Project Final

Kimberly Sultan

12/16/2018

Cars for 2017:

Ford Escape Honda CRV Hyundai Santa Fe Toyota Rav4

Ford Escape

```
#This url gives us the car basic dataset
#make, model, other, and price for each model allowed.

#This returns the Ford Escape

url_Ford <- "https://www.cars.com/for-sale/searchresults.action/?rd=99999&mkId=20015&mdId=21088&searchS

#Reading the HTML code from the website
webpage <- read_html(url_Ford)

#Using CSS selectors to scrap the data
make_data_html <- html_nodes(webpage,'.listing-row__title')

price_data_html <- html_nodes(webpage,'.listing-row__price')

#Converting the car data to text or numeric (or in PowerBI)
yr_make_mod_data <- html_text(make_data_html)
yr_make_mod_data <- gsub("\n", "", yr_make_mod_data)
trimws(yr_make_mod_data)
```

```
##  [1] "2017 Ford Escape Titanium" "2017 Ford Escape Titanium"
##  [3] "2017 Ford Escape Titanium" "2017 Ford Escape Titanium"
##  [5] "2017 Ford Escape Titanium" "2017 Ford Escape Titanium"
##  [7] "2017 Ford Escape Titanium" "2017 Ford Escape SE"
##  [9] "2017 Ford Escape Titanium" "2017 Ford Escape Titanium"
## [11] "2017 Ford Escape Titanium" "2017 Ford Escape S"
## [13] "2017 Ford Escape SE"       "2017 Ford Escape SE"
## [15] "2017 Ford Escape Titanium" "2017 Ford Escape SE"
## [17] "2017 Ford Escape SE"       "2017 Ford Escape Titanium"
## [19] "2017 Ford Escape SE"       "2017 Ford Escape Titanium"
## [21] "2017 Ford Escape Titanium" "2017 Ford Escape Titanium"
## [23] "2017 Ford Escape SE"       "2017 Ford Escape Titanium"
## [25] "2017 Ford Escape Titanium" "2017 Ford Escape SE"
## [27] "2017 Ford Escape Titanium" "2017 Ford Escape SE"
## [29] "2017 Ford Escape Titanium" "2017 Ford Escape Titanium"
## [31] "2017 Ford Escape Titanium" "2017 Ford Escape SE"
## [33] "2017 Ford Escape Titanium" "2017 Ford Escape Titanium"
## [35] "2017 Ford Escape Titanium" "2017 Ford Escape Titanium"
```

```
## [37] "2017 Ford Escape Titanium" "2017 Ford Escape Titanium"
## [39] "2017 Ford Escape Titanium" "2017 Ford Escape Titanium"
## [41] "2017 Ford Escape Titanium" "2017 Ford Escape Titanium"
## [43] "2017 Ford Escape Titanium" "2017 Ford Escape Titanium"
## [45] "2017 Ford Escape Titanium" "2017 Ford Escape Titanium"
## [47] "2017 Ford Escape Titanium" "2017 Ford Escape Titanium"
## [49] "2017 Ford Escape Titanium" "2017 Ford Escape Titanium"
```

```
#Fix issues with price, text to numeric etc
price_data <- html_text(price_data_html)
price_data <- gsub(",", "", price_data)
price_data <- gsub("\\$", "", price_data)
price_data <- as.numeric(price_data)
price_data
```

```
