Lucille Barnett

DS-160-01

Individual Project 3

11/7/2024

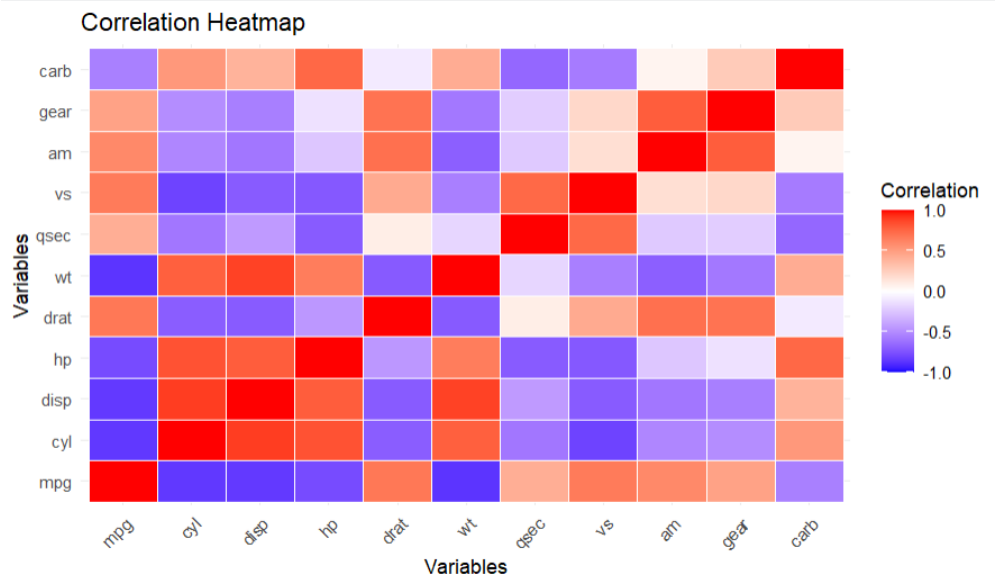
The mtcars dataset originates from the 1974 Motor Trend US magazine and contains fuel consumption and performance data for 32 different cars. This dataset includes 11 variables, each representing different characteristics of the cars: mpg (miles per gallon), cyl (number of cylinders), disp (displacement in cubic inches), hp (gross horsepower), drat (rear axle ratio), wt (weight in 1000 lbs), qsec (1/4 mile time), vs (engine type: V-shaped or straight), am (transmission: automatic or manual), gear (number of forward gears), and carb (number of carburetors). These variables offer a comprehensive view of various attributes that affect car performance and fuel efficiency.

The exploratory data analysis (EDA) revealed several interesting relationships among the variables. The histogram of mpg showed that most cars in the dataset have moderate fuel efficiency, with fewer cars on the extreme ends of low and high miles per gallon. A boxplot of hp grouped by number of cylinders indicated that cars with more cylinders tend to have higher horsepower, suggesting a positive relationship between cylinder count and engine power. Furthermore, a scatter plot of weight versus mpg showed a negative correlation, indicating that heavier cars generally have lower fuel efficiency. Overall, these visualizations and summary statistics highlight the trade-offs between car weight, horsepower, and fuel efficiency.

Based on the EDA, we can conclude that there are strong associations among several variables in the mtcars dataset. Heavier cars tend to be less fuel-efficient, likely due to the greater energy required to move larger masses. Similarly, cars with more cylinders and higher horsepower tend to consume more fuel, as they are typically designed for performance rather than efficiency. This dataset underscores the performance-versus-efficiency trade-off in car design, where increased power and weight tend to reduce fuel efficiency.

To deepen this analysis, we recommend exploring predictive models to quantify the impact of variables like weight, horsepower, and cylinder count on fuel efficiency (mpg). A multiple regression analysis could help estimate the contribution of each factor and provide a formula to predict mpg based on car specifications. Additionally, segmenting the data by transmission type (automatic vs. manual) could reveal whether transmission has a significant effect on fuel efficiency and performance characteristics.

