

Korea Stock Market, centered on Samsung Electronics by Daniel Jang

Samsung Electronics, the company of Galaxy series cell-phone manufacturer in Korea, <https://finance.yahoo.com/quote/005935.KS?p=005935.KS> (<https://finance.yahoo.com/quote/005935.KS>), is said to account for 20% of total Korea stock market - the market capitalization of Samsung Electronics is 329 B dollars and entire of Korea is 1933 B dollars. In case of Apple, it is less than 3% - Apple's market capitalization is 869 B dollars, which is 2.9 % of total US stock market, on November 30, 2017. Accordingly, the stock value of Samsung Electronics has significant weight in KOSPI, the KOrea composite Stock Price Index. Even small fluctuation of Samsung Electronics solely can impact the whole market index. In this project we will try to understand the characteristics of Korea stock market, highlighting Samsung stock.

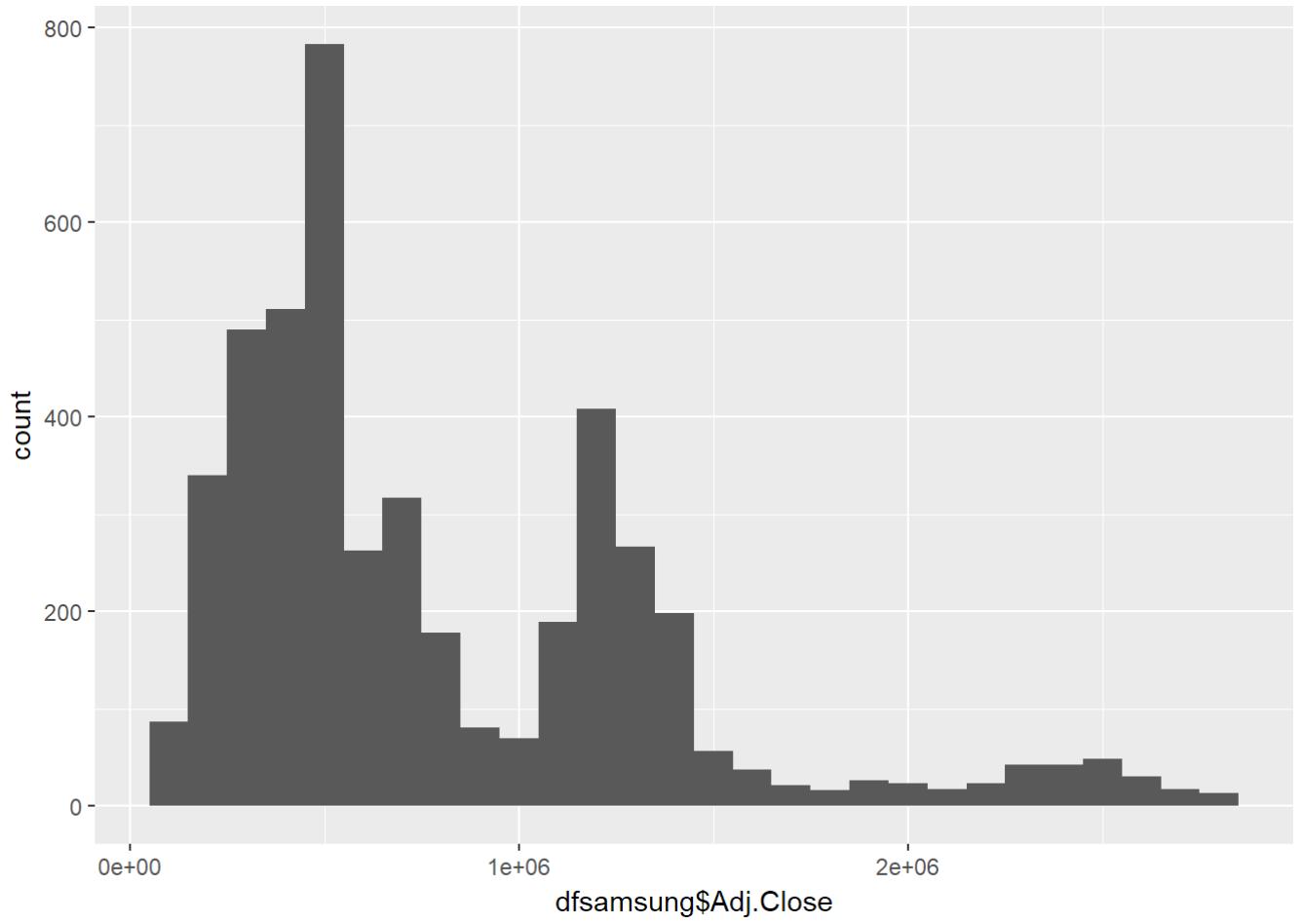
- Market Capitalization or MarketCap is the total price of a company, calculated by the product of the count of the stock and the price of the stock.
- The source of the data - Yahoo Finance, Kosis(Korea Staticstical Information Institute) and KRX(Korea Stock Exchange)
- 3 CSV files were formed by calling API and downloading through Python codes.
 - 1. `samsung_electronics_20000104_20180321.csv`
 - 4592 observations x 7 variables
 - Samsung Electronics stock price since 2000-01-04 using Yahoo Finance API
 - 2. `df_gross_market_cap_trend.csv`
 - 145 observations x 9 variables
 - KOSPI market capitalization monthly trend since 2006-01, from Kosis
 - 3. `dfallcompanies.csv`
 - 2312 observations x 7 variables
 - Company and ticker information in Korea stock market via KRX API.
 - Samsung Electronics is just a single item of 2312 tickers but we can see how big its market capitalization is with this data.

Univariate Plots Section

Let's look into each 3 CSV file.

Daily stock price and transaction volume of Samsung Electronics.

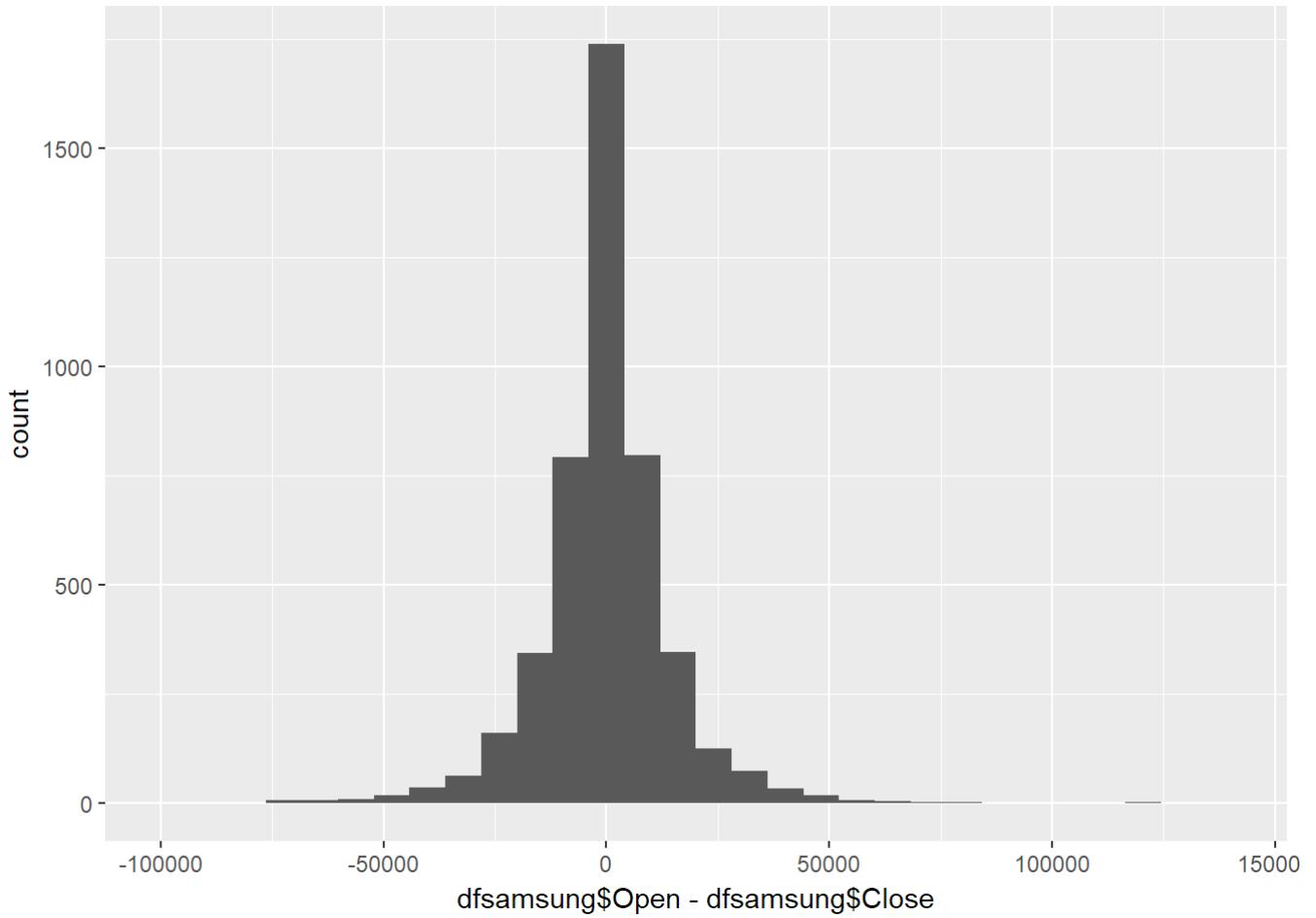
```
##      Date  Open  High   Low Close Adj.Close  Volume
## 1 2000-01-04 300000 305500 283000 305500  264865.0 1483900
## 2 2000-01-05 290000 303000 276000 279000  241889.9 1493600
## 3 2000-01-06 287500 289000 279000 281000  243623.8 1087800
## 4 2000-01-07 278000 283500 268000 277000  240155.9  806100
## 5 2000-01-10 280000 288500 279000 288500  250126.2  937600
```



