Lab 2A

08/02/2021

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v tibble  3.0.3     v dplyr   1.0.0
## v tidyr   1.1.1     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
## v purrr   0.3.4

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
library(colorspace)
library(Metrics)
```

```
## Warning: package 'Metrics' was built under R version 4.0.3
```

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.0.3
```

```
data1<-read_excel("C:/Users/Siddharth S Chandran/Downloads/data.xlsx")
```

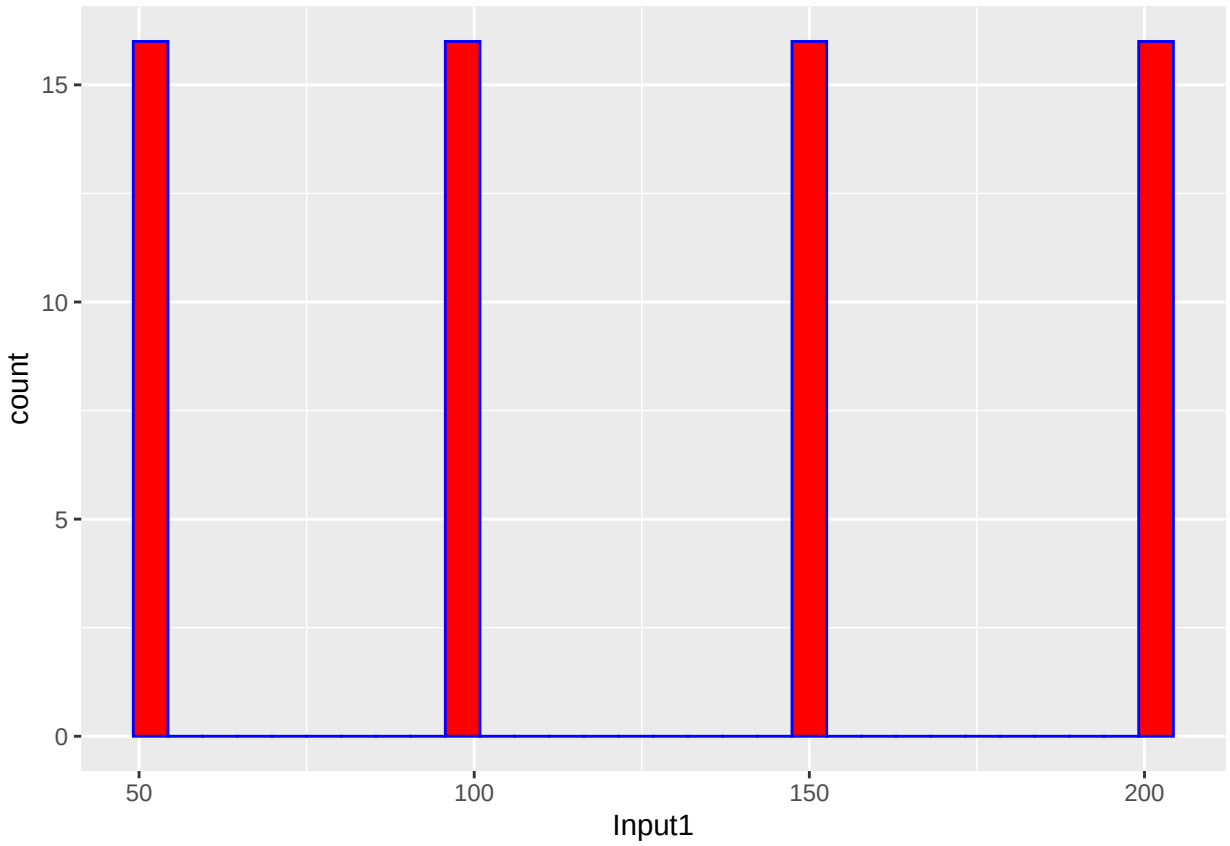
```
str(data1)
```

```
## tibble [64 x 4] (S3: tbl_df/tbl/data.frame)
##  $ Input1: num [1:64] 50 50 50 50 50 50 50 50 50 50 ...
##  $ Input2: num [1:64] 6 6 6 6 8 8 8 8 10 10 ...
##  $ Input3: num [1:64] 30 45 60 75 30 45 60 75 30 45 ...
##  $ Output: num [1:64] 4.86 4.97 5.79 6.28 5.06 ...
```

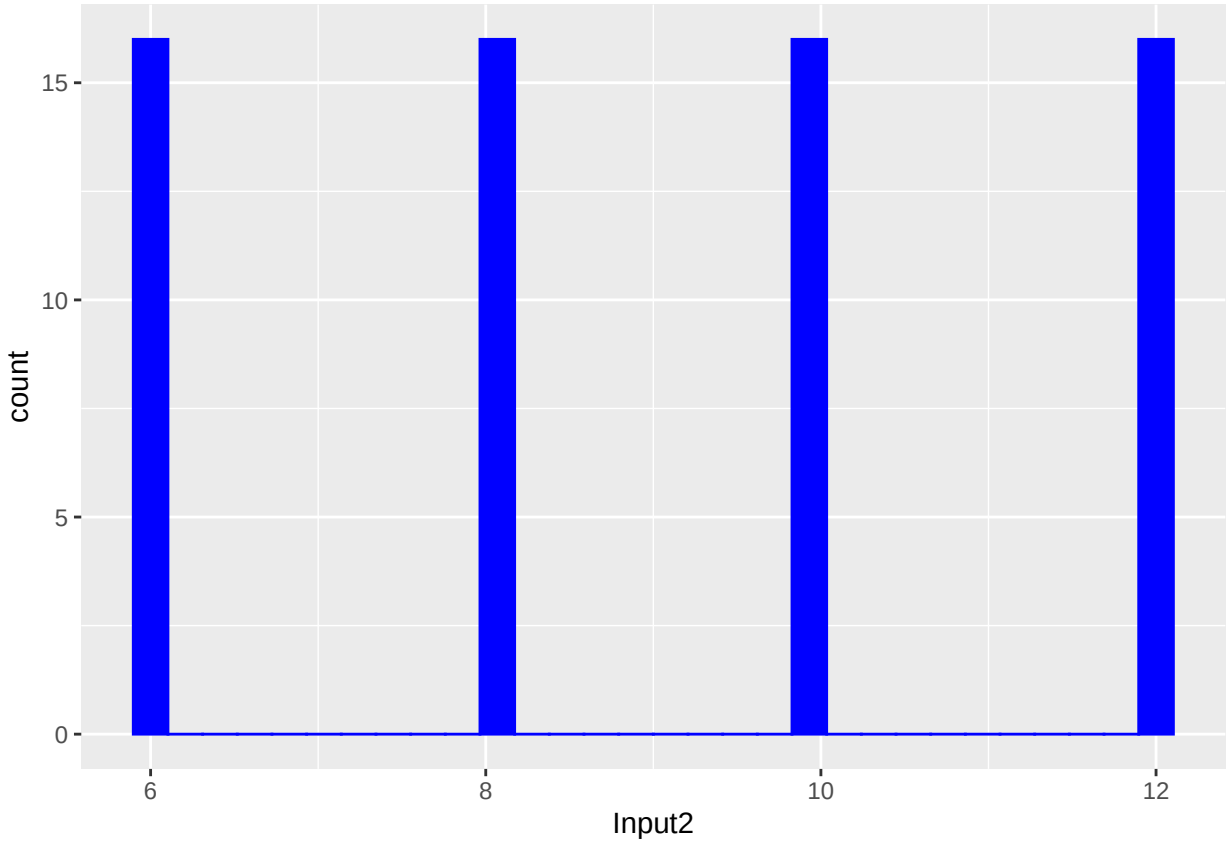
```
summary(data1)
```

```
##      Input1          Input2          Input3          Output      
##  Min.   : 50.0   Min.   : 6.0   Min.   :30.00   Min.   : 4.855  
##  1st Qu.: 87.5   1st Qu.: 7.5   1st Qu.:41.25   1st Qu.: 6.946  
##  Median :125.0   Median : 9.0   Median :52.50   Median : 7.208  
##  Mean   :125.0   Mean   : 9.0   Mean   :52.50   Mean   : 7.532  
##  3rd Qu.:162.5   3rd Qu.:10.5   3rd Qu.:63.75   3rd Qu.: 8.703  
##  Max.   :200.0   Max.   :12.0   Max.   :75.00   Max.   :10.127
```

