```
library(RColorBrewer)
library(ggplot2)
library(patchwork)
```

Car information dataset analysis.

The dataset contains 399 rows of 9 features, which contains some general properties of cars. These 9 features are the following:

1. Name: Unique identifier for each automobile.
2. MPG: Fuel efficiency measured in miles per gallon.
3. Cylinders: Number of cylinders in the engine.
4. Displacement: Engine displacement, indicating its size or capacity.
5. Horsepower: Power output of the engine.
6. Weight: Weight of the automobile.
7. Acceleration: Capability to increase speed, measured in seconds.
8. Model Year: Year of manufacture for the automobile model.
9. Origin: Country or region of origin for each automobile.

The dataset can be found via this *link*

Data exploration

```
setwd("/media/sf_SF/Fedora/R_course/Assignment")
car_data <- read.csv("Automobile.csv")
head(car_data)
```

```
##                       name mpg cylinders displacement horsepower weight
## 1 chevrolet chevelle malibu  18         8          307        130   3504
## 2         buick skylark 320  15         8          350        165   3693
## 3        plymouth satellite  18         8          318        150   3436
## 4             amc rebel sst  16         8          304        150   3433
## 5               ford torino  17         8          302        140   3449
## 6          ford galaxie 500  15         8          429        198   4341
##   acceleration model_year origin
## 1         12.0         70    usa
## 2         11.5         70    usa
## 3         11.0         70    usa
## 4         12.0         70    usa
## 5         10.5         70    usa
## 6         10.0         70    usa
```

```
str(car_data)
```

```
## 'data.frame':    398 obs. of  9 variables:
##  $ name        : chr  "chevrolet chevelle malibu" "buick skylark 320" "plymouth satellite" "amc rebel
##  $ mpg         : num  18 15 18 16 17 15 14 14 14 15 ...
##  $ cylinders   : int  8 8 8 8 8 8 8 8 8 8 ...
```

```
##  $ displacement: num  307 350 318 304 302 429 454 440 455 390 ...
##  $ horsepower  : int  130 165 150 150 140 198 220 215 225 190 ...
##  $ weight      : int  3504 3693 3436 3433 3449 4341 4354 4312 4425 3850 ...
##  $ acceleration: num  12 11.5 11 12 10.5 10 9 8.5 10 8.5 ...
##  $ model_year  : int  70 70 70 70 70 70 70 70 70 70 ...
##  $ origin      : chr  "usa" "usa" "usa" "usa" ...
```

The dataset contains of 3 categorical variables (name, model_year, cylinders and origin) and 6 numerical variables (mpg, cylinders, displacement, horsepower, weight and acceleration). IMPORTANT NOTE: The amount of cylinders also falls under categorical variable

```
car_data$model_year <- as.character(car_data$model_year)
car_data$model_year <- paste0("19",car_data$model_year)
cat_var <- c("name", "model_year", "origin", "cylinders")
num_var <- c("mpg", "displacement", "horsepower", "weight", "acceleration")
```

Let's take a look at the frequencies of each categorical variable. Because of the huge amount of unique car models, no representative barplot can be generated.