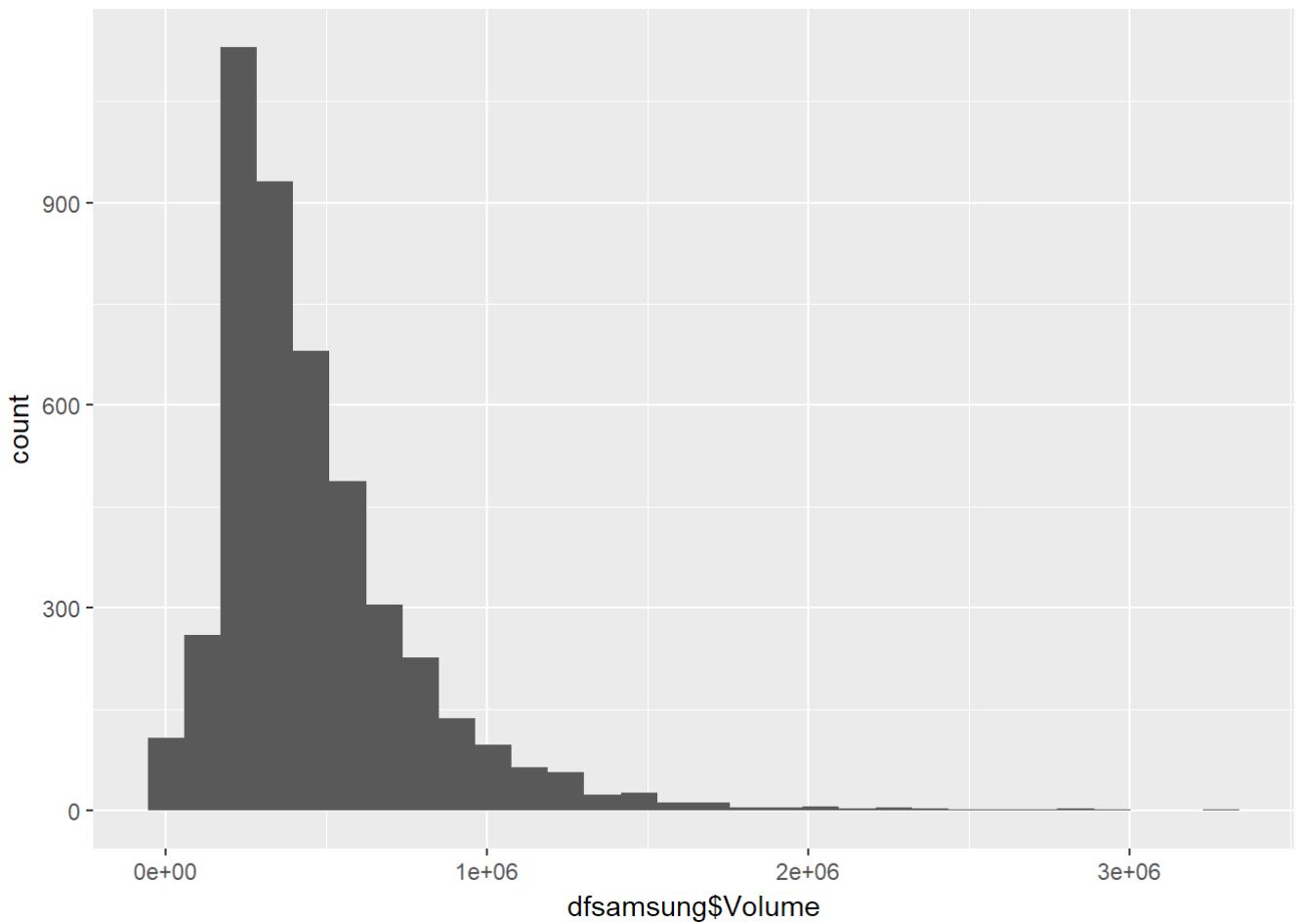
The histogram says there are two groups of the stock price in history.

However, because we already know it is stock price over time, the 2 groups would reveal its true meaning in the flow of the time.

Next. Check the price gap between open and close.



The gaps converge to zero. It might mean that the stock price of this company has been stable over time. However, the judge has to consider the time flow, so we would need bivariate analysis. Let's see the volume traded daily.



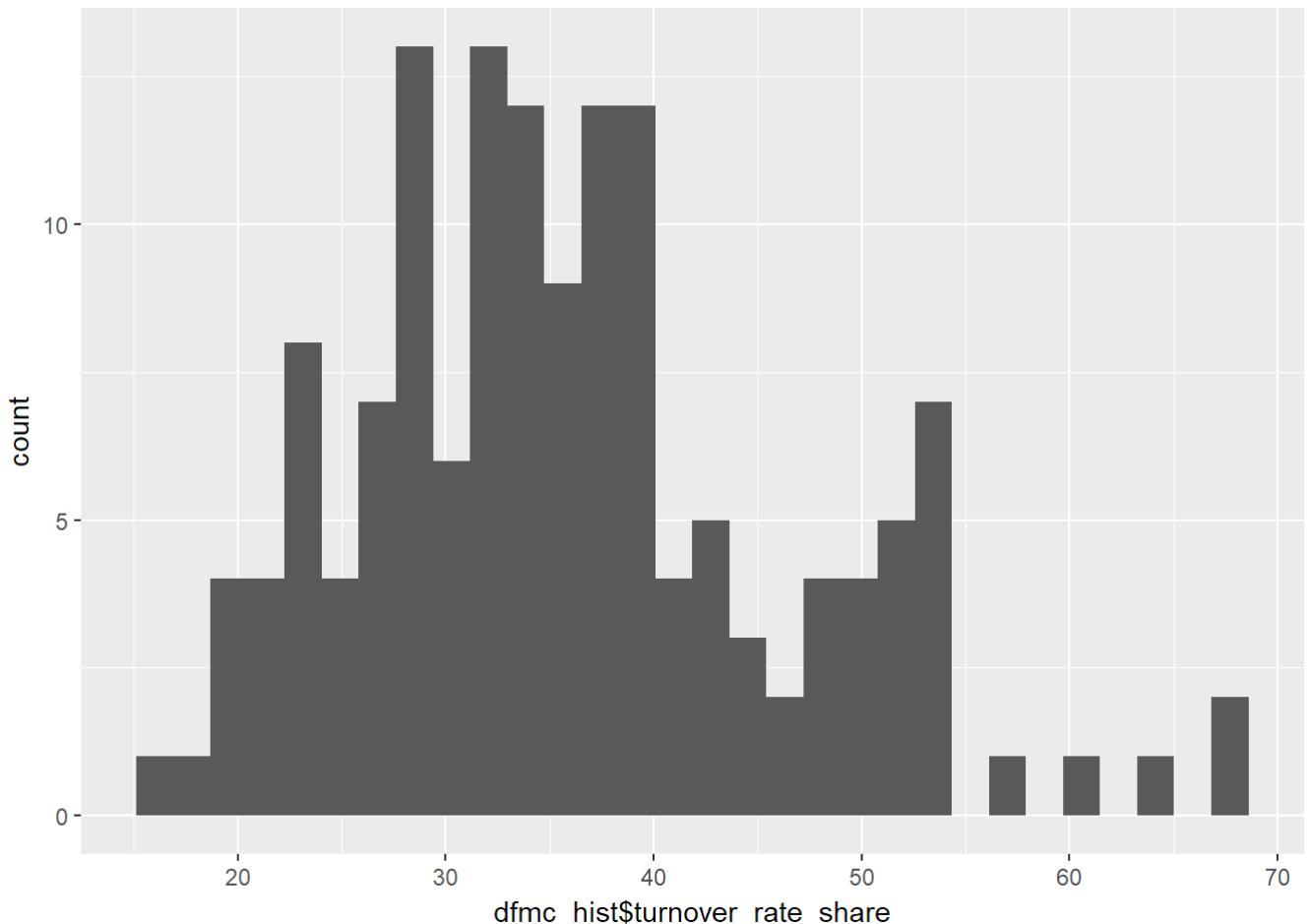
We can see there is a peak at volume = 226503, meaning the most frequently traded amount is about 226,503 daily. When we consider the total number of the stock 128,386,494, we can understand that the price of stock is set by the price of limited transactions, less than 0.18 % of total stock number here.

