The original data set containing 1516 crypto currencies is too large for R to process it. So in Python we decrease the dimensionality , using the 5 currencies (smallerDataSet.csv).

Reading the data set:

> coin.df <- read.csv(paste("smallerDataSet.csv"),sep=",")

> dim(coin.df)

[1] 6329 13

List of all the crypto currencies:

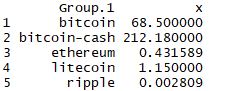
> table(coin.df$slug)

bitcoin bitcoin-cash ethereum litecoin ripple

1761 214 930 1761 1663

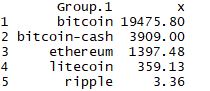
Minimum open value of all the currencies:

> aggregate(coin.df$open,by=list(coin.df$slug),min)



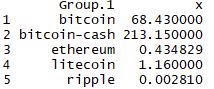
Maximum open value of all the currencies:

> aggregate(coin.df$open,by=list(coin.df$slug),max)



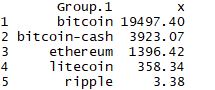
Minimum close value of all the currencies:

> aggregate(coin.df$close,by=list(coin.df$slug),min)



Maximum close value of all the currencies:

> aggregate(coin.df$close,by=list(coin.df$slug),max)



One way contingency table:

> table(coin.df$slug)

bitcoin bitcoin-cash ethereum litecoin ripple

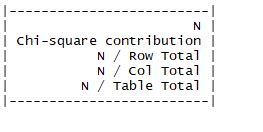
1761 214 930 1761 1663

Two way contingency table:

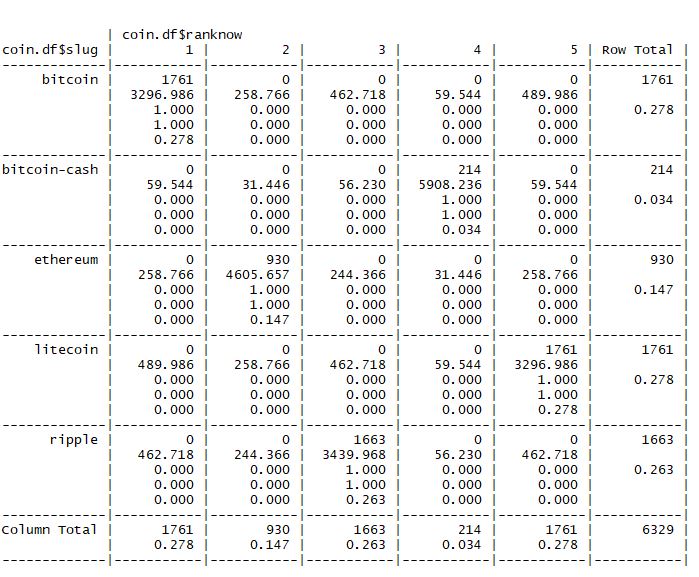
> library(gmodels)

> CrossTable(coin.df$slug,coin.df$ranknow)

Cell Contents

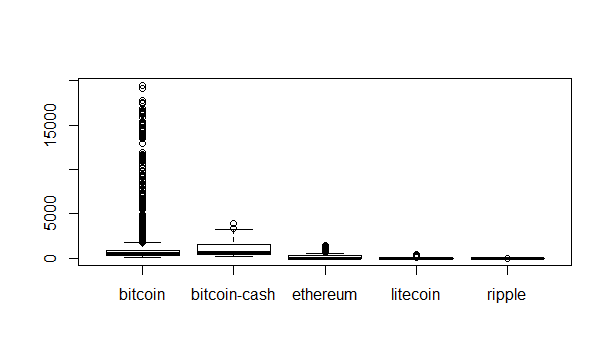


Total Observations in Table: 6329



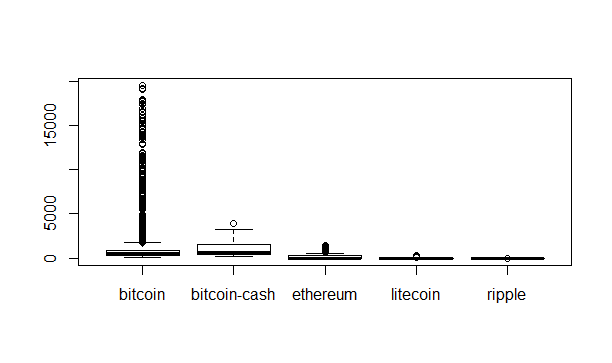
Boxplot of open prices of currencies:

> boxplot(coin.df$open~coin.df$slug)



