# SUGAR PRICE VOLATILITY

## Sugar Price Volatility

by Matthew Aronowitz

## Abstract

Sugar is one of the most heavily traded commodity in the world. Since the price of sugar is relatively small compared with the finished good, the volatility of this commodity is usually ignored. That being said, in percentage terms, sugar prices are quite volatile. In a given year, it is not uncommon to see prices moves of 30% or more. In this analysis we are going to look at the prices of daily sugar futures contracts traded on The ICE Exchange from 2010-2018. The goal of this exploratory data analysis is to determine if certain variables could help predict future volatility of this commodity.

## Dataset and Preperations, Remove NA's

```
##        date        open        high         low       close        move 
##      "Date"   "numeric"   "numeric"   "numeric"   "numeric"   "numeric" 
##   max_range close_range     average      volume     fut_vol 
##   "numeric"   "numeric"   "numeric"   "numeric"   "numeric"
```

## Summary of Data Set

For each day we get the "date", opening price of that day "open", high price of that day "high", low price of that day "low" and closing price of that day "close". The "move" is the percent change from previous day's close to the current day's close. The "move" could be positive (sugar prices go up) or the "move" could be negative (sugar price went down). The "max_range" represents the maximum movement in absolute terms from the previous day's close to either that day's high or low price. The "close_range" is the absolute value of the "move". The "average" variable represents the mean of the "close_range" and "max_range".

The "fut_vol" is the 20 day forward mean of the "average" variable, then translated into a yearly volatility percentage. This represents the 20 day future volatility of the commodity. The "fut_vol" is the variable we would like to predict as it represents the future volatility of sugar at that moment in time.

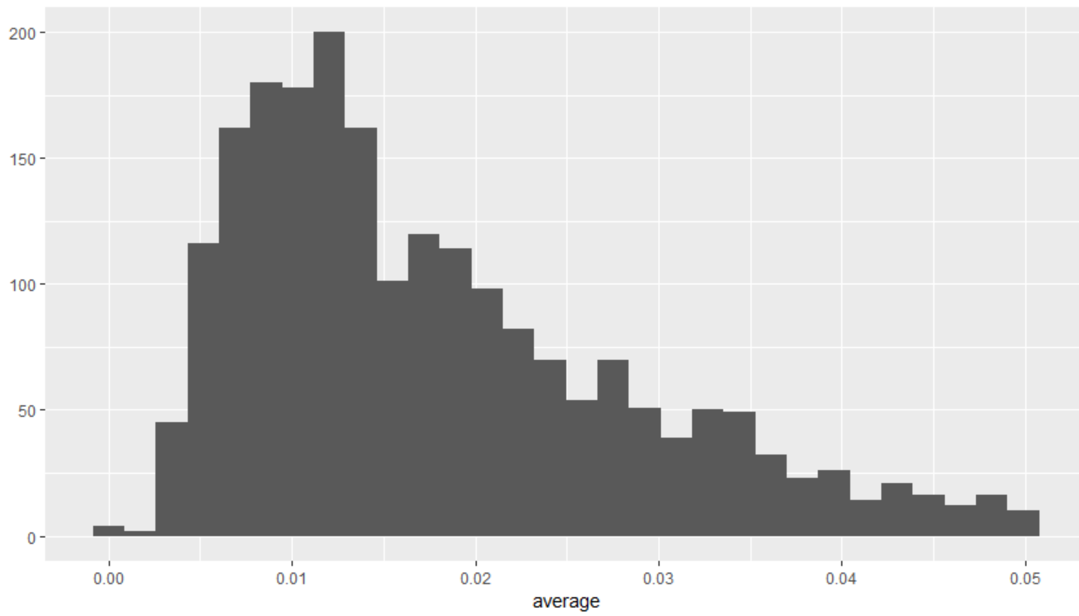
All these prices are in cents per pound or .01/lb, so a value of 12.95 is 12.95 cents per pound.

## First Look Into The Volatility of the Sugar Prices

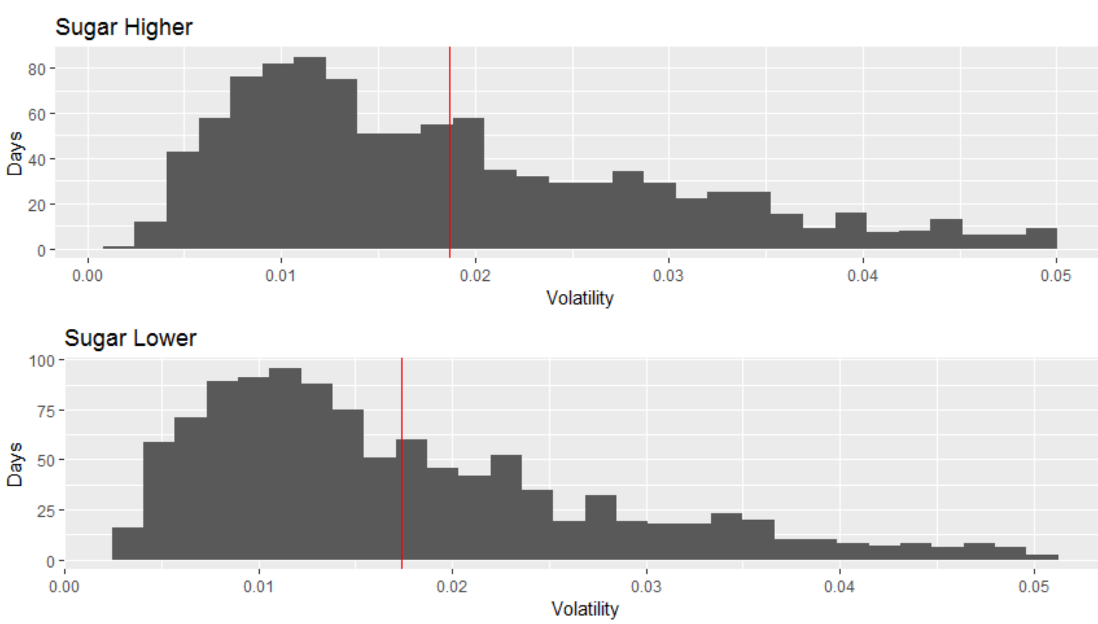
Lets look at a summary of our average daily volatility "average"