Obviously, the whole stock is not sellable and buyable in one day, however, it is stunning finding to me that this small portion of the total stock determines the price.

Let's move on to the next file, the trend of whole Korea stock market.

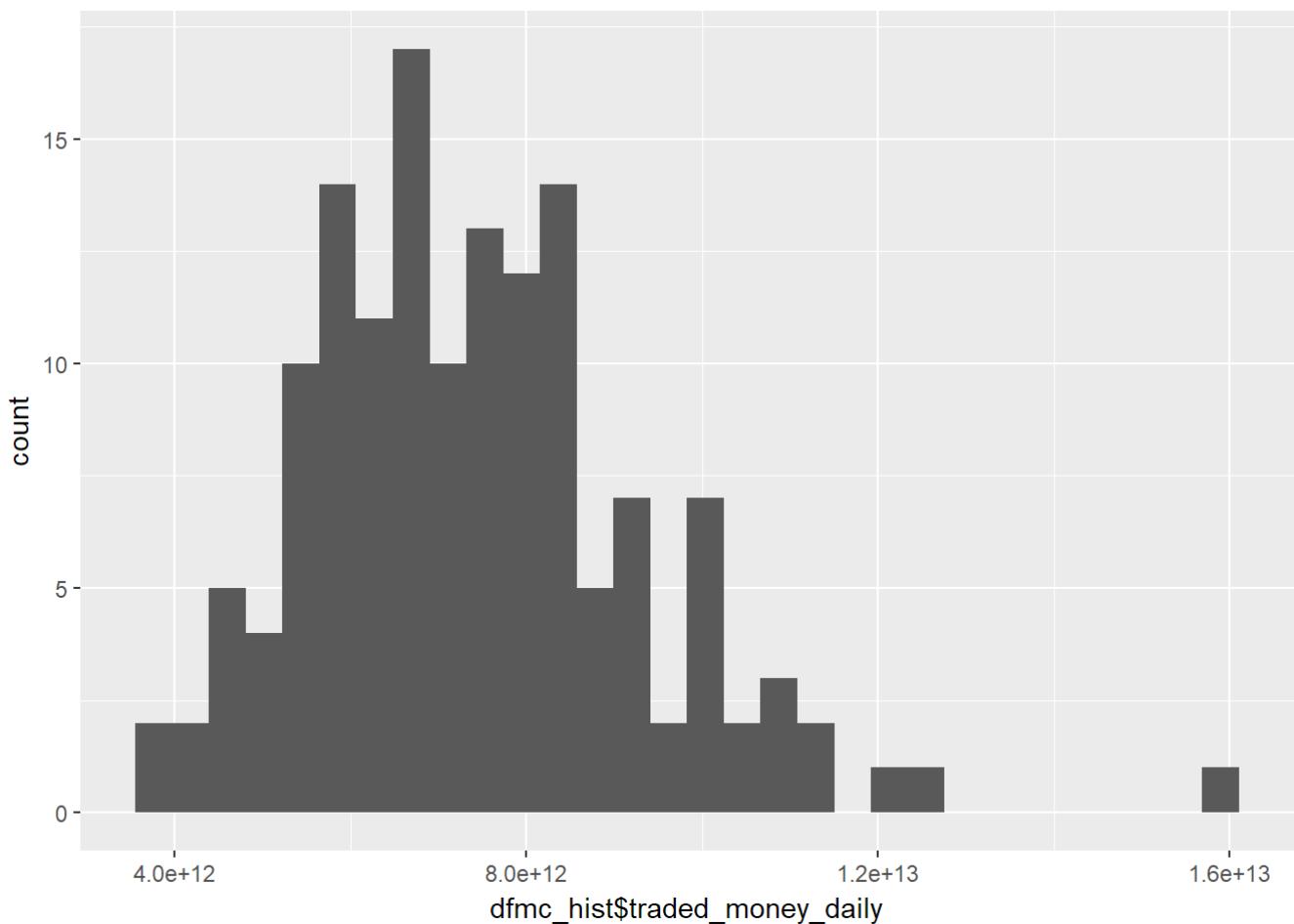
When it is univariate analysis, one thing I have interest most here is turnover rate, how much portion of the stock changed the ownership. That figure would provide some insight how much the market is in people's attention and interest.

```
## X year_month ticker_count gross_market_cap traded_volume_monthly
## 1 0 2006-01 1793 7.370863e+14 25534982227
## 2 1 2006-02 1800 7.370793e+14 20162998131
## 3 2 2006-03 1803 7.327310e+14 20529536788
## 4 3 2006-04 1800 7.654961e+14 19082705266
## 5 4 2006-05 1806 7.087395e+14 14706137416
## traded_volume_daily traded_money_monthly traded_money_daily
## 1 1215951535 1.791798e+14 8.532373e+12
## 2 1008149907 1.304636e+14 6.523182e+12
## 3 933160763 1.161253e+14 5.278422e+12
## 4 954135263 1.225890e+14 6.129448e+12
## 5 735306871 1.087598e+14 5.437990e+12
## turnover_rate_share turnover_rate_marketcap
## 1 68.38 24.62
## 2 53.51 18.25
## 3 53.69 16.20
## 4 49.35 16.04
## 5 37.72 14.52
```



Notice that the time basis of the data is a month. Every month, about 35% of the total stock in Korea market has changed the ownership. Would that figure mean the ownership is way too rapidly changing? no idea for now. We would need to compare that with US and Japan later.

Now, let's see daily average flow of money between sellers and buyers.



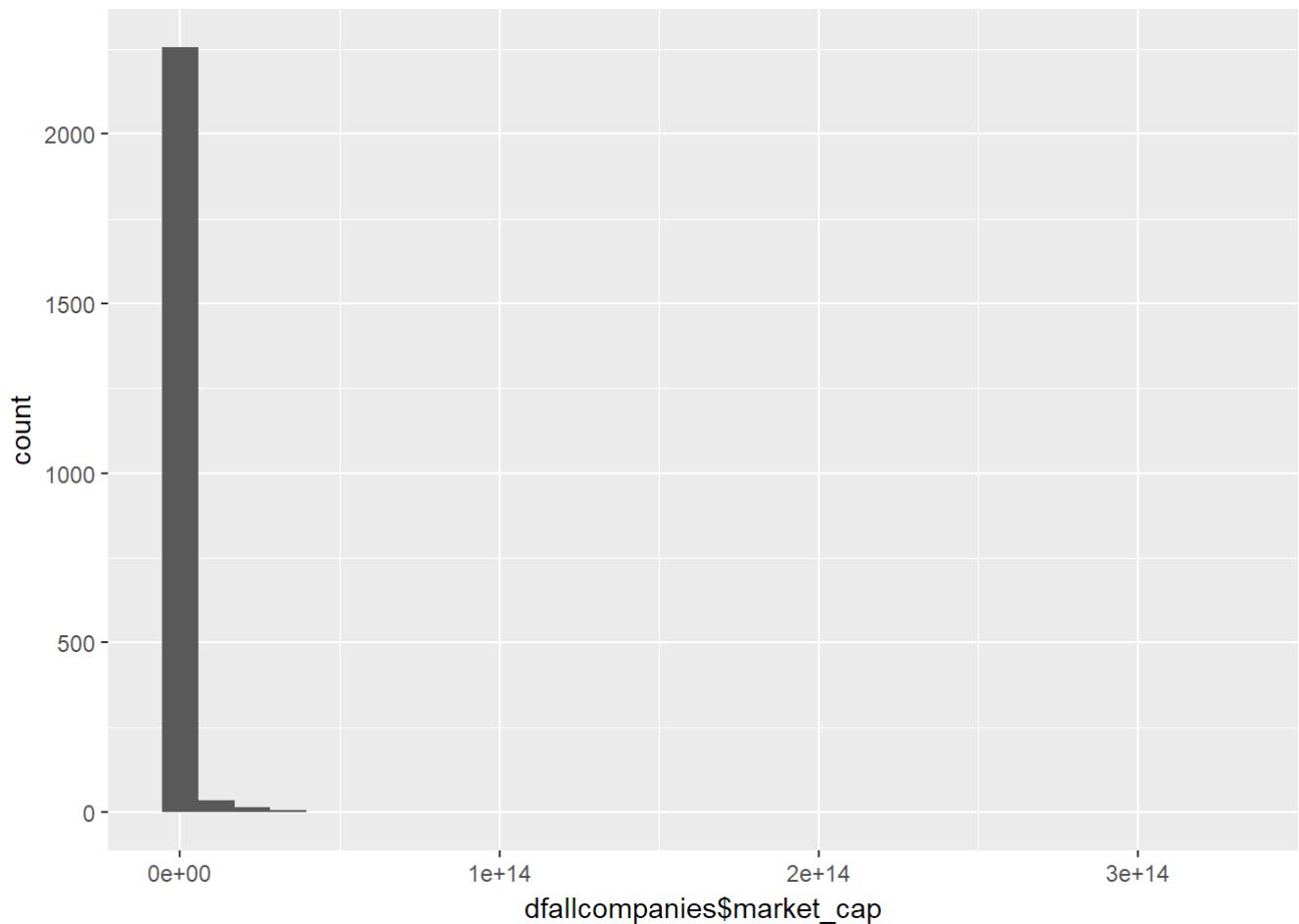
Mean value is roughly 7 B USD. When I consider that with 1933 B USD, the whole value of Korea stock market, we can notice just 0.4 % of total market share plays daily and determines the price. It looks consistent with the previous finding that Samsung stock price was also determined by only 0.18 % of total share daily. There are some occasions where daily money traded was way much like 16,000,000,000,000 KRW (= 16 B USD roughly), more than twice of the mean value. When I see the date of the many transactions, 2009-05, 2011-07, 2018-01 do not remind me of much for now.