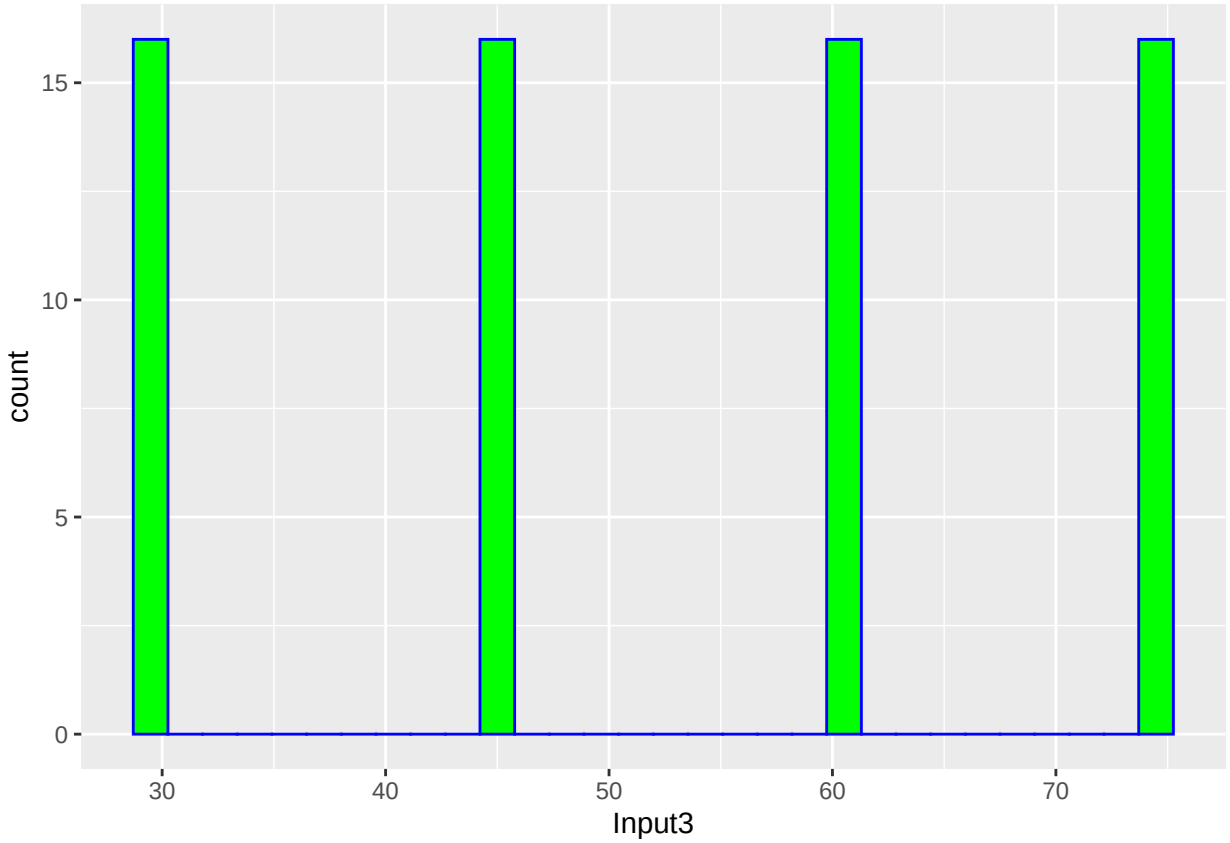
```
ggplot(data1, aes(x = Input1)) + geom_histogram(fill="red", color="blue")
```