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
## 0.00000 0.00990 0.01550 0.01982 0.02550 0.18470
```

Looking at the summary, it seems some of the averages could be bad data as it seems extreme. It could be a result of when the data changes futures contract of reference. Lets subset the data and exclude any “average” above 5% and take a look at the distribution. The “average”" is a good representation of the daily volatility.

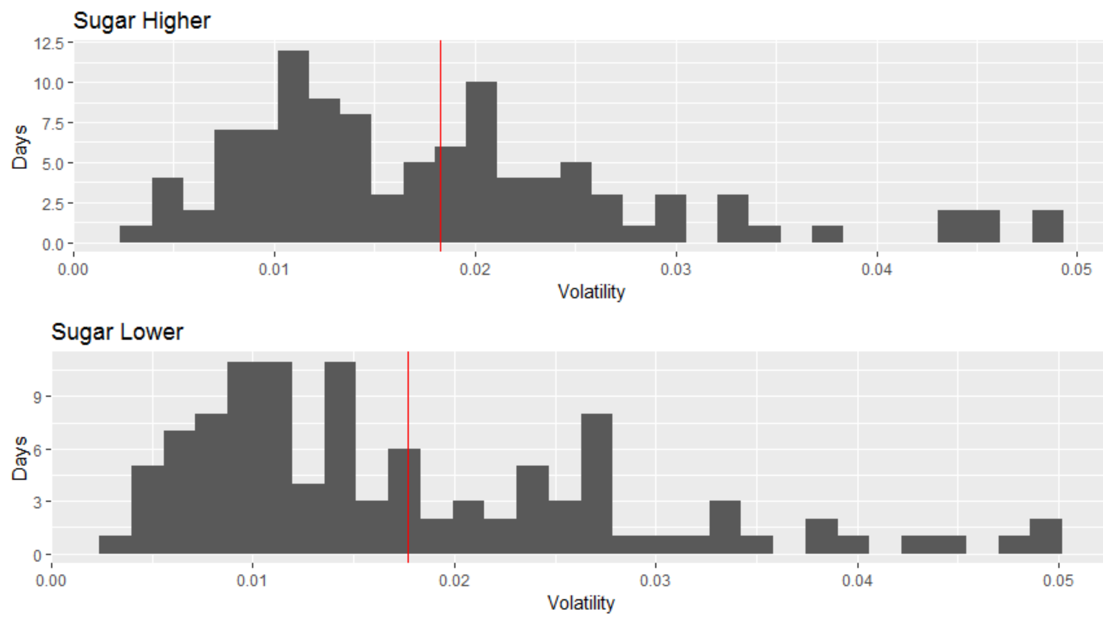



The data definetely looks skewed, which is probably why the options prices usually prices in a skew when looking at the prices in terms of volatility. Lets see if the volatility and skew is more apparent on up days versus down days. Is there more volatility on days when sugar prices go down versus when prices go up.



The red lines represent the mean of the averages. It seems that when sugar goes higher, it is more volatile with a higher average. It also seems to have more skew. Curious if over a more current period, it is still more volatile to the upside than the downside.

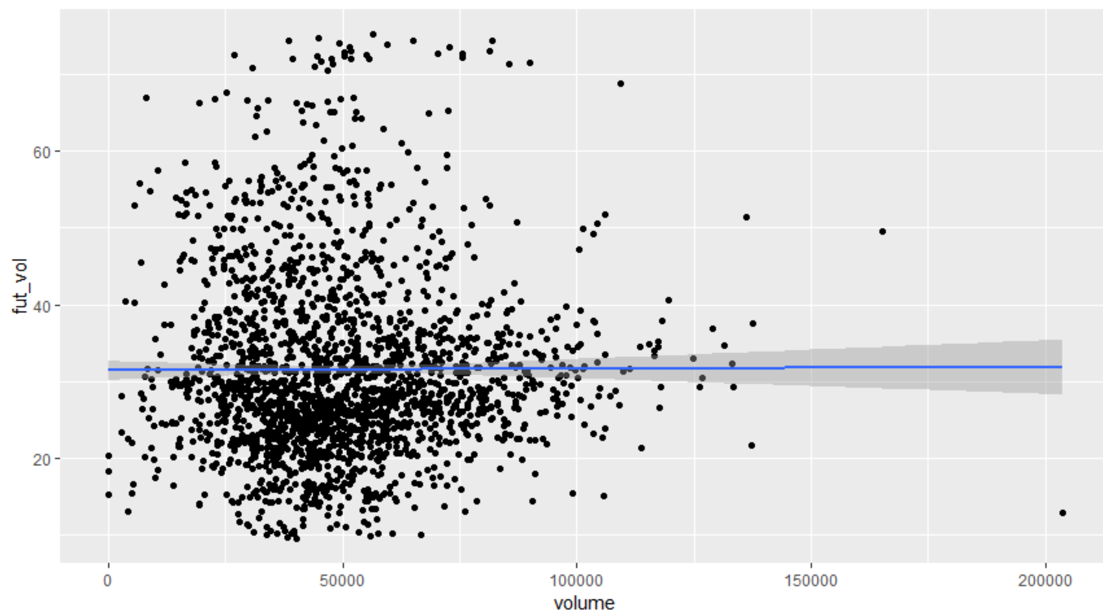
```
## [1] 105

