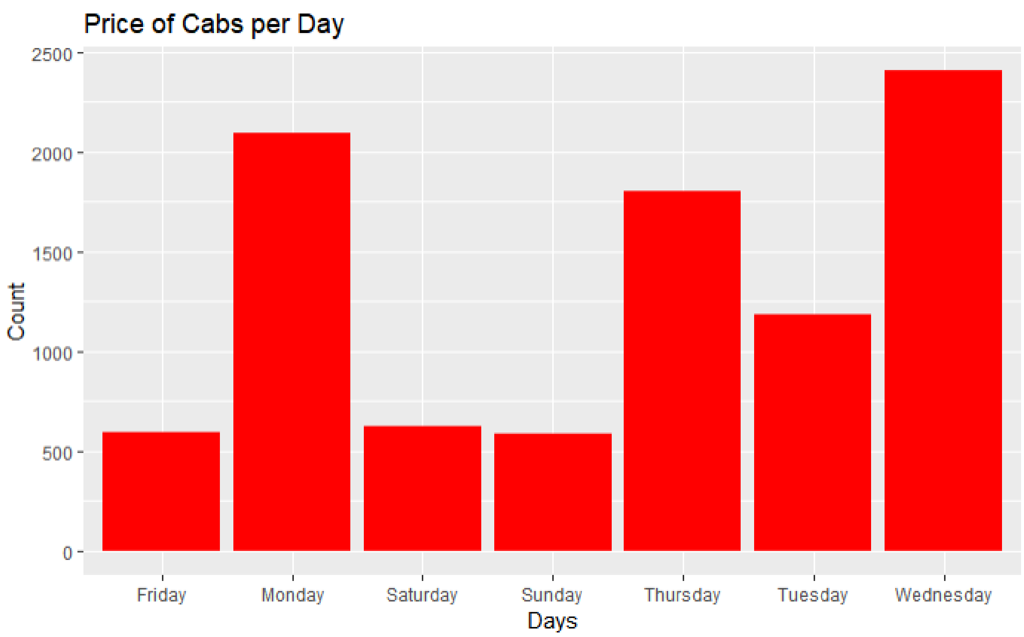
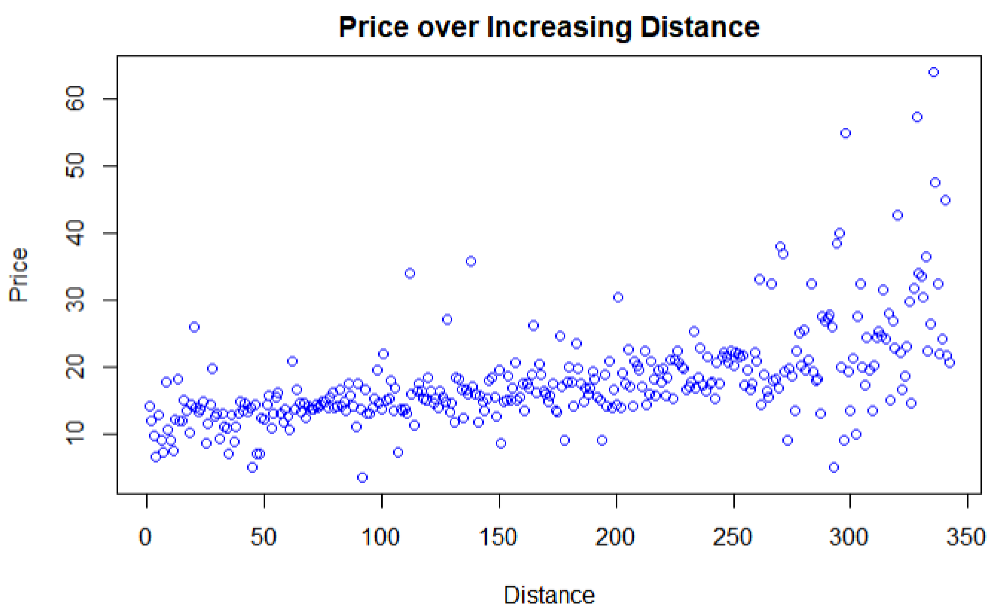
Lyft-Uber Price Prediction

```
##EXPLORING VARIABLES AND THEIR RELATIONSHIPS
```

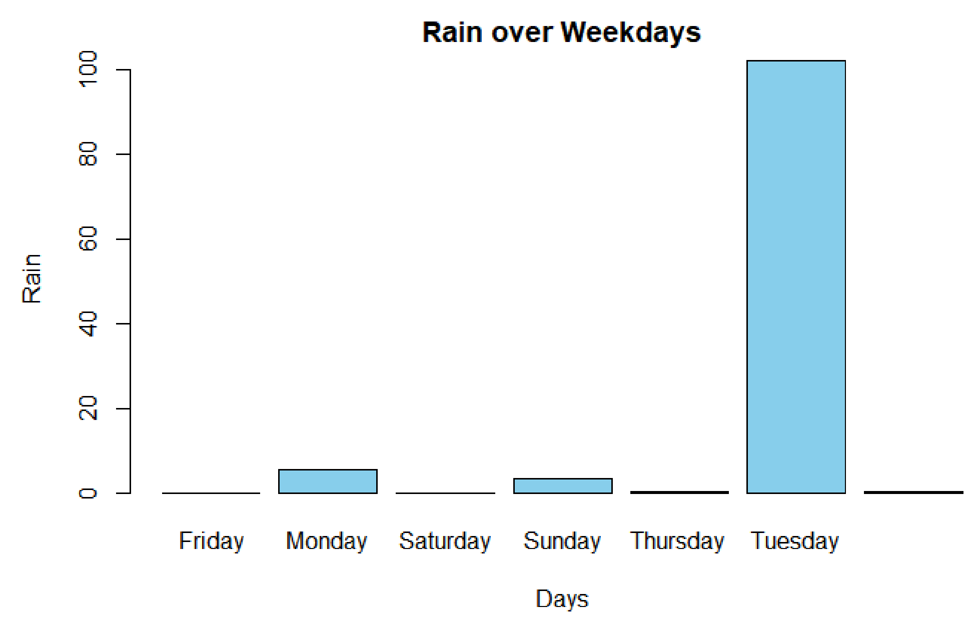
```{r pressure, echo=FALSE}
library(ggplot2)
ggplot(data = merge_data, aes(x = merge_data$day, fill= merge_data$price))+
  geom_bar(fill = "red", size = 2 ) +
  labs(y="Count", x= "Days", title="Price of Cabs per Day")
```



