Correlation and Linear Regression

Thileepan Paulraj

21 December 2018

Correlation

Two variables are said to be correlated when they vary together. An unit increase in one variable should result in an unit increase or decrease in another variable for the variables to be correlated.

Let's take the cars data set from R and see how it's mileage and weight variables are correlated. Let's use pearson's correlation for that.

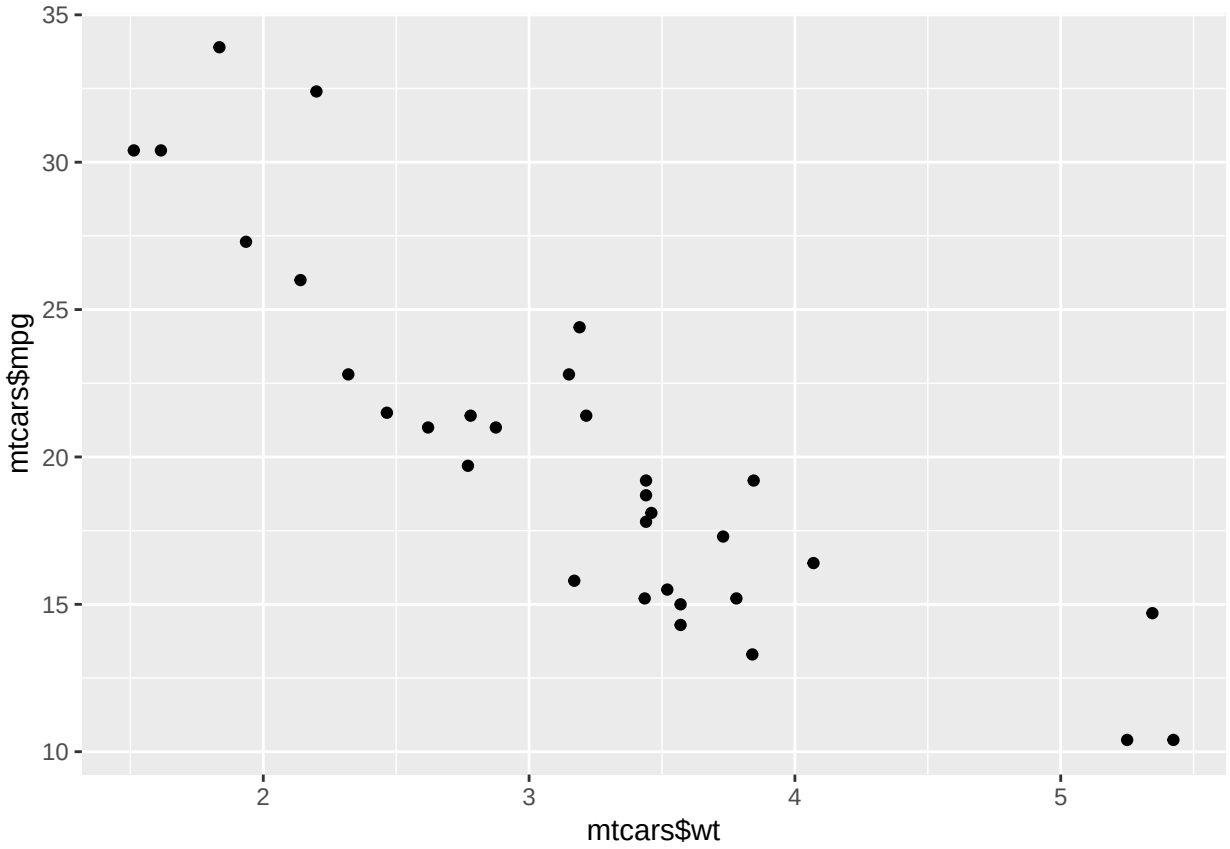
```
cor.test(mtcars$mpg, mtcars$wt)
```

```
##
##  Pearson's product-moment correlation
##
## data:  mtcars$mpg and mtcars$wt
## t = -9.559, df = 30, p-value = 1.294e-10
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.9338264 -0.7440872
## sample estimates:
##        cor
## -0.8676594
```

Correlation coefficients are usually between -1 and +1. Here our correlation coefficient is high, which indicates high correlation between the variables. the negative sign indicates a negative correlation which means as the weighht of the car increases it's mileage decreases.

Let's visualize the correlation also.

```
ggplot(data =mtcars, aes(x = mtcars$wt, y = mtcars$mpg)) + geom_point()
```



From the plot we can clearly see that as the weight of the car increases it's mileage decreases.