## [1] 105
```




It seems that in the current year, sugar prices continue to be more volatile on averagewhen the market rises.
In addition, the distribution is also more skewed to the right.

## Let’s Examine the Potential Impact of Volume on Sugar Price Volatility

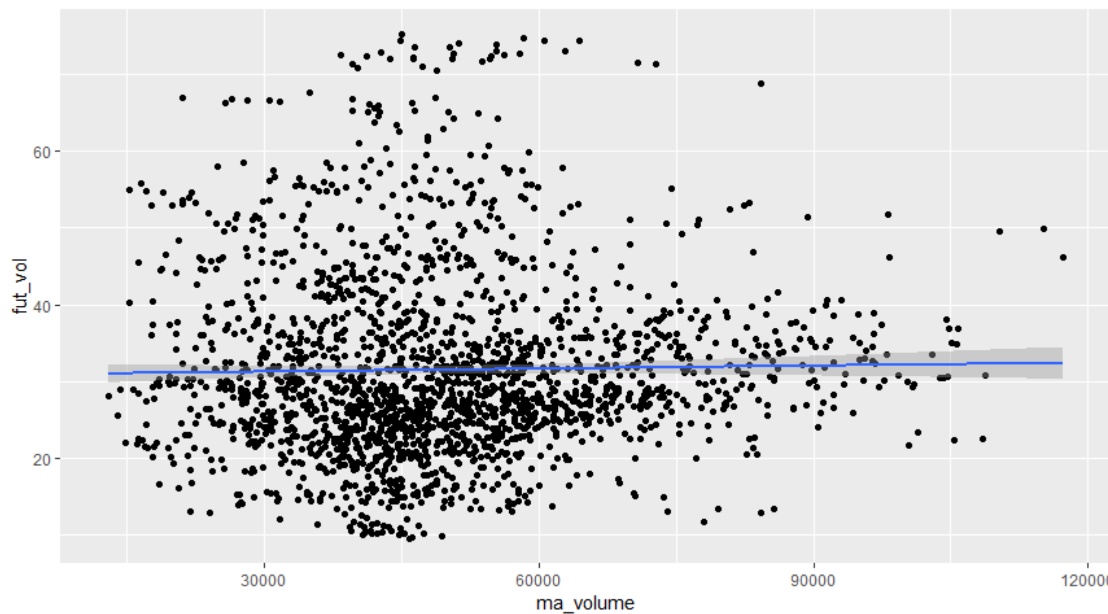
```
## 
##  Pearson's product-moment correlation
## 
## data:  new_sb$fut_vol and new_sb$volume
## t = 0.18764, df = 2115, p-value = 0.8512
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.03852884  0.04667427
## sample estimates:
##         cor 
## 0.004080122
```




Initial indication is that there is no correlation or relationship in future volatility and volume. Maybe if we smoothed the volume data by creating a new column which has the moving average of the volume.

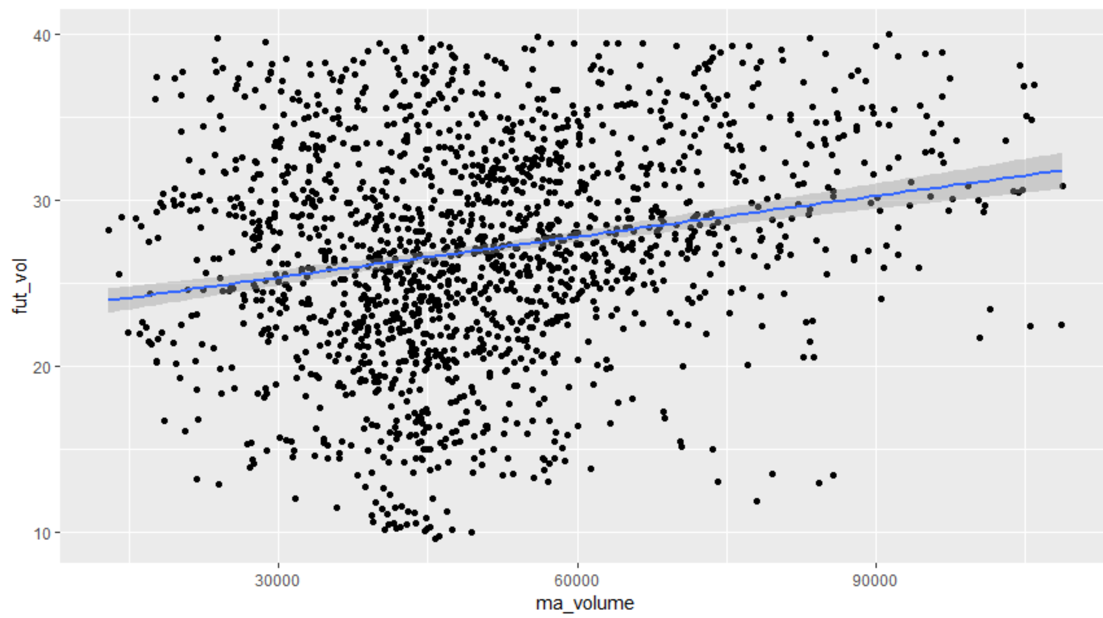
Lets run a scatter plot of the new smoother volume variable against fut_vol and see if there is any correlation.

```
##
##  Pearson's product-moment correlation
##
## data:  new_sb_reverse$fut_vol and new_sb_reverse$ma_volume
## t = 0.86494, df = 2111, p-value = 0.3872
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.02383967  0.06141532
## sample estimates:
##        cor
## 0.01882204
```




At first glance it does not seem like there is much correlation or relationship between volume and future volatility. Some of these fut_vols seem extreme and might. Lets remove some of the more extreme fut_vol data, those above 40 and re-draw the relationship.

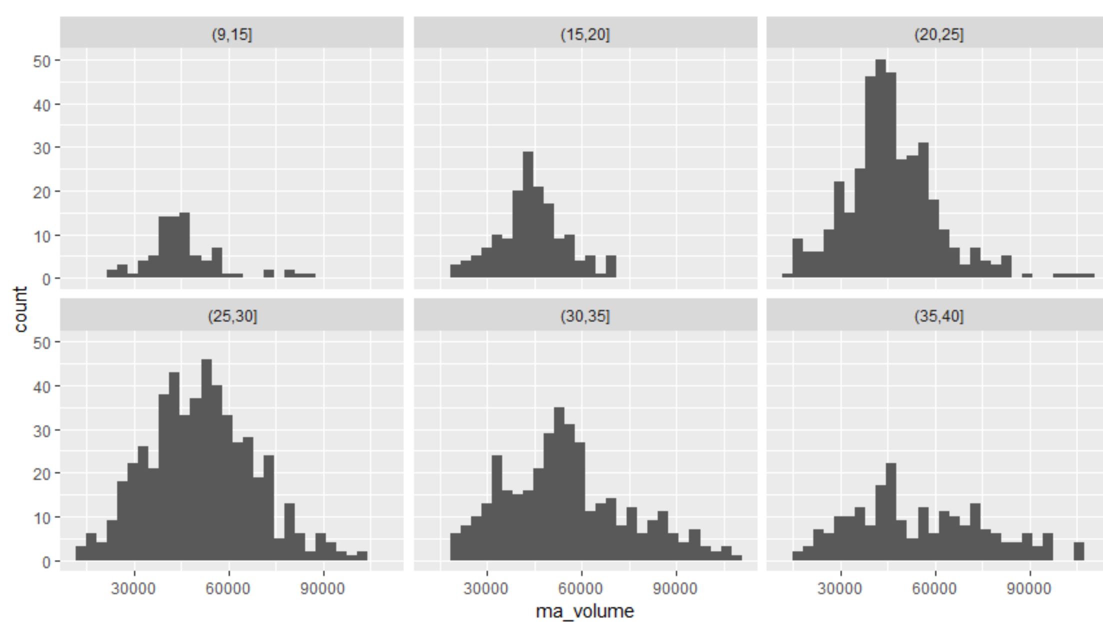
```
##
##  Pearson's product-moment correlation
##
## data:  new_sb_lower_fut_vol$fut_vol and new_sb_lower_fut_vol$ma_volume
## t = 9.1969, df = 1716, p-value < 0.00000000000000022
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1712007 0.2613520
## sample estimates:
##       cor
## 0.2167384
```




Here it shows a positive relationship - so maybe, volumes could be a predictor of fut_vol. Lets create categorical categories by bucketing the fut_vol and volumes so we can quantify some of the relationships between variables.

## Fut_Vol Buckets and Volumes

Set Fut_vol buckets

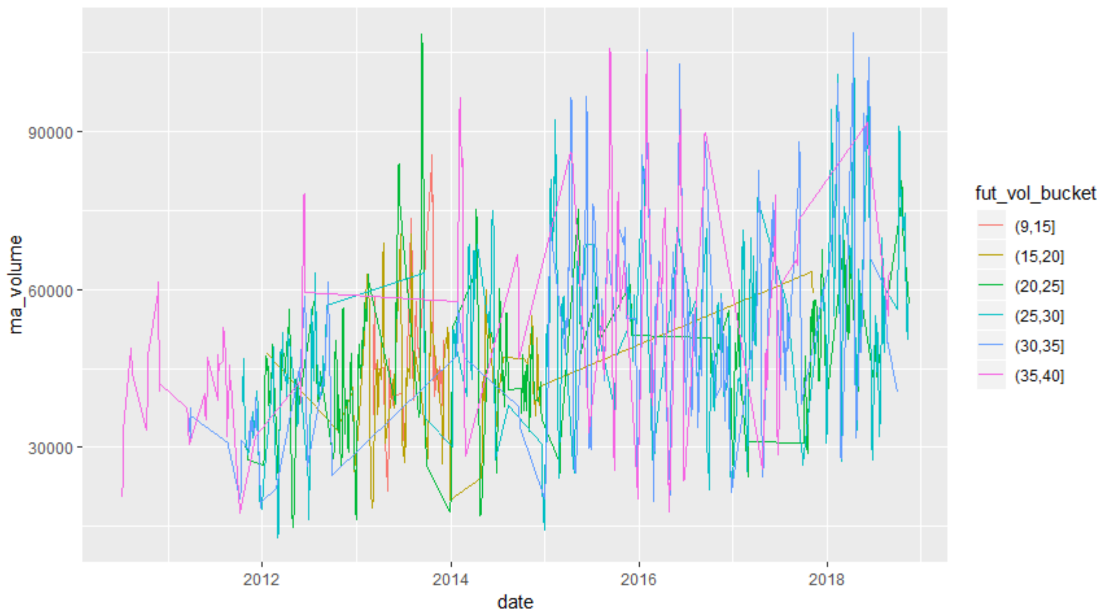



