Marcus Crowder

Data Mining 3920

Homework Assignment 2

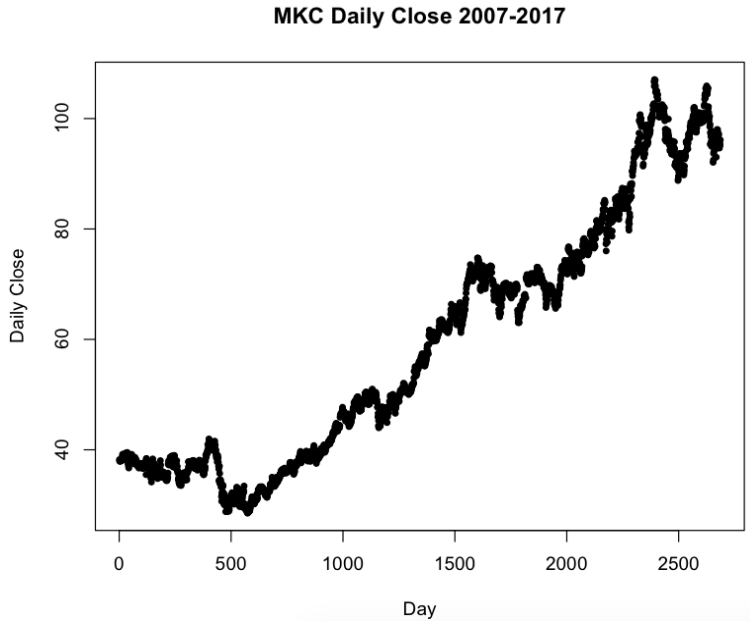
1. My First name, Middle name and last name can be abbreviated to spell out MKC. I was fortunate enough to find a company with a stock symbol directly corresponding to my initials. The name of the company is McCormick & Company a fortune 1000 company that manufactures spices, herbs and flavorings for a wide variety of markets. I was satisfied because this is a Company whose products I am familiar with. This company basically helped raised me. I grew up in a Household where McCormick black pepper and Chili Powder were always at hand. Basically, I am a big fan of their spices and while Researching them I came across other spices and products I would love to try at home. Specifically the Italian Seasoning and Crushed Red Pepper mixed with eggs.

2. Download your data as instructed and import it into R. Delete any rows with missing values. Then, show

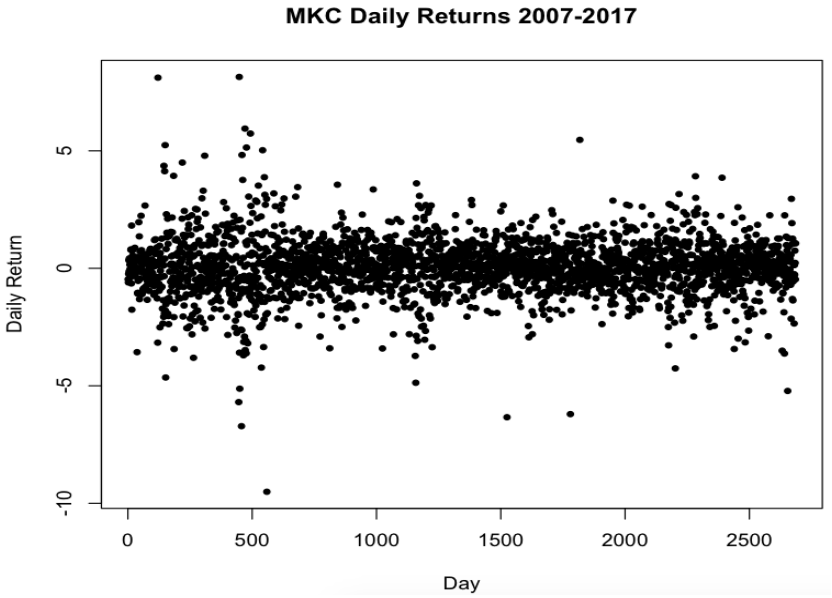
(a) The time series plot of your closing daily prices, showing proper titles;

```
> MKC = MKC[-1, ]
> MKC[1:4, ]
    Date  Open  High   Low Close Adj.Close Volume    MKC_ret LGF_range
2 1/4/07 38.42 38.49 38.00 38.22  30.16636 585200 -0.4687310  0.490002
3 1/5/07 38.19 38.39 37.91 37.99  29.98484 464800 -0.6017464  0.479999
4 1/8/07 37.94 38.05 37.44 37.90  29.91379 683600 -0.2369331  0.610000
5 1/9/07 38.01 38.16 37.93 38.03  30.01640 595100  0.3429923  0.230000
```