Now let's look into the 3rd CSV, all companies in Korea stock market.

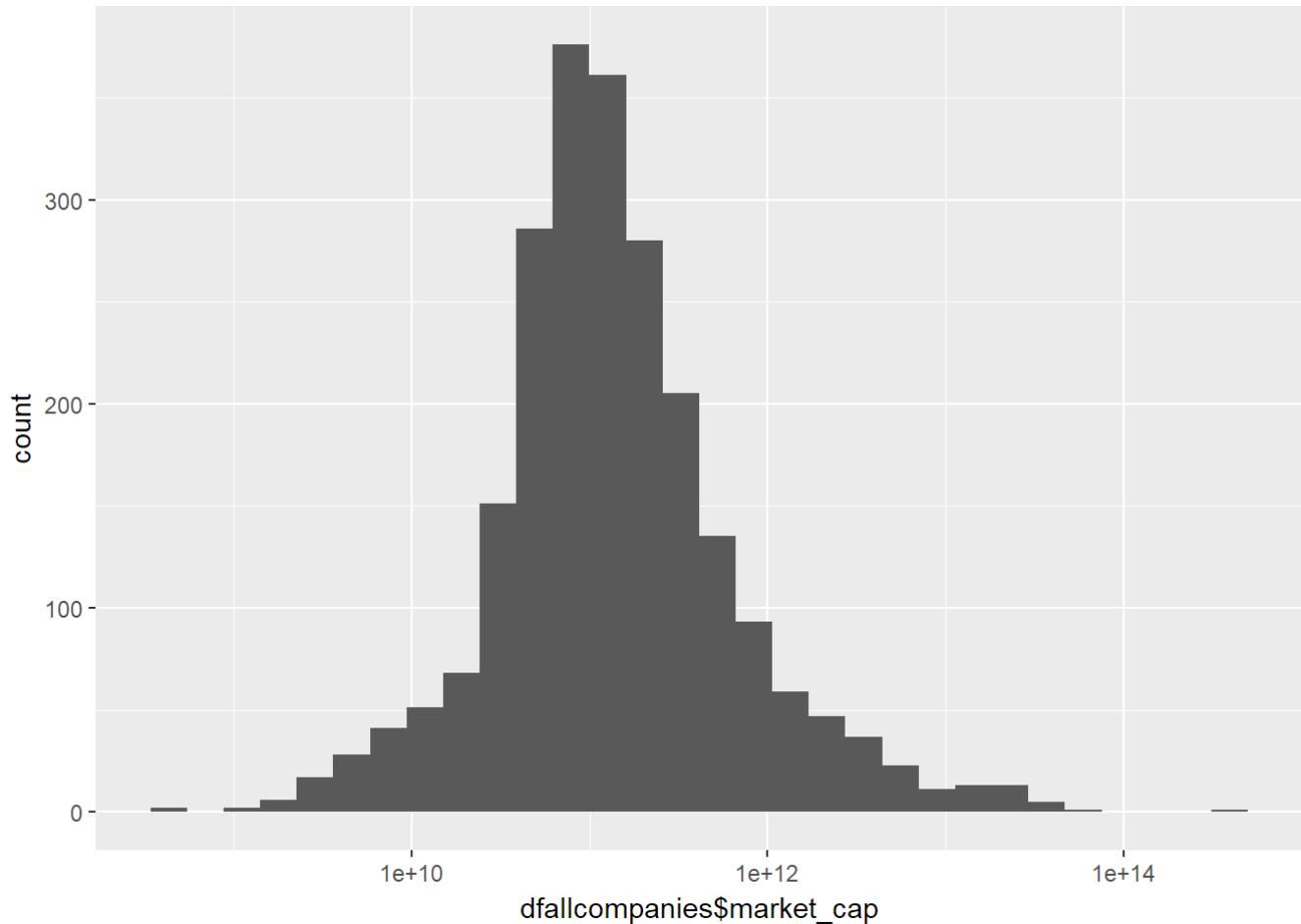
```
##   ticker   ticker_nm   market_cap   market_cap_ratio   share_count
## 1 005930 삼성전자 3.277707e+14      0.1688    128386494
## 2 000660 SK하이닉스 6.435541e+13      0.0331    728002365
## 3 068270 셀트리온 3.821059e+13      0.0197    122666424
## 4 005935 삼성전자우 3.806085e+13      0.0196    18072580
## 5 005380 현대차 3.403272e+13      0.0175    220276479
##   foreign_ratio           industry
## 1             NA   통신 및 방송 장비 제조업
## 2             NA   반도체 제조업
## 3             NA   기초 의약물질 및 생물학적 제제 제조업
## 4             NA
## 5             NA   자동차용 엔진 및 자동차 제조업
```

Samsung Electronics sits at the top. The raw data from KOSIS already calculated the market capitalization and its ratio to Korea entire. For the later use, the stock count of Samsung Electronics has been noted - 128386494.

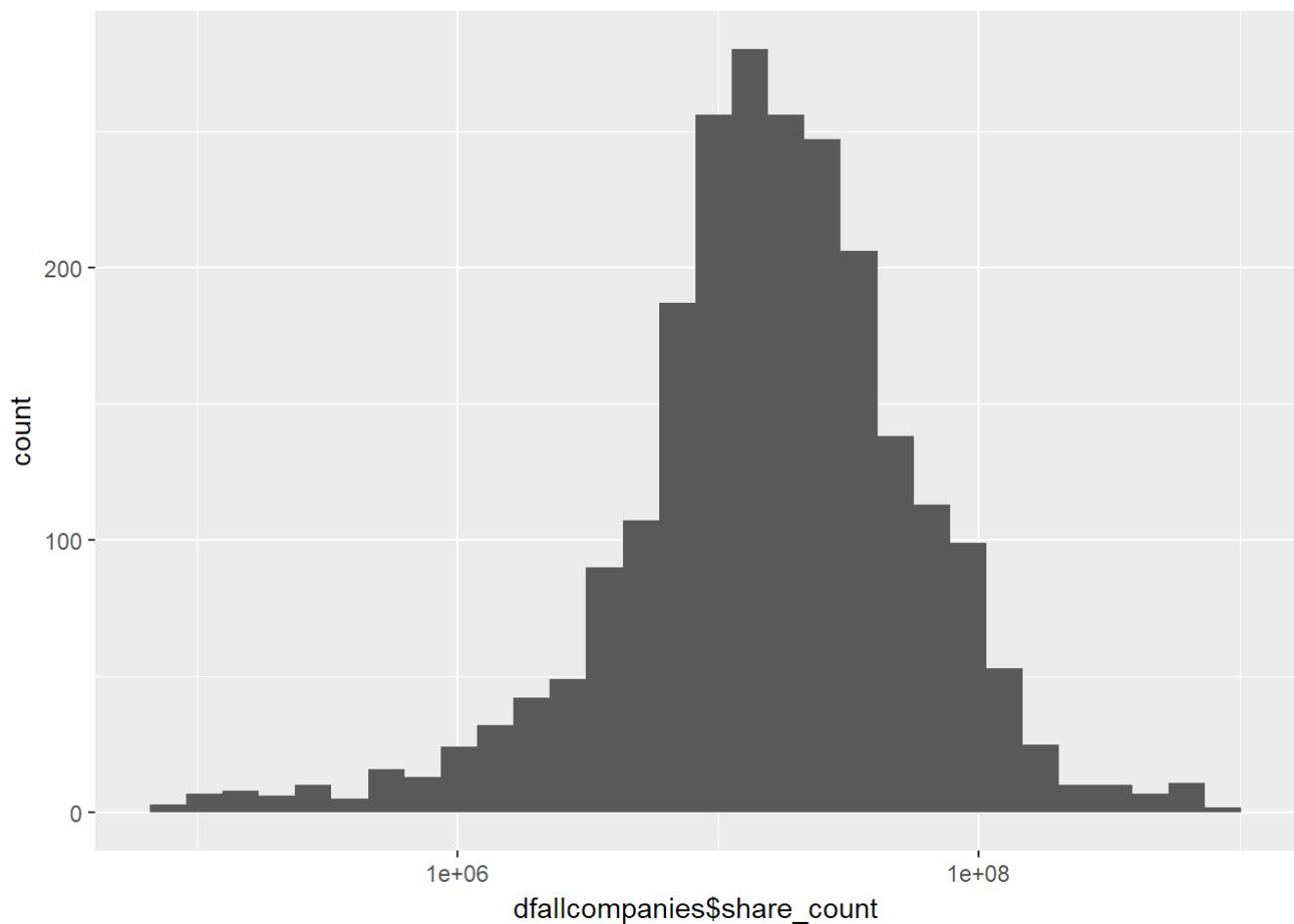
Histogram for market capitalization.