```
##   (9,15]  (15,20]  (20,25]  (25,30]  (30,35]  (35,40]
## 12425.23 10681.86 14659.02 16608.51 19484.92 20985.61
```

```
##    (9,15]  (15,20]  (20,25]  (25,30]  (30,35]  (35,40]
## 45939.82 44174.00 46174.03 51364.73 55093.22 55112.91
```

As we can see, higher volatility buckets are associated with higher rolling means. There are less instances of higher volatilities but the central tendency or mean is higher and more skewed to the right. That being said, the standard deviation of the rolling mean volumes are quite large and therefore this variable by itself might not be statistically significant to predict future volatility.

Lets look at this relationship over time and see if that central tendency of higher median volumes being associated with higher fut_vols is consistent over time.



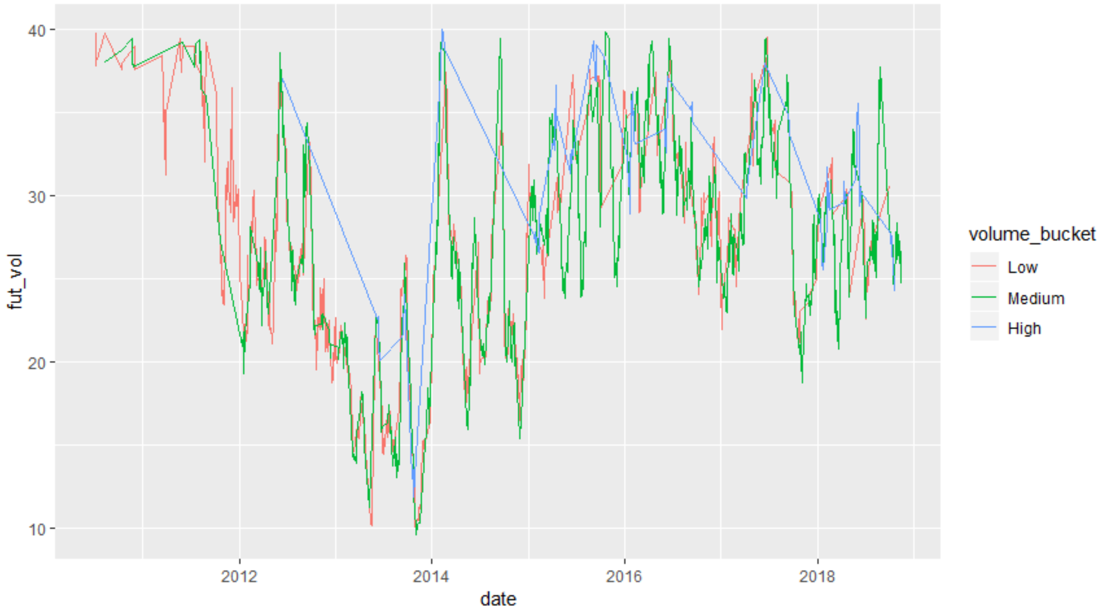

It looks like this chart is all over the place. Cannot determine any clear trend over time with respect to fut_vol and volumes. Maybe it would be more clear if we bucket the rolling_averages.

Lets create a new categorical variable of volume buckets and look at the range of ma_volumes.

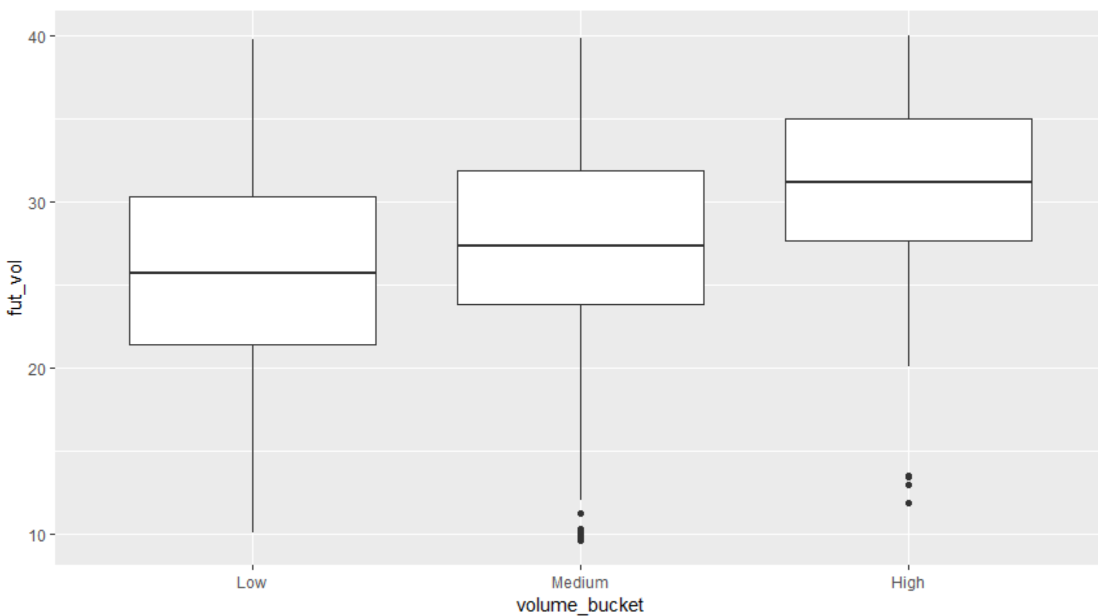
Now lets create the buckets and summarize that data.