##  [1] 28081 31049 29664 34558 36627 35677 36627 19987 28991 27840 28950
## [12] 19995 24990 31260 34385 16795 21242 27260 21879 27250 29700 28500
## [23] 31794 28145 27350 21997 27990 27563 29708 32505 28995 27987 26329
## [34] 29851 27075 31018 28257 30369 31649 29975 27964 20384 26998 28890
## [45] 33330 32999 29999 34038 32266 32548
```

```
data_all <- data.frame(cbind(yr_make_mod_data,price_data))
#data_all$price_data <- as.numeric(data_all$price_data)
data_all
```

```
##                                                yr_make_mod_data
## 1                                2017 Ford Escape Titanium
## 2                                2017 Ford Escape Titanium
## 3                                2017 Ford Escape Titanium
## 4                                2017 Ford Escape Titanium
## 5                                2017 Ford Escape Titanium
## 6                                2017 Ford Escape Titanium
## 7                                2017 Ford Escape Titanium
## 8                                      2017 Ford Escape SE
## 9                                2017 Ford Escape Titanium
## 10                               2017 Ford Escape Titanium
## 11                               2017 Ford Escape Titanium
## 12                                      2017 Ford Escape S
## 13                                     2017 Ford Escape SE
## 14                                     2017 Ford Escape SE
## 15                               2017 Ford Escape Titanium
## 16                                     2017 Ford Escape SE
## 17                                     2017 Ford Escape SE
## 18                               2017 Ford Escape Titanium
## 19                                     2017 Ford Escape SE
## 20                               2017 Ford Escape Titanium
## 21                               2017 Ford Escape Titanium
## 22                               2017 Ford Escape Titanium
## 23                                     2017 Ford Escape SE
## 24                               2017 Ford Escape Titanium
## 25                               2017 Ford Escape Titanium
## 26                                     2017 Ford Escape SE
## 27                               2017 Ford Escape Titanium
## 28                                     2017 Ford Escape SE
## 29                               2017 Ford Escape Titanium
```

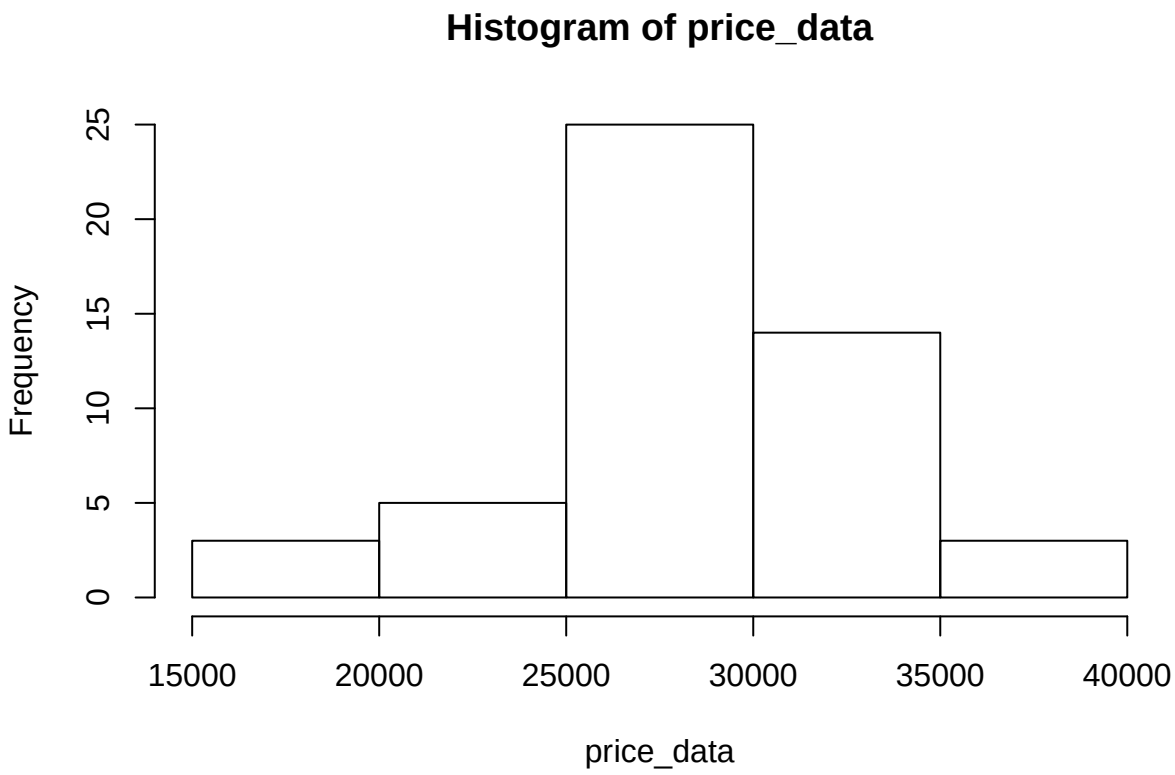
```
