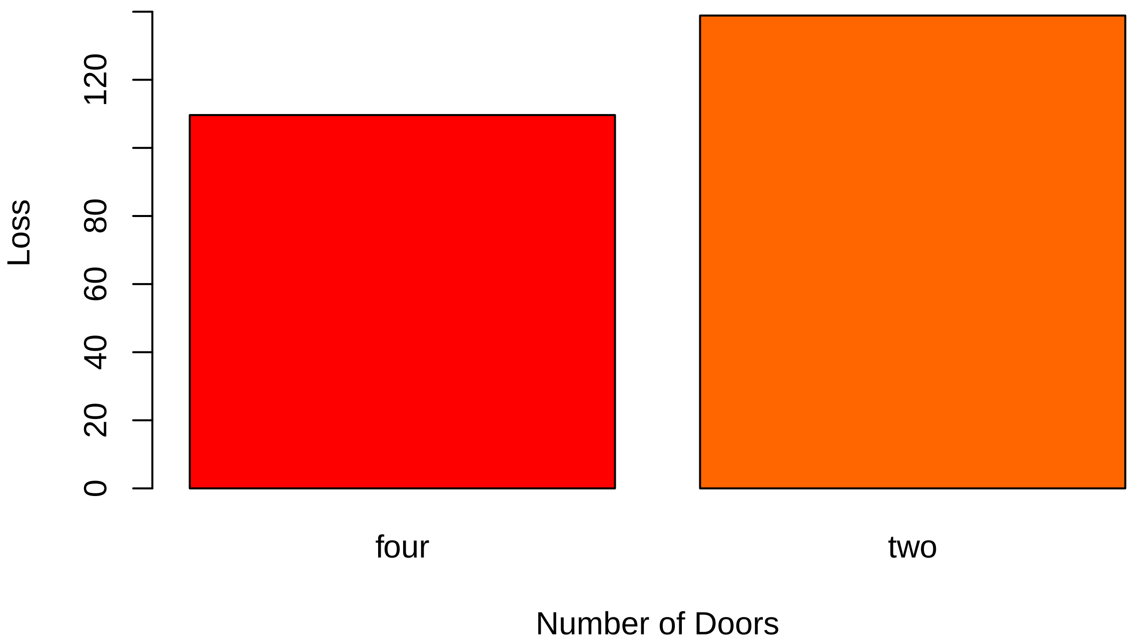
# Auto Losses and Bikeshare Analysis - Shahin Shakeri

I use R notebooks in RStudio cloud and the data is stored in the clound on my google drive. I submitted this notebook in original format which can be simply be opened in R desktop and run.

## 1- Auto Losses

### 1.a) Two door cars are correlated with more losses

```
library('lattice')
Autoloss <- read.csv(url("https://drive.google.com/uc?export=download&id=1cbk_KGdo5sfY0c4_ALB5ULTeYQ87u
Autoloss <-na.omit (Autoloss)
ByDoors=tapply(Autoloss$Losses, Autoloss$NumDoors, mean)
barplot(ByDoors, col = rainbow(15),xlab="Number of Doors",ylab = "Loss",ylim=c(0,150))
```




### 1.b) Sedans have the lowest losses among the least costly losses. This doesn't mean the entire dataset follows this pattern.

```
Lowest<- Autoloss[order(Autoloss$Losses,decreasing=FALSE),][1:10,]
Lowest<-tapply(Lowest$Losses,Lowest$BodyStyle , mean)
Lowest
```