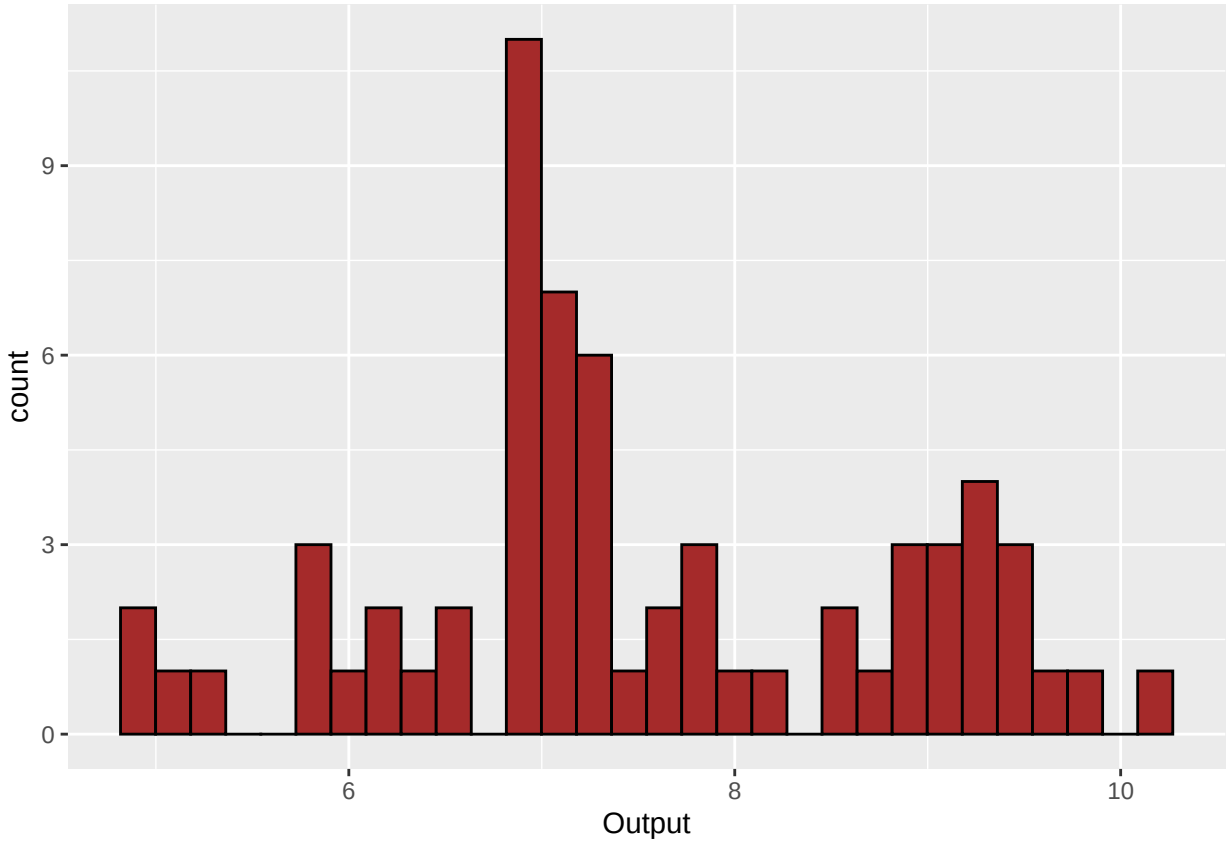

```
ggplot(data1, aes(x = Input2)) + geom_histogram(fill="blue", color="blue")
```



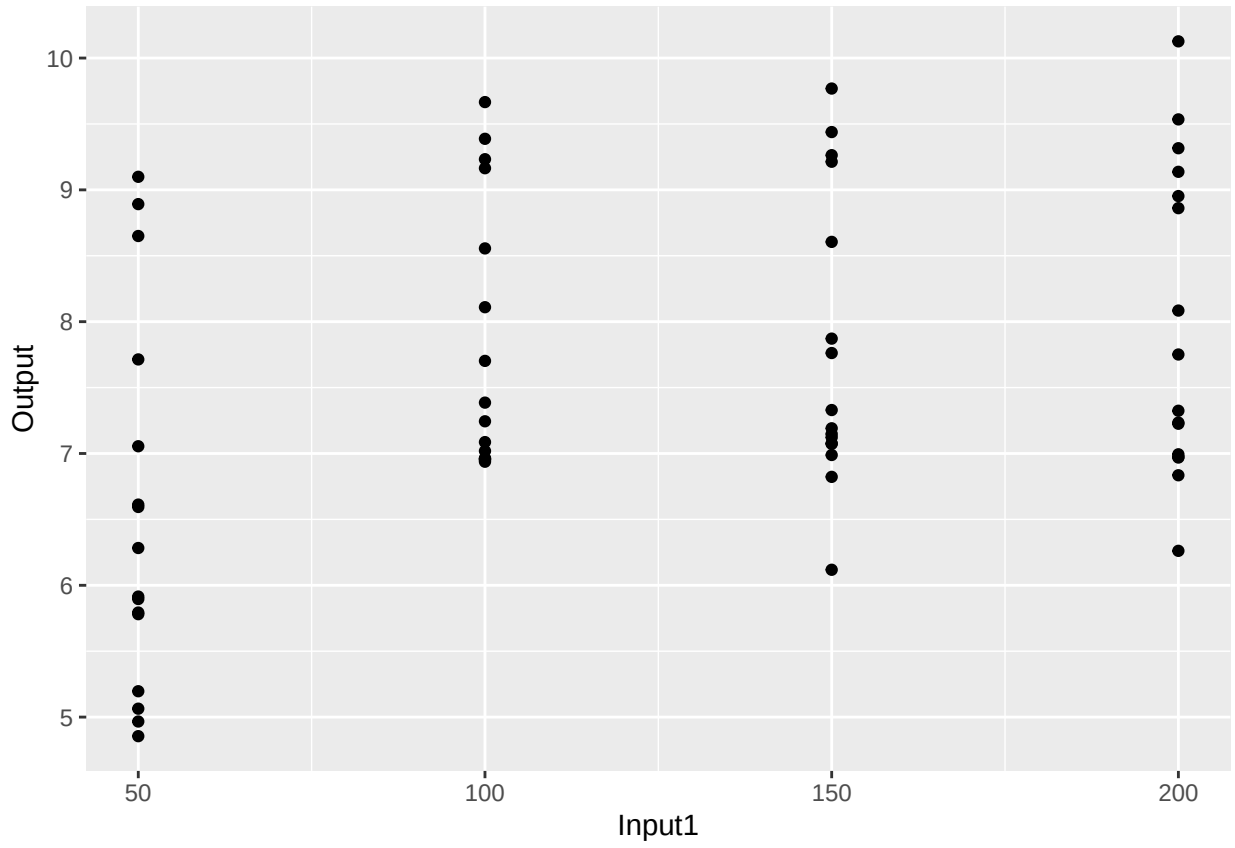

```
ggplot(data1, aes(x = Input3)) + geom_histogram(fill="green", color="blue")
```