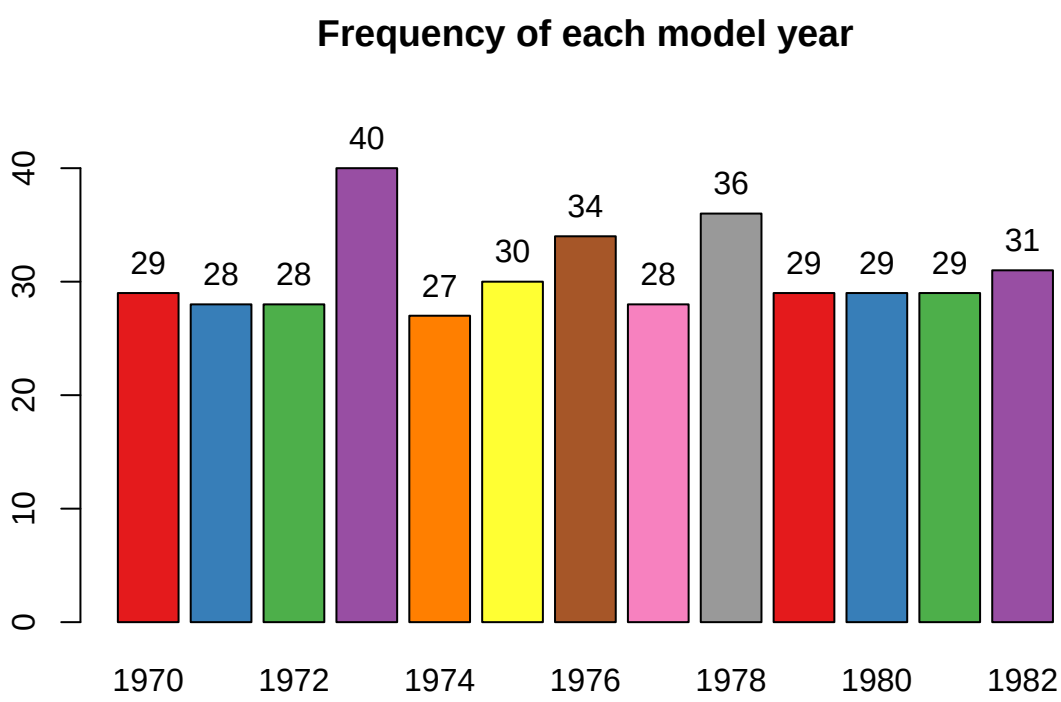
```
paste("Unique car models:", length(unique(car_data$name)))
```

```
## [1] "Unique car models: 305"
```

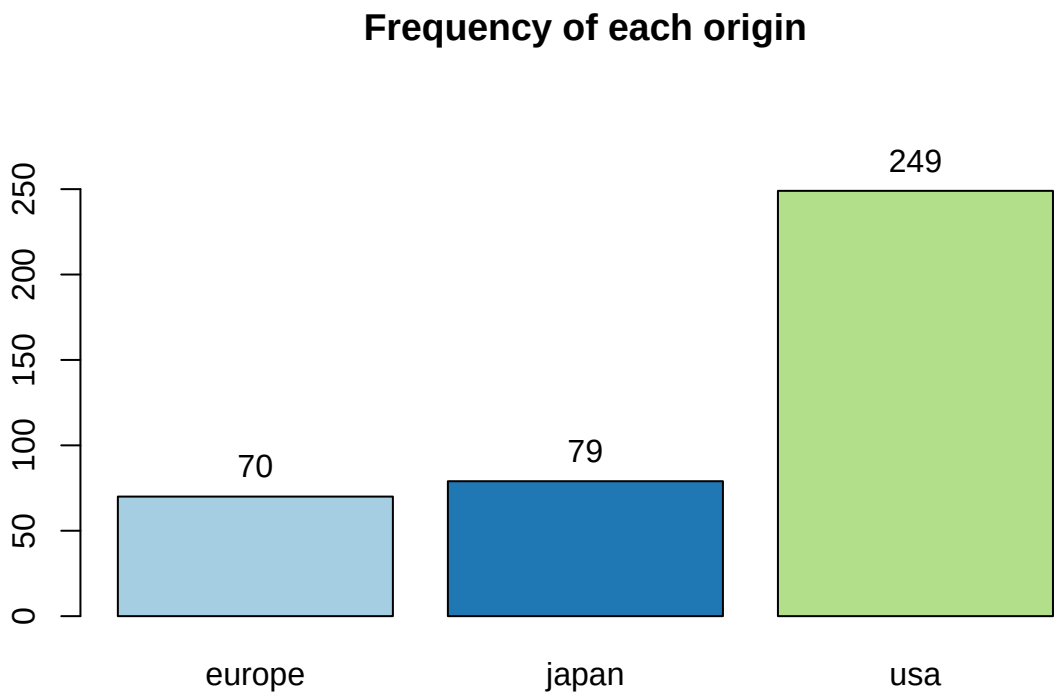
```
bp <- barplot(table(car_data$model_year),
        main = "Frequency of each model year",
        ylim = c(0,45),
        col = brewer.pal(12, "Set1"))
```

```
## Warning in brewer.pal(12, "Set1"): n too large, allowed maximum for palette Set1 is 9
## Returning the palette you asked for with that many colors
```