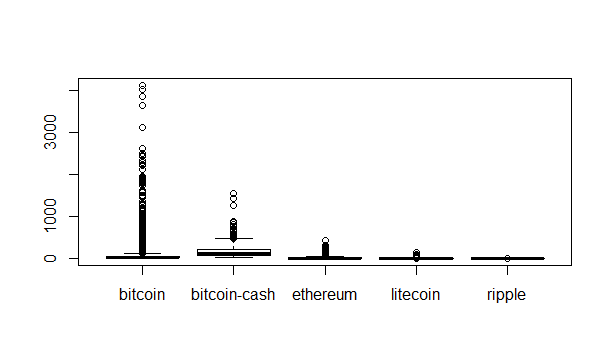
Boxplot of closing prices of currencies:

> boxplot(coin.df$close~coin.df$slug)



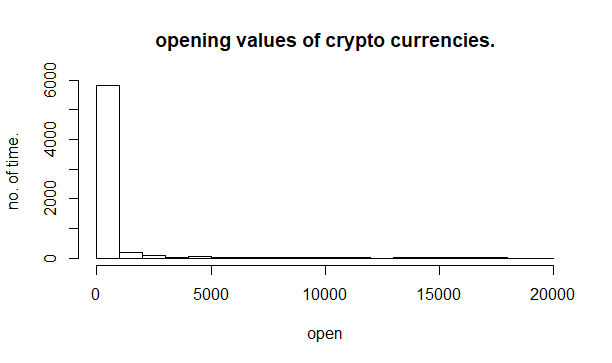
Let’s have a look on the difference between high and low of the day for a crypto coin:

> boxplot(coin.df$spread~coin.df$slug)

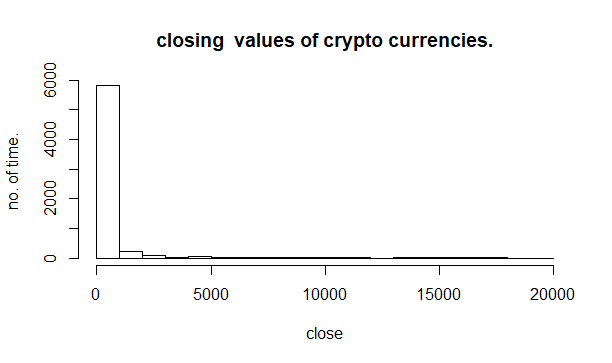


Histograms for suitable data fields:

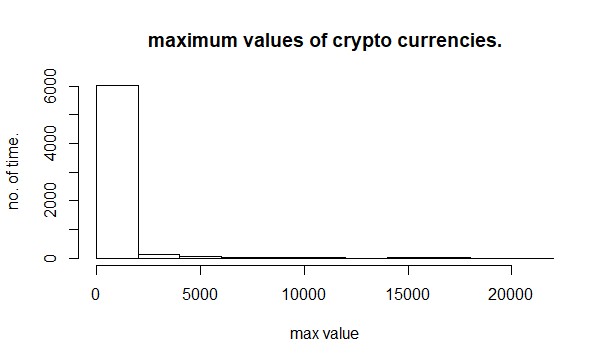
> hist(coin.df$open,main="opening values of crypto currencies.",xlab="open",ylab = "no. of time.")



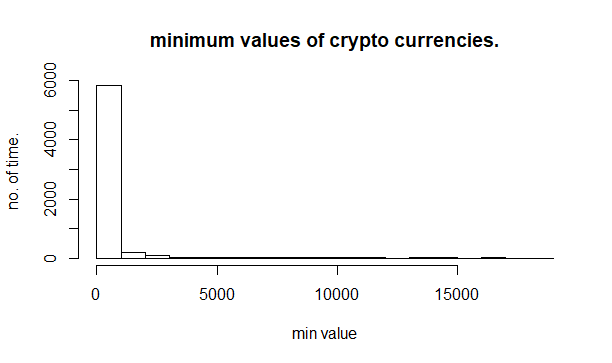
> hist(coin.df$close,main="closing values of crypto currencies.",xlab="close",ylab = "no. of time.")



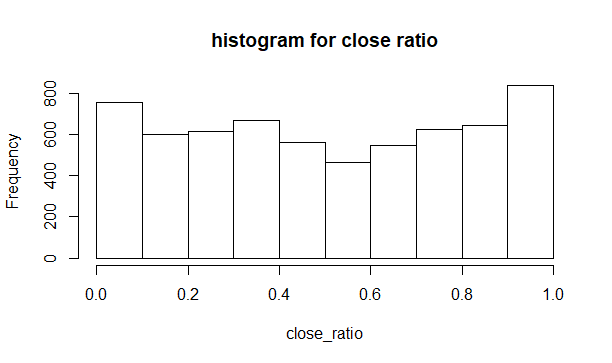
> hist(coin.df$high,main="maximum values of crypto currencies.",xlab="max value",ylab = "no. of time.")



> hist(coin.df$low,main="minimum values of crypto currencies.",xlab="min value",ylab = "no. of time.")