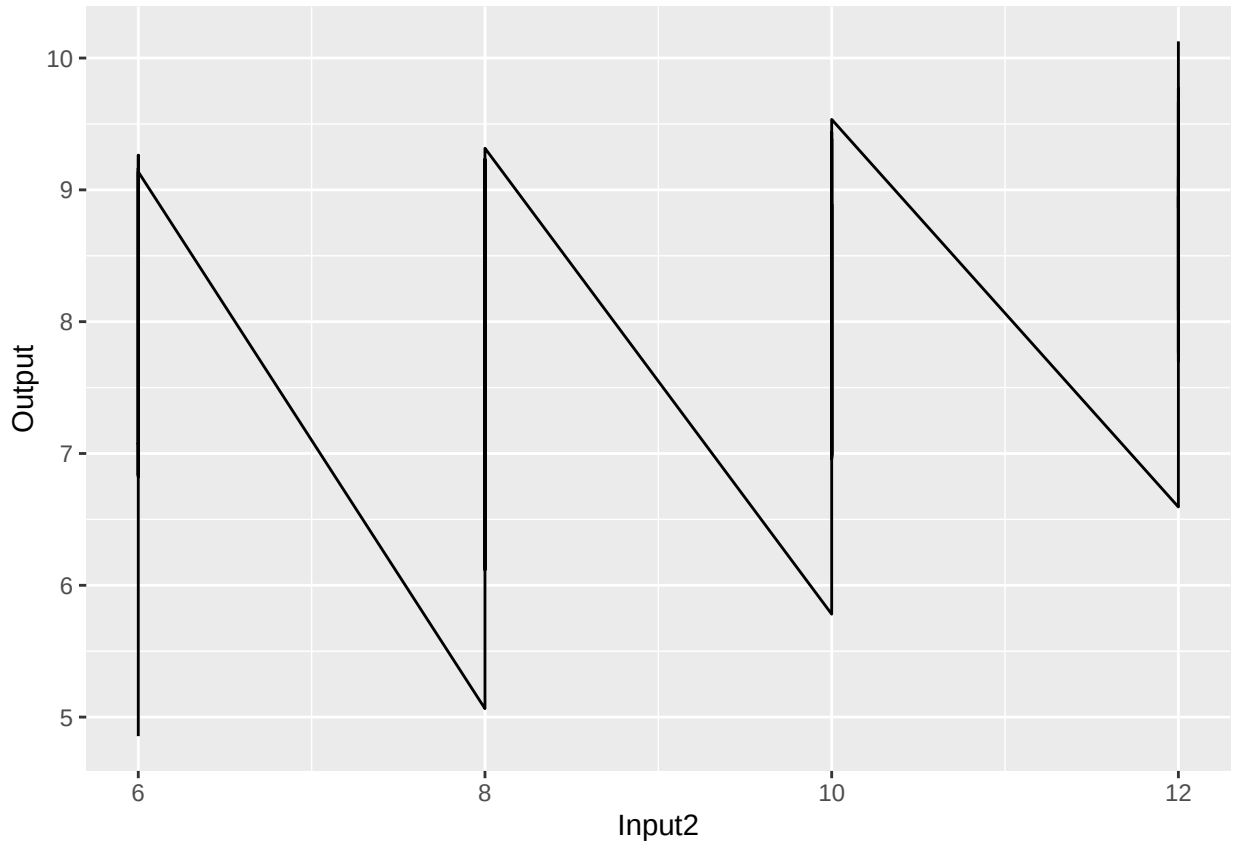

```
ggplot(data1, aes(x = Output)) + geom_histogram(fill="brown", color="black")
```



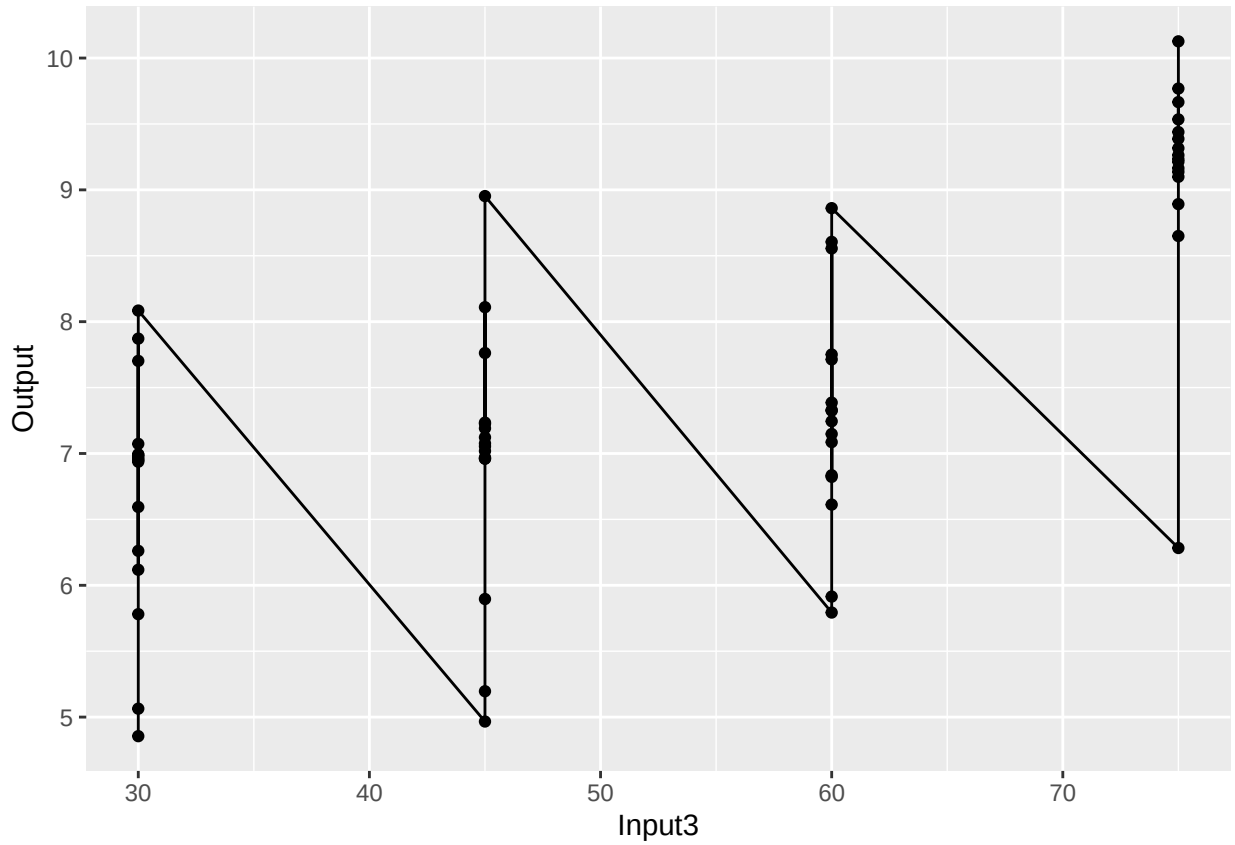

```
ggplot(data1, aes(x = Input1, y = Output)) + geom_point()
```




```
ggplot(data1, aes(x = Input2, y = Output)) + geom_line()
```




```
ggplot(data1, aes(x = Input3, y = Output)) + geom_line() + geom_point()
```