## 30                              2017 Ford Escape Titanium
## 31                              2017 Ford Escape Titanium
## 32                                    2017 Ford Escape SE
## 33                              2017 Ford Escape Titanium
## 34                              2017 Ford Escape Titanium
## 35                              2017 Ford Escape Titanium
## 36                              2017 Ford Escape Titanium
## 37                              2017 Ford Escape Titanium
## 38                              2017 Ford Escape Titanium
## 39                              2017 Ford Escape Titanium
## 40                              2017 Ford Escape Titanium
## 41                              2017 Ford Escape Titanium
## 42                              2017 Ford Escape Titanium
## 43                              2017 Ford Escape Titanium
## 44                              2017 Ford Escape Titanium
## 45                              2017 Ford Escape Titanium
## 46                              2017 Ford Escape Titanium
## 47                              2017 Ford Escape Titanium
## 48                              2017 Ford Escape Titanium
## 49                              2017 Ford Escape Titanium
## 50                              2017 Ford Escape Titanium
##    price_data
## 1       28081
## 2       31049
## 3       29664
## 4       34558
## 5       36627
## 6       35677
## 7       36627
## 8       19987
## 9       28991
## 10      27840
## 11      28950
## 12      19995
## 13      24990
## 14      31260
## 15      34385
## 16      16795
## 17      21242
## 18      27260
## 19      21879
## 20      27250
## 21      29700
## 22      28500
## 23      31794
## 24      28145
## 25      27350
## 26      21997
## 27      27990
## 28      27563
## 29      29708
## 30      32505
## 31      28995
## 32      27987
```

```
## 33      26329
## 34      29851
## 35      27075
## 36      31018
## 37      28257
## 38      30369
## 39      31649
## 40      29975
## 41      27964
## 42      20384
## 43      26998
## 44      28890
## 45      33330
## 46      32999
## 47      29999
## 48      34038
## 49      32266
## 50      32548
```

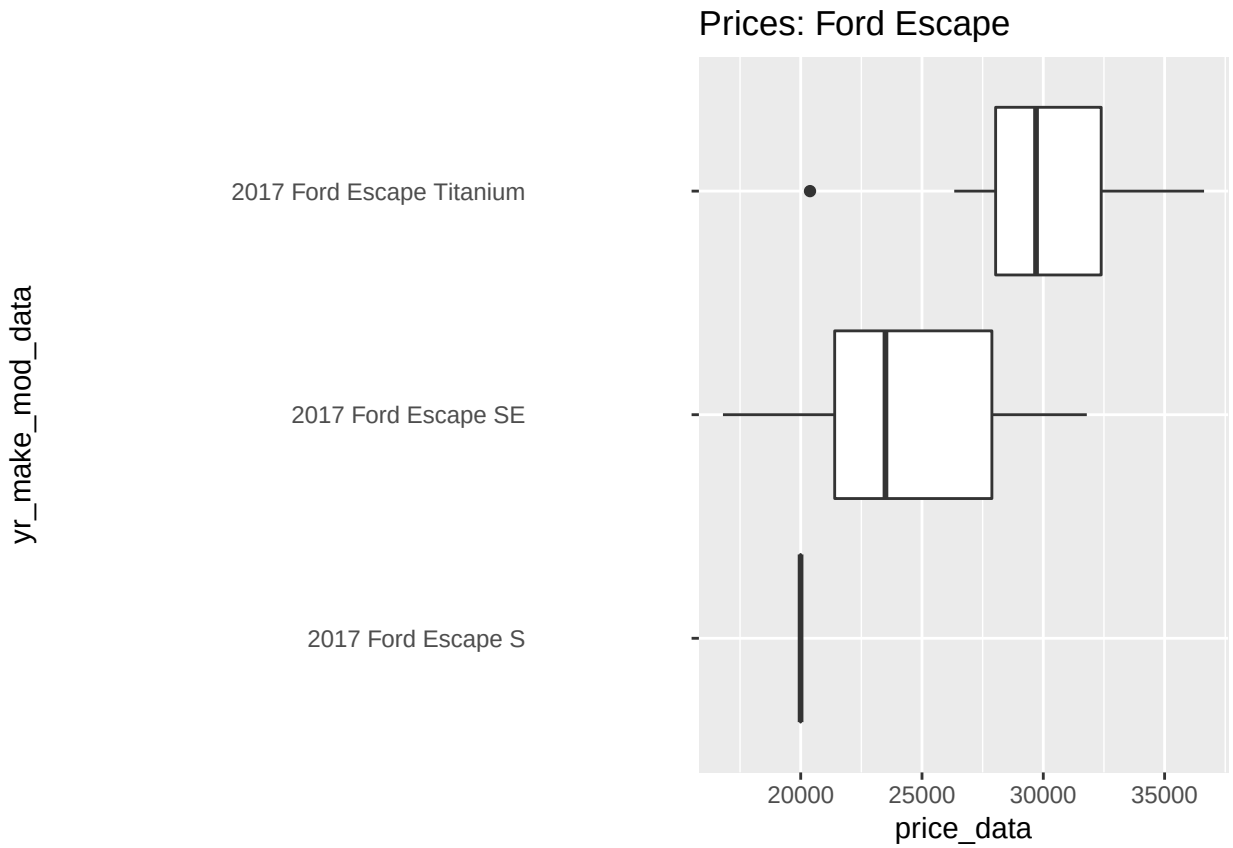
```
attach(data_all)
```

```
## The following objects are masked _by_ .GlobalEnv:
##
##     price_data, yr_make_mod_data
```

```
colnames(data_all) <- c("CarInfo","Price")

write.csv(data_all, file = "/Users/Kimberly/Desktop/School Maryville U/ford.csv", row.names = FALSE)

hist(price_data)
```



```
ggplot(data_all, aes(x = yr_make_mod_data , y = price_data))+
  geom_boxplot() +
  coord_flip()+
  ggtitle("Prices: Ford Escape")
```




```
df1 <- data_all
```

Honda CVR

```
library(rvest)
library(tidyverse)
library(stringr)
library(dplyr)
library(tidyr)
```

```
# code for Honda
#This returns the Honda CRV

url_Honda <- "https://www.cars.com/for-sale/searchresults.action/?mdId=20762&mkId=20017&page=1&perPage=

#Reading the HTML code from the website
webpage <- read_html(url_Honda)

#Using CSS selectors to scrap the data
