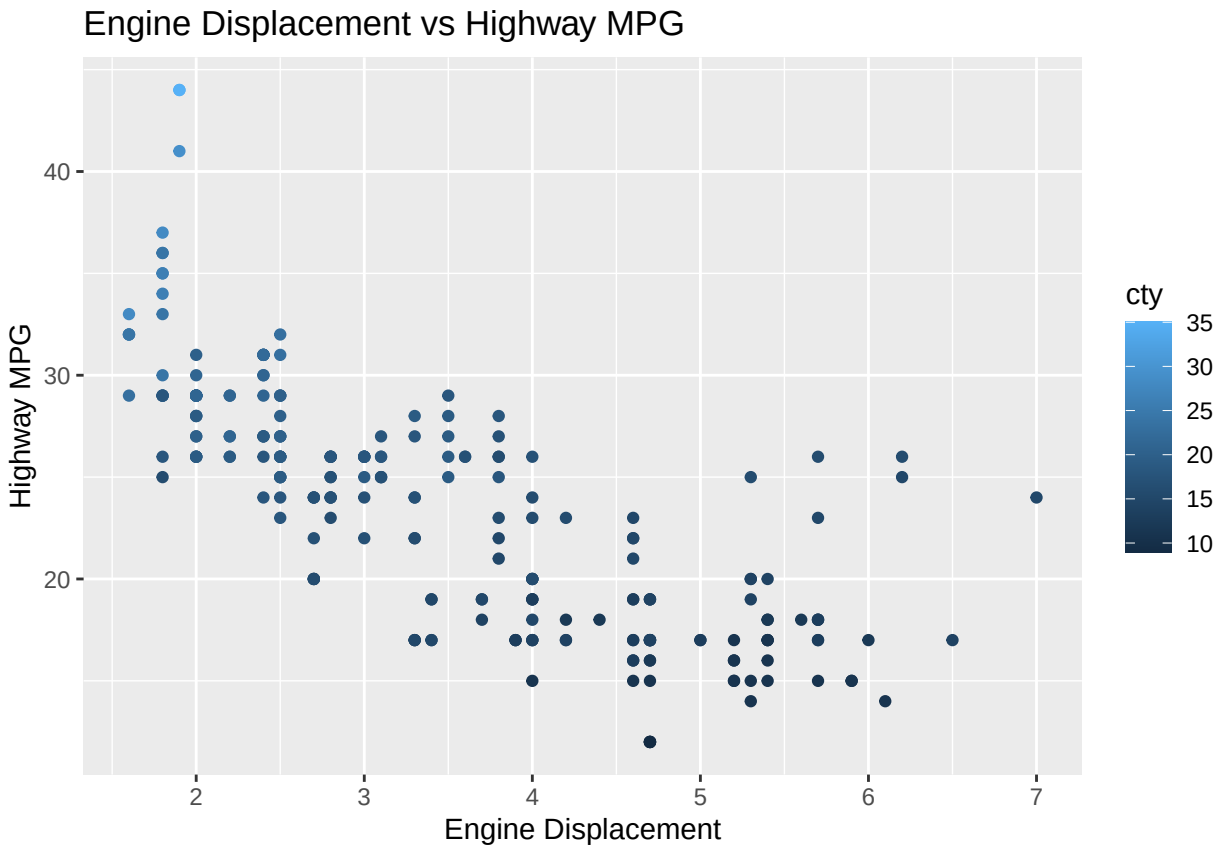
#5

```
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")
```



#6 #a

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



#7 #a

```
library(readxl)

alexa <- read_excel("alexa_file.xlsx")
```

#b

```
dim(alexa)
```

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

#c

```
variation_count <- alexa %>%
  group_by(variation) %>%
  summarise(total = n())

variation_count
```

```
## # 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
```