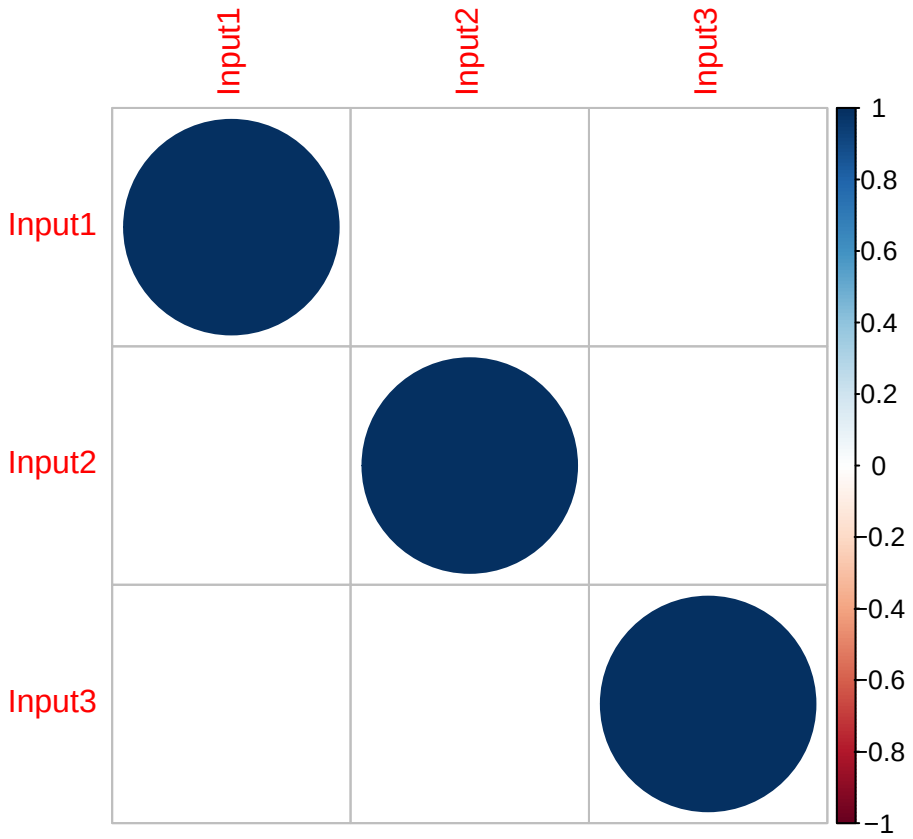



```
data1<- log(data1)
```

```
num1 <- data1[c(1,2,3)]
cor(num1)
```

```
##               Input1        Input2        Input3
## Input1  1.000000e+00 -3.143324e-20 -8.310578e-21
## Input2 -3.143324e-20  1.000000e+00 -1.910709e-20
## Input3 -8.310578e-21 -1.910709e-20  1.000000e+00
```

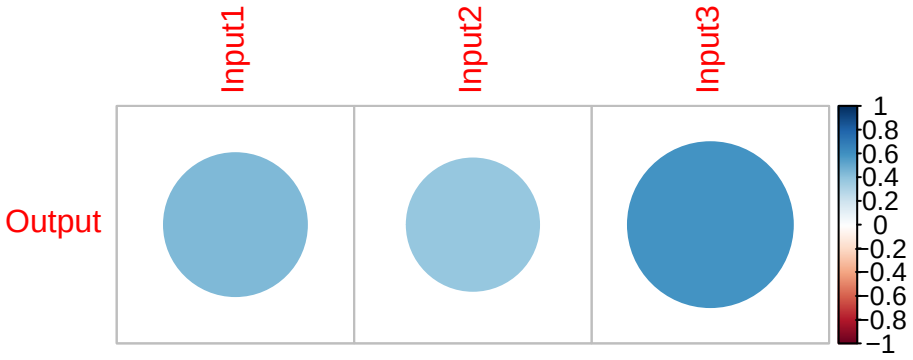
```
corrplot(cor(num1))
```




```
num1 <- data1[c(1,2,3)]
cor(data1[c(4)],num1)
```

```
##           Input1    Input2    Input3
## Output 0.4468511 0.3828276 0.5947629
```