```
##
##     Low Medium   High
##     710    866    142
```

Lets graph this new variable as lines over time against fut_vol.




It is hard from this graph to see any real trend over time with respect to volumes. It seems that higher fut_vol are associated with higher volume means but we are unsure of this relationship over time.




Here you can see clearer that higher volume are associated with higher future volatility. Conversely, lower volumes are associated with lower volatilities.

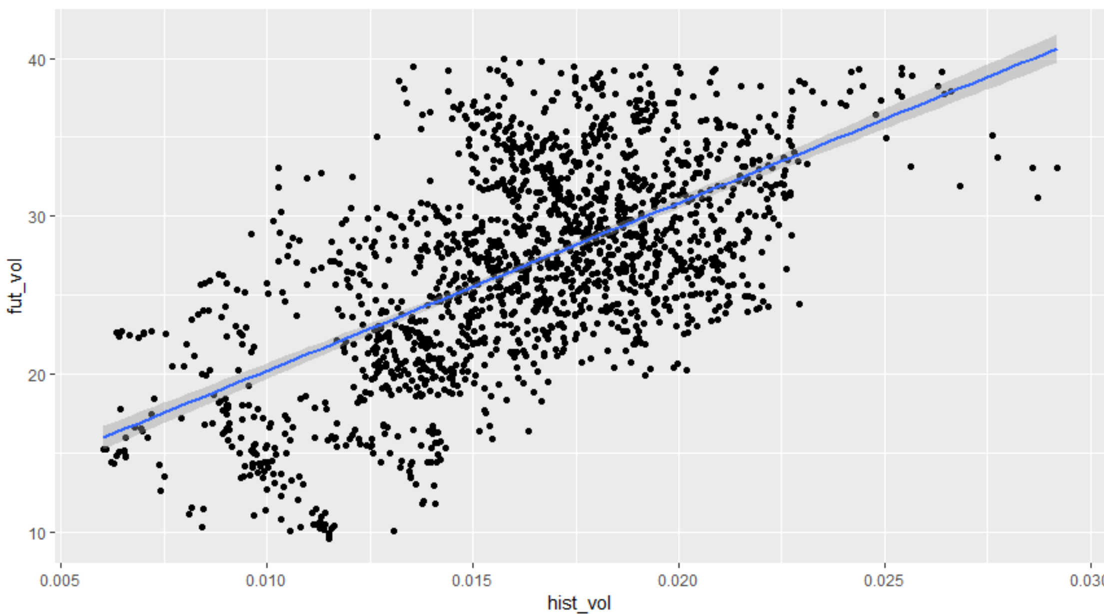
## Past Predicting the Future

In finance, we often look at the past to predict the future. Lets look to see if past volatility could be a good predictor of future volatility.

Averages represent sugar's daily volatility. Lets create a new variable which is the 20 day rolling mean of the average variable.

Let's look at a scatter plot and correlation of this variable against fut_vol.

```
##
##  Pearson's product-moment correlation
##
## data:  new_sb_lower_fut_vol$fut_vol and new_sb_lower_fut_vol$hist_vol
## t = 31.627, df = 1697, p-value < 0.00000000000000022
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5781625 0.6380527
## sample estimates:
##       cor
## 0.6089748
```

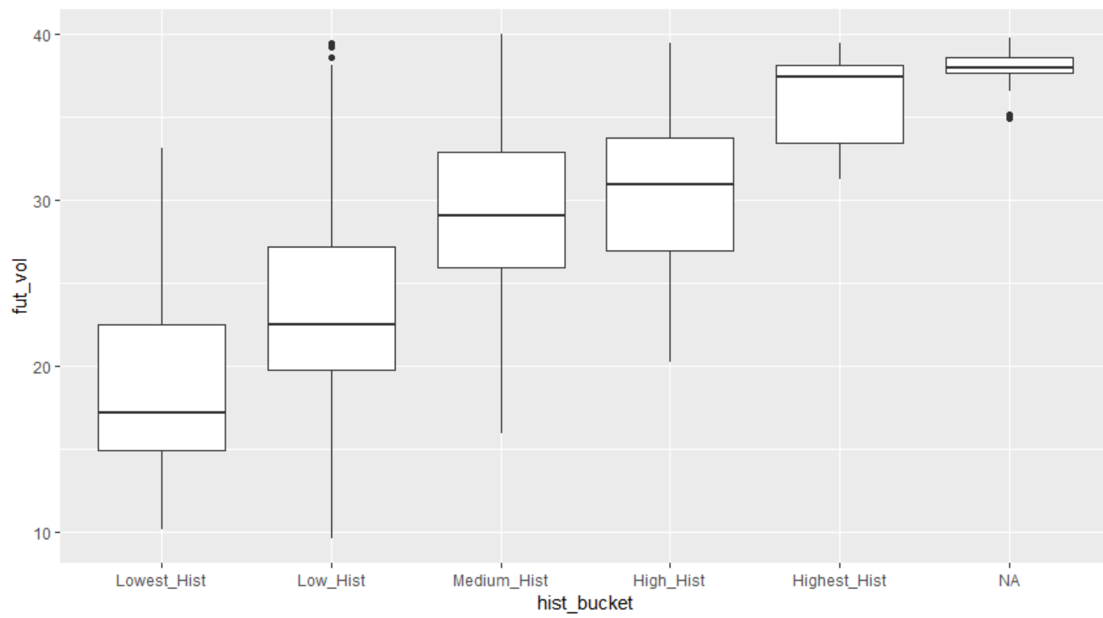



So, there seems to be a relatively large, positive relationship between historical volatility and future volatility. This historical volatility might be a good predictor of future volatility.

Lets look at the historical volatility means across fut_vol buckets