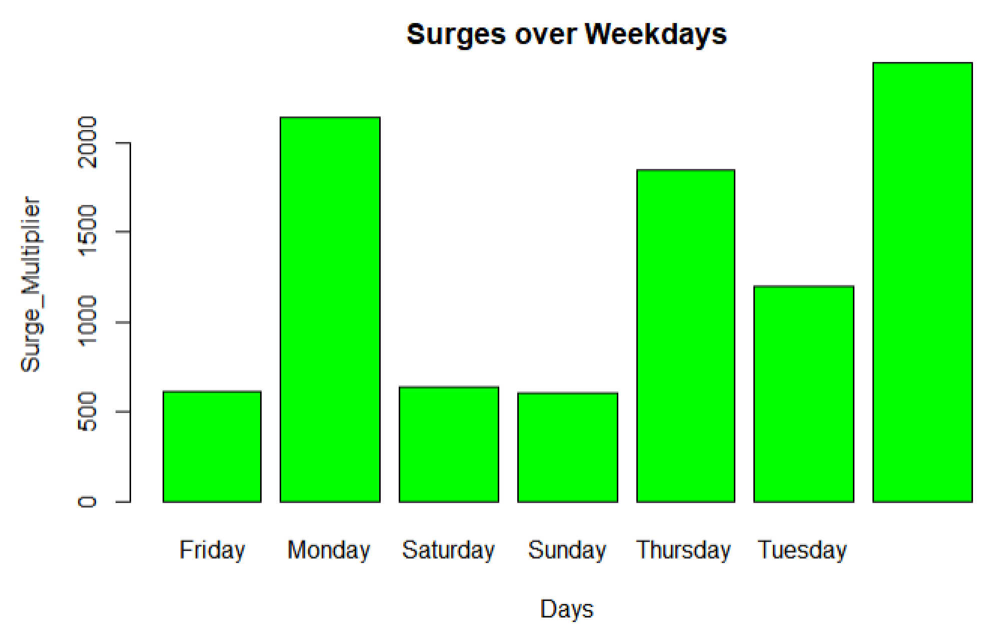

```{r pressure, echo=FALSE}
#Price over Distance
tapply(merge_data$price, merge_data$distance, FUN=mean)
plot(tapply(merge_data$price, merge_data$distance, FUN=mean),xlab = "Distance",ylab="Price",
     main = "Price over Increasing Distance", col="blue")
```



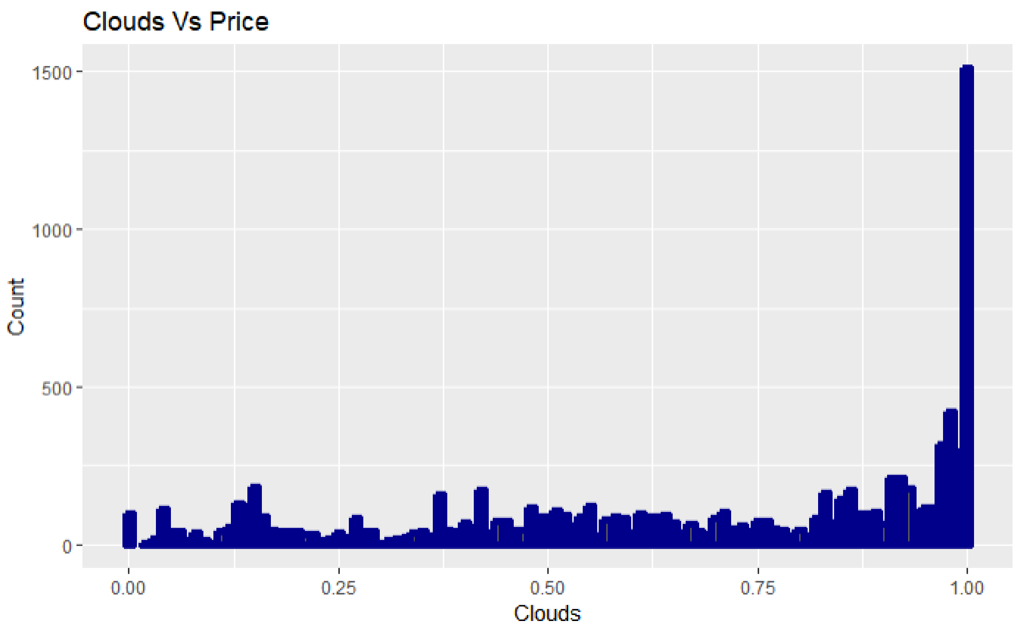
```{r pressure, echo=FALSE}
#Rain over days
table(merge_data$rain)
tapply(merge_data$rain, merge_data$day, FUN=sum)
barplot(tapply(merge_data$rain, merge_data$day, FUN=sum),xlab = "Days",ylab="Rain",
        main = "Rain over Weekdays", col="sky blue")
```




```{r pressure, echo=FALSE}
#Surge_Multiplier over Days
tapply(merge_data$surge_multiplier, merge_data$day, FUN=mean)
barplot(tapply(merge_data$surge_multiplier, merge_data$day, FUN=sum),xlab = "Days",
        ylab="Surge_Multiplier",main = "Surges over Weekdays", col = "green")
```