In ordinary X-axis, the plot does not reveal much information. It is positively skewed extremely.



After applying log transformation, We can get market capitalization mostly sits between roughly 10 M USD and 1 B USD. Samsung and Hyundai are a few outliers at the right most, reaching around 300 B USD. Then, I want to check how diverse the share count.



X axis is also log transformed because it is positively skewed like market cap.

We can see the stock count of each company ranges 1 M ~ 100 M mostly.

Samsung is 128 M in the stock count positioning at the right-most.

Univariate Analysis

What is the structure of your dataset?

1. samsung_electronics_20000104_20180321.csv

- 4592 observations x 7 variables

2. df_gross_market_cap_trend.csv

- 145 observations x 9 variables

3. dfallcompanies.csv

- 2312 observations x 7 variables

What is/are the main feature(s) of interest in your dataset?

1. samsung_electronics_20000104_20180321.csv

- 'Adj.Close' : Daily end price at closure. Market capitalization is calculated by crossing this value and stock count.

2. df_gross_market_cap_trend.csv

- 'gross_market_cap' : the market capitalization of the whole Korea stock market.
- We will watch Samsung's value in the flow of total market, so this data reveal how the main stream is flowing.

3. dfallcompanies.csv

- 'market_cap_ratio' : the ratio of each company's market cap against the whole Korea market.

What other features in the dataset do you think will help support your investigation into your feature(s) of interest?

1. samsung_electronics_20000104_20180321.csv

- 'Volume' : the traded volume per a day, which might have association with the price. For example, if the volume increases how it would impact the price?

2. df_gross_market_cap_trend.csv

- 'turnover_rate_marketcap' : the proportion of the total stock that changed the ownership might have some connection with the market cap trend.

3. dfallcompanies.csv

- 'industry' : industry the company belongs to. This is the only categorical variable in my datasets. Though we already have Korea total market trends, industry segment that Samsung belongs to might be showing more granular understanding.

Did you create any new variables from existing variables in the dataset?

Though not used for univariate analysis, Many variables were created afterward.

The first two datasets are time series datasets. When I imported the data into Rstudio, the 'date' was just recognized as string. So, I created a new variable that converted string date to date type. To examine seasonality, year and month were separated, to repeatedly show the stock price on the same 1 ~ 12 month axis. To normalize the value range, new variable was created to put standardized value in it.

One challenge was that the 2nd dataset is monthly record of the past. To convert '2011-01' to date type, a day has to be specified. I inserted '-15' as the center of the month and converted the string to date type.

Of the features you investigated, were there any unusual distributions?

Did you perform any operations on the data to tidy, adjust, or change the form

of the data? If so, why did you do this?

There was no unusual distribution according to univariate exploration. There are many outliers like Samsung and Hyundai, but they are the main subjects of this document.

During the data preparation, the API at Korea government organization returned Korean variable name like '종목코드' and so on. I translated that to English.

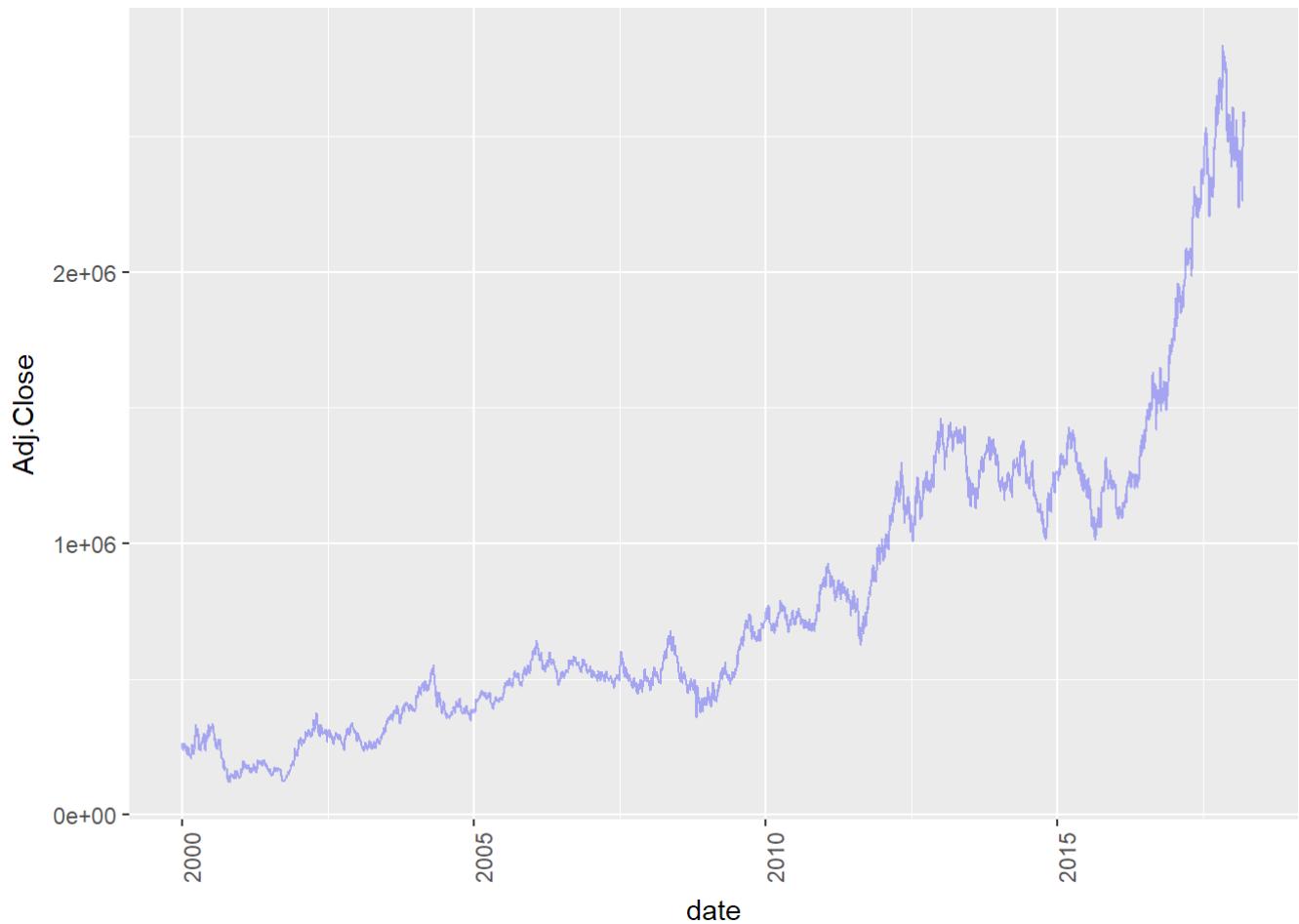
One tricky part was dfallcompanies.csv, because the data was originally from 2 data files according to the market - KOSPI and KOSDAQ in Korea. I concatenated the dataset to see the whole company landscape. For that, ticker code was reformatted as '005390' from '5390' for example, to support consistent code name.