make_data_html <- html_nodes(webpage,'.listing-row__title')
```

```
price_data_html <- html_nodes(webpage,'.listing-row__price')

#Converting the car data to text or numeric (or in PowerBI)
yr_make_mod_data <- html_text(make_data_html)
yr_make_mod_data <- gsub("\n", "", yr_make_mod_data)
trimws(yr_make_mod_data)
```

```
##  [1] "2017 Honda CR-V EX"      "2017 Honda CR-V EX"
##  [3] "2017 Honda CR-V LX"      "2017 Honda CR-V EX-L"
##  [5] "2017 Honda CR-V EX-L"    "2017 Honda CR-V LX"
##  [7] "2017 Honda CR-V EX-L"    "2017 Honda CR-V LX"
##  [9] "2017 Honda CR-V EX"      "2017 Honda CR-V EX-L"
## [11] "2017 Honda CR-V LX"      "2017 Honda CR-V LX"
## [13] "2017 Honda CR-V LX"      "2017 Honda CR-V EX-L"
## [15] "2017 Honda CR-V Touring" "2017 Honda CR-V EX-L"
## [17] "2017 Honda CR-V LX"      "2017 Honda CR-V Touring"
## [19] "2017 Honda CR-V LX"      "2017 Honda CR-V LX"
## [21] "2017 Honda CR-V EX-L"    "2017 Honda CR-V LX"
## [23] "2017 Honda CR-V LX"      "2017 Honda CR-V EX-L"
## [25] "2017 Honda CR-V EX-L"    "2017 Honda CR-V EX-L"
## [27] "2017 Honda CR-V EX-L"    "2017 Honda CR-V EX-L"
## [29] "2017 Honda CR-V EX"      "2017 Honda CR-V EX"
## [31] "2017 Honda CR-V EX-L"    "2017 Honda CR-V EX-L"
## [33] "2017 Honda CR-V Touring" "2017 Honda CR-V EX-L"
## [35] "2017 Honda CR-V EX-L"    "2017 Honda CR-V EX-L"
## [37] "2017 Honda CR-V EX"      "2017 Honda CR-V EX"
## [39] "2017 Honda CR-V LX"      "2017 Honda CR-V EX-L"
## [41] "2017 Honda CR-V EX-L"    "2017 Honda CR-V EX-L"
## [43] "2017 Honda CR-V EX-L"    "2017 Honda CR-V EX-L"
## [45] "2017 Honda CR-V Touring" "2017 Honda CR-V"
## [47] "2017 Honda CR-V EX"
```

```
#Fix issues with price, text to numeric etc
price_data <- html_text(price_data_html)
price_data <- gsub(",", "", price_data)
price_data <- gsub("\\$", "", price_data)
price_data <- as.numeric(price_data)
```

```
## Warning: NAs introduced by coercion
```

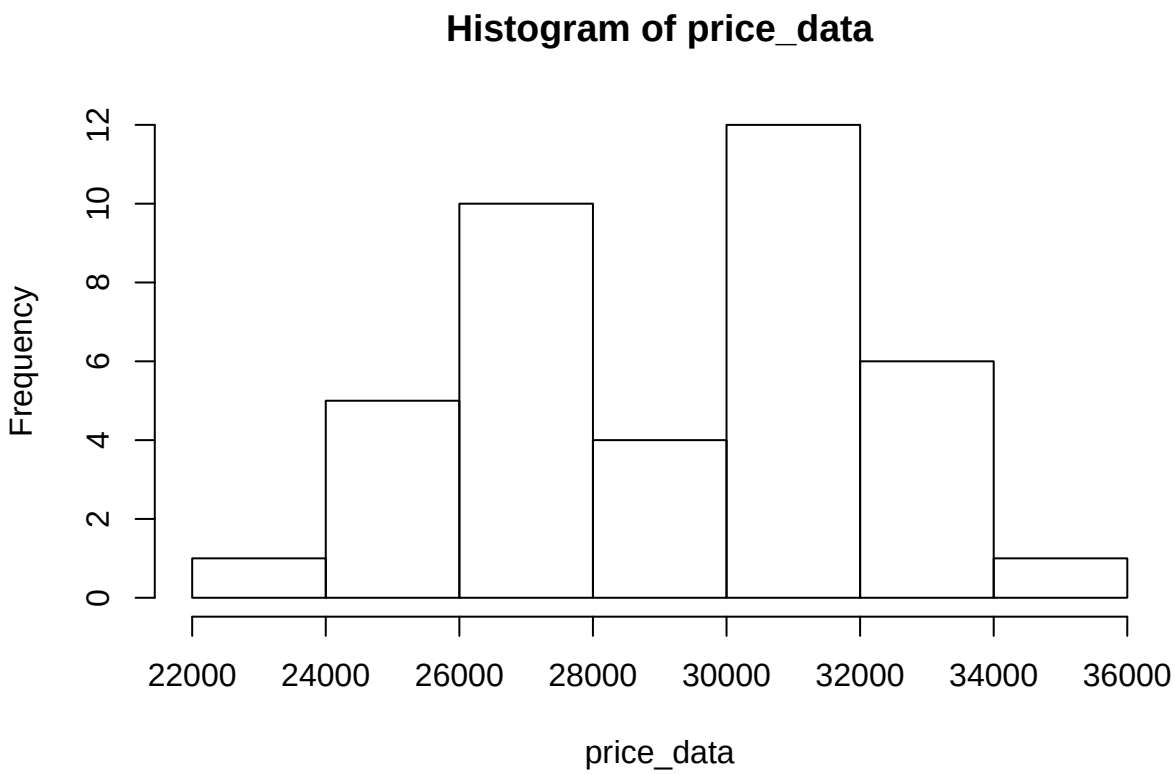
```
data_all <- data.frame(cbind(yr_make_mod_data,price_data))
#data_all$price_data <- as.numeric(data_all$price_data, length=5)

attach(data_all)
```

```
## The following objects are masked _by_ .GlobalEnv:
##
##     price_data, yr_make_mod_data
```

```
## The following objects are masked from data_all (pos = 3):
##
##     price_data, yr_make_mod_data
```

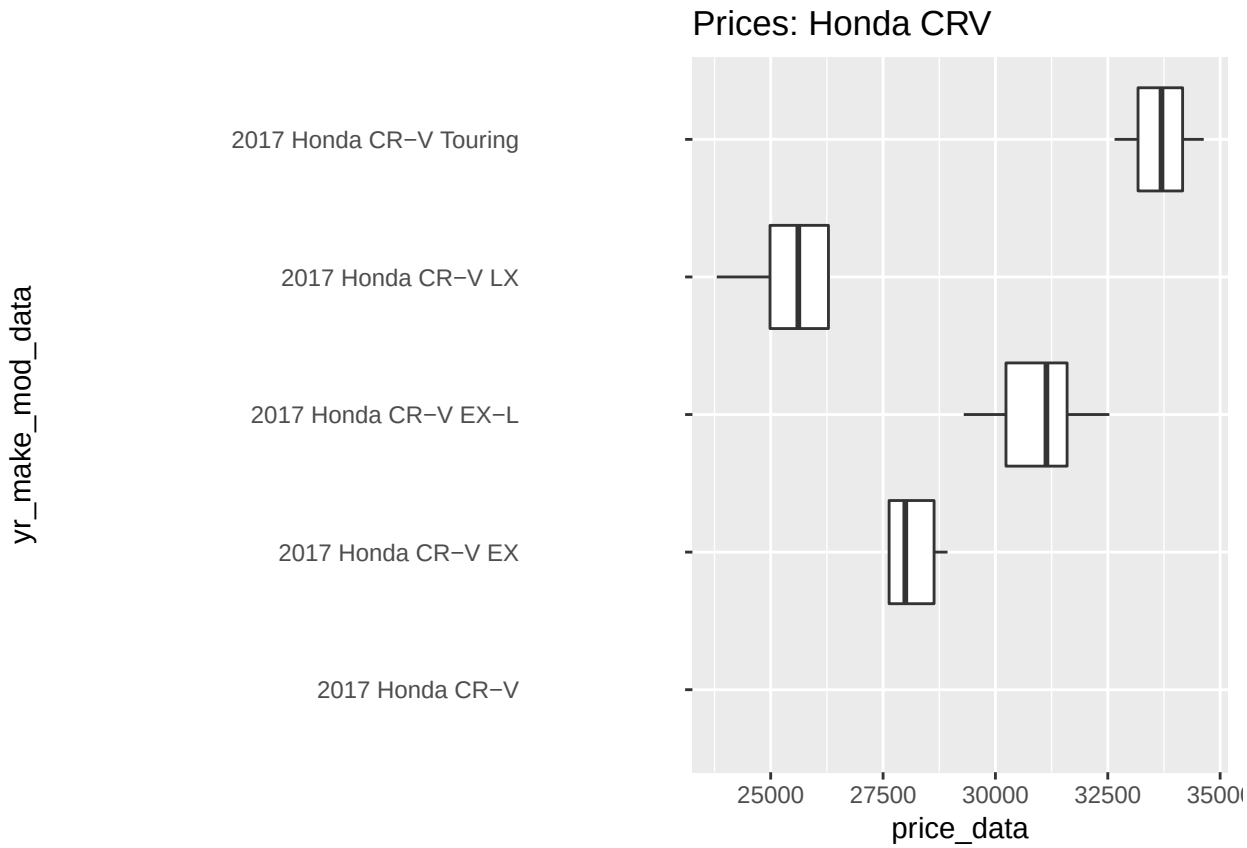
```
colnames(data_all) <- c("CarInfo","Price")
write.csv(data_all, file = "/Users/Kimberly/Desktop/School Maryville U/honda.csv", row.names = FALSE)
```

```
hist(price_data)
```




```
ggplot(data_all, aes(x = yr_make_mod_data, y = price_data))+
  geom_boxplot()+
  coord_flip()+
  ggtitle("Prices: Honda CRV")
```