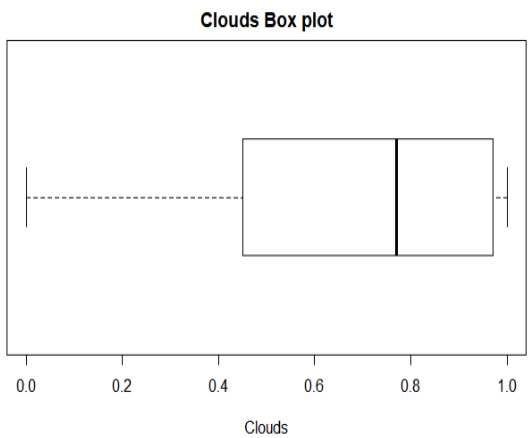


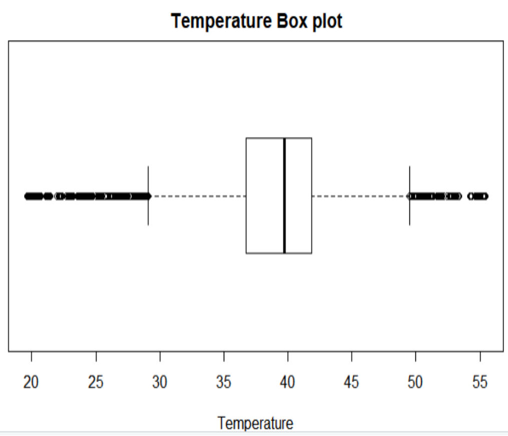
~~~~
```{r pressure, echo=FALSE}
#Cloud vs Price (We can see an increase in price with increase in clouds)
ggplot(data = merge_data, aes(x = merge_data$clouds, fill = merge_data$price))+
  geom_bar(color = "dark blue", size = 2)+labs(y="Count", x= "Clouds",title="Clouds Vs Price")
```
~~~~

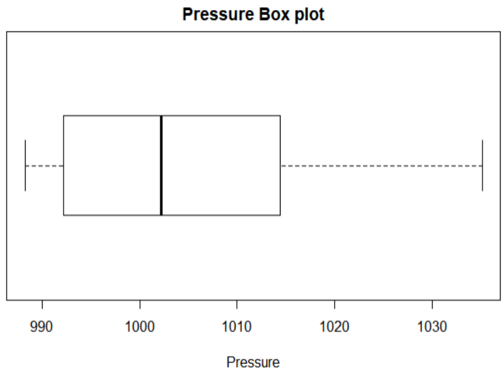


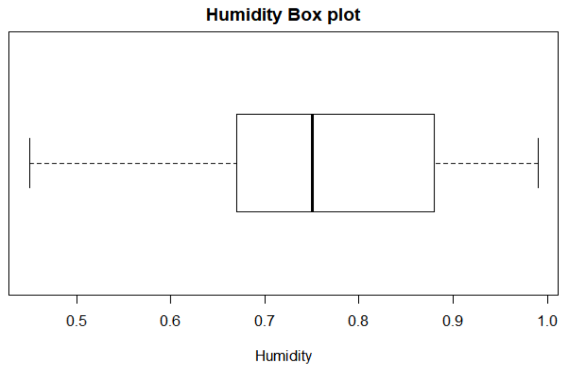
#BOXPLOT--#Outliers??

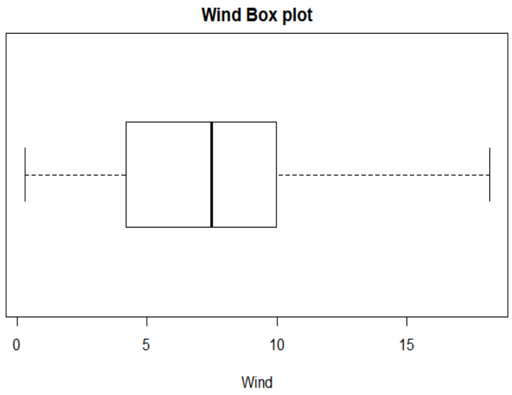
~~~~
```{r pressure, echo=FALSE}
#BOXPLOT
boxplot(df$clouds, main="Clouds Box plot",yaxt="n", xlab="Clouds", horizontal=TRUE)
boxplot(df$temp, main="Temperature Box plot",yaxt="n", xlab="Temperature", horizontal=TRUE)
boxplot(df$pressure, main="Pressure Box plot",yaxt="n", xlab="Pressure", horizontal=TRUE)
boxplot(df$humidity, main="Humidity Box plot",yaxt="n", xlab="Humidity", horizontal=TRUE)
boxplot(df$wind, main="Wind Box plot",yaxt="n", xlab="Wind", horizontal=TRUE)
boxplot(df$distance, main="Distance Box plot",yaxt="n", xlab="Distance", horizontal=TRUE)
```
~~~~