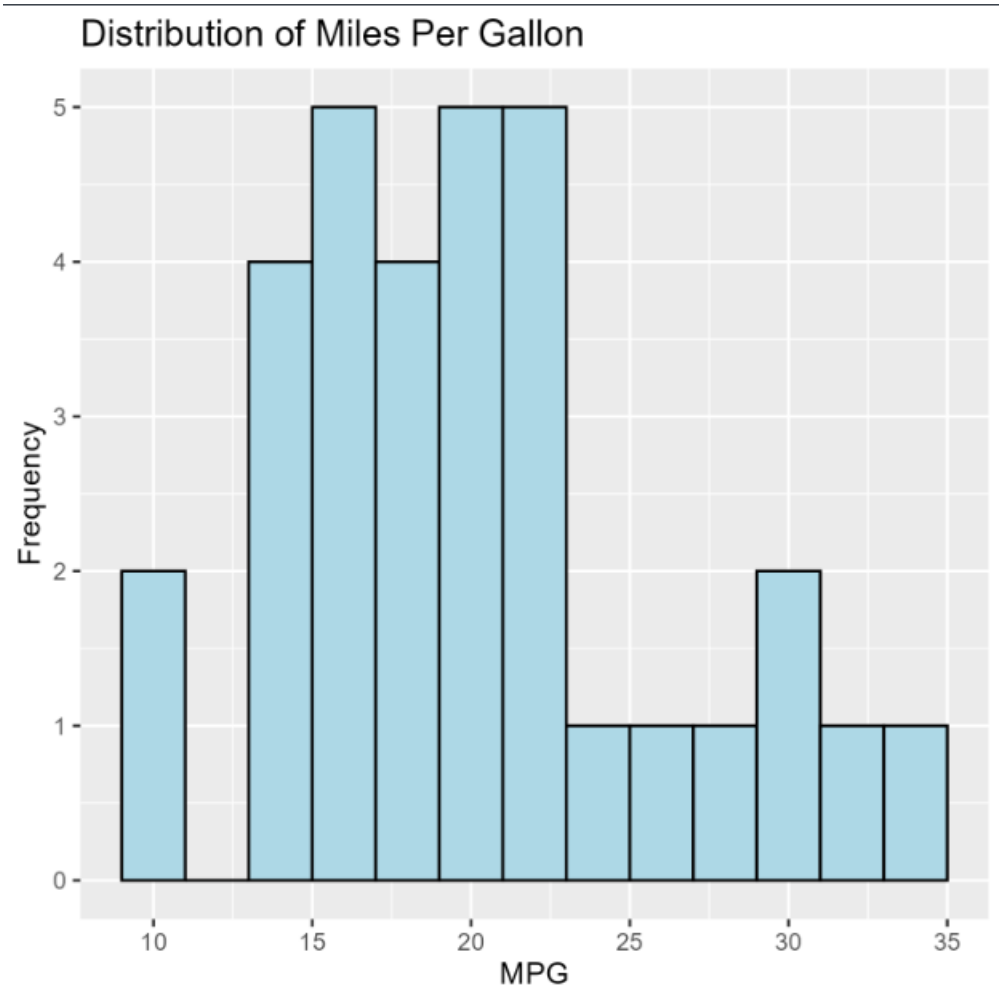
```
data("mtcars")

str(mtcars)

summary(mtcars)


library(ggplot2)


#Correlation Heatmap

cor_matrix <- cor(mtcars)

library(reshape2)

cor_data <- melt(cor_matrix)

heatmap<-ggplot(cor_data, aes(x = Var1, y = Var2, fill = value)) +

 geom_tile(color = "white") +

 scale_fill_gradient2(low = "blue", high = "red", mid = "white", midpoint = 0, limit = c(-1, 1), name = "Correlation") +

 theme_minimal() +

 theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +

 labs(title = "Correlation Heatmap", x = "Variables", y = "Variables")
```



#Shows a strong positive correlation between cly and hp variables, as well as gear and am variables. More cylinders = more horsepower.