```
## Warning: Removed 8 rows containing non-finite values (stat_boxplot).
```




```
df2 <- data_all
```

```
# code for Hyundai

#This returns the Hyundai Santa Fe

url_Hyundai <- "https://www.cars.com/for-sale/searchresults.action/?mdId=21899&mkId=20064&page=1&perPag

#Reading the HTML code from the website
webpage <- read_html(url_Hyundai)

#Using CSS selectors to scrap the data
make_data_html <- html_nodes(webpage,'.listing-row__title')
price_data_html <- html_nodes(webpage,'.listing-row__price')

#Converting the car data to text or numeric (or in PowerBI)
yr_make_mod_data <- html_text(make_data_html)
yr_make_mod_data <- gsub("\n", "", yr_make_mod_data)
trimws(yr_make_mod_data)
```

```
##  [1] "2017 Hyundai Santa Fe SE Ultimate"
##  [2] "2017 Hyundai Santa Fe Limited Ultimate"
##  [3] "2017 Hyundai Santa Fe SE"
##  [4] "2017 Hyundai Santa Fe Limited Ultimate"
##  [5] "2017 Hyundai Santa Fe SE"
##  [6] "2017 Hyundai Santa Fe SE"
##  [7] "2017 Hyundai Santa Fe 2.4L"
##  [8] "2017 Hyundai Santa Fe 2.4L"
```

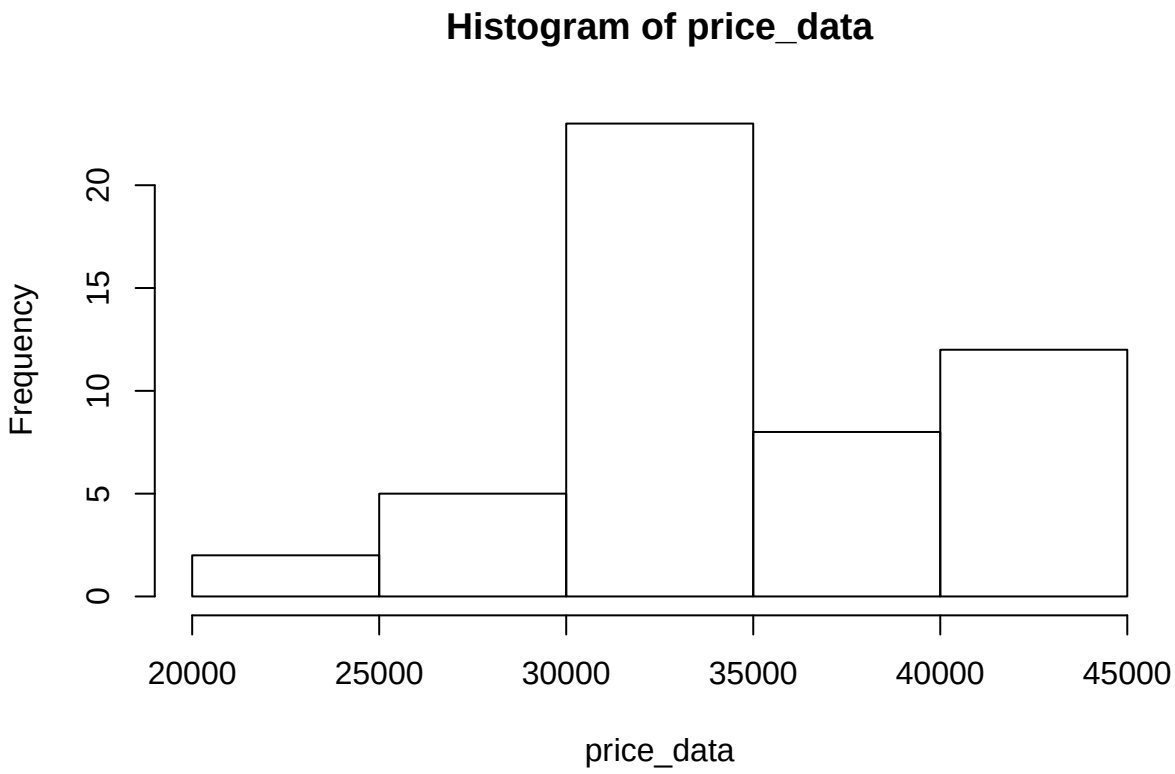
```
##  [9] "2017 Hyundai Santa Fe 2.0L Turbo Ultimate"
## [10] "2017 Hyundai Santa Fe SE Ultimate"
## [11] "2017 Hyundai Santa Fe SE Ultimate"
## [12] "2017 Hyundai Santa Fe SE Ultimate"
## [13] "2017 Hyundai Santa Fe SE Ultimate"
## [14] "2017 Hyundai Santa Fe Limited Ultimate"
## [15] "2017 Hyundai Santa Fe Limited Ultimate"
## [16] "2017 Hyundai Santa Fe SE"
## [17] "2017 Hyundai Santa Fe Limited Ultimate"
## [18] "2017 Hyundai Santa Fe SE"
## [19] "2017 Hyundai Santa Fe SE"
## [20] "2017 Hyundai Santa Fe Limited"
## [21] "2017 Hyundai Santa Fe Limited Ultimate"
## [22] "2017 Hyundai Santa Fe SE Ultimate"
## [23] "2017 Hyundai Santa Fe SE Ultimate"
## [24] "2017 Hyundai Santa Fe SE"
## [25] "2017 Hyundai Santa Fe SE"
## [26] "2017 