Math 189: Homework 2

Alan Lui, Derek So, Xiangyu Wei

2021-01-17

Introduction

The Motor Trend Car Road Tests dataset contains information from the 1974 Motor Trend US magazine, of 32 observations on 11 variables. Our study explores the relationships between weight, miles per gallon, and number of cylinders of a car. Using variations of scatter plots, we explore the relationships of all the variables to each other and the implications of the number of cylinders to the relationship of weight and miles per gallon.

Tasks & Analysis

Dataset

The Motor Trend Car Road Tests dataset was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models). It was also used in Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391–411. We extracted the dataset from https://github.com/tuckermcelroy/ma189/blob/main/Data/mtcars.csv at 2021-01-14 20:18:53 PST.

```
mtcars <- read.csv('Data/mtcars.csv')
head(mtcars)
```

```
##               model  mpg cyl disp  hp drat    wt  qsec vs am gear carb
## 1         Mazda RX4 21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## 2     Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## 3        Datsun 710 22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## 4    Hornet 4 Drive 21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## 5 Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## 6           Valiant 18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

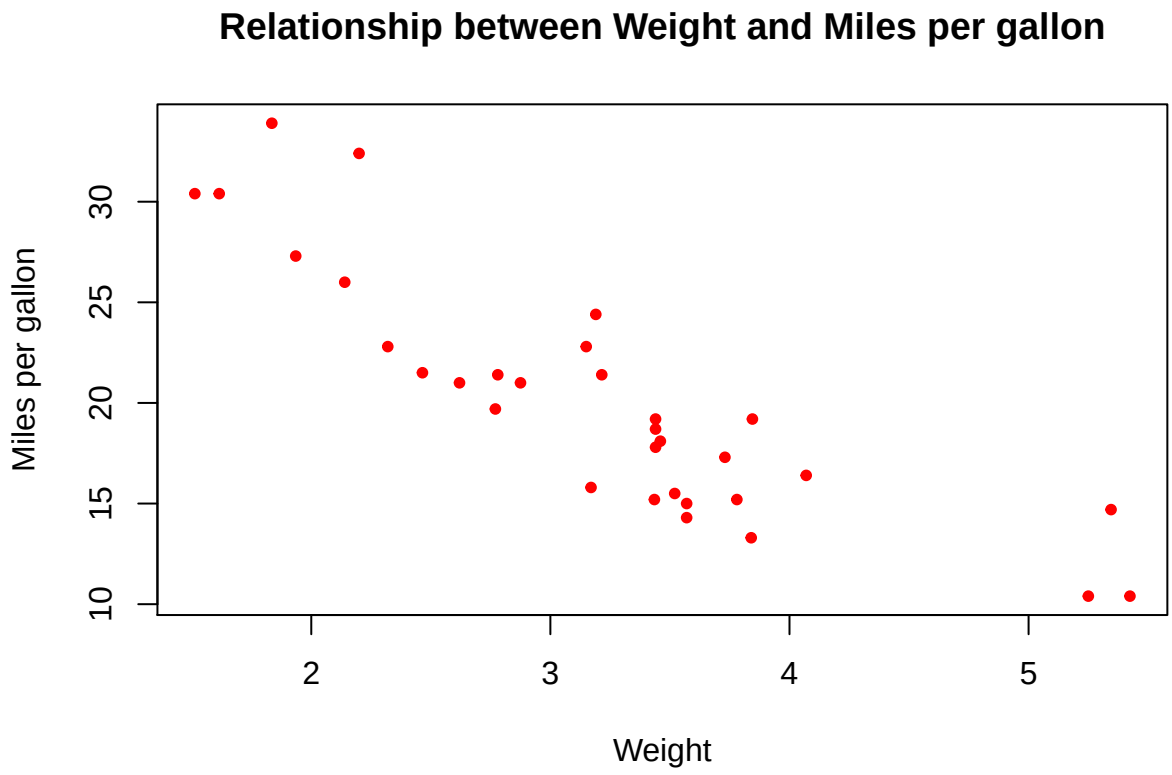
Method

Our methods were primarily data representations of relevant variables using scatter plots. Our primary method was 2d, in initial exploration and in pairwise, with one 3d plot to see the interactions of all variables. This allows us to see the relationships (or lack thereof) between the relevant variables.

Analysis

1. Draw a scatter plot between wt (Weight) and mpg (Miles per gallon).