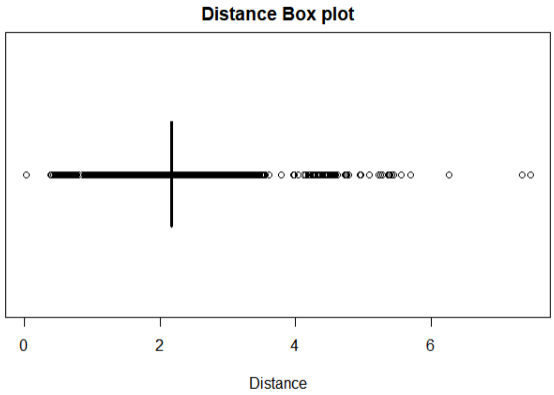






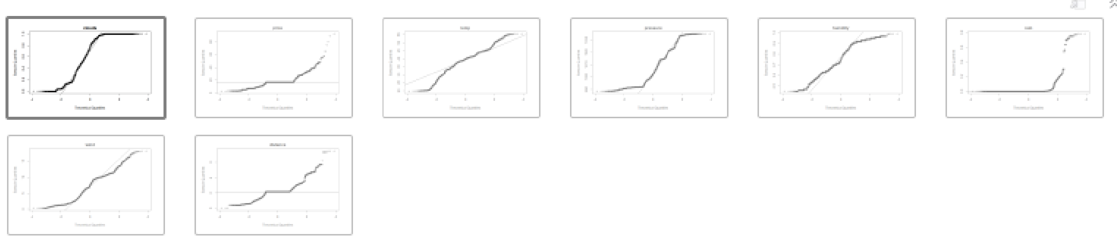


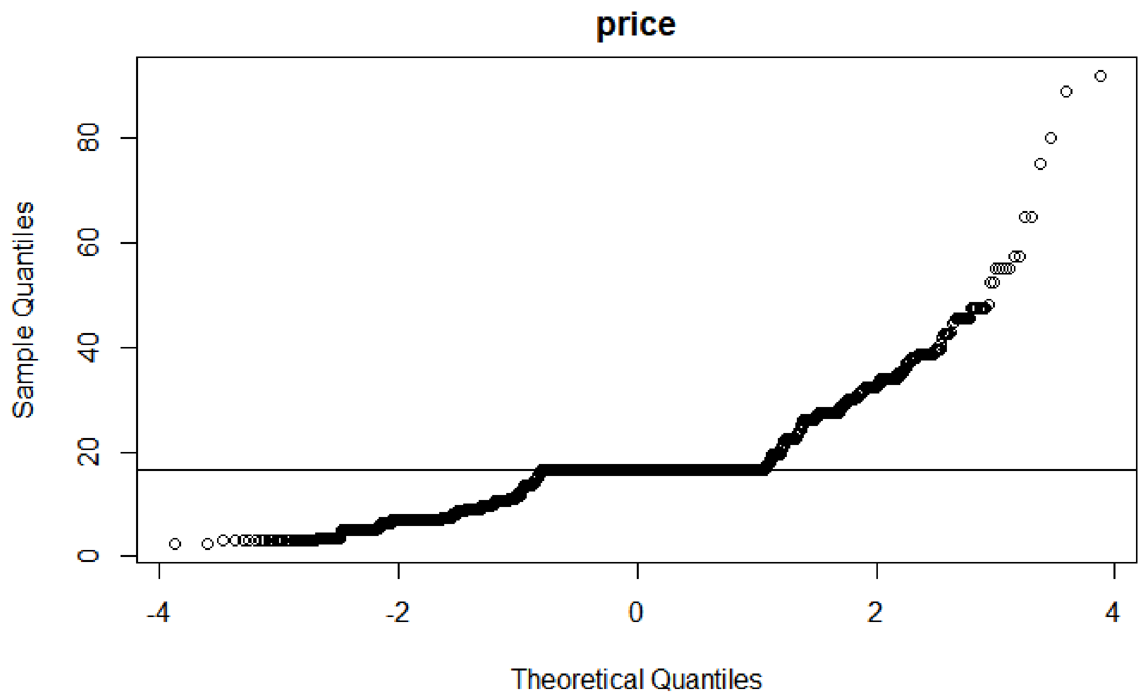


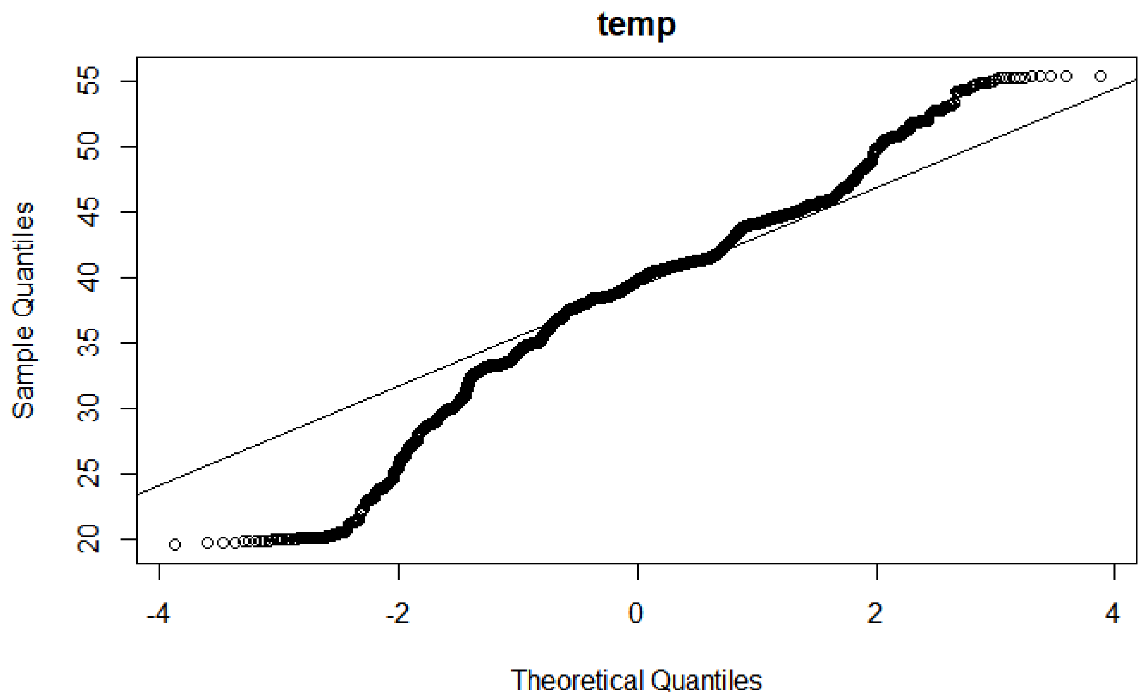


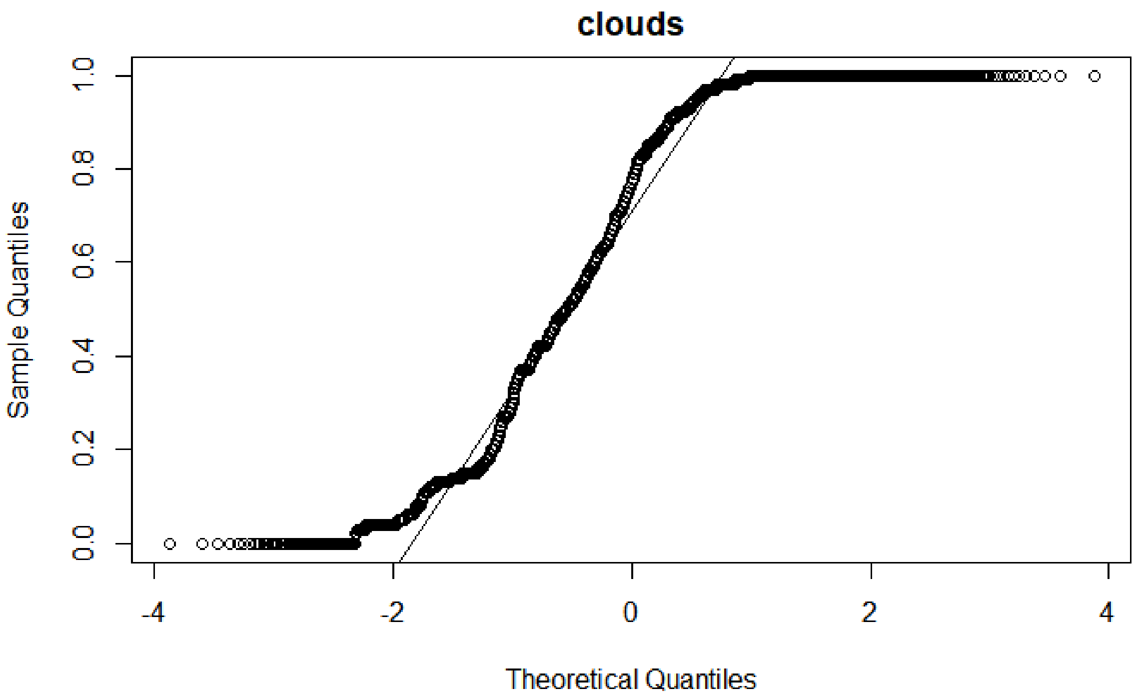

```
#Normal Probability plots for clouds, price, temp, pressure, humidity, rain, wind, distance
```

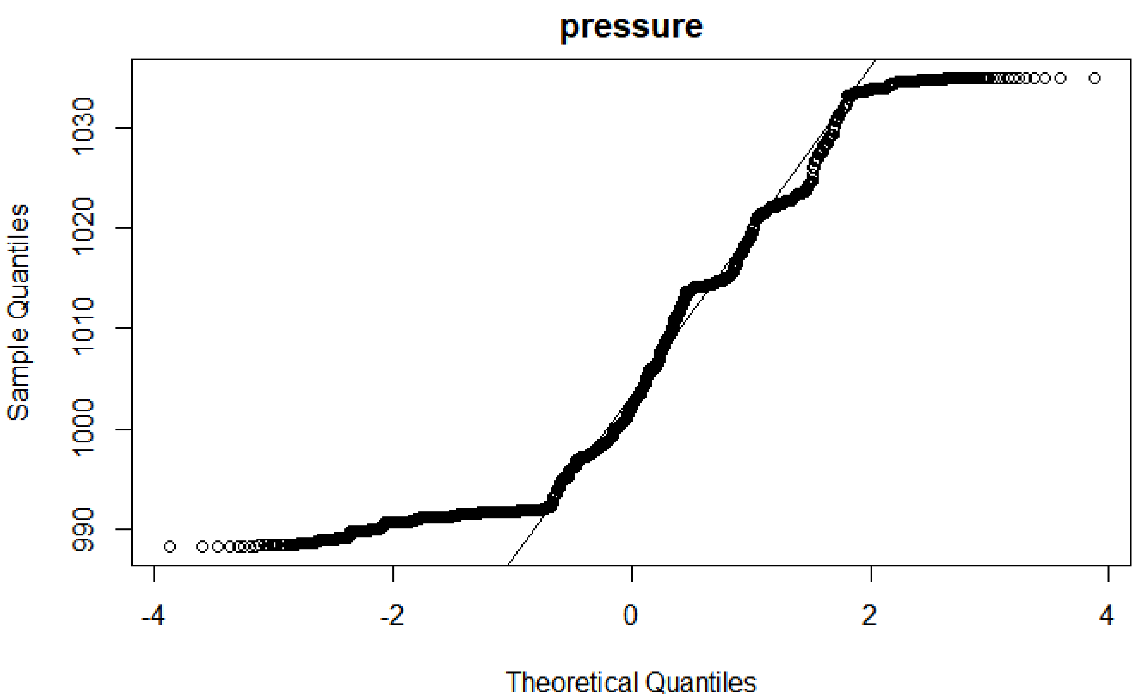
````
```{r pressure, echo=FALSE}
qqnorm(df[,"clouds"], main = "clouds"); qqline(df[,"clouds"])
qqnorm(df[,"price"], main = "price"); qqline(df[,"price"])
qqnorm(df[,"temp"], main = "temp"); qqline(df[,"temp"])
qqnorm(df[,"pressure"], main = "pressure"); qqline(df[,"pressure"])
qqnorm(df[,"humidity"], main = "humidity"); qqline(df[,"humidity"])
qqnorm(df[,"rain"], main = "rain"); qqline(df[,"rain"])
qqnorm(df[,"wind"], main = "wind"); qqline(df[,"wind"])
qqnorm(df[,"distance"], main = "distance"); qqline(df[,"distance"])
```
````