Hyundai Santa Fe SE Ultimate"
## [27] "2017 Hyundai Santa Fe SE Ultimate"
## [28] "2017 Hyundai Santa Fe Limited Ultimate"
## [29] "2017 Hyundai Santa Fe Limited Ultimate"
## [30] "2017 Hyundai Santa Fe SE"
## [31] "2017 Hyundai Santa Fe SE Ultimate"
## [32] "2017 Hyundai Santa Fe SE"
## [33] "2017 Hyundai Santa Fe SE"
## [34] "2017 Hyundai Santa Fe Limited Ultimate"
## [35] "2017 Hyundai Santa Fe Limited Ultimate"
## [36] "2017 Hyundai Santa Fe Limited Ultimate"
## [37] "2017 Hyundai Santa Fe Limited Ultimate"
## [38] "2017 Hyundai Santa Fe Limited Ultimate"
## [39] "2017 Hyundai Santa Fe Limited Ultimate"
## [40] "2017 Hyundai Santa Fe Limited Ultimate"
## [41] "2017 Hyundai Santa Fe 2.4L"
## [42] "2017 Hyundai Santa Fe SE Ultimate"
## [43] "2017 Hyundai Santa Fe SE Ultimate"
## [44] "2017 Hyundai Santa Fe Limited Ultimate"
## [45] "2017 Hyundai Santa Fe Limited Ultimate"
## [46] "2017 Hyundai Santa Fe Limited Ultimate"
## [47] "2017 Hyundai Santa Fe SE"
## [48] "2017 Hyundai Santa Fe SE"
## [49] "2017 Hyundai Santa Fe SE Ultimate"
## [50] "2017 Hyundai Santa Fe Limited Ultimate"
```

```
#Fix issues with price, text to numeric etc
price_data <- html_text(price_data_html)
price_data <- gsub(",", "", price_data)
price_data <- gsub("\\$", "", price_data)
price_data <- as.numeric(price_data)

data_all <- data.frame(cbind(yr_make_mod_data,price_data))
#data_all$price_data <- as.numeric(data_all$price_data, length=5)

attach(data_all)
```

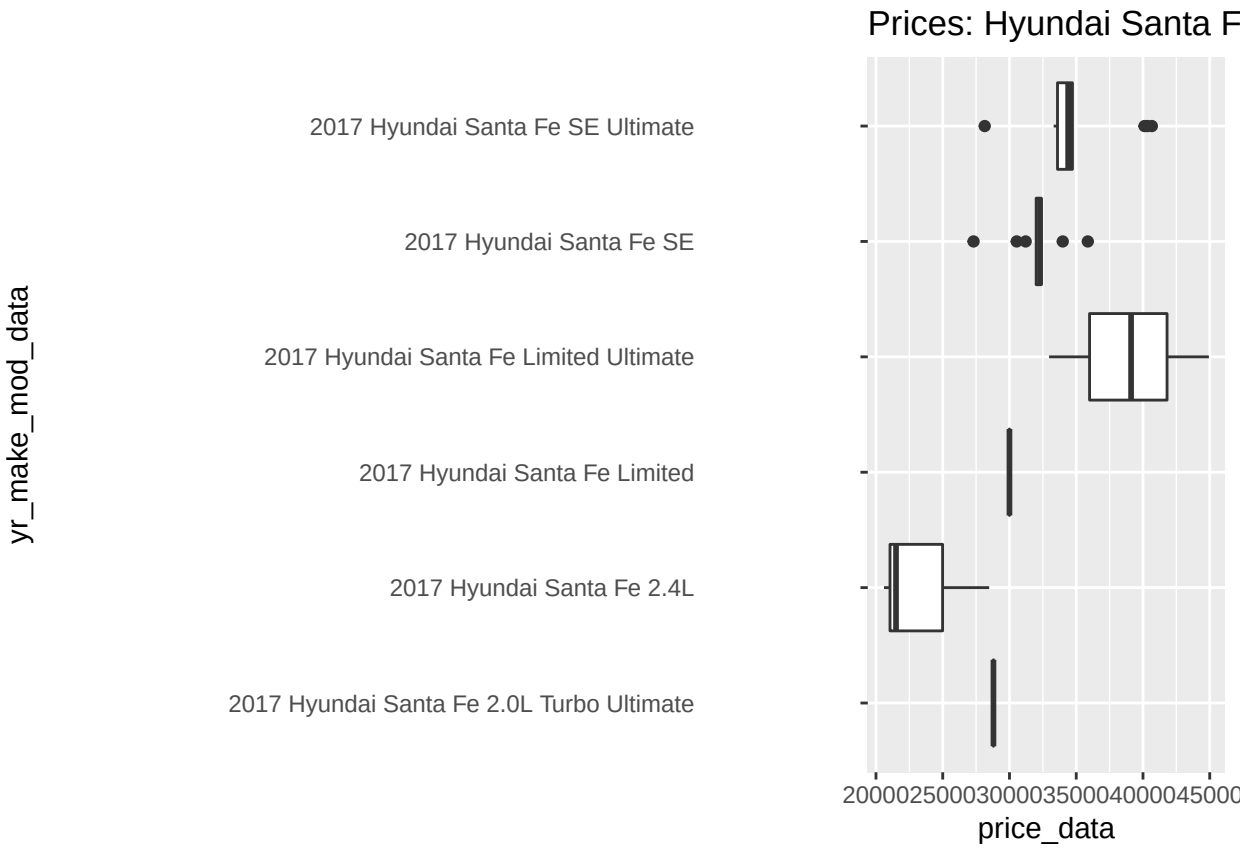
```
## The following objects are masked _by_ .GlobalEnv:
```

```
##
##     price_data, yr_make_mod_data

## The following objects are masked from data_all (pos = 3):
##
##     price_data, yr_make_mod_data

## The following objects are masked from data_all (pos = 4):
##
##     price_data, yr_make_mod_data
```

```