a. This is the Time Series plot of MKC between the years 2007 to 2017 Using the R plot function commands.



```
> plot(MKC$MKC_ret, main="MKC Daily Returns 2007-2017", xlab="Day", ylab="Daily Return", pch=20)
> plot(MKC$Close, main="MKC Daily Close 2007-2017", xlab="Day", ylab="Daily Close", pch=20)
```

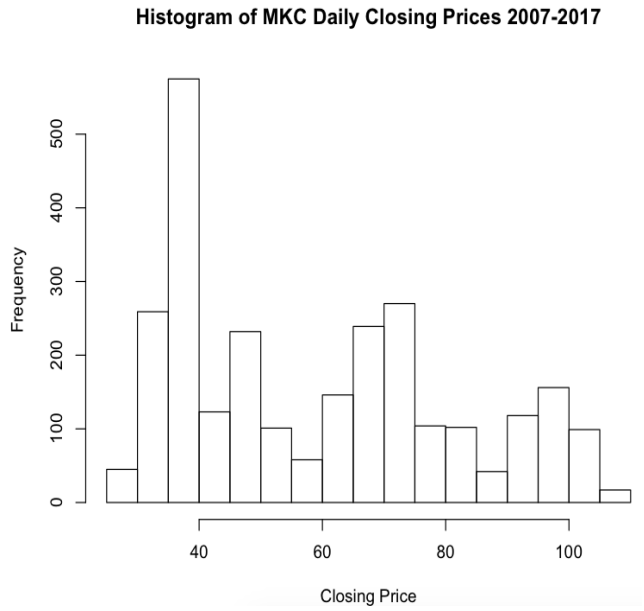


(b) This is a time series plot of my daily return using the above R command.

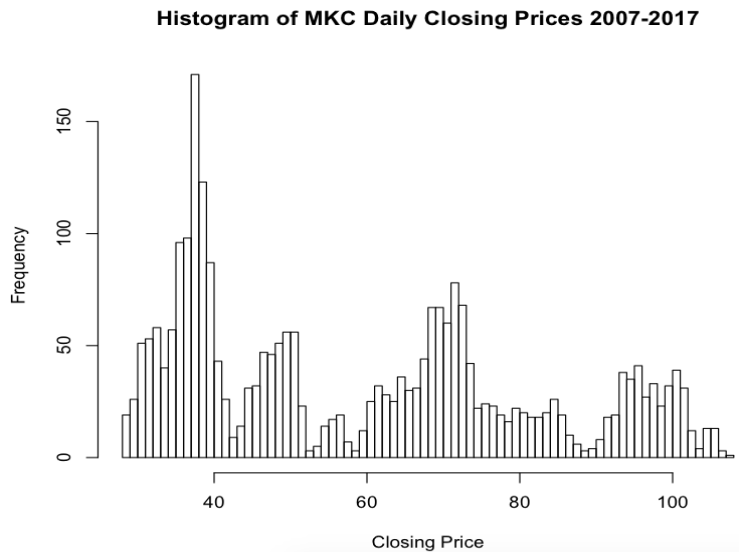
(c) By looking at these two plots there seems to be a major difference in stability and change. Over time Daily returns seem to stay around zero with a straight line and variations. On the other hand the plot info for daily close seems to be continually rising and falling with no real pattern. (except for the rise in price for the long term between 2007-2017).

(d) Is your daily percentage return data coming from a source with a stable population standard deviation? The standard deviation of the daily returns seems to shift over time. For instance the scatter of plots for MKC Daily returns seems unpredictable around day 500 compared to day 1000 which has less variances in Daily Returns.