> hist(coin.df$close\_ratio,main = "histogram for close ratio",xlab="close\_ratio")



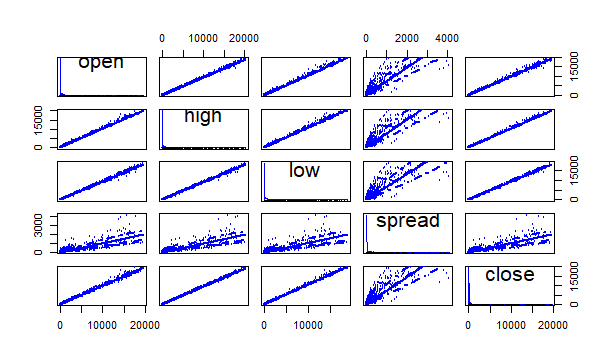
We can see the there is no significant differences between the opening and closing prices of crypto currencies.

Scatter plot:

> library(car)

Loading required package: carData

> scatterplotMatrix(formula= ~open+high+low+spread+close, data=coin.df,diagonal="histogram",cex=0.1)

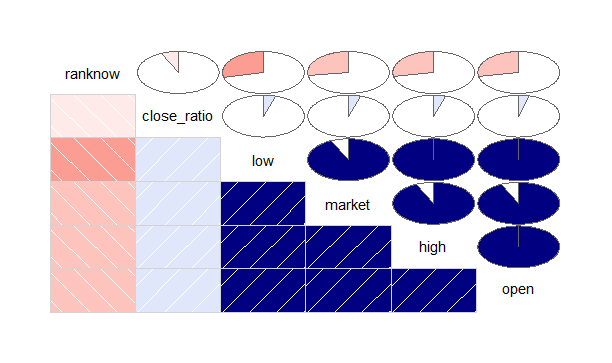


Correlation matrix:

> install.packages("corrgram")

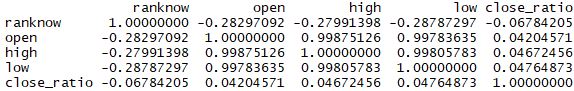
> library(corrgram)

> corrgram(coin.df[,c(5,6,7,8,11,12)],order = TRUE, lower.panel=panel.shade,upper.panel=panel.pie, text.panel=panel.txt)



Generating correlation matrix :

> cor(coin.df[,c(5,6,7,8,12)])



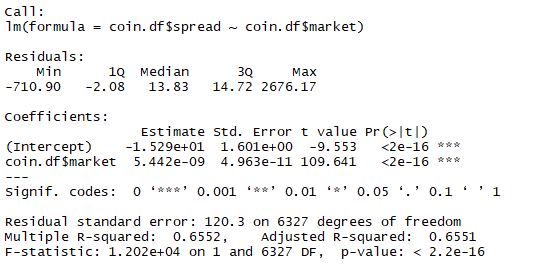
**Model**

we proposed the following null hypothesis :

null hypothesis : there is no significant relation between change in daily price and number of corresponding coins in market.

> fit <- lm(coin.df$spread~ coin.df$market)

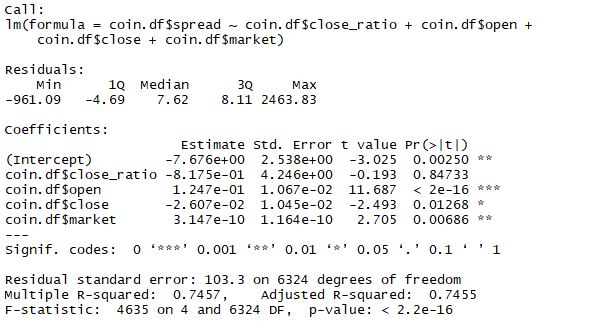
> summary(fit)



We can see that since the p-value is less than 0.05 we can reject the null hypothesis. This means that the change in price of crypto-currencies is highly correlated on the number of crypto currencies floating in the market.

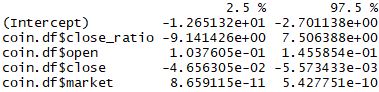
> fit1 <- lm(coin.df$spread~ coin.df$close\_ratio+coin.df$open+coin.df$close+coin.df$market)

> summary(fit1)



Getting confidence interval :

> confint(fit1)



Getting coefficients :

> coefficients(fit1)

con2.JPG

**Results**

From the above coefficients we can formulate the regression model for the given data , “spread” being the dependent variable and rest being the independent variable.

**spread = -7.67 -0.8175 x close\_ratio + 0.125x open -0.026 x close + 3.15 x x10^-10 x market**

**Conclusion**

-we can see from the given data set that price of crypto currencies are highly dependent on the volume floating in the market.

-the price of crypto currencies is also dependent on the opening and closing of the price of the day.

> summary(coin.df)