colnames(data_all) <- c("CarInfo","Price")
write.csv(data_all, file = "/Users/Kimberly/Desktop/School Maryville U/hyundai.csv", row.names = FALSE)
hist(price_data)
```




```
ggplot(data_all, aes(x = yr_make_mod_data, y = price_data))+
  geom_boxplot()+
  coord_flip()+
  ggtitle("Prices: Hyundai Santa Fe")
```




```
df3 <- data_all
```

```
# code for Toyota
#This returns the Toyota RAV4

url_Toyota <- "https://www.cars.com/for-sale/searchresults.action/?mdId=21780&mkId=20088&page=1&perPage=

#Reading the HTML code from the website
webpage <- read_html(url_Toyota)

#Using CSS selectors to scrap the data
make_data_html <- html_nodes(webpage,'.listing-row__title')
price_data_html <- html_nodes(webpage,'.listing-row__price')

#Converting the car data to text or numeric (or in PowerBI)
yr_make_mod_data <- html_text(make_data_html)
yr_make_mod_data <- gsub("\n", "", yr_make_mod_data)
trimws(yr_make_mod_data)
```

```
##  [1] "2017 Toyota RAV4 SE"          "2017 Toyota RAV4 XLE"
##  [3] "2017 Toyota RAV4 SE"          "2017 Toyota RAV4 Limited"
##  [5] "2017 Toyota RAV4 Limited"     "2017 Toyota RAV4 XLE"
##  [7] "2017 Toyota RAV4 Platinum"    "2017 Toyota RAV4 Limited"
##  [9] "2017 Toyota RAV4 Limited"     "2017 Toyota RAV4 Platinum"
## [11] "2017 Toyota RAV4"             "2017 Toyota RAV4 XLE"
## [13] "2017 Toyota RAV4 Platinum"    "2017 Toyota RAV4 LE"
## [15] "2017 Toyota RAV4 Platinum"    "2017 Toyota RAV4 XLE"
## [17] "2017 Toyota RAV4 XLE"         "2017 Toyota RAV4 XLE"
```

```
## [19] "2017 Toyota RAV4 LE"         "2017 Toyota RAV4 Platinum"
## [21] "2017 Toyota RAV4 Limited"    "2017 Toyota RAV4 LE"
## [23] "2017 Toyota RAV4 LE"         "2017 Toyota RAV4 Platinum"
## [25] "2017 Toyota RAV4 Platinum"   "2017 Toyota RAV4 XLE"
## [27] "2017 Toyota RAV4 LE"         "2017 Toyota RAV4 XLE"
## [29] "2017 Toyota RAV4 XLE"        "2017 Toyota RAV4 LE"
## [31] "2017 Toyota RAV4 LE"         "2017 Toyota RAV4 LE"
## [33] "2017 Toyota RAV4 Platinum"   "2017 Toyota RAV4 SE"
## [35] "2017 Toyota RAV4 EN FE SC 2" "2017 Toyota RAV4 SE"
```