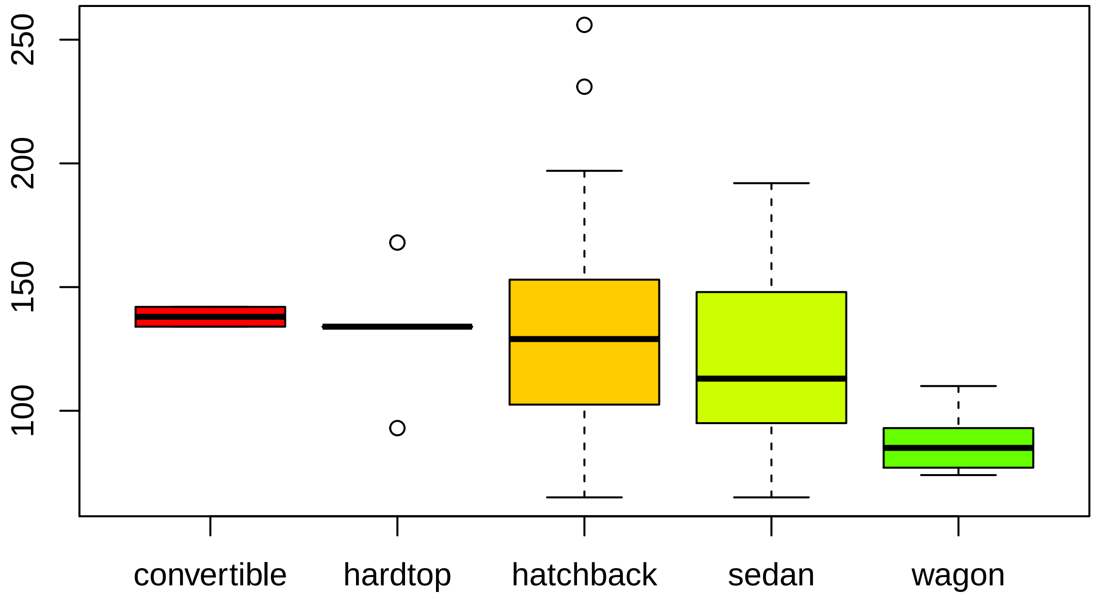
```
## convertible     hardtop   hatchback       sedan       wagon
##          NA          NA          68          65          74
```

**1.c) Sedans and Hatchbacks have the widest spread of losses. The losses in sedan are skewed towards the lower end where in the other styles are more or less evenly distributed.**

```
ByWheels<-tapply(Autoloss$Losses,Autoloss$BodyStyle,mean)
ByWheels
```

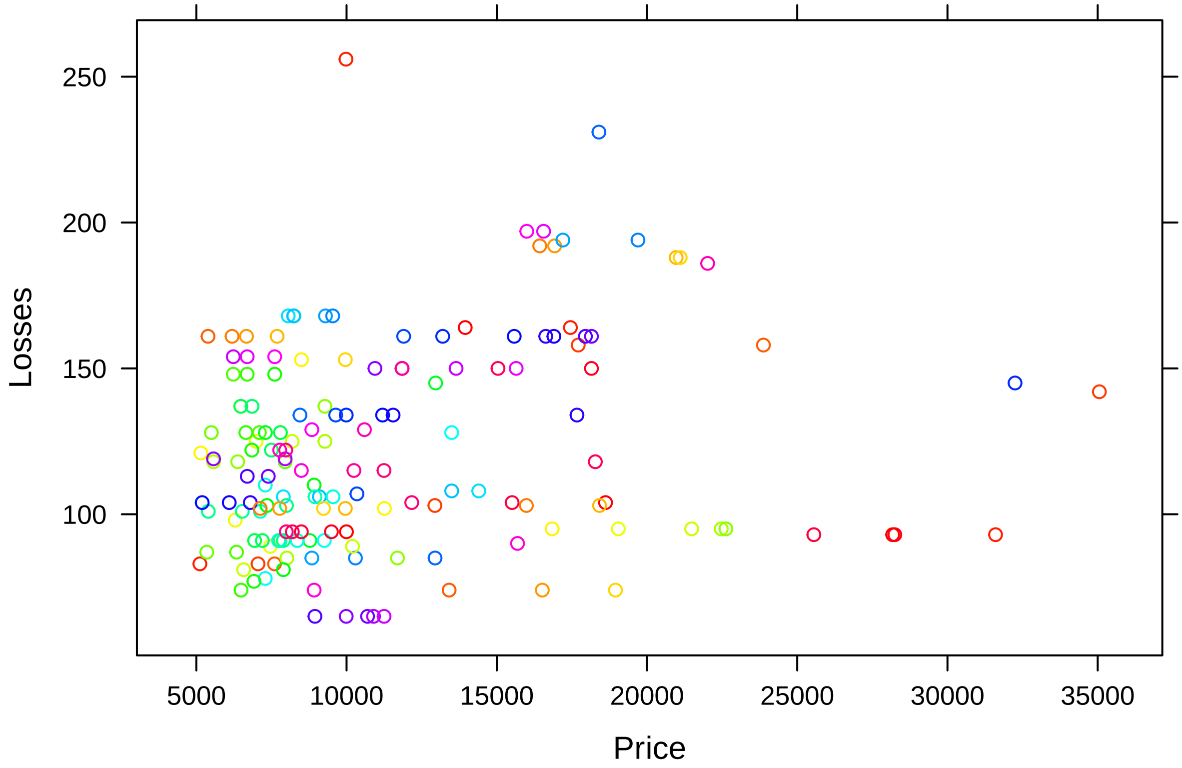
```
## convertible     hardtop   hatchback       sedan       wagon
##   138.00000   132.60000   132.08333   120.35443    87.52941
```

```
boxplot(Losses~BodyStyle,data=Autoloss ,col = rainbow(15))
```