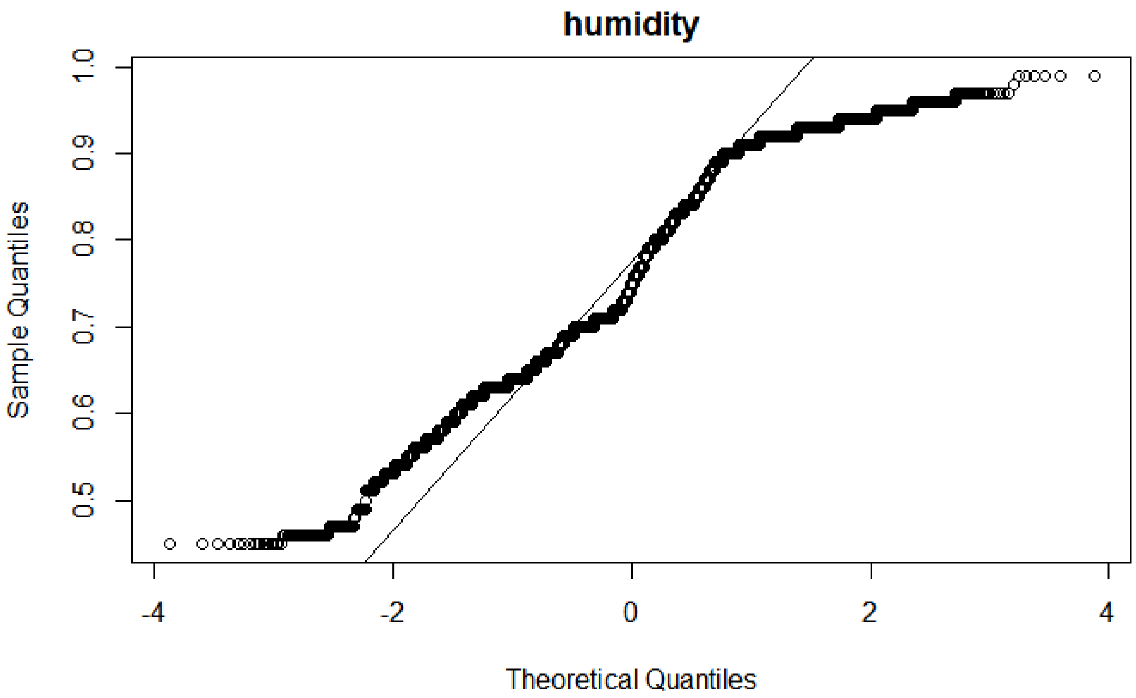



price
Sample Quantiles
0
20
40
60
80
-4
-2
0
2
4
Theoretical Quantiles


temp
Sample Quantiles
20
25
30
35
40
45
50
55
-4
-2
0
2
4
Theoretical Quantiles

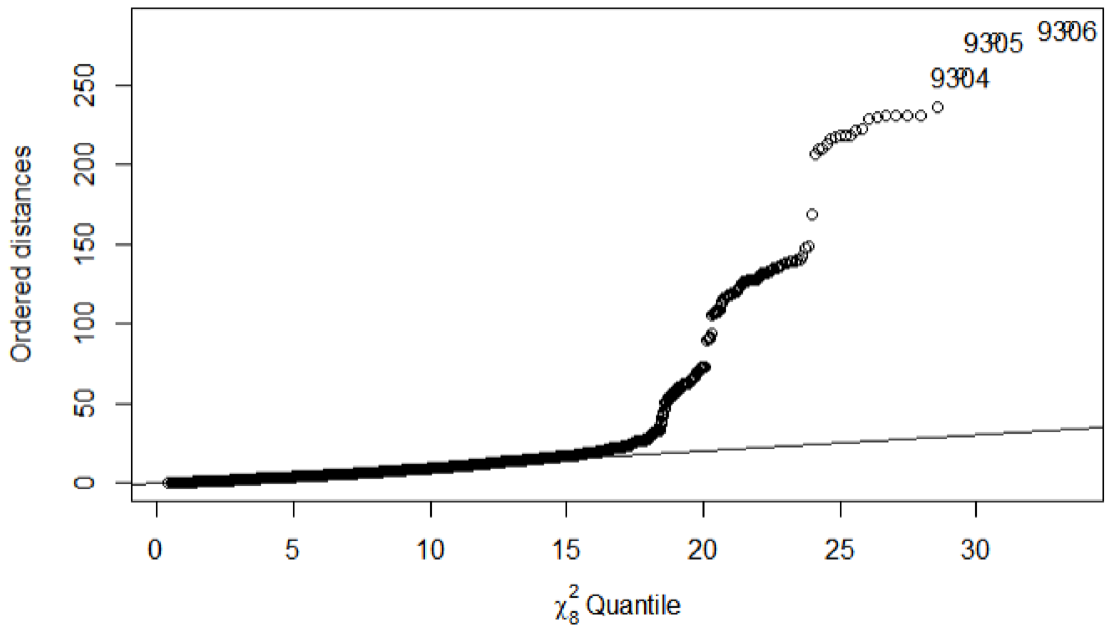

clouds
Sample Quantiles
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Theoretical Quantiles


pressure
Sample Quantiles
990
1000
1010
1020
1030
-4
-2
0
2
4
Theoretical Quantiles



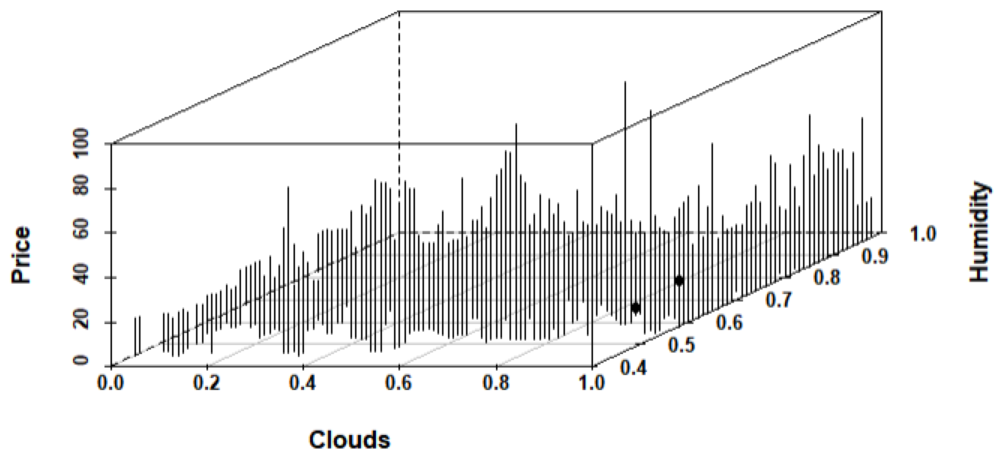

```
#we saw that invidual variables somewhat deviate from linearity. Now let's check multivarite
linearity. we calculated the genwralized distance of each observation and they should be appromiately
have chi-square distribution if all observations are linearly distributed..