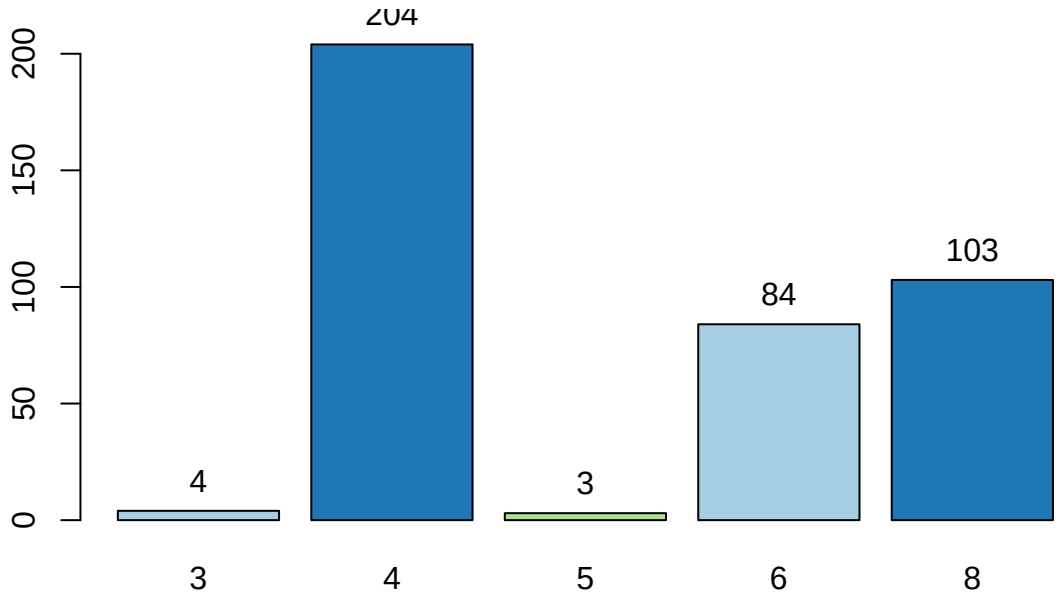
```
text(x=bp, y=table(car_data$model_year),label=table(car_data$model_year),pos=3)
```




```
bp <- barplot(table(car_data$origin),
        main = "Frequency of each origin",
        ylim = c(0,max(table(car_data$origin))+50),
        col = brewer.pal(3, "Paired"))
 text(x=bp, y=table(car_data$origin),label=table(car_data$origin),pos=3)
```