#Strong negative correlation between wt and mpg. Sp... heavier cars generally have lower fuel efficiency.

```
#Histogram of mpg
histogram<-ggplot(mtcars, aes(x = mpg)) +
 geom_histogram(binwidth = 2, fill = "lightblue", color = "black") +
 labs(title = "Distribution of Miles Per Gallon", x = "MPG", y = "Frequency")
```



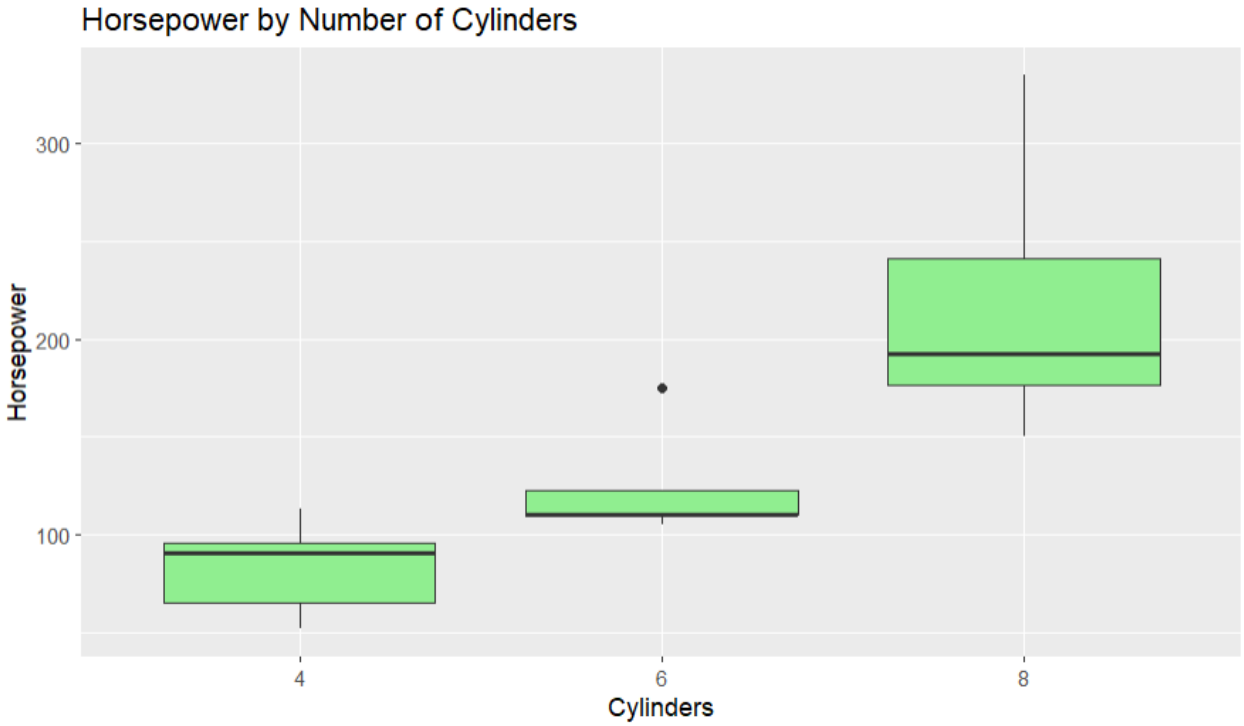
#Slightly skewed fuel efficiency in the cars in this dataset, but most have a “good” fuel efficiency

Boxplot of horsepower by cylinder count

boxplot<-ggplot(mtcars, aes(x = factor(cyl), y = hp)) +

geom_boxplot(fill = "lightgreen") +

labs(title = "Horsepower by Number of Cylinders", x = "Cylinders", y = "Horsepower")



#Closer look into the horsepower vs # of cylinders. Though there is one outlier in the '6 cylinders' column, this graph is consistent with the hypothesis that more cylinders = more horsepower.