#d

```
ggplot(alexa, aes(x = date, y = verified_reviews)) +
  geom_line(color = "blue") +
  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

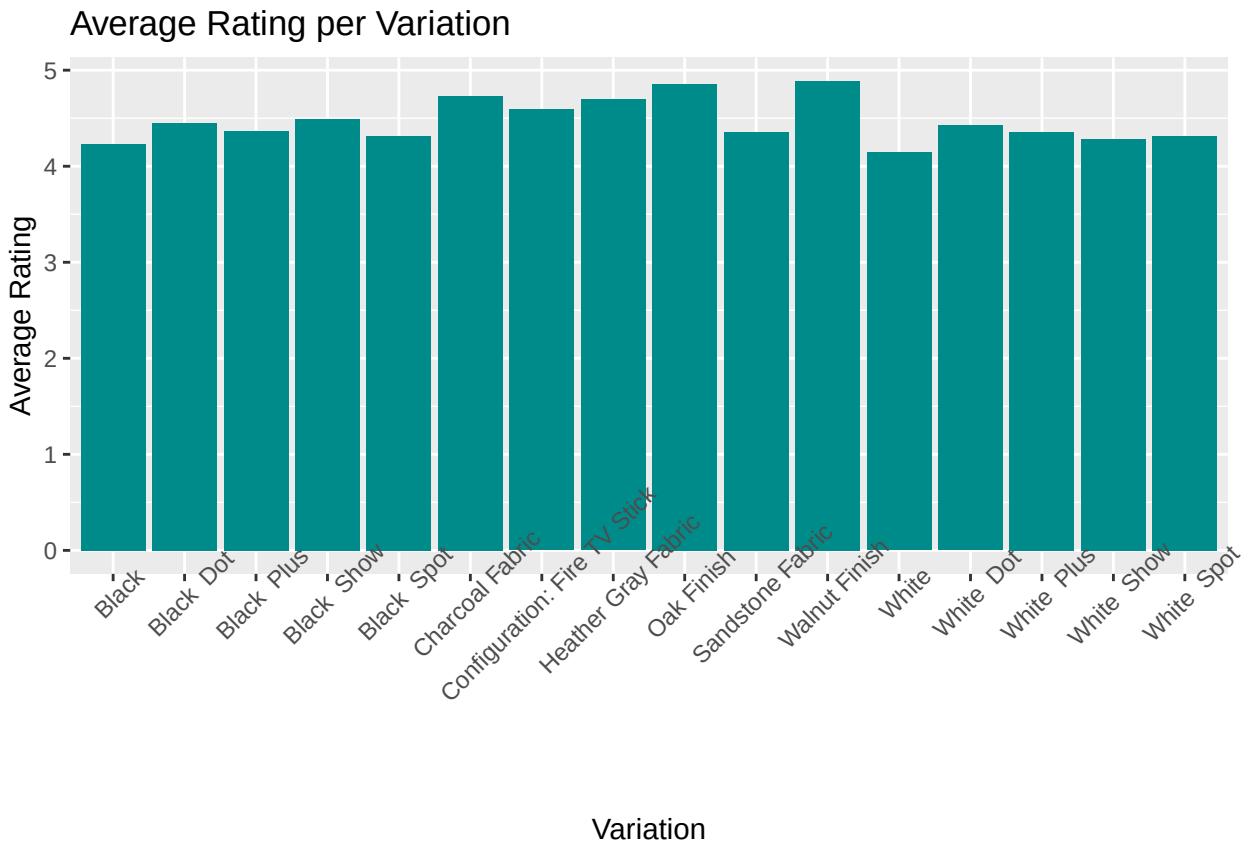
expensive alternative option to fill the gap. Ordered the Amazon Fire Stick from Best Buy. Instructions were short and

one of the lights by saying "Alexa, turn off the second light". In the Alexa app, I created a 'Group' with &#34
, but lately I've been getting terrible support. The guy that took my call just rambled off a (completely unhelpful) script a

nting to add this both to my Alexa Echo Plus. Everything I tried ended in a Discovery Failed message. I tried to set
gh multiple pages. The one thing that I am not a fan of is the home screen cards do not really rotate that much - they

#e

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

ggplot(rating_variation, aes(x = variation, y = avg_rating)) +
  geom_col(fill = "darkcyan") +
  labs(title = "Average Rating per Variation",
       x = "Variation",
       y = "Average Rating") +
  theme(axis.text.x = element_text(angle = 45))
```


Average Rating per Variation
Average Rating
5
4
3
2
1
0
Black
Black Dot
Black Plus
Black Show
Black Spot
Charcoal Fabric
Configuration: Fire TV Stick
Heather Gray Fabric
Oak Finish
Sandstone Fabric
Walnut Finish
White
White Dot
White Plus
White Show
White Spot
Variation