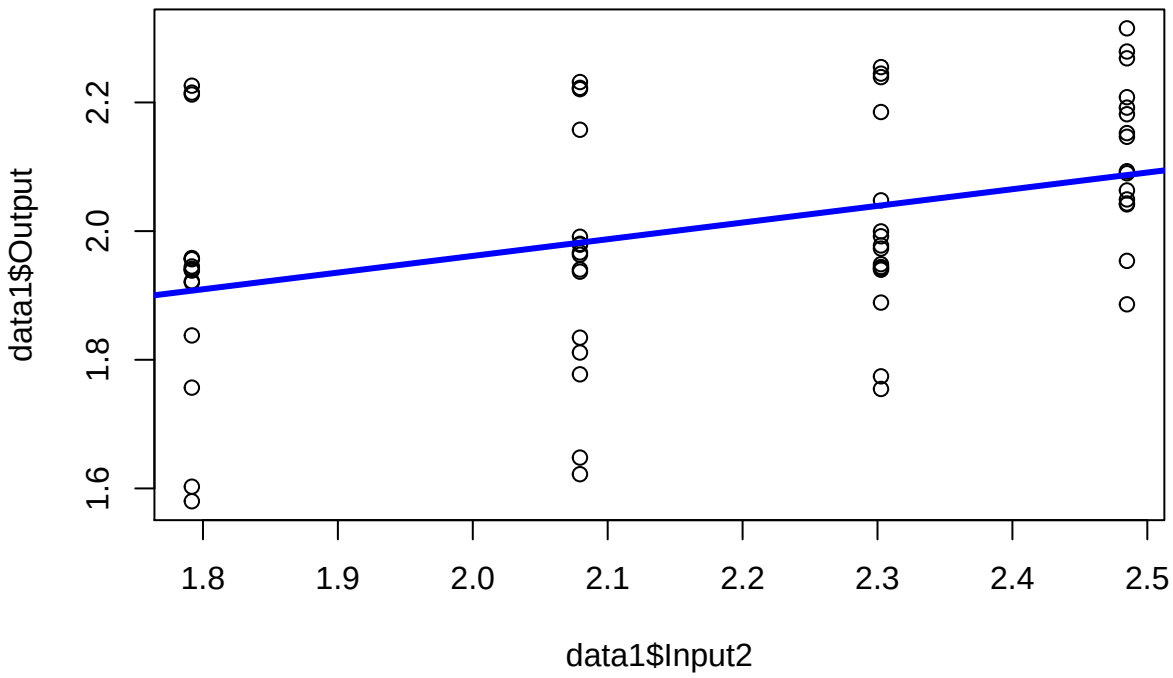
```
corrplot(cor(data1[c(4)],num1))
```



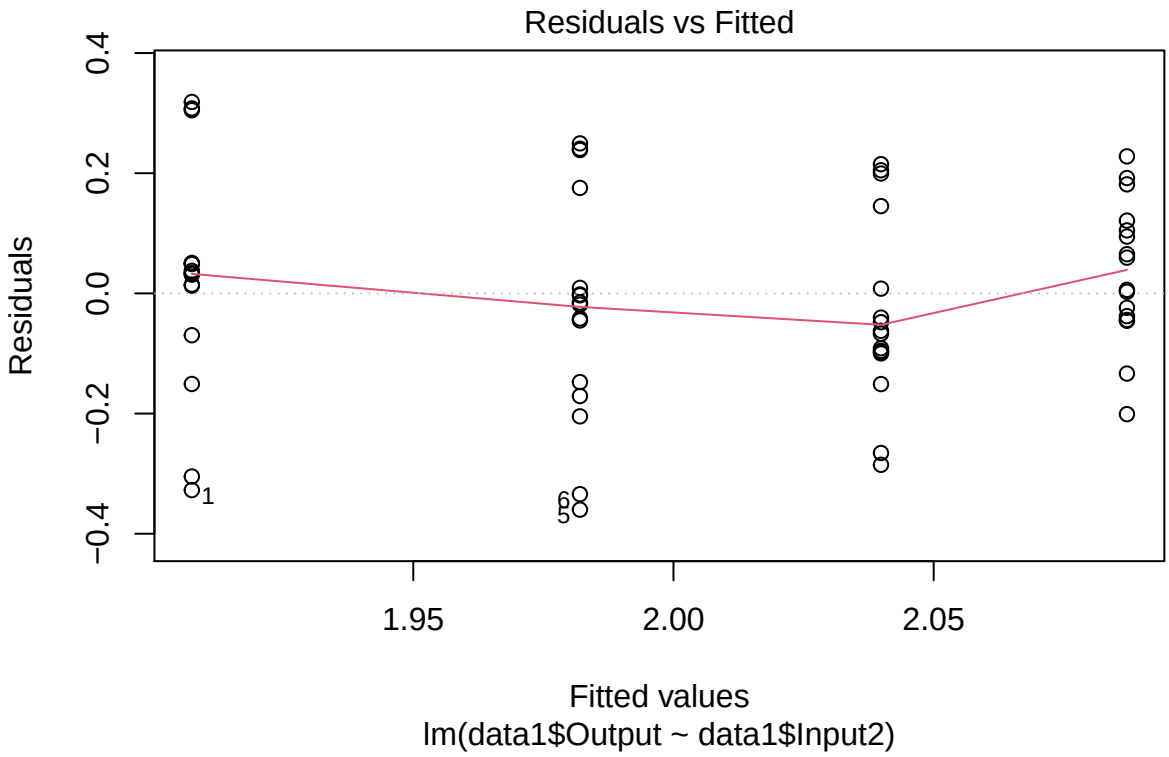

```
modelA<- lm(data1$Output~data1$Input2)
modelA
```

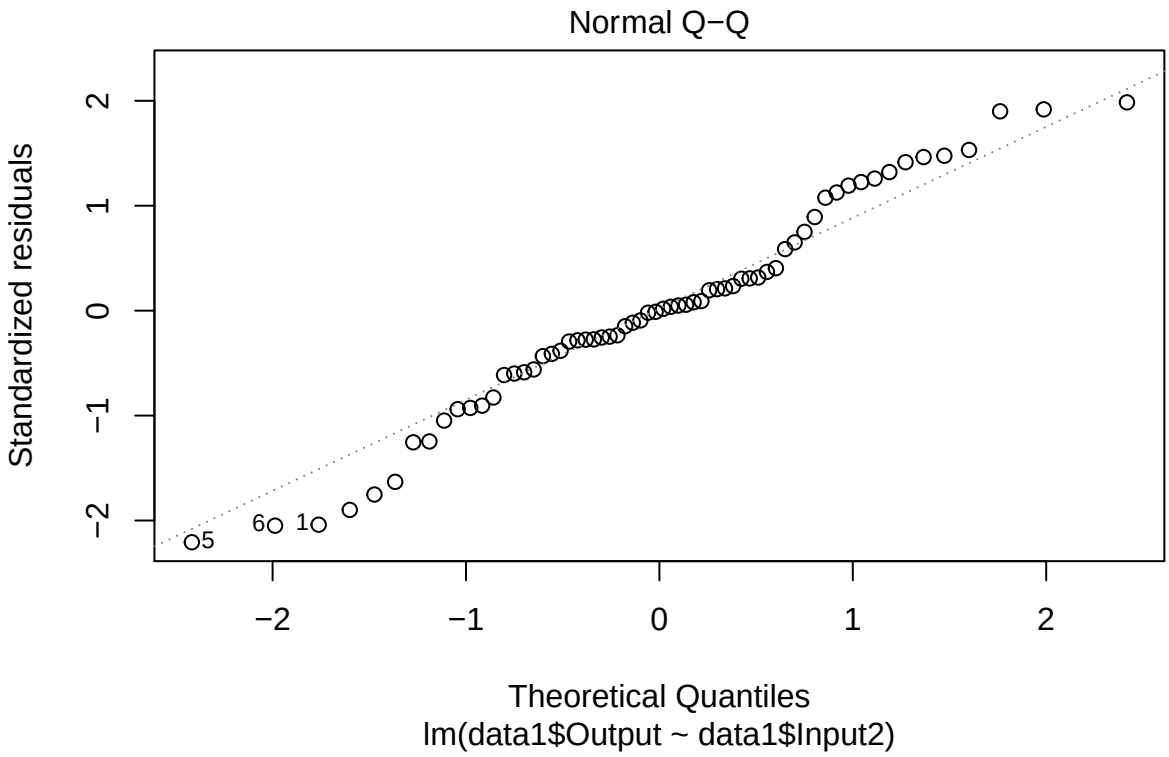
```
##
## Call:
## lm(formula = data1$Output ~ data1$Input2)
##
## Coefficients:
##  (Intercept)  data1$Input2
##       1.4430        0.2592
```