3. Show
   (a) Histogram of Closing Daily prices with Default Bins



b. R histogram of closing daily price but with 100 bins instead of default

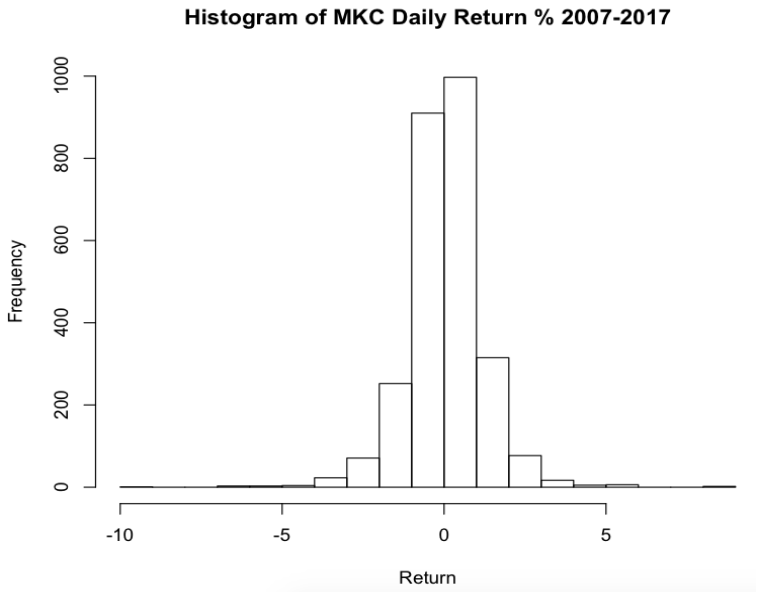


These were the commands used for creating both Histograms.

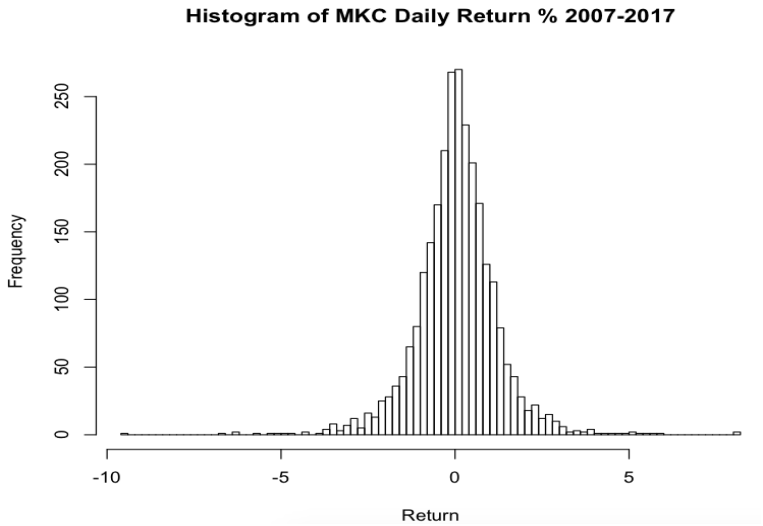
```
> hist(MKC$Close,main ="Histogram of MKC Daily Closing Prices 2007-2017", xlab= "Closing Price")
> hist(MKC$Close, breaks=100, main ="Histogram of MKC Daily Closing Prices 2007-2017", xlab= "Closing Price")
```

c. Using more Bins does not show that the data contains stable features an argument can be even made for the opposite. By using more bins you can see the distribution has a lot of ups and downs with no real shape.

4. (a). R histogram of daily returns with default breaks



b. R histogram of daily returns with 100 breaks.