```{r pressure, echo=FALSE}
plot(qc <- qchisq((1:nrow(x) - 1/2) / nrow(x), df = 8), sd <- sort(d),xlab =
expression(paste(chi[8]^2, " Quantile")),ylab = "Ordered distances")
oups <- which(rank(abs(qc - sd), ties = "random") > nrow(x) - 3)
text(qc[oups], sd[oups] - 1.5,oups)
abline(a = 0, b = 1)
```
```



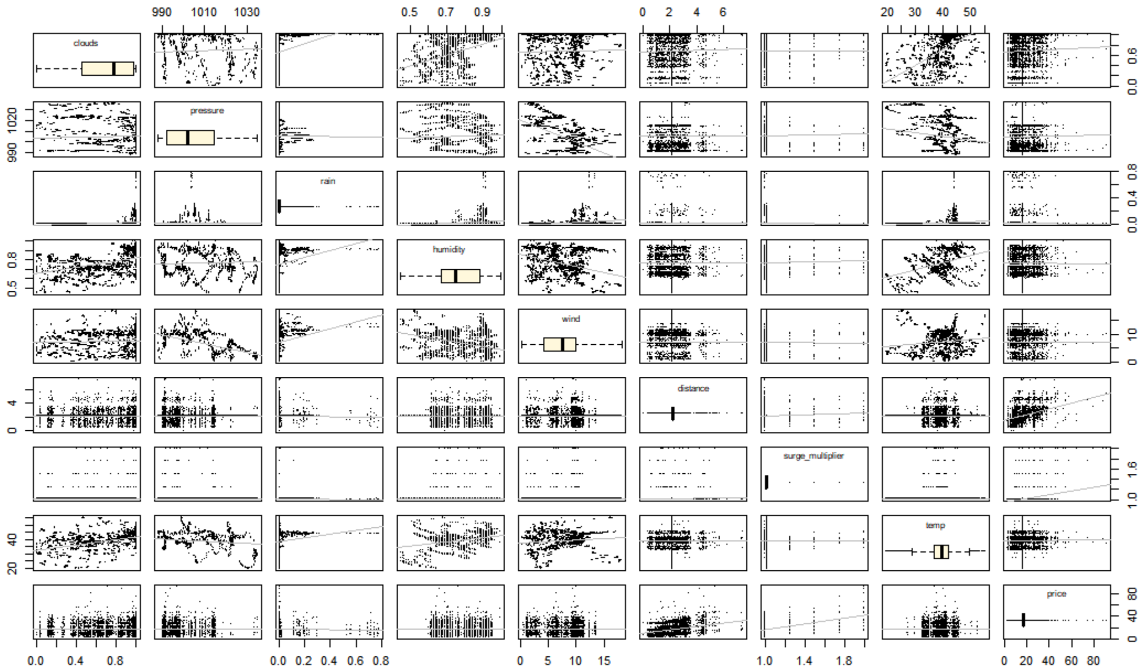
Creating 3D plot

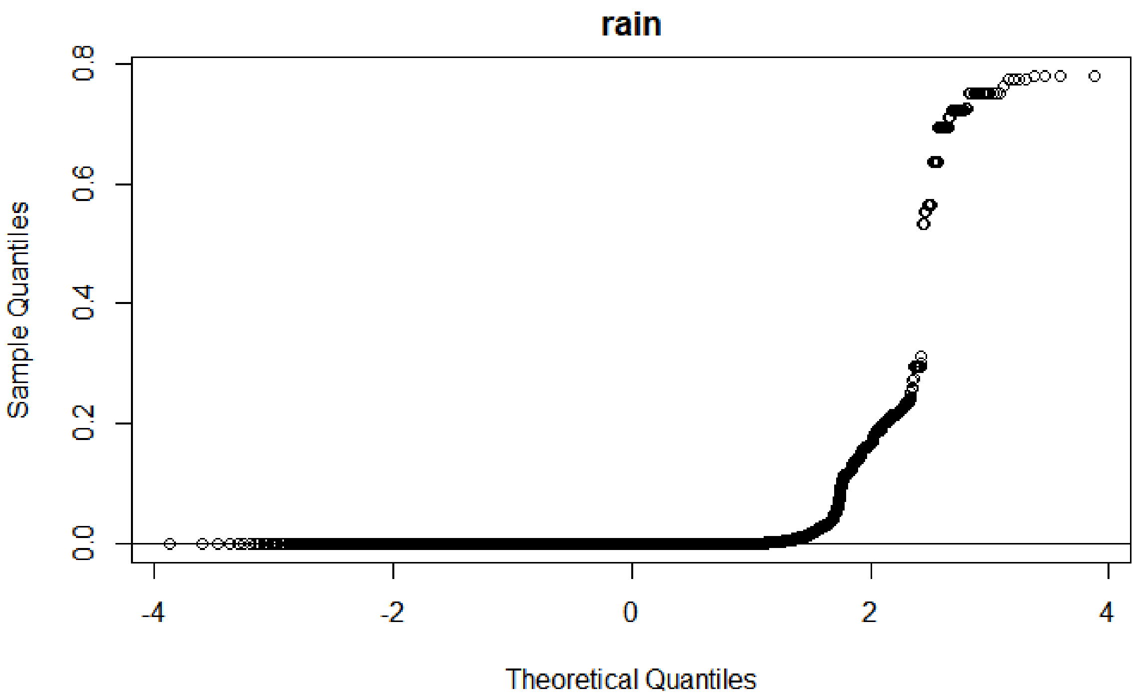
```{r}
#3d scatterplot
library(scatterplot3d)
s3d <- scatterplot3d(x$clouds,x$humidity,x$price,pch=c(1,16)[x$price],
                     xlab="Clouds", ylab="Humidity", angle=45,zlab="Price",
                     lty.hide=2,type="h",y.margin.add=0.1,font.axis=2,font.lab=2)
```