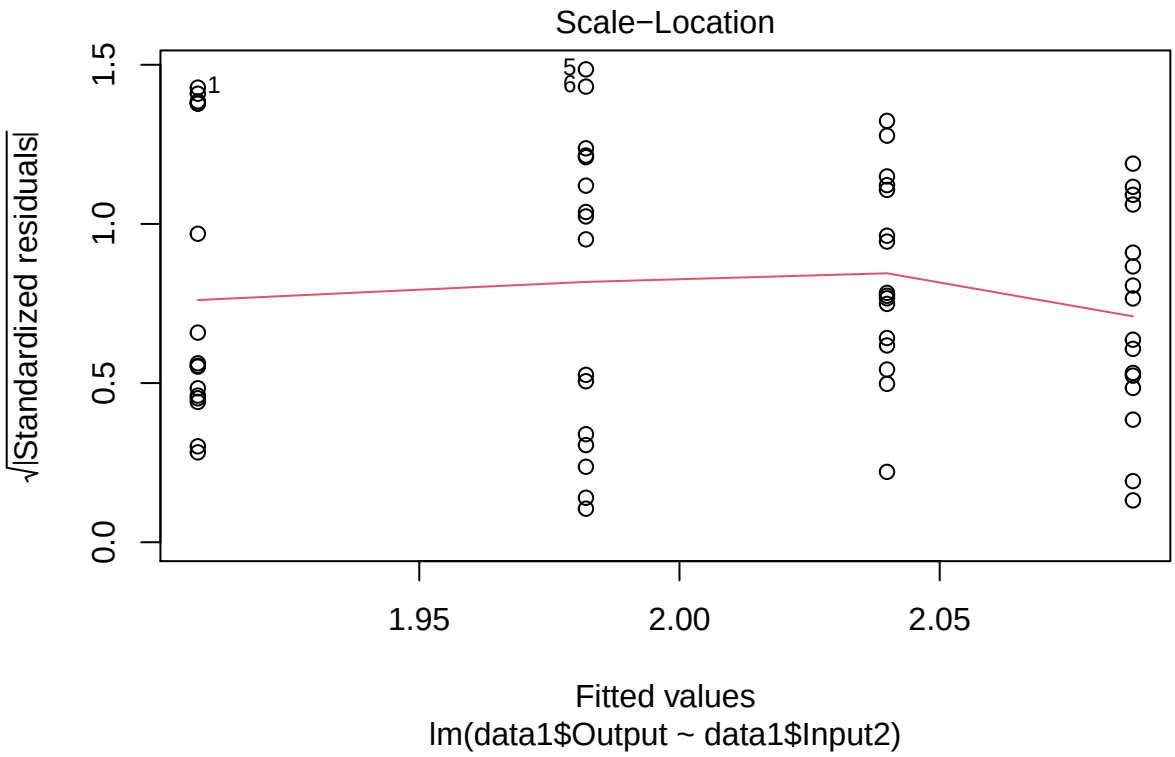
```
plot(data1$Output~data1$Input2)
abline(modelA, col="blue", lwd=3)
```

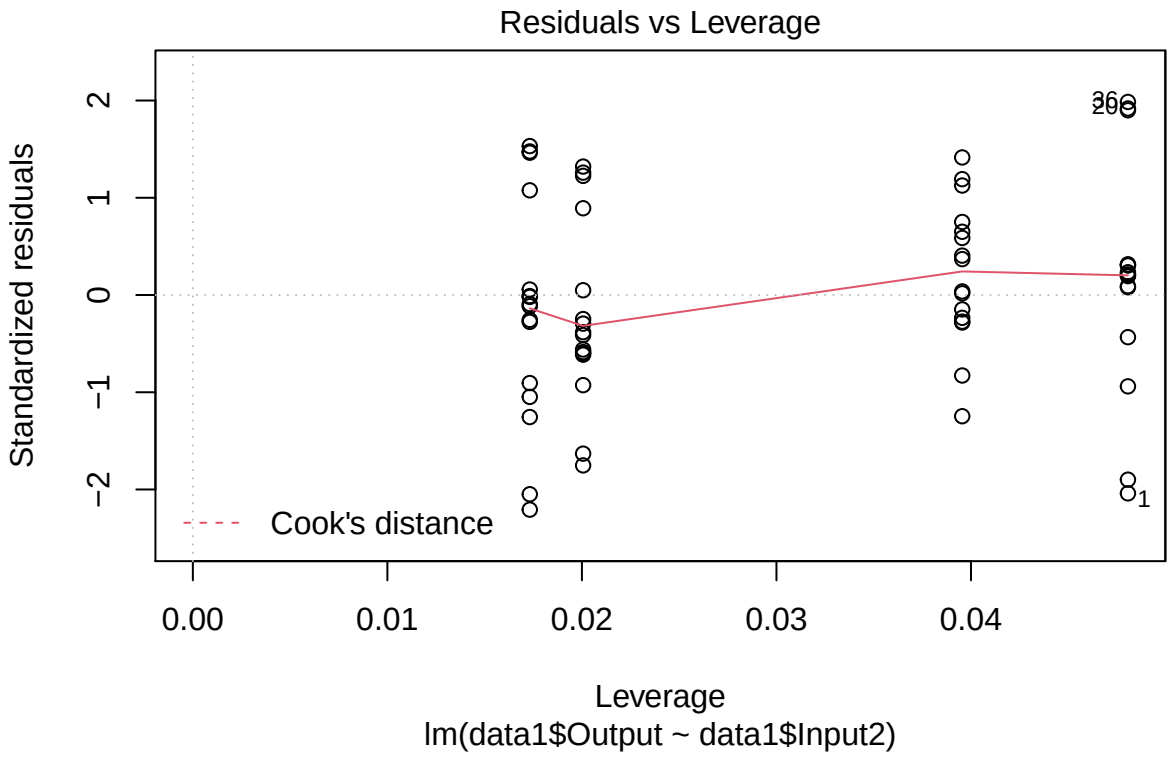



```
plot(modelA)
```


Residuals vs Fitted
Residuals
0.4
0.2
0.0
−0.2
−0.4
1
6
5
1.95
2.00
2.05
Fitted values
lm(data1$Output ~ data1$Input2)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(data1$Output ~ data1$Input2)
5
6
1
−2
−1
0
1
2


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
1.95
2.00
2.05
Fitted values
lm(data1$Output ~ data1$Input2)




```
new <- data.frame(Input2 = data1$Input2)
a<- predict(modelA,new)
b<-c(data1$Output)
mae(a,b)
```

```
## [1] 0.1245307
```

```
mse(a,b)
```

```
## [1] 0.02620837
```

```
rmse(a,b)
```

```
## [1] 0.16189
```

```
summary(modelA)
```