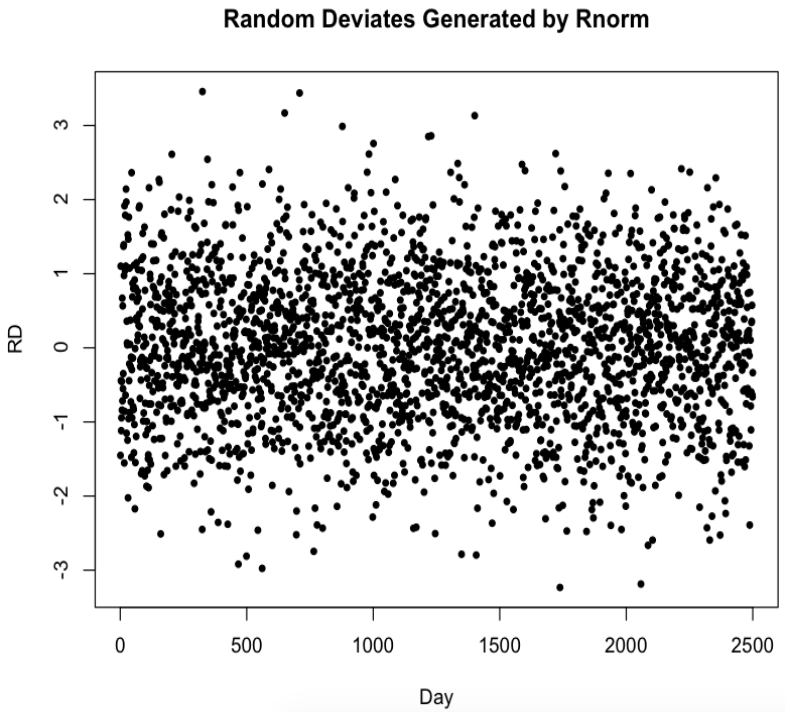


R Commands used to generate Histograms.

```
> hist(MKC$MKC_ret,main ="Histogram of MKC Daily Return % 2007-2017", xlab= "Return")
> hist(MKC$MKC_ret, breaks = 100, main ="Histogram of MKC Daily Return % 2007-2017", xlab= "Return")
```

c. Increasing the number of bins does show signs of stable data. The Bell-curve is solid with a long tail on the left.

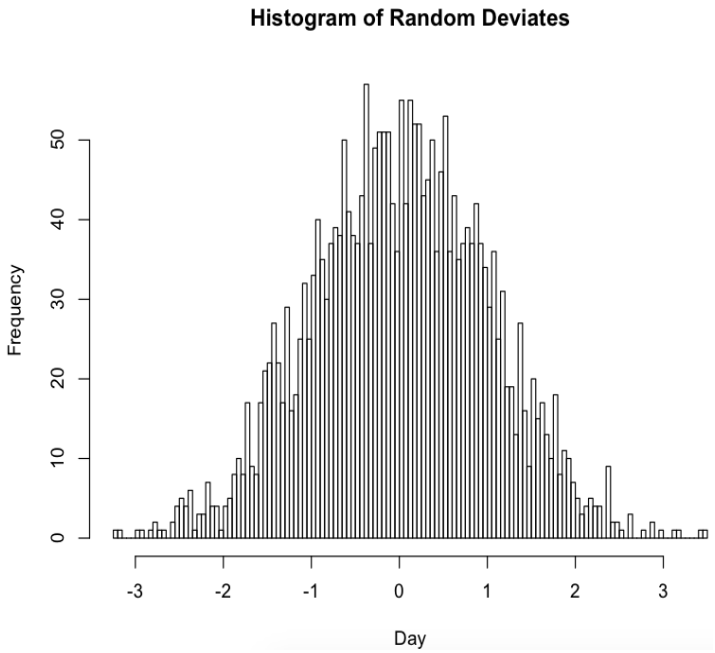
4. Stock returns are often modeled in Finance as coming from an underlying Normal distribution. Generate 2500 standard Normal deviates.



a. Generation of 2500 standard Normal Deviates.
b. Comparing this model to the time series model of daily returns in question 2 I can understand why people use the Standard Normal Distribution to model Stock returns. Both seem to congregate around 0 and in between 1 to -1. However the Normal distribution does have less variation but still great to use.

c. Histogram with 100 breaks using R.

hist(RD, breaks = 100, main ="Histogram of Random Deviates", xlab= "Day")