Bivariate Plots Section

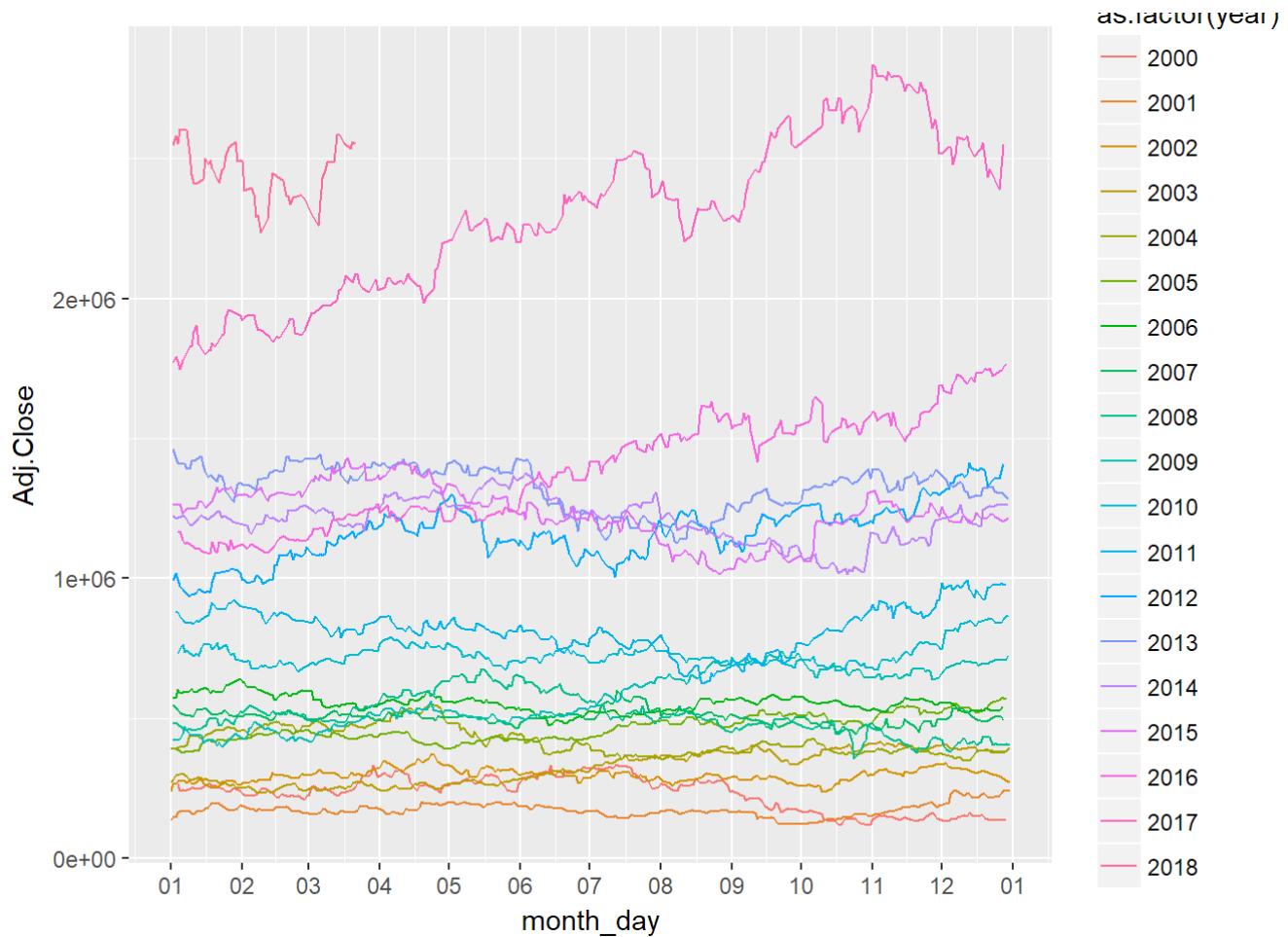
Because the first two datasets are time-series data. Univariate analysis might not be informative. Now we can put x-axis the date.

First, stock price trends of Samsung Electronics.

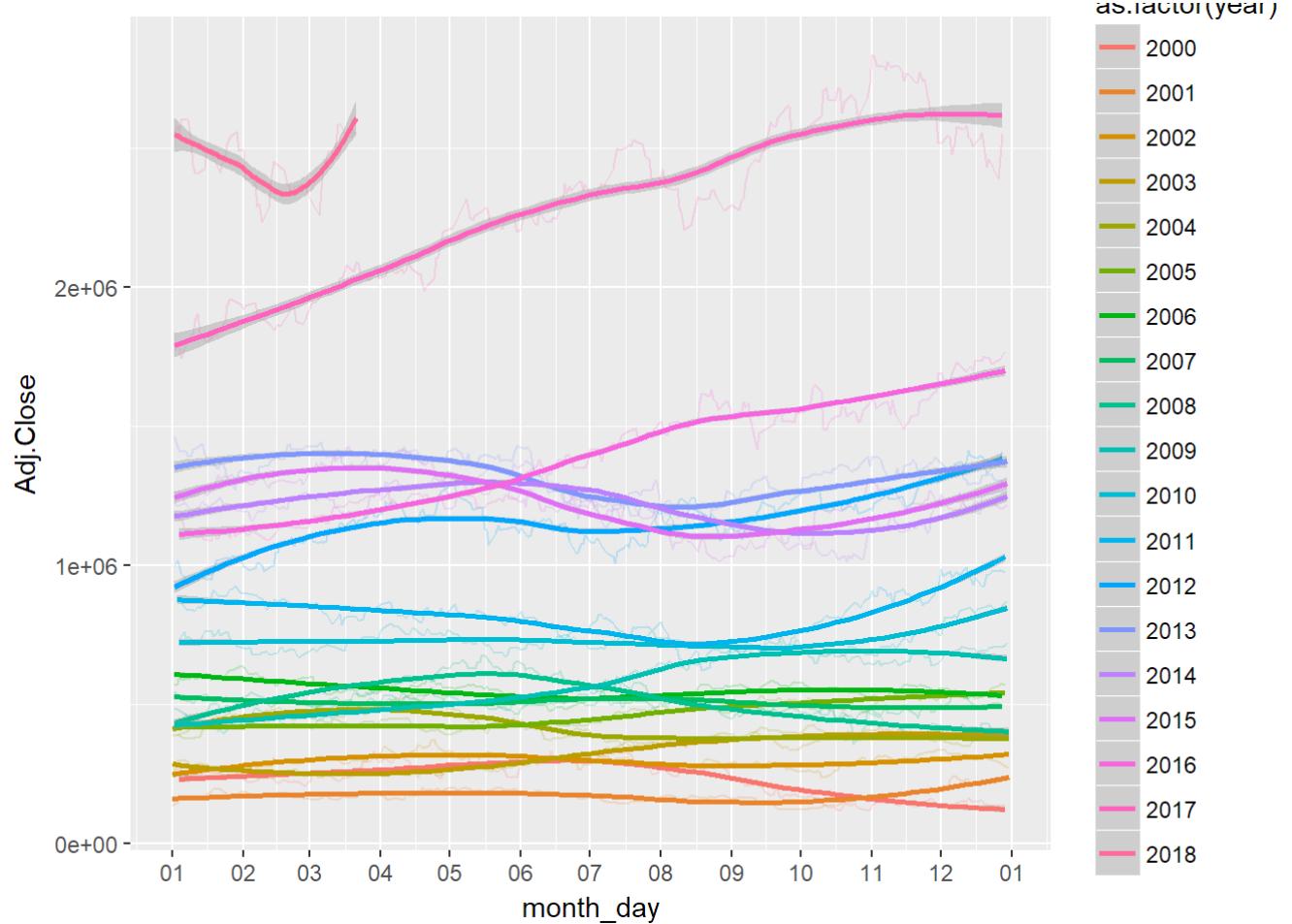
* 'Adj.Close' is adjusted price at close, usually same as 'Close'



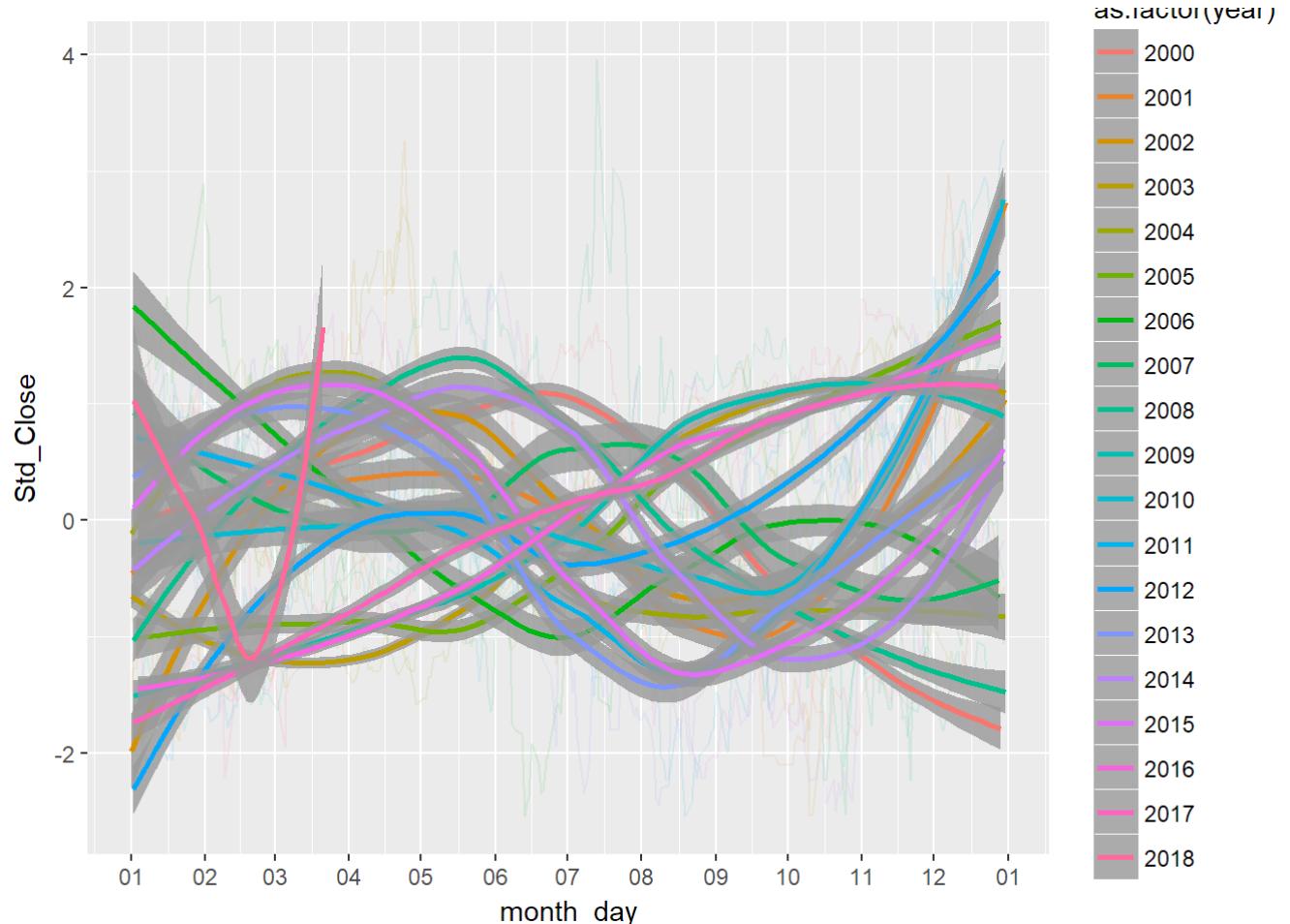
I get that due to the good sales of Galaxy series, the stock price skyrocketed in recent years.
It is time-series data. There might be seasonality.