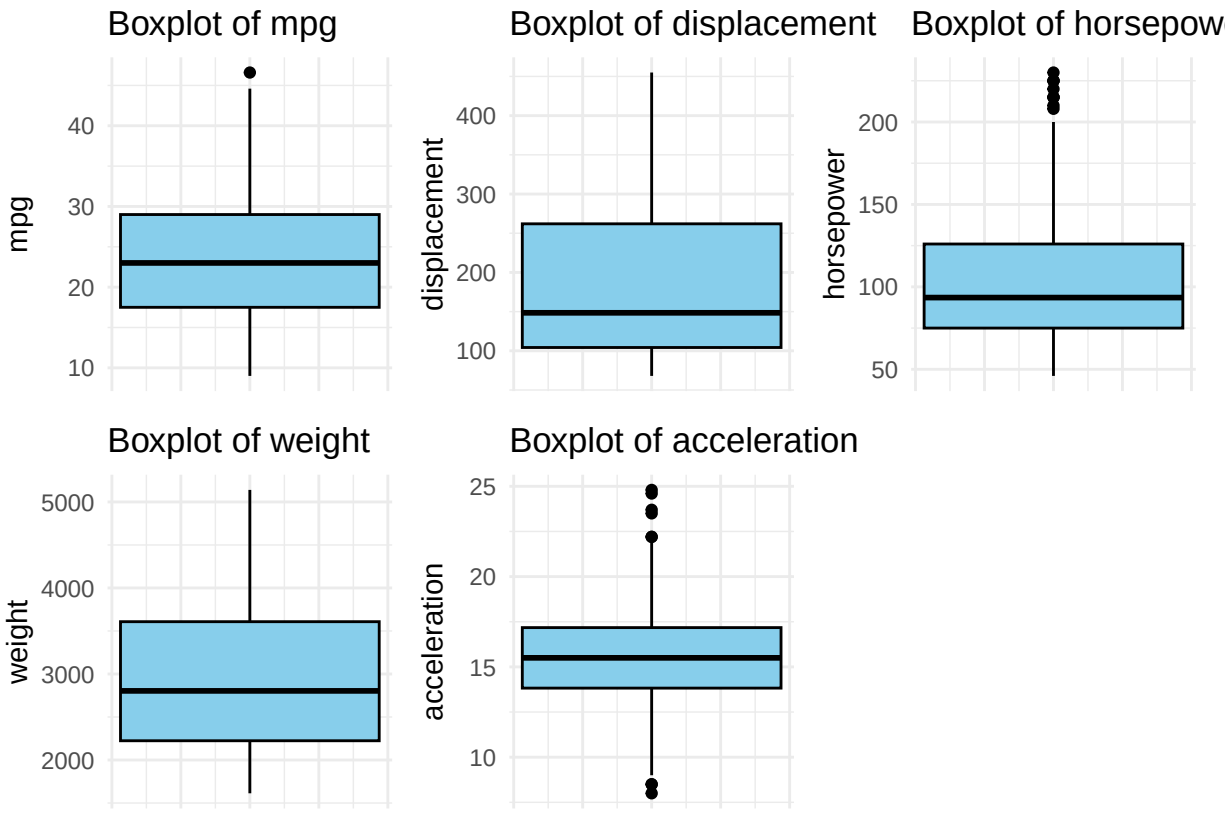

```
bp <- barplot(table(car_data$cylinders),
        main = "Frequency of amount of cylinders",
        ylim = c(0,max(table(car_data$cylinders)+15)),
        col = brewer.pal(3, "Paired"))
 text(x=bp, y=table(car_data$cylinders),label=table(car_data$cylinders),pos=3)
```



Let's take a look at the numerical variables now.

```
plot_list <- list()
for (var in num_var) {
  # Create boxplot
  p <- ggplot(car_data, aes(y = !!sym(var))) +
    geom_boxplot(fill = "skyblue", color = "black") +
    labs(title = paste("Boxplot of", var), y = var) +
    theme_minimal() +
    theme(axis.title.x = element_blank(),
          axis.text.x = element_blank())
  plot_list[[var]] <- p
}
combined_plots <- wrap_plots(plotlist = plot_list, ncol = 3)
print(combined_plots)
```

```
## Warning: Removed 6 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```


Boxplot of mpg
mpg
40
30
20
10
Boxplot of displacement
displacement
400
300
200
100
Boxplot of horsepow
horsepower
200
150
100
50
Boxplot of weight
weight
5000
4000
3000
2000
Boxplot of acceleration
acceleration
25
20
15
10