```
##     (9,15]    (15,20]    (20,25]    (25,30]    (30,35]    (35,40]
## 0.01077750 0.01194242 0.01516965 0.01694414         NA         NA
```

Let's bucket the historical volatility and graph a box plot of fut_vol




Here we can clearly see that the higher the historical volatility, the higher the central tendencies of the future volatility.

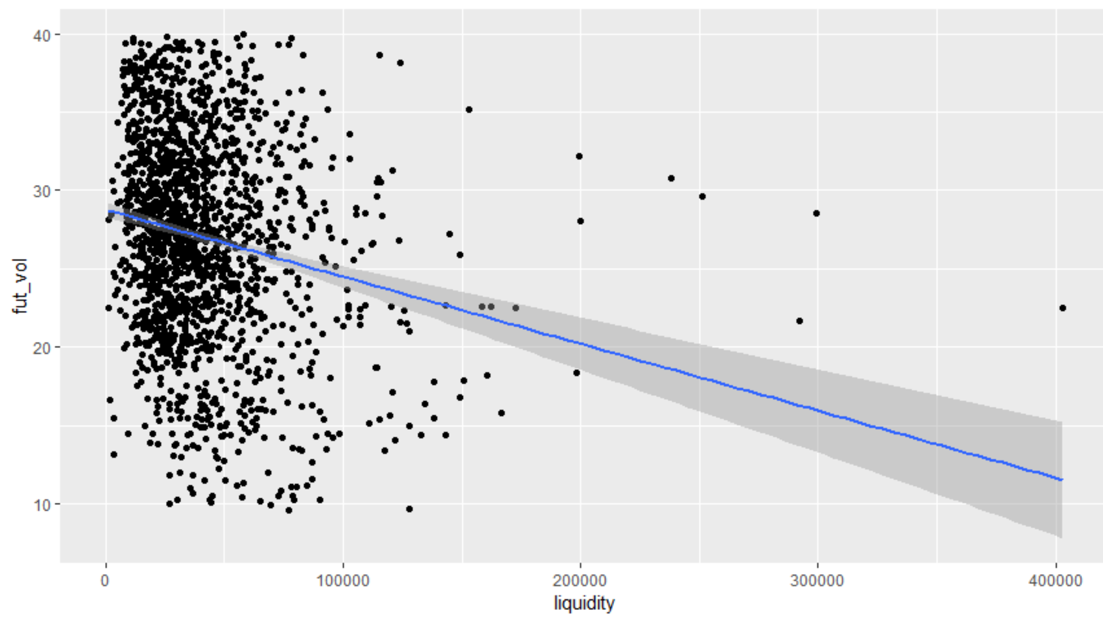
## Liquidity as a Predictor of Future Volatility

Curious if liquidity could predict future volatility. We can define liquidity as volume / average. It represents how much volume it takes to move prices. Lower liquidity should lead to more volatility, everything else being equal. Therefore liquidity should have a negative correlation with future volatility.

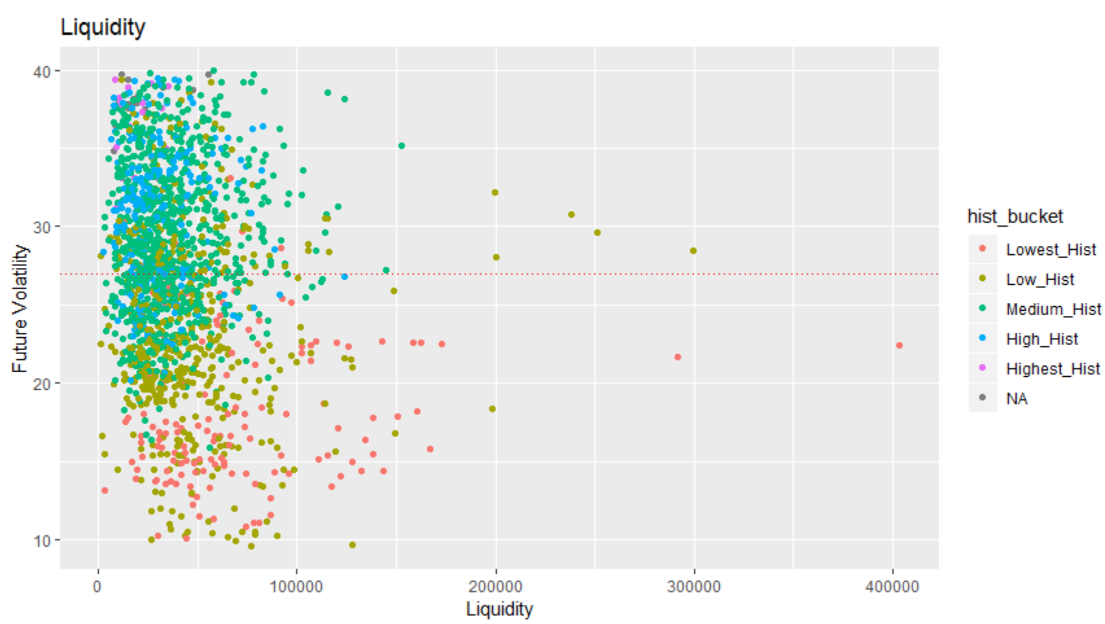
Lets create a new variable, liquidity

Let's calculate the correlation and graph the scatter plot of this variable against fut_vol.