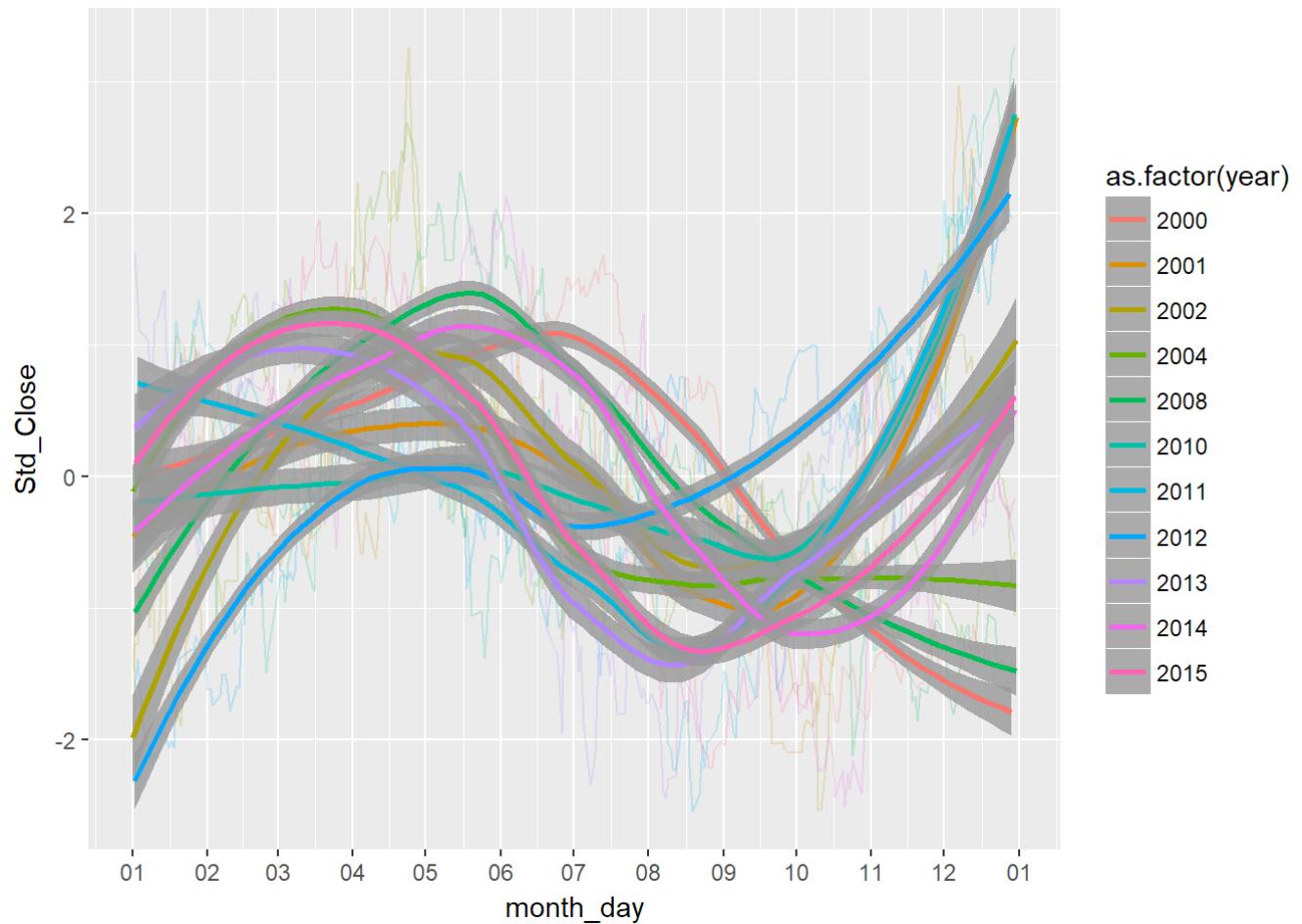
It looks too noisy. Adding Smoothing.



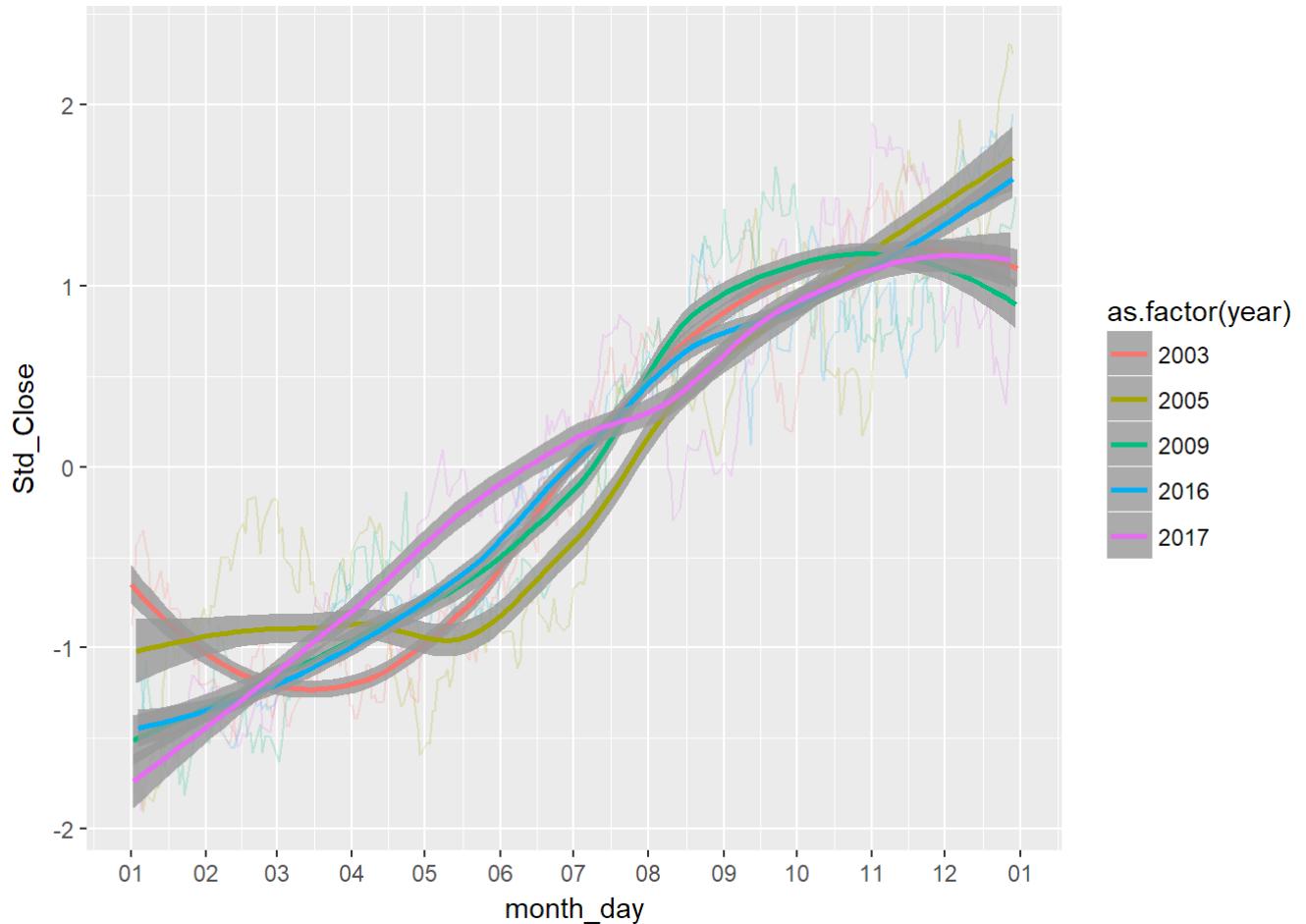
It is hard to detect any pattern because y range is wide. Applying standardization.