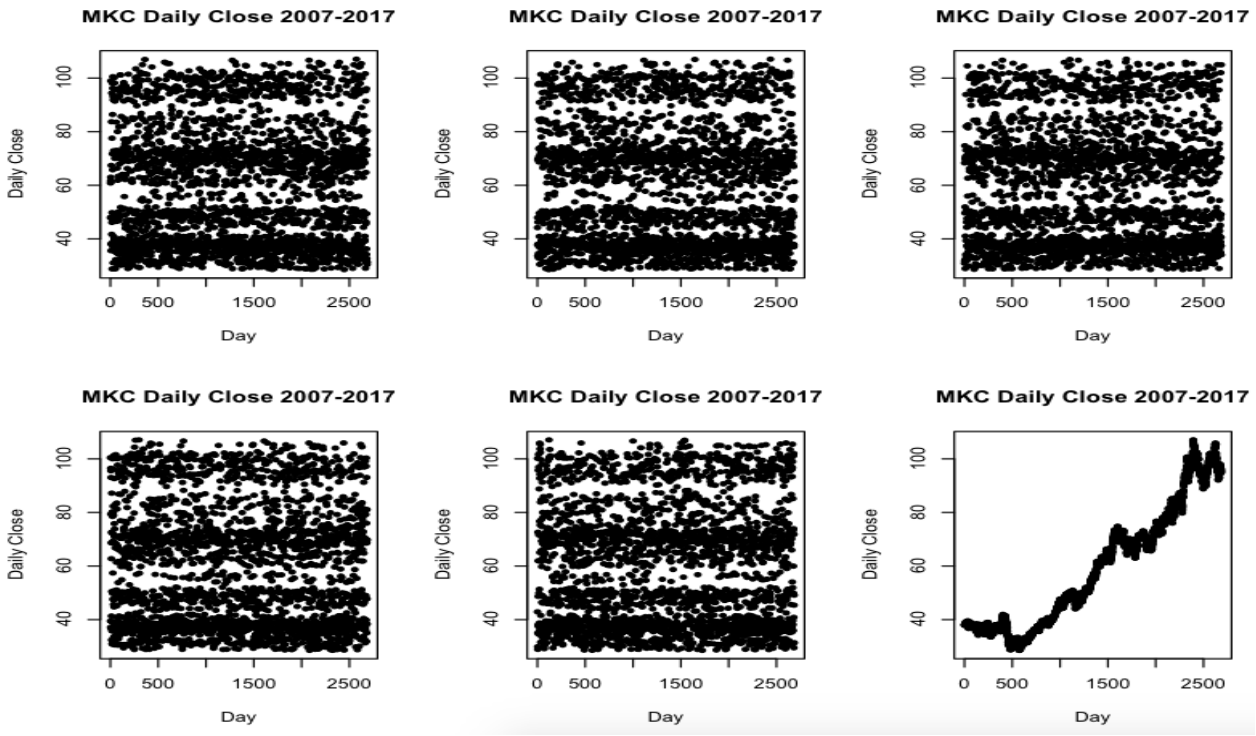
R commands used to generate Histograms.

```
> RD = rnorm(2500)
> plot(RD, main="Random Deviates Generated by Rnorm", xlab="Day", pch=20)
> hist(RD, breaks = 100, main ="Histogram of Random Deviates", xlab= "Day")
>
```

D. Comparing the shape of both histograms It is noticeable that the shape of the Normal distribution is more stable than our Histogram for returns. The frequencies of the Normal distribution histogram seem to be spread out evenly. While our return histogram does not display the same equality across the center or mean. Both histograms do however contain the similar shape which makes it a great model for comparisons.

5. One way to test for stability of a time series is to randomize the order of the data several times, say five times. Then, put those five plots together on one page and insert the original time series plot as a sixth. Then, ask people to try to pick which is "different" from the other five. If people can do that, then the time series was not stable! Try this for (a) the closing prices and (b) the percentage change. You ask, "But how do I randomize the order of my data?" The answer is use the "sample" function, our favorite. The command y = sample(IBM$Close) will produce a randomized series. Repeating the command will generate a different randomized order.