```
plot(x = mtcars$wt, y = mtcars$mpg,
     xlab = "Weight", ylab = "Miles per gallon",
     main = "Relationship between Weight and Miles per gallon",
     col = "red", pch = 20)
```




```
cor(mtcars$wt, mtcars$mpg)
```

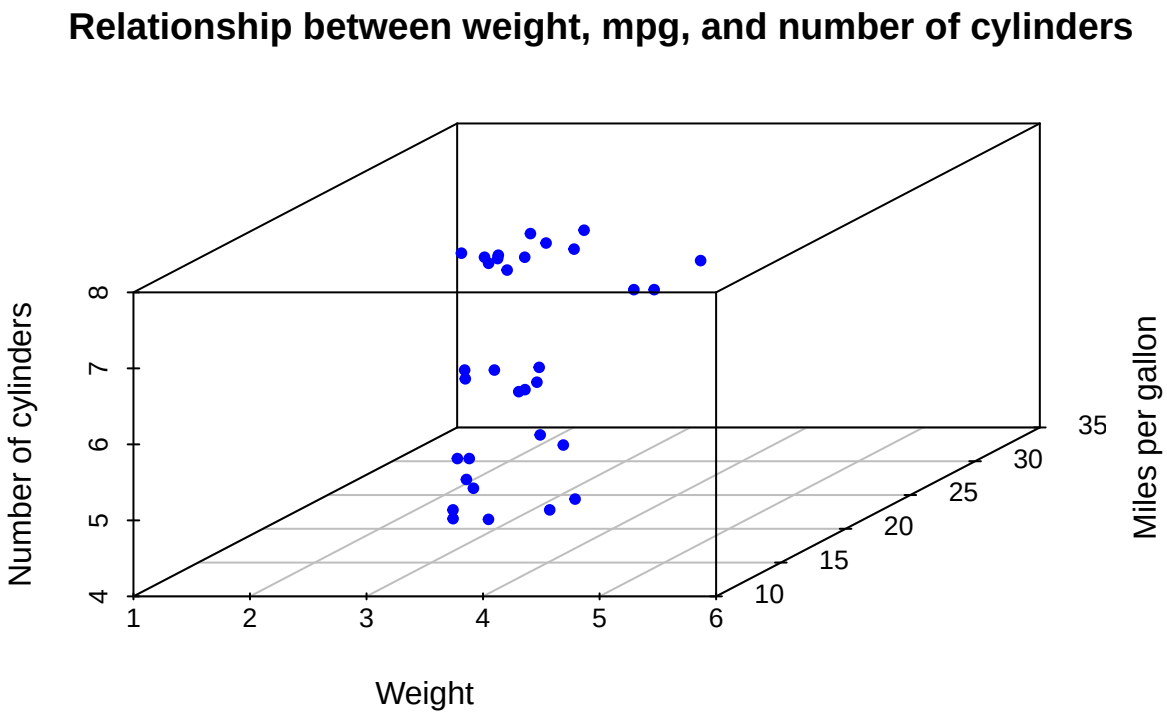
```
## [1] -0.8676594
```

We used simple *plot()* function to plot the scattered relationship between weight and miles per gallon. We then used the *cor()* function to calculate the correlation coefficient between those two variables. Looking at the scatter plot, we see that there is a moderately strong, negative relationship between weight and mpg with a correlation of r = -.86. There are no apparent outliers and the scatter plot seems to be homoscedastic.

2. Draw a scatter plot to show the relationship between wt (Weight), mpg (Miles per gallon) and cyl (Number of cylinders). Use a 3D scatter plot.