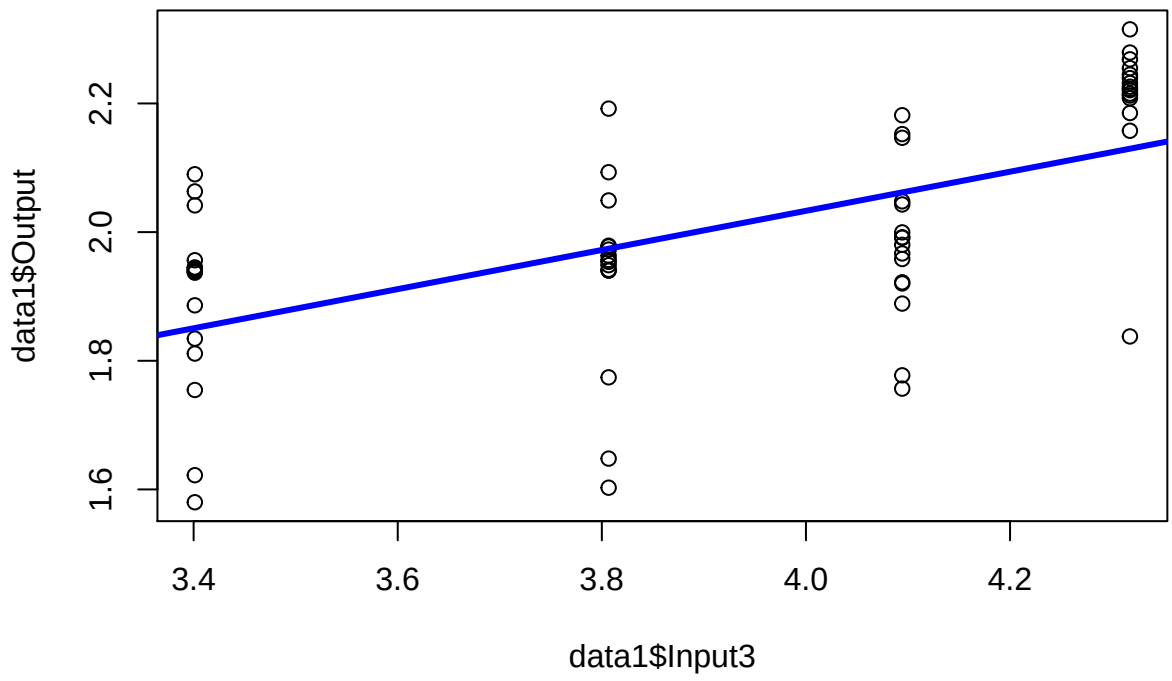
```
##
## Call:
## lm(formula = data1$Output ~ data1$Input2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -0.35987 -0.09237  0.00049  0.09711  0.31858
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.44298    0.17320   8.331 1.05e-11 ***
## data1$Input2  0.25923    0.07945   3.263   0.0018 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1645 on 62 degrees of freedom
## Multiple R-squared:  0.1466, Adjusted R-squared:  0.1328
## F-statistic: 10.65 on 1 and 62 DF,  p-value: 0.001796
```

```
modelB<- lm(data1$Output~data1$Input3)
modelB
```

```
##
## Call:
## lm(formula = data1$Output ~ data1$Input3)
##
## Coefficients:
##  (Intercept)  data1$Input3
##       0.8161        0.3042
```

```
plot(data1$Output~data1$Input3)
abline(modelB, col="blue", lwd=3)
```



```r
new <- data.frame(Input3 = data1$Input3)
a<- predict(modelB,new)
b<-c(data1$Output)
mae(a,b)
```

```
## [1] 0.1117756
```

```r
mse(a,b)
```

```
## [1] 0.0198459
```

```r
rmse(a,b)
```

```
## [1] 0.1408755
```

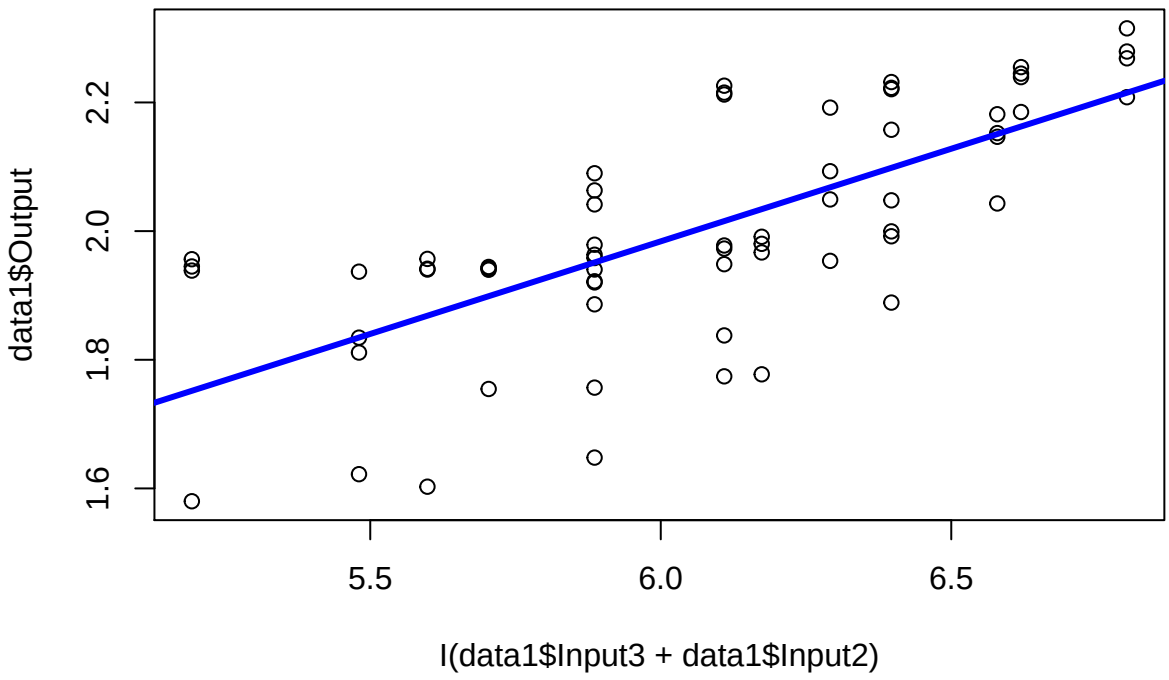
```r
summary(modelB)
```

```
##
## Call:
## lm(formula = data1$Output ~ data1$Input3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.37151 -0.07000  0.01626  0.09357  0.23905
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.81614    0.20471   3.987 0.000179 ***
## data1$Input3  0.30423    0.05222   5.826 2.19e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1431 on 62 degrees of freedom
## Multiple R-squared:  0.3537, Adjusted R-squared:  0.3433
## F-statistic: 33.94 on 1 and 62 DF,  p-value: 2.191e-07
```

```r
modelC<- lm(data1$Output~I((data1$Input3 + data1$Input2)))
modelC
```

```
##
## Call:
## lm(formula = data1$Output ~ I((data1$Input3 + data1$Input2)))
##
## Coefficients:
##                       (Intercept)  I((data1$Input3 + data1$Input2))
##                            0.2568                            0.2879
```

```r
plot(data1$Output~I(data1$Input3+ data1$Input2))
abline(modelC, col="blue", lwd=3)
```




```
new <- data.frame(Input3 = data1$Input3,Input2=data1$Input2)
a<- predict(modelC,new)
b<-c(data1$Output)
mae(a,b)
```

```
## [1] 0.09864359
```

```
mse(a,b)
```

```
## [1] 0.01543162
```

```
rmse(a,b)
```

```
## [1] 0.1242241
```

```
summary(modelC)
```

```
##
## Call:
## lm(formula = data1$Output ~ I((data1$Input3 + data1$Input2)))
##
## Residuals:
##       Min        1Q    Median        3Q       Max
```

```
## -0.303370 -0.065569  0.009415  0.088785  0.210498
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      0.25682    0.22358   1.149    0.255
## I((data1$Input3 + data1$Input2))  0.28788    0.03674   7.835 7.67e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1262 on 62 degrees of freedom
## Multiple R-squared:  0.4975, Adjusted R-squared:  0.4894
## F-statistic: 61.38 on 1 and 62 DF,  p-value: 7.665e-11
```