```
#Fix issues with price, text to numeric etc
price_data <- html_text(price_data_html)
price_data <- gsub(",", "", price_data)
price_data <- gsub("\\$", "", price_data)
price_data <- as.numeric(price_data)
```

```
## Warning: NAs introduced by coercion
```

```
data_all <- data.frame(cbind(yr_make_mod_data,price_data))
#data_all$price_data <- as.numeric(data_all$price_data, length=5)

attach(data_all)
```

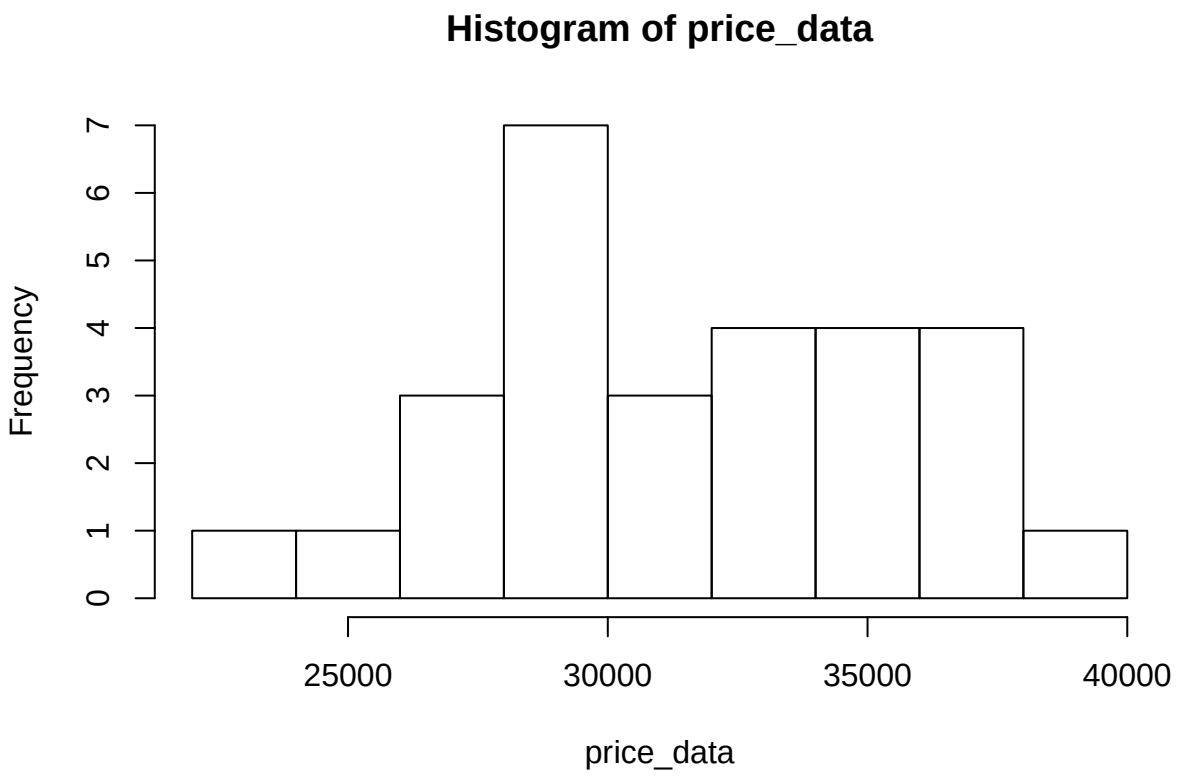
```
## The following objects are masked _by_ .GlobalEnv:
##
##     price_data, yr_make_mod_data
```

```
## The following objects are masked from data_all (pos = 3):
##
##     price_data, yr_make_mod_data
```

```
## The following objects are masked from data_all (pos = 4):
##
##     price_data, yr_make_mod_data
```

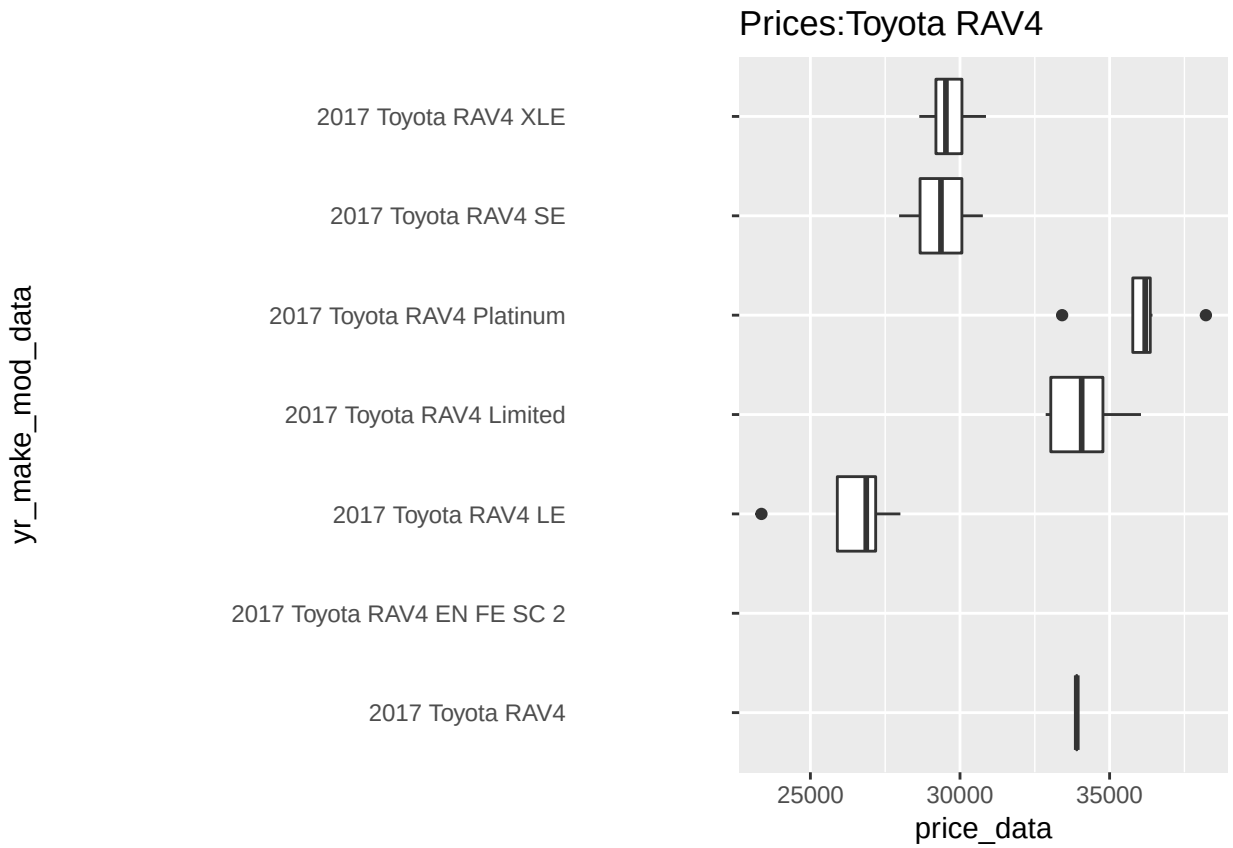
```
## The following objects are masked from data_all (pos = 5):
##
##     price_data, yr_make_mod_data
```

```
colnames(data_all) <- c("CarInfo","Price")
write.csv(data_all, file = "/Users/Kimberly/Desktop/School Maryville U/toyota.csv", row.names = FALSE)

hist(price_data)
```




```
ggplot(data_all, aes(x = yr_make_mod_data, y = price_data))+
  geom_boxplot()+
  coord_flip() +
  ggtitle("Prices:Toyota RAV4")
```

```
## Warning: Removed 8 rows containing non-finite values (stat_boxplot).
```




```
df4 <- data_all
```

```
df_new <- merge(df1,df2, all = TRUE)
df_new <- merge(df_new, df3, all = TRUE)
df_new <- merge(df_new, df4, all = TRUE)
write.csv(df_new, file = "/Users/Kimberly/Desktop/School Maryville U/allcars.csv", row.names = FALSE)
```