```
library("scatterplot3d")
scatterplot3d(x = mtcars$wt, y = mtcars$mpg, z = mtcars$cyl,
              xlab = "Weight", ylab = "Miles per gallon", zlab = "Number of cylinders",
```

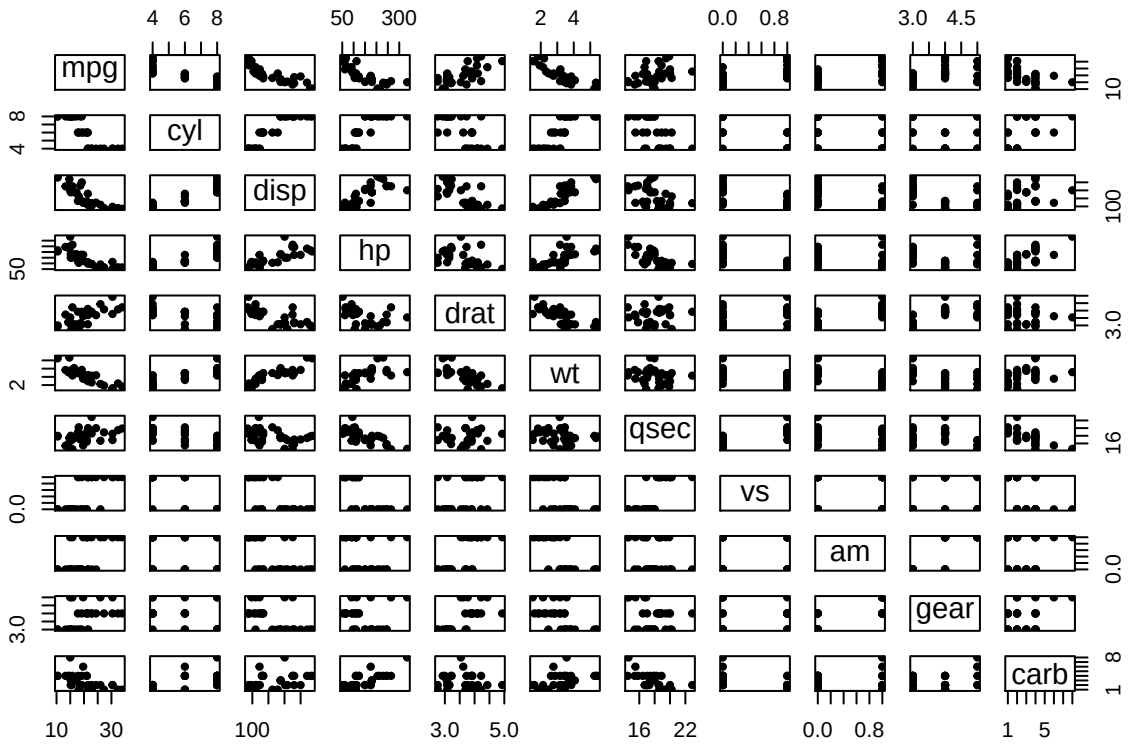
```
        main = "Relationship between weight, mpg, and number of cylinders",
        color = "blue", pch = 20)
```



We used *scatterplot3d()* to plot a 3D scattered relationship between weight, miles per gallon, and number of cylinders. Looking at the 3d plot of weight, miles per gallon, and number of cylinders, it is difficult to determine if there are any underlying patterns among these three variables. We will explore these three variables in deeper context through pairwise scatter plots.

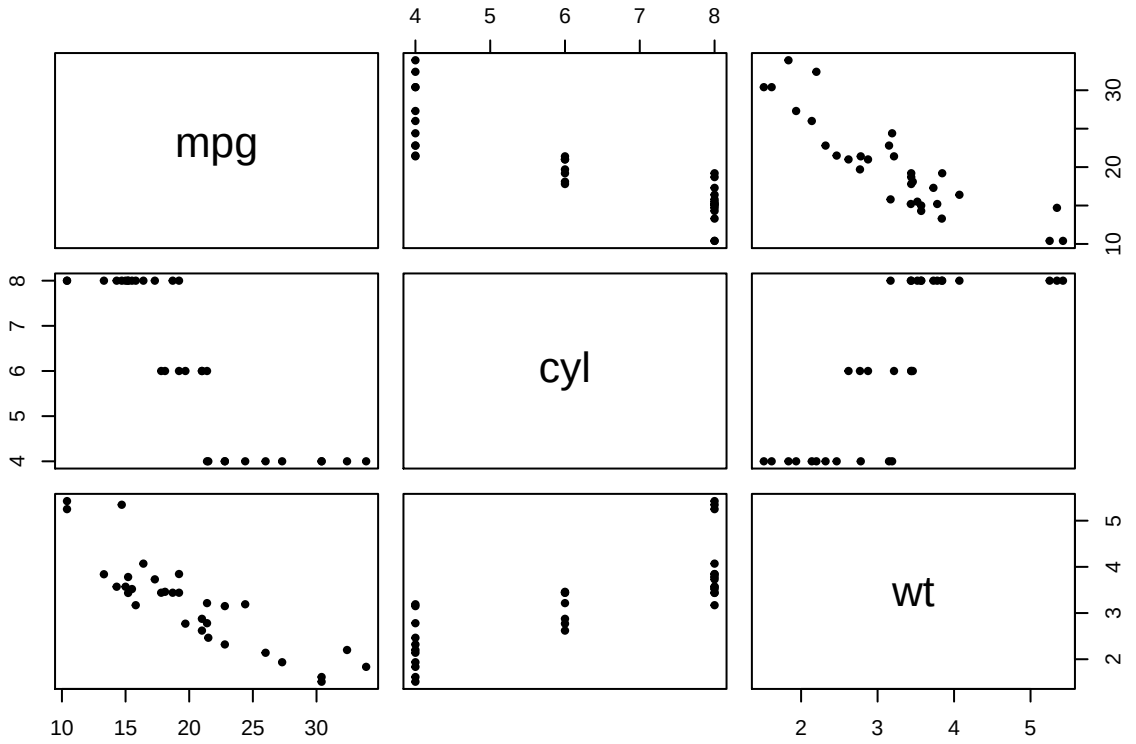
3. Draw pairwise scatter plot for all variables.

