

# RWorksheet\_Camarista#4c

John Lyxton Camarista

2024-11-01

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

```
library(readr)
```

```
mpg_data <- read_csv("E:/Github/Data Science Worksheets/DataScience_Worksheets_Camarista/Worksheet#4/mpg_data.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'
```

```
head(mpg_data)
```

```
## # 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~
```

- b. Which variables from mpg dataset are categorical?
  - The categorical variables are: manufacturer, model, year, trans, drv, fl, and class.
- c. Which are continuous variables?
  - \*The continous variables are: displ, cyl, cty, and hwy.

##2. Which manufacturer has the most models in this data set? Which model has the most variations? Show your answer. - a. Group the manufacturers and find the unique models. Show your codes and result.

```
# Load necessary library
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
```

```
# Get unique models for each manufacturer
unique_models <- mpg_data %>%
  select(manufacturer, model) %>%
  distinct() %>%
  arrange(manufacturer)

# Display the result
unique_models
```

```
## # A tibble: 38 x 2
##   manufacturer model
##   <chr>         <chr>
## 1 audi          a4
## 2 audi          a4 quattro
## 3 audi          a6 quattro
## 4 chevrolet    c1500 suburban 2wd
## 5 chevrolet    corvette
## 6 chevrolet    k1500 tahoe 4wd
## 7 chevrolet    malibu
## 8 dodge         caravan 2wd
## 9 dodge         dakota pickup 4wd
## 10 dodge        durango 4wd
## # i 28 more rows
```

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

```
library(dplyr)

model_count <- mpg_data %>%
  select(manufacturer, model) %>%
  distinct() %>%
  group_by(manufacturer) %>%
  summarise(model = n())

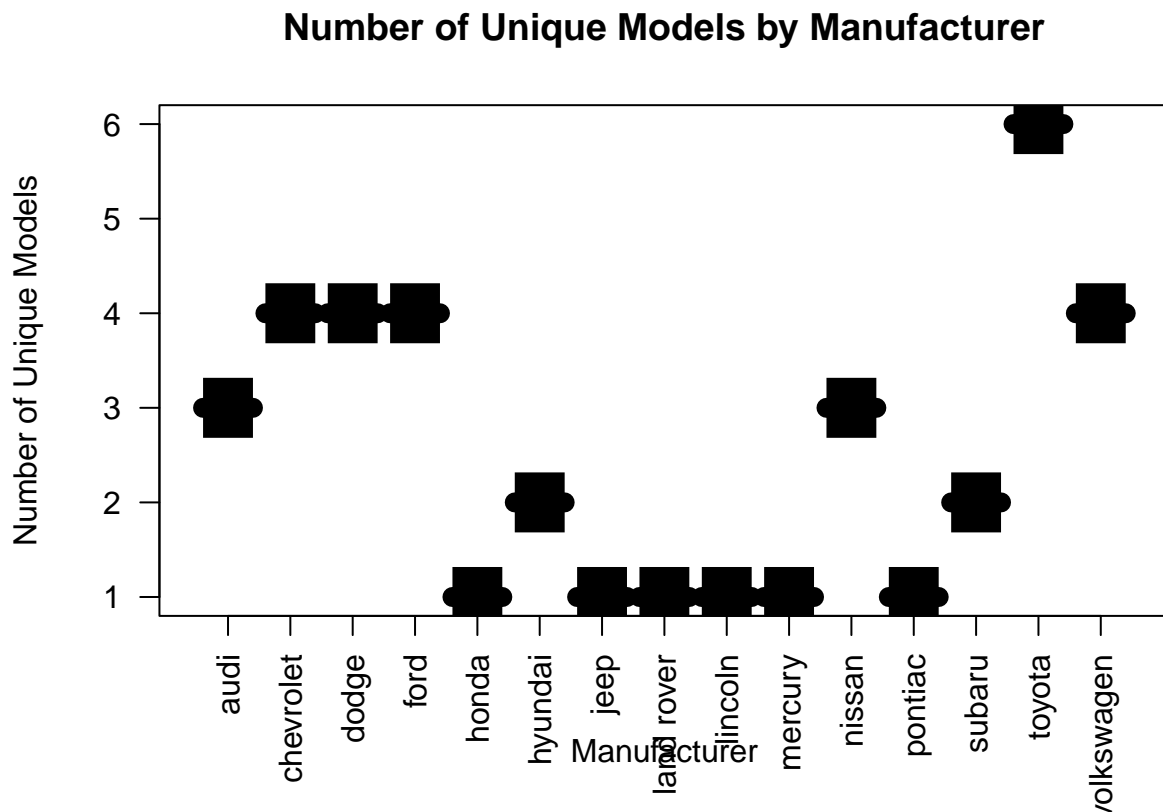
model_count$manufacturer <- factor(model_count$manufacturer, levels = unique(model_count$manufacturer))

plot(
```

```

model_count$manufacturer,
model_count$model,
type = "h", lines,
main = "Number of Unique Models by Manufacturer",
xlab = "Manufacturer",
ylab = "Number of Unique Models",
col = "skyblue",
las = 2,
lwd = 10
)