Still it looks random because of too many streams. So I picked years showing similar pattern.

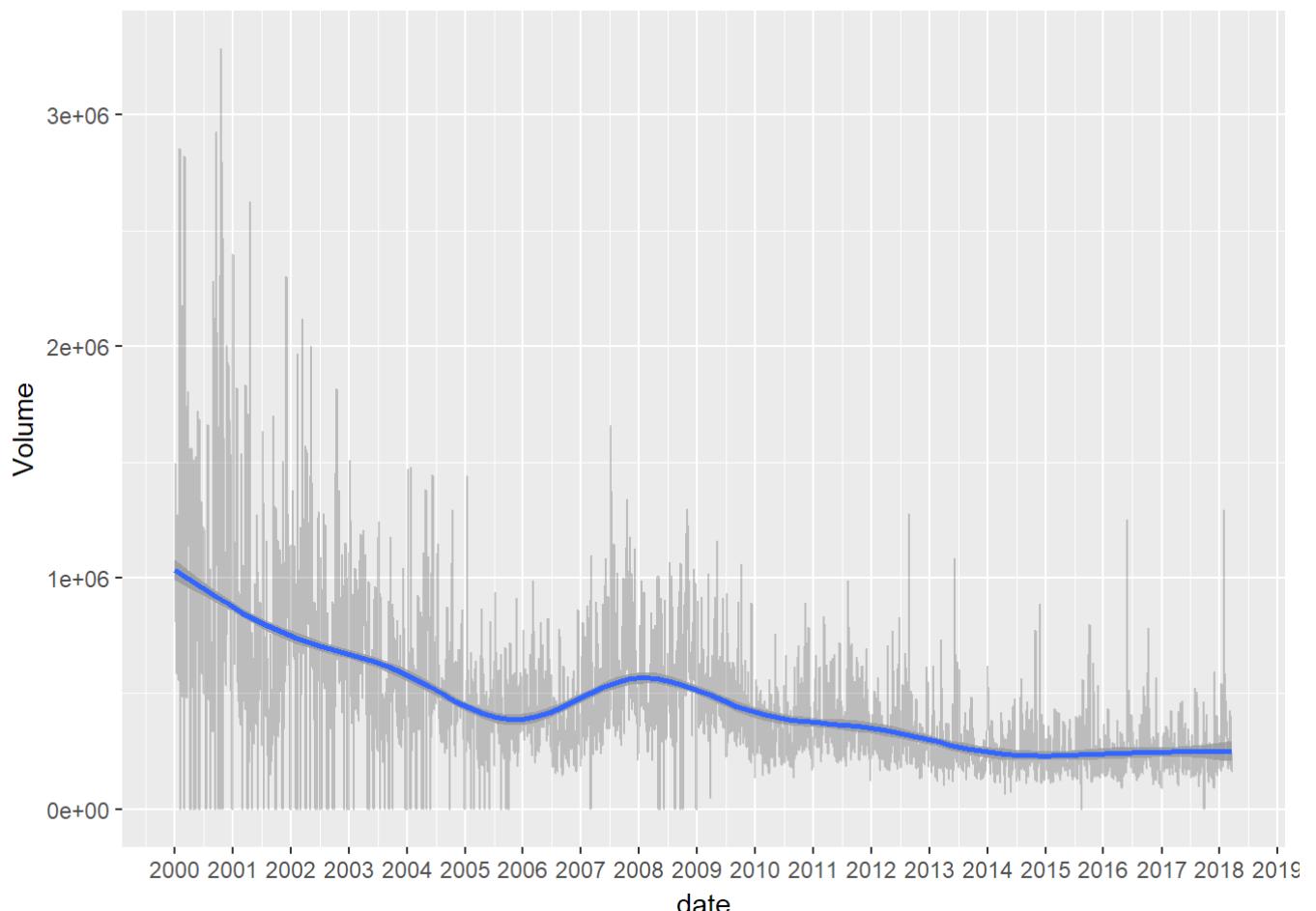


Pattern 1 - decrease in summer and rise in winter, in year 2000,2001,2002, 2004,2008,2010,2011,2012,2013,2014,2015. I guess this cycle basically is the consistent power that boosted Samsung stock price. There might be some event repeatedly takes place every year. For example, it might be spree of new cell phone release. We would need more historical reference to find a clue.



Patter 2 - just up and up. These years need another research on why this acceleration took place.

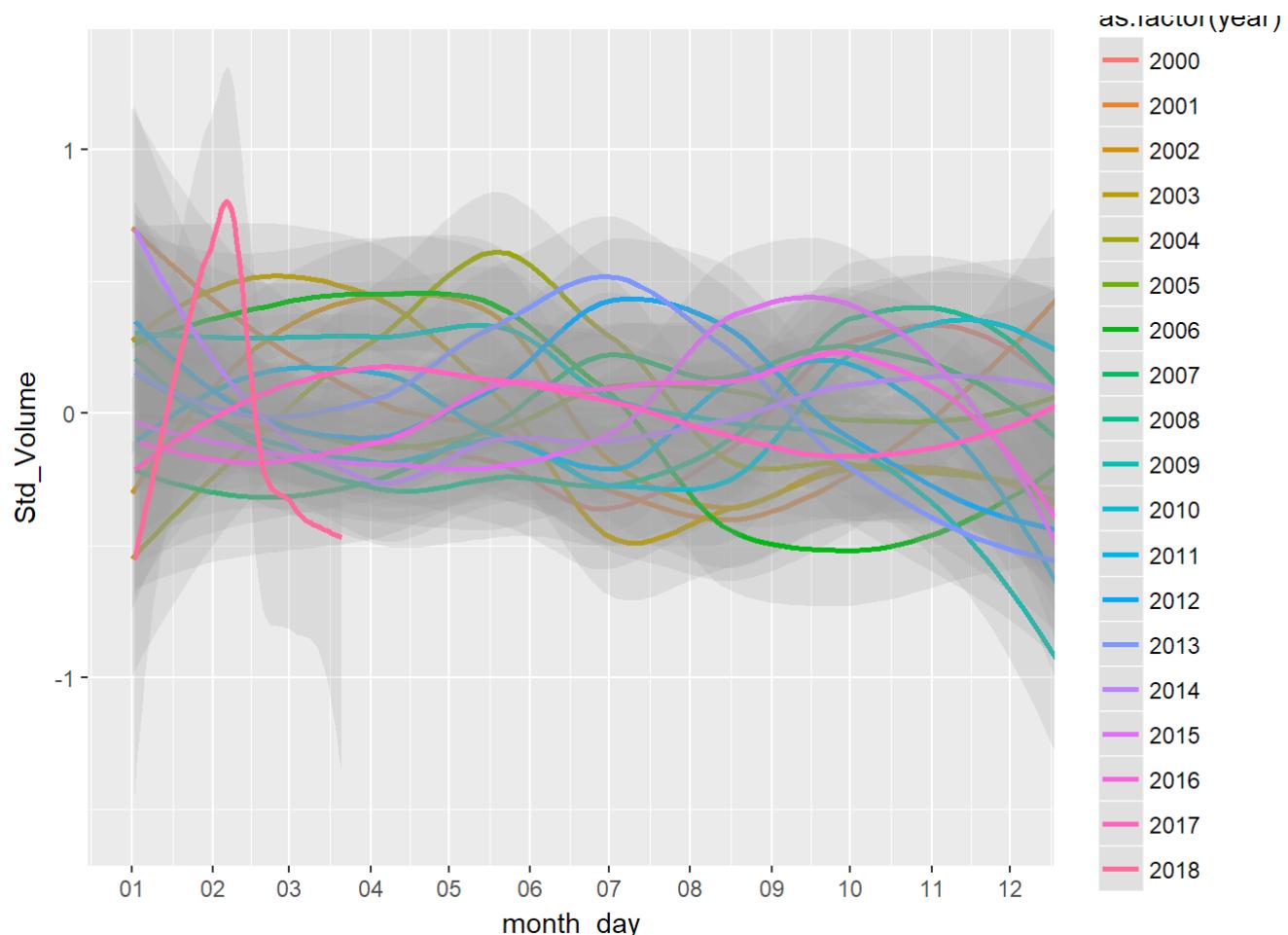
Next