**1.d) Lower priced car have costlier loses than higher priced cars on average**

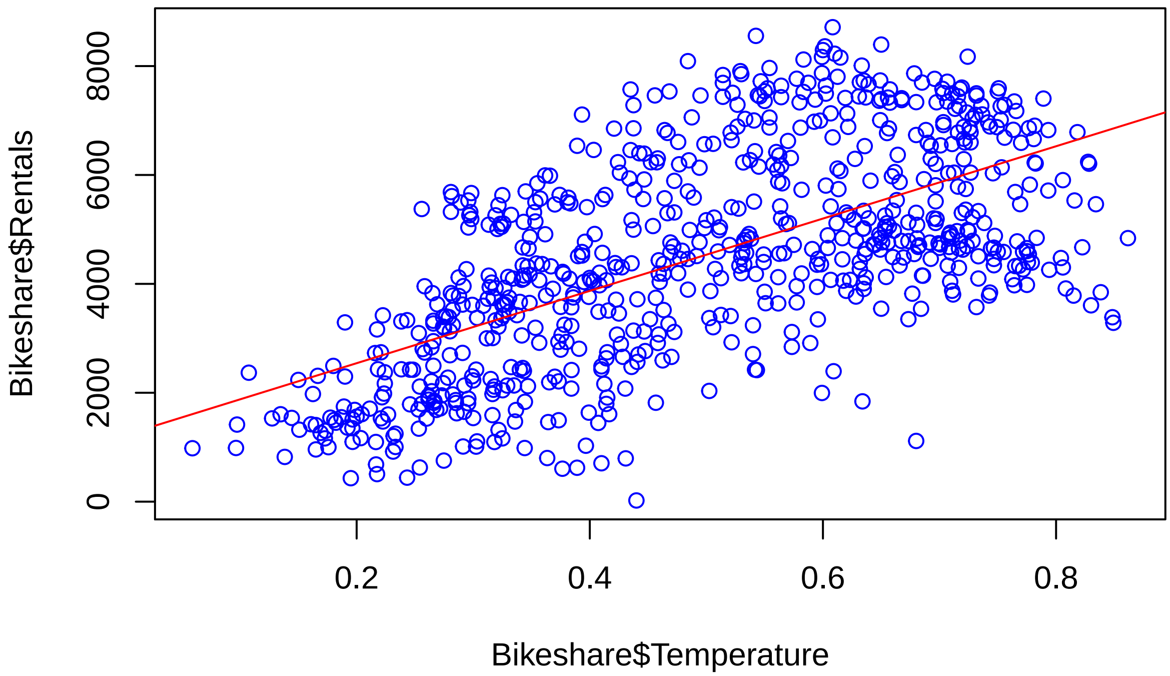
```
xyplot(Losses~Price,data=Autoloss, col = rainbow(50))
```