```
mtcars.3a <- mtcars[,-c(1)]
pairs(mtcars.3a, pch = 20)
```



We used *pairs()* to plot the pairwise scatter plot between all variables. Above is the pairwise scatterplot for all variables. It seems like it's really hard to see the relationship between weight, miles per gallon, and number of cylinders, so we decided to have a pairwise scatter plot for only those three variables.

```
mtcars.3b <- mtcars[,c(2,3,7)]
pairs(mtcars.3b, pch = 20)
```



We then used *pairs()* to plot only the weight, miles per gallon, and number of cylinders. Looking at the pairwise plots, we have a better understanding of the underlying patterns. The more specific relationships between those three variables are described below.

4. One engineer suggests that the relationship between wt and mpg is subject to the number of cylinders. According to the plot you draw in step 2, what is your opinion towards this suggestion?

From the pariwise plots in Q3, we noticed that as the number of cylinders increases, the miles per gallon decreases. On the contrary, as the number of cylinders increases, the weight increases. From these two points, we see that the negative correlation between mpg and weight as seen in plot may be explained by the change in number of cylinders. So we would agree that the relationship between weight and miles per gallon might be subject to the number of cylinders.

Conclusion

In summary, we can see a clear relationship between the number of cylinders to weight and miles per gallon individually. From this, we can also conclude that the number of cylinders affects the relationship between weight and miles per gallon as well, or that the number of cylinders is the crucial factor in the relationship between weight and miles per gallon.

Discussion

Logically, it follows that a heavier car needs more cylinders in the engine to power, and the increased cylinders lead to higher fuel consumption and thus lower miles per gallon. Further exploration of the other variables

in the original dataset could uncover more potential factors. Moreover, as this data is several decades old, it would be worthwhile to study a similar dataset over time, perhaps every decade, to see whether this relationship holds true and if correlations become stronger or weaker.

Though 3d scatterplot is a robust tool, it may not always be the best way to analyze the data since it is hard to initially recognize underlying patterns of the data. This is mostly exacerbated by the simplicity of R's representation, and an interactive 3d scatterplot would be more effective in both visualizing and understanding trends.