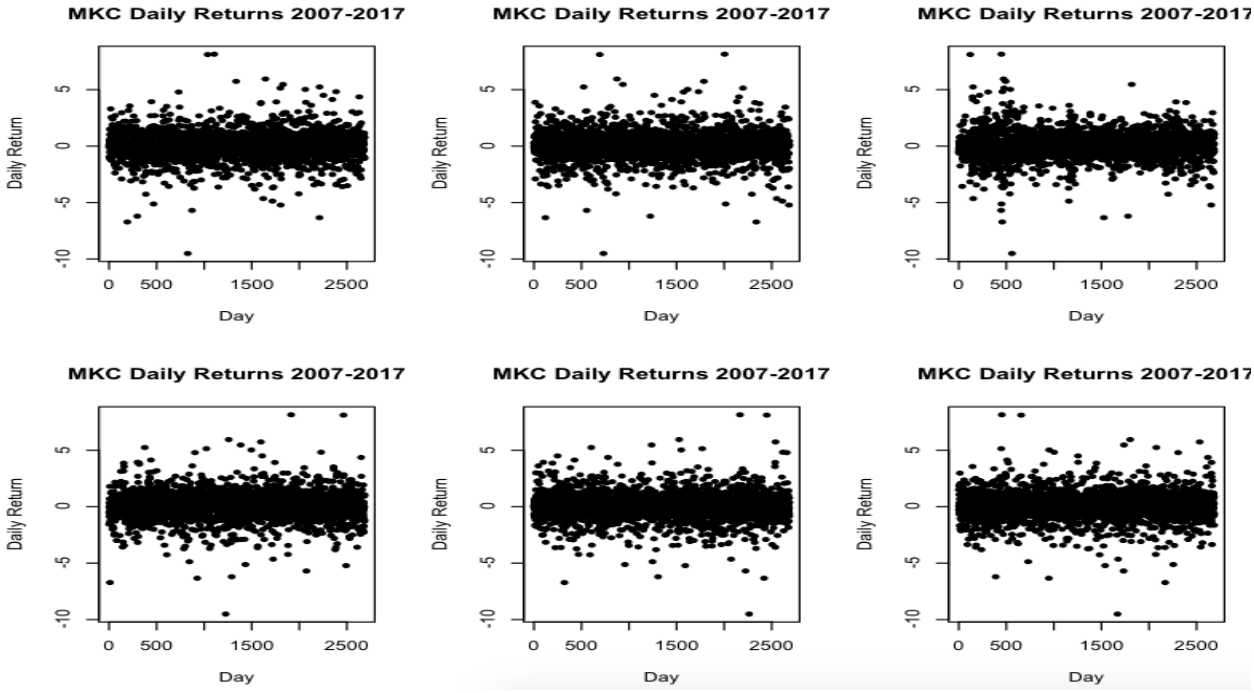
   a. This was a fun experiment because I got to use my nephew to help with my Homework. In the first experiment with 5 random samples from daily close and the original he quickly pointed out the difference. Which is obvious and shows the time series was not stable.



commands used

```
> plot(sample(MKC$Close), main="MKC Daily Close 2007-2017", xlab="Day", ylab="Daily Close", pch=20)
> par(mfrow=c(2,3))
> plot(sample(MKC$Close), main="MKC Daily Close 2007-2017", xlab="Day", ylab="Daily Close", pch=20)
> plot(sample(MKC$Close), main="MKC Daily Close 2007-2017", xlab="Day", ylab="Daily Close", pch=20)
> plot(sample(MKC$Close), main="MKC Daily Close 2007-2017", xlab="Day", ylab="Daily Close", pch=20)
> plot(sample(MKC$Close), main="MKC Daily Close 2007-2017", xlab="Day", ylab="Daily Close", pch=20)
> plot(sample(MKC$Close), main="MKC Daily Close 2007-2017", xlab="Day", ylab="Daily Close", pch=20)
> plot(MKC$Close, main="MKC Daily Close 2007-2017", xlab="Day", ylab="Daily Close", pch=20)
```

b. The second experiment was a lot more fun because I had my nephew exit the room multiple times and come back in. I told him I switched which one was different (but didn't). Each time he came into the room he had a different answer to the question. Which implies that the time series is stable. Special thanks to my sharp 6 year old nephew Xavier.