## 2 -Bikeshare

### 2.a) We run a simple regression for Rentals~Temperature

```
Bikeshare <- read.csv(url("https://drive.google.com/uc?export=download&id=1jtfh-qmyDM3l_nU3AkPP_ZsepL-j
LModel=lm(Rentals~Temperature,data=Bikeshare)
{plot(Bikeshare$Temperature,Bikeshare$Rentals,col='blue')
abline(LModel,col="red" )}
```



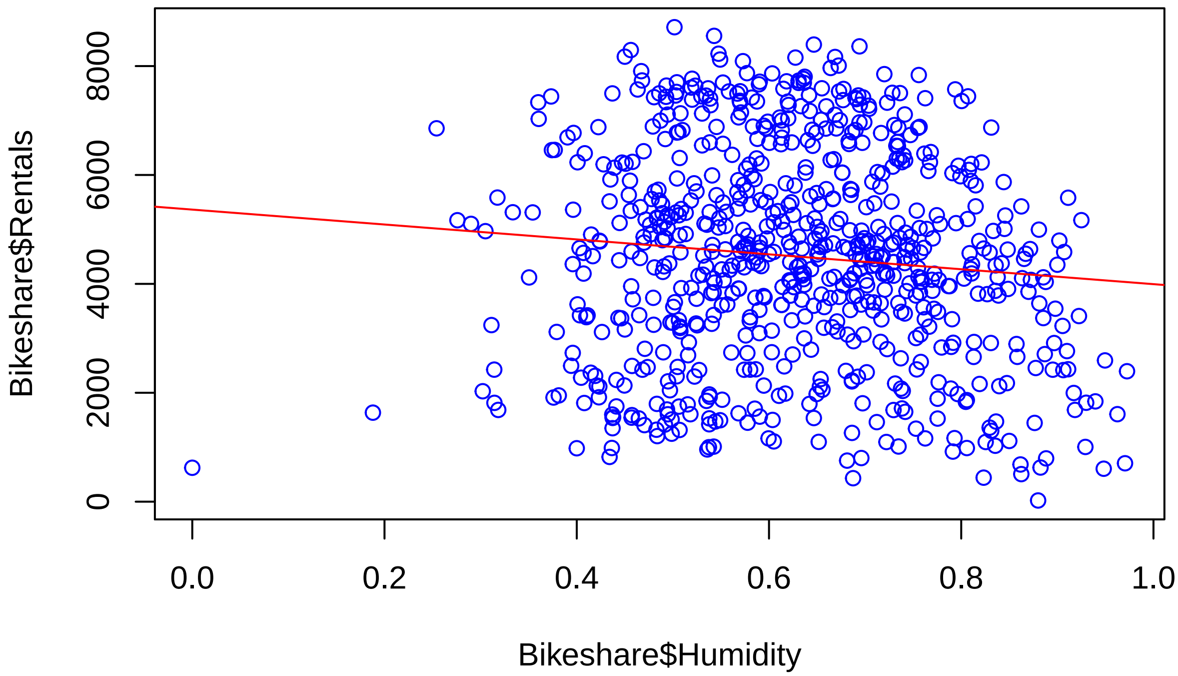

```r
summary(LModel)
```

```
##
## Call:
## lm(formula = Rentals ~ Temperature, data = Bikeshare)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4615.3 -1134.9  -104.4  1044.3  3737.8
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1214.6      161.2   7.537 1.43e-13 ***
## Temperature   6640.7      305.2  21.759  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1509 on 729 degrees of freedom
## Multiple R-squared:  0.3937, Adjusted R-squared:  0.3929
## F-statistic: 473.5 on 1 and 729 DF,  p-value: < 2.2e-16
```

**For every 1 change to normalized Temperature, Rentals change by 6640.7. Accoring to small p $<.05$ this is significant**

### 2.b) We run a simple regression for Rentals~Humidity

```r
LModel=lm(Rentals~Humidity,data=Bikeshare)
{plot(Bikeshare$Humidity,Bikeshare$Rentals,col='blue')
abline(LModel,col="red" )}
```