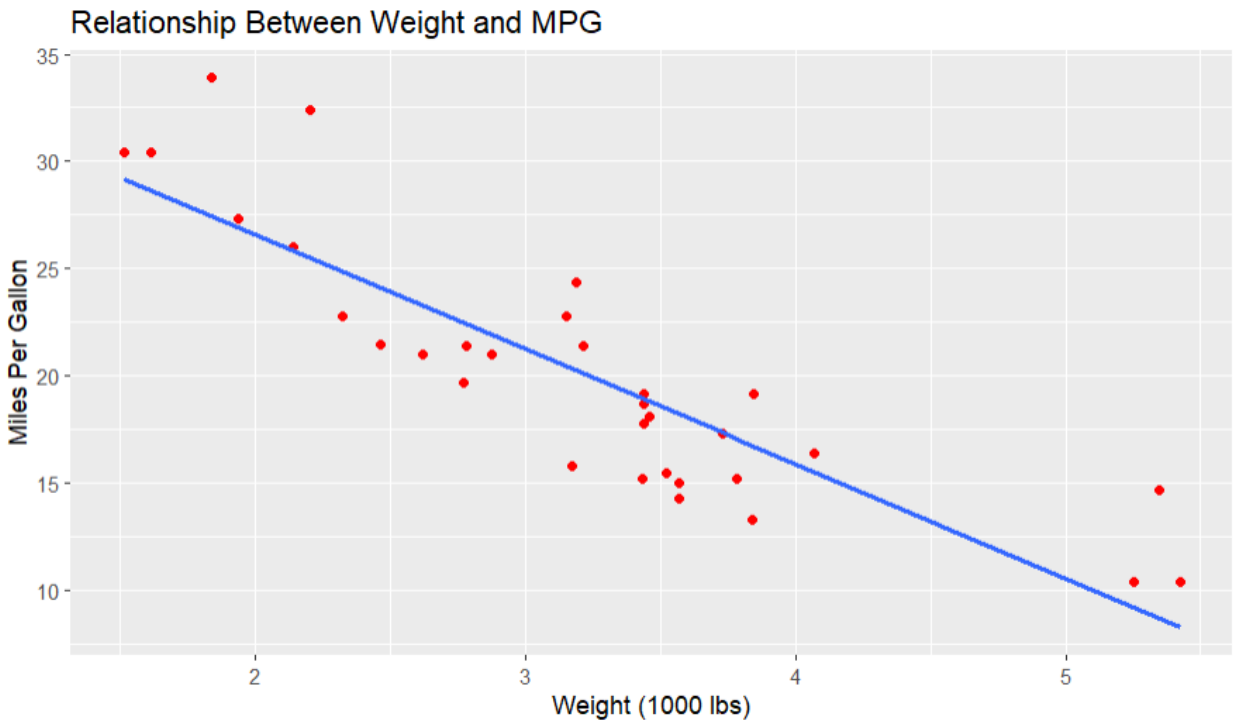
Scatter plot of weight vs MPG

scatter<-ggplot(mtcars, aes(x = wt, y = mpg)) +

geom_point(color = "red") +

geom_smooth(method = "lm", se = FALSE) +

labs(title = "Relationship Between Weight and MPG", x = "Weight (1000 lbs)", y = "Miles Per Gallon")



#Closer look into the relationship between weight and miles per gallon. The lighter a car is, the more miles per gallon they are likely to get. This is a relatively linear relationship, though some cars in the dataset were able to get more gallons per mile, despite their weight.

#Saving the plots

ggsave("plot_mpg_heatmap.png", plot = heatmap, width = 5, height = 5)

ggsave("plot_mpg_histogram.png", plot = histogram, width = 5, height = 5)

ggsave("plot_mpg_boxplot.png", plot = boxplot, width = 5, height = 5)

ggsave("plot_mpg_scatter.png", plot = scatter, width = 5, height = 5)