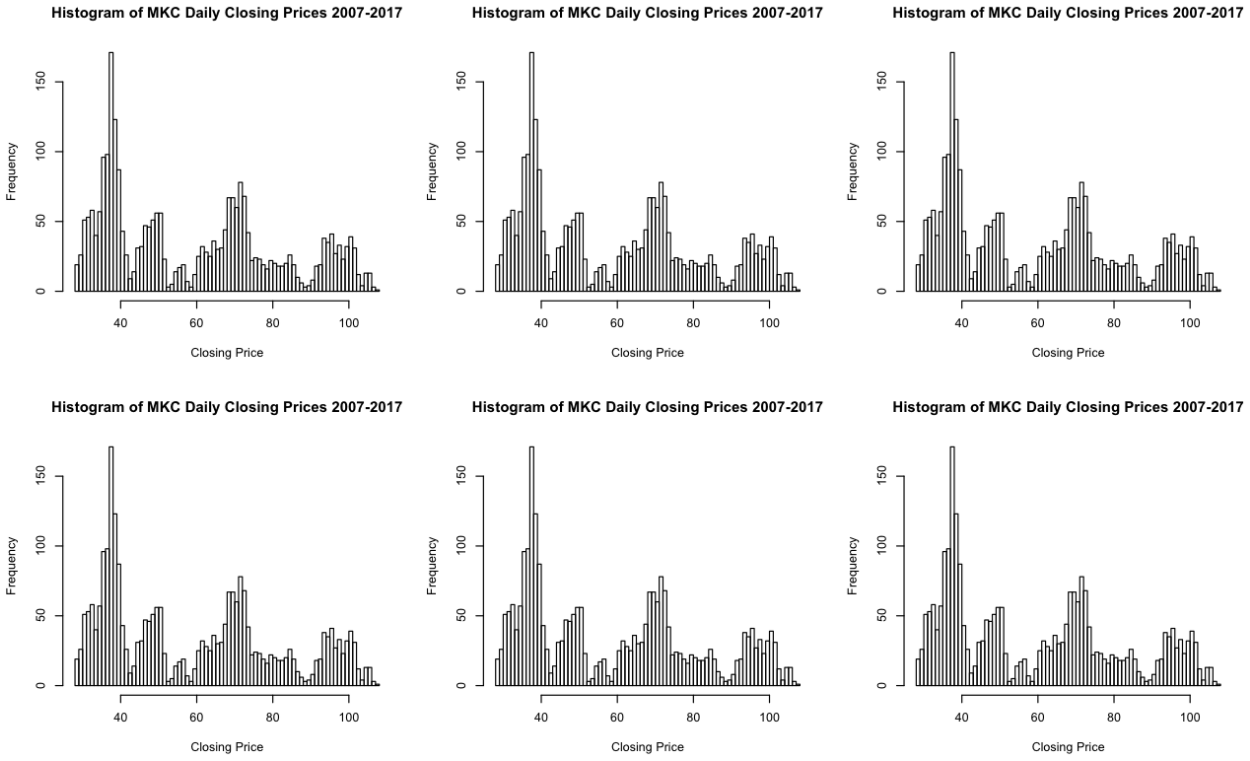


commands used.

```
> par(mfrow=c(2,3))
> plot(sample(MKC$MKC_ret), main="MKC Daily Returns 2007-2017", xlab="Day", ylab="Daily Return", pch=20)
> plot(sample(MKC$MKC_ret), main="MKC Daily Returns 2007-2017", xlab="Day", ylab="Daily Return", pch=20)
> plot(MKC$MKC_ret, main="MKC Daily Returns 2007-2017", xlab="Day", ylab="Daily Return", pch=20)
> plot(sample(MKC$MKC_ret), main="MKC Daily Returns 2007-2017", xlab="Day", ylab="Daily Return", pch=20)
> plot(sample(MKC$MKC_ret), main="MKC Daily Returns 2007-2017", xlab="Day", ylab="Daily Return", pch=20)
> plot(sample(MKC$MKC_ret), main="MKC Daily Returns 2007-2017", xlab="Day", ylab="Daily Return", pch=20)
>
```

6. (a) Would randomizing the order of a time series change the histogram? Change the sample average? The sample standard deviation?

A. By sampling the Histogram of MKC daily returns I noticed no difference in the sample average or sample standard deviation. All of the graphs look the same.