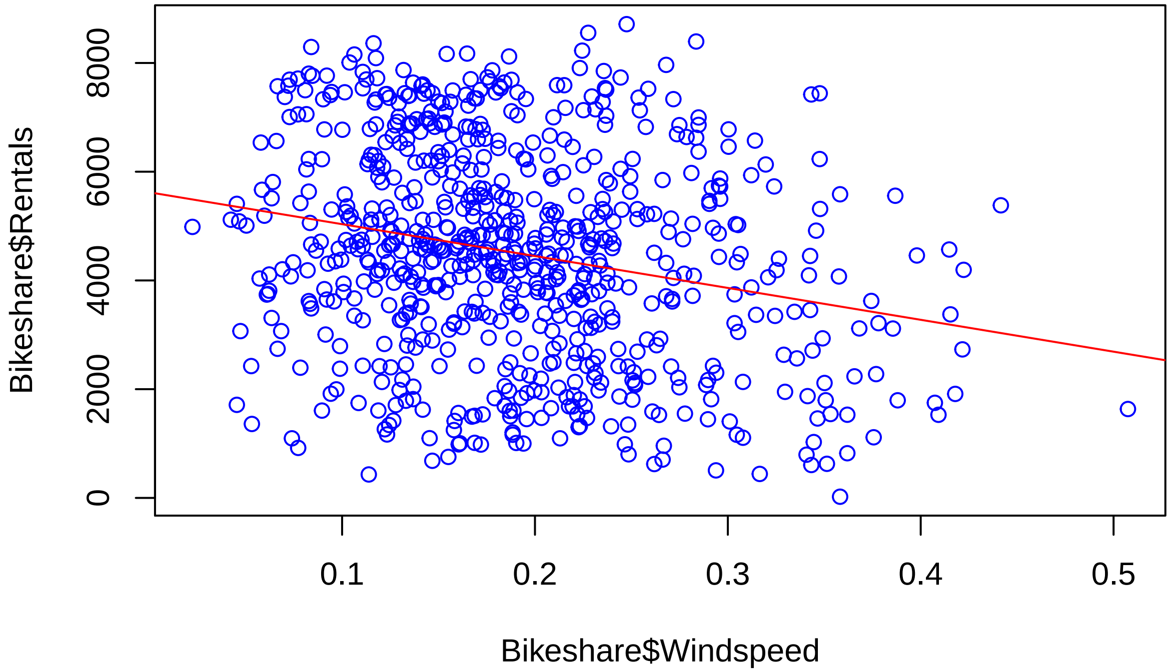

```
summary(LModel)
```

```
##
## Call:
## lm(formula = Rentals ~ Humidity, data = Bikeshare)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4741.0 -1386.9    50.3  1439.3  4036.8
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5364.0      322.7  16.623  < 2e-16 ***
## Humidity     -1369.1      501.2  -2.732  0.00645 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1929 on 729 degrees of freedom
## Multiple R-squared:  0.01013,    Adjusted R-squared:  0.008774
## F-statistic: 7.462 on 1 and 729 DF,  p-value: 0.006454
```

**For every 1 increase to normalized Humidity, Rentals change by -1369.1 . Accoring to small p <.05 this is significant but the R-squared: 0.01013 shows this is not a good model**

### 2.c) We run a simple regression for Rentals~Windspeed

```
LModel=lm(Rentals~Windspeed,data=Bikeshare)
{plot(Bikeshare$Windspeed,Bikeshare$Rentals,col='blue')
abline(LModel,col="red" )}
```




```r
summary(LModel)
```

```
##
## Call:
## lm(formula = Rentals ~ Windspeed, data = Bikeshare)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4522.7 -1374.7   -74.6  1461.8  4544.0
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5621.2      185.1  30.374  < 2e-16 ***
## Windspeed    -5862.9      900.0  -6.514 1.36e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1884 on 729 degrees of freedom
## Multiple R-squared:  0.05501,    Adjusted R-squared:  0.05372
## F-statistic: 42.44 on 1 and 729 DF,  p-value: 1.36e-10
```

**For every 1 increase to normalized Windspeed, Rentals change by -5862.9. Accoring to small p <.05 this is significant but R-squared: 0.05501 shows this is not significant**

#3 Multiple Linear Regression

```
pairs(Rentals~Temperature+ Humidity +Windspeed ,data=Bikeshare)
```