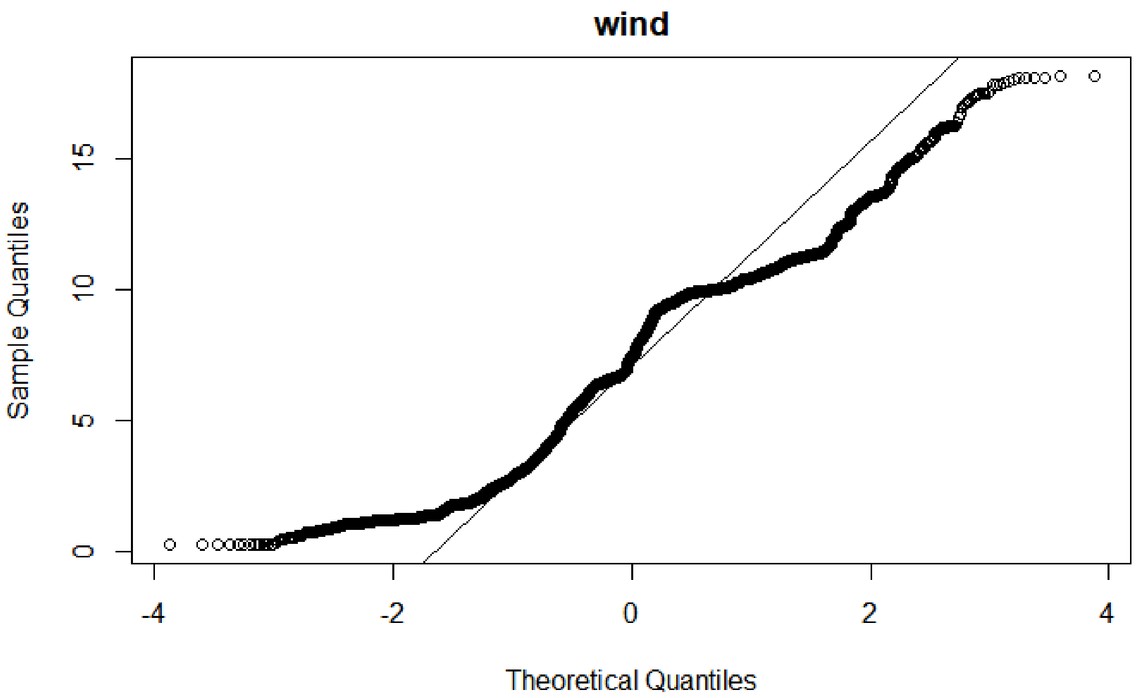


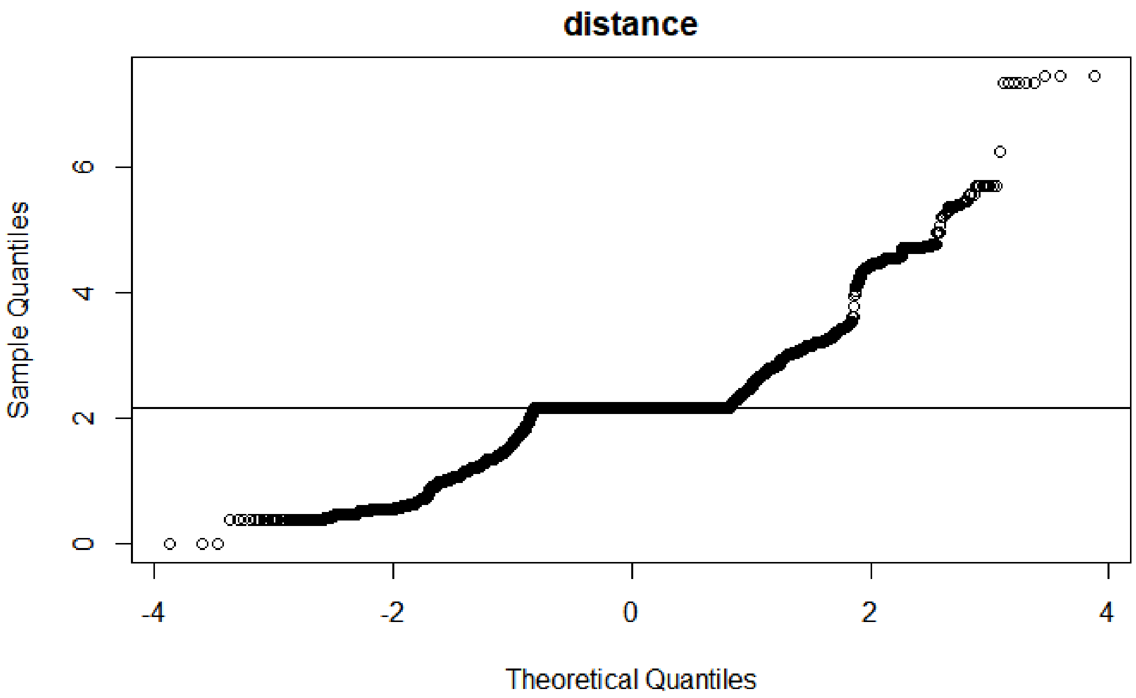
#Scatterplot Matrix

```{r}
library(sciViews)
pairs(x,
      panel=function(x,y,...){
        points(x,y,...)
        abline(lm(y~x),col="grey")
      }, diag.panel=panel.boxplot, pch=".",cex=1.5)
```


clouds
pressure
rain
humidity
wind
distance
surge_multiplier
temp
price
990 1010 1030
0.5 0.7 0.9
0 2 4 6
20 30 40 50
0.0 0.6
990 1020
0.0 0.4 0.8
0.5 0.8
0 10
0 4
1.0 1.6
20 40
0 40 80
0.0 0.4 0.8
0.0 0.2 0.4 0.6 0.8
0 5 10 15
1.0 1.4 1.8
0 20 40 60 80


rain
Sample Quantiles
0.0
0.2
0.4
0.6
0.8
-4
-2
0
2
4
Theoretical Quantiles


wind
Sample Quantiles
0
5
10
15
-4
-2
0
2
4
Theoretical Quantiles


distance
Sample Quantiles
0
2
4
6
-4
-2
0
2
4
Theoretical Quantiles