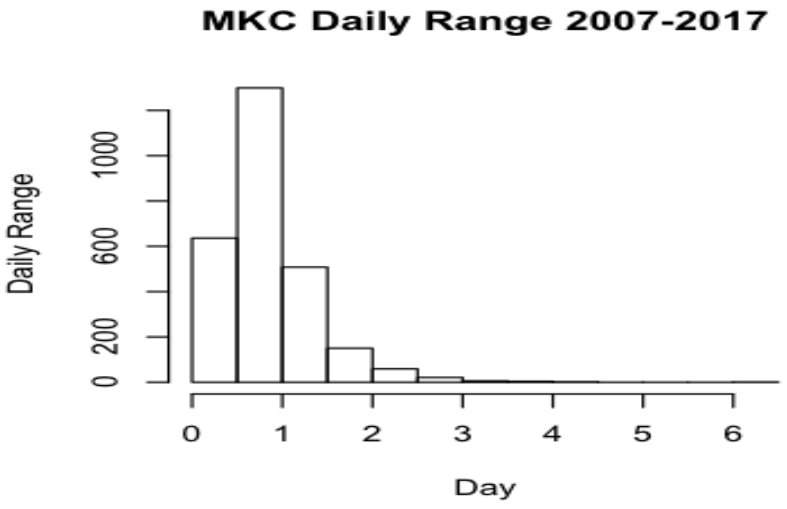


commands used:

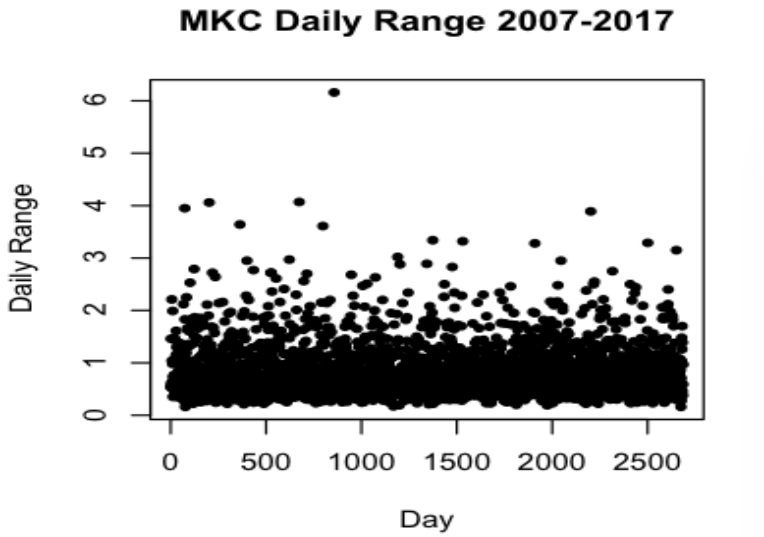
```
> par(mfrow=c(2,3))
> hist(MKC$Close, breaks=100, main ="Histogram of MKC Daily Closing Prices 2007-2017", xlab= "Closing Price")
> hist(sample(MKC$Close), breaks=100, main ="Histogram of MKC Daily Closing Prices 2007-2017", xlab= "Closing Price")
> hist(sample(MKC$Close), breaks=100, main ="Histogram of MKC Daily Closing Prices 2007-2017", xlab= "Closing Price")
> hist(sample(MKC$Close), breaks=100, main ="Histogram of MKC Daily Closing Prices 2007-2017", xlab= "Closing Price")
> hist(sample(MKC$Close), breaks=100, main ="Histogram of MKC Daily Closing Prices 2007-2017", xlab= "Closing Price")
> hist(sample(MKC$Close), breaks=100, main ="Histogram of MKC Daily Closing Prices 2007-2017", xlab= "Closing Price")
```

B. The assumption we are making about time series data when using Histograms is that the order does not really matter because we fit the data into bins that has nothing to do with order. It is the same when computing sample average and standard deviation.

7. When plotting the histogram of the daily range between high and low closing prices the Ranges look stable. There is not too many variations of extreme price changes throughout the day. The Histogram is skewed a bit to the right but nothing extreme. It would be interesting however to learn on which days the graph got skewed the way it did and if It was a result of the market or company.



The same results can be found with the time series graph. There are extreme deviations but the dots are mostly in line and stable.



Commands used to generate graphs

```
> par(mfrow=c(2,2))
> plot(sample(MKC$LGF_range), main="MKC Daily Range 2007-2017", xlab="Day", ylab="Daily Range", pch=20)
> hist(sample(MKC$LGF_range), main="MKC Daily Range 2007-2017", xlab="Day", ylab="Daily Range", pch=20)
>
```