```
##
##  Pearson's product-moment correlation
##
## data:  new_sb_lower_fut_vol$fut_vol and new_sb_lower_fut_vol$liquidity
## t = -8.2005, df = 1712, p-value = 0.0000000000000004646
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.2395530 -0.1484291
## sample estimates:
##        cor
## -0.1944105
```

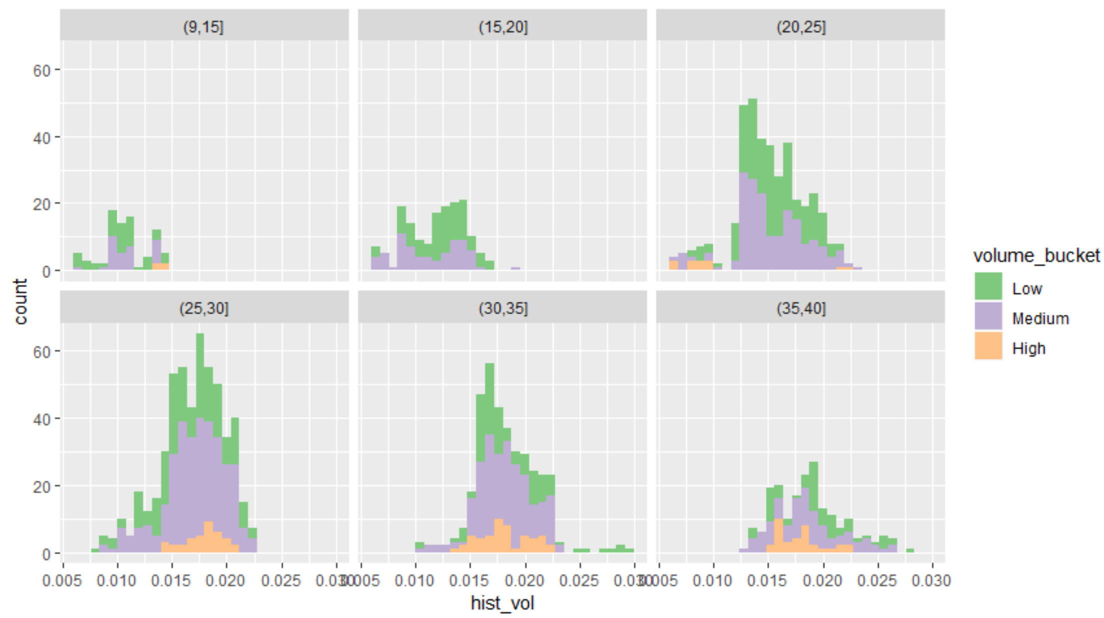


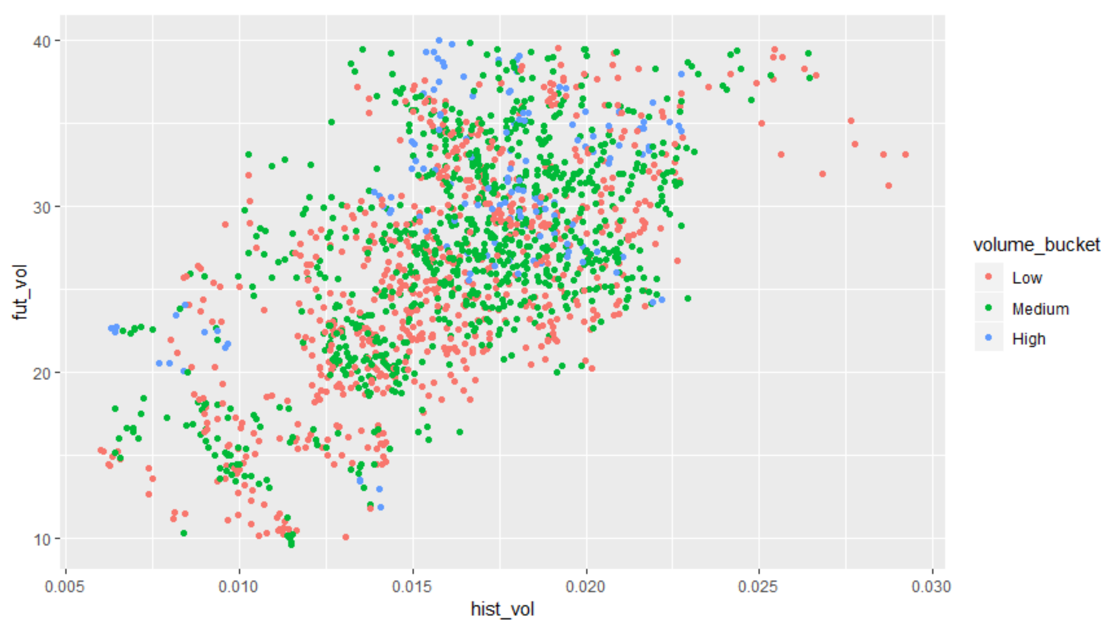

There seems to be a negative relationship between liquidity and fut_vol. With a correlation of just .19, we are unsure how strong that relationship would be in predicting future vol. Volume has a similar correlation with fut_vol and is simpler to define than liquidity.




## Historical Volatility and Volume as a Predictor of Future Volatility

Let's look at some graphs that show both of these variables



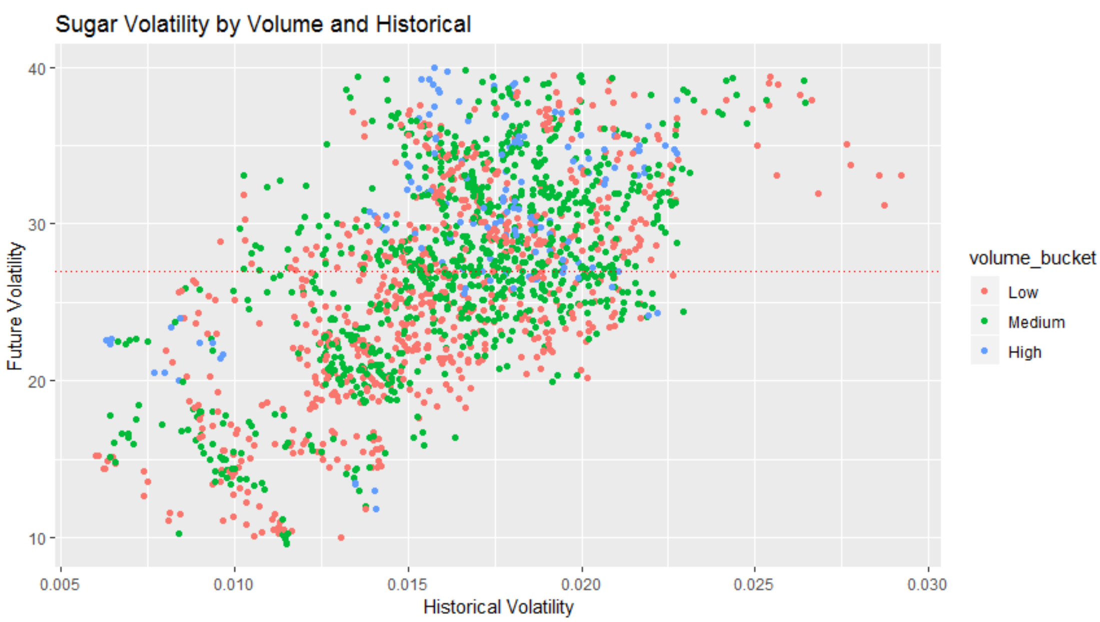

Most of the red and green in lower left and most of purple and blue in upper half.