```

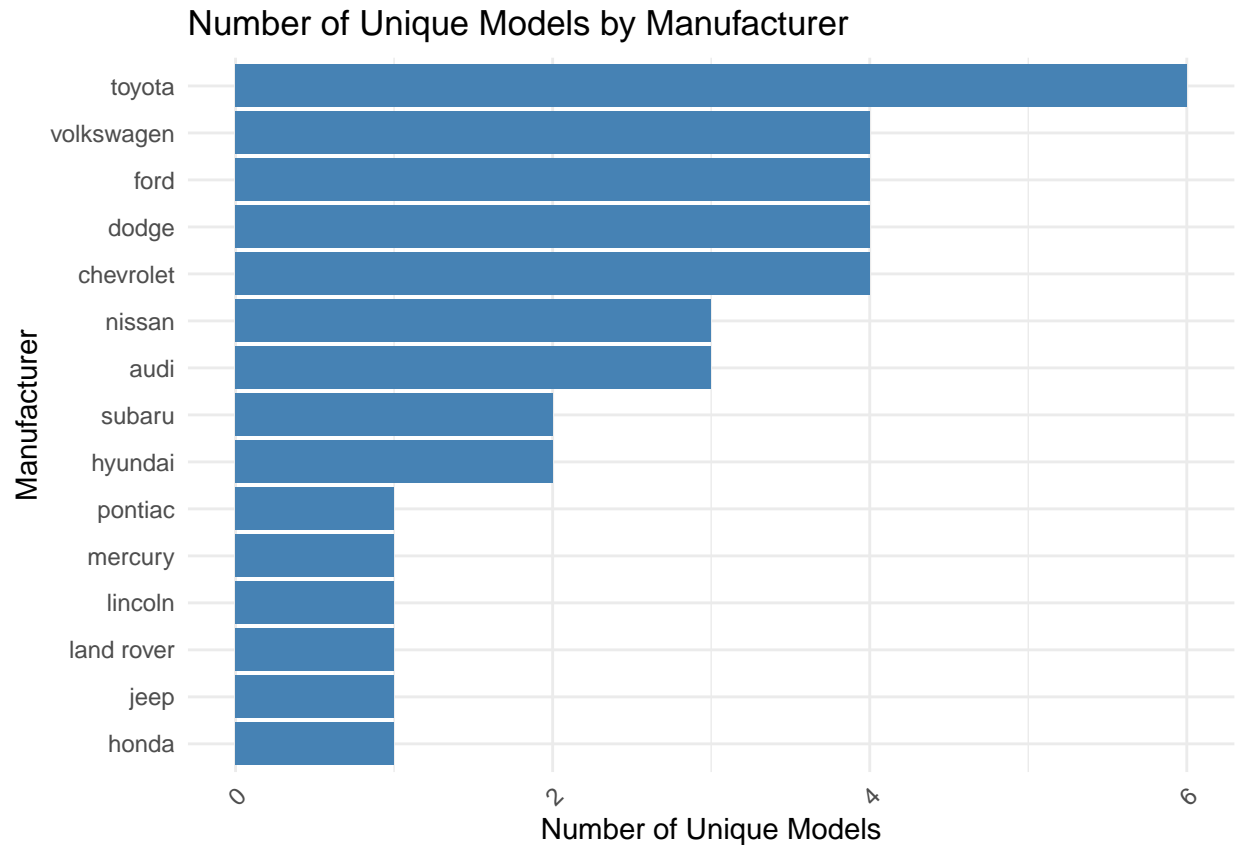


```

library(ggplot2)

ggplot(model_count, aes(x = reorder(manufacturer, model), y = model)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  coord_flip() +
  labs(title = "Number of Unique Models by Manufacturer",
       x = "Manufacturer",
       y = "Number of Unique Models") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```



##3. Same dataset will be used. You are going to show the relationship of the model and the manufacturer.

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

##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".

##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?

##7. Import the `traffic.csv` onto your R environment.

##8. From `alexa_file.xlsx`, import it to your environment