```
## # A tibble: 17 x 5
## # Groups:   fut_vol_bucket [?]
##    fut_vol_bucket volume_bucket hist_vol hist_med     n
##    <fct>          <fct>            <dbl>    <dbl> <int>
##  1 (9,15]         Low             0.0104   0.0104    47
##  2 (9,15]         Medium          0.0109   0.0109    31
##  3 (9,15]         High            0.0138   0.0138     4
##  4 (15,20]        Low             0.0123   0.0123    89
##  5 (15,20]        Medium          0.0115   0.0115    70
##  6 (20,25]        Low             0.0154   0.0154   193
```

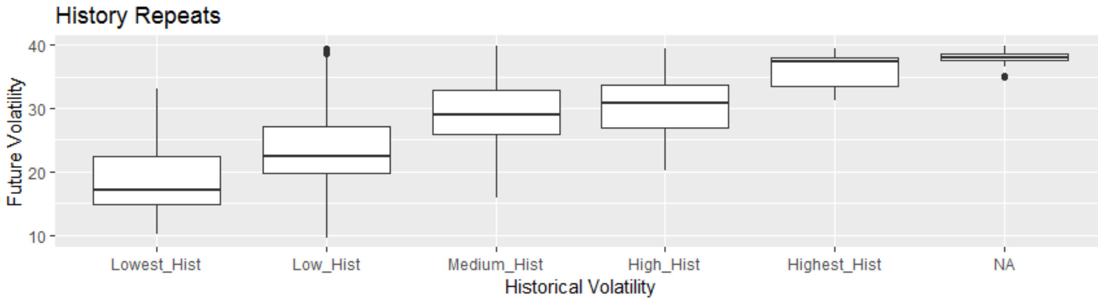
```
##  7 (20,25]        Medium          0.0153   0.0153   180
##  8 (20,25]        High            0.0101   0.0101    14
##  9 (25,30]        Low             0.0165   0.0165   191
## 10 (25,30]        Medium          0.0171   0.0171   290
## 11 (25,30]        High            0.0179   0.0179    37
## 12 (30,35]        Low            NA       NA        111
## 13 (30,35]        Medium          0.0182   0.0182   198
## 14 (30,35]        High            0.0181   0.0181    52
## 15 (35,40]        Low            NA       NA         79
## 16 (35,40]        Medium         NA       NA         97
## 17 (35,40]        High            0.0178   0.0178    35
```

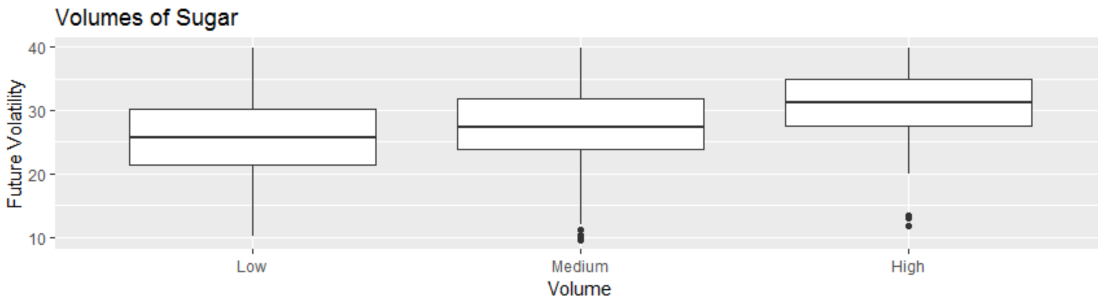
This summarise table breaks down the information in the multivariable plot, shown above.

## Final Plots and Summary

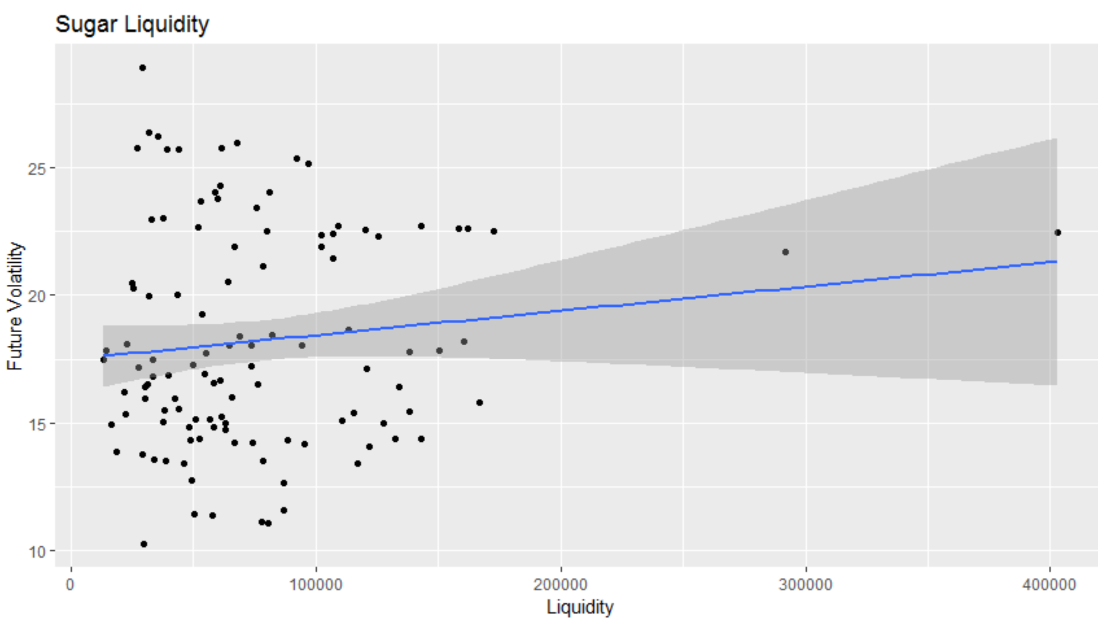



Sugar Volatility by Volume and Historical plot is a scatter of the historical volatility and future volatility with the scatter plots colored by volume. The y dotted line drawn is the mean future volatility. This graph shows the upward slope and positive correlation between historical volatility and future volatility. This indicates that the two is related and therefore historical volatility is probably useful in predicting future volatility. The colored dots indicate that volume could further enhance the predictive power of future volatility. Most of the blue dots, indicating high volume sit above the mean value of future volatility. When they sit below the dotted line, they are usually to the left of the center. This shows that when historical volatilities are around the center, higher volumes, could indicate that sugar is entering a period of above average volatility. That being said, I do not think medium and low volumes send as strong of a signal of future volatility (or lack thereof) as historical volatility. It seems that in cases of low to medium volumes, the more relevant signal is historical volatility and the volumes do not seem to help predict future volatility.




With these boxplots, we can see that as historical volatility and volume increase so do their central tendencies in relation to future volatility. That being said, in this comparison, the differences in central tendencies seem larger for historical volatility. This is another confirmation that historic volatility is the better variable in predicting future volatility.




One of the things I noticed in the "Liquidty" section is that when I plotted the scatter and colored the dots by the historical volatility buckets, it looked like the correlation of the red (lowest historical volatility) was positive. Here I created a scatter and subset the data with the lowest historical volatility. As you can see, the data takes on a completely different shape when it includes only the lowest historical volatilities. It is suprising to see that as sugar prices are presenting low historical volatilities, MORE liquidity lead to higher volatilities.

## Reflection

History seems to repeat itself but do not be lulled into sleep with low sugar trading volumes. In this data set sugar historical price volatilties show the strongest relationship to future sugar price volatility. Larger volumes could be another good indicator of future sugar price volatility but do not assume that lower volumes lead to lower future volatilities especially when they are accompanied by periods of high historical volatilities. In these instances, the sugar market could be taking a collective trading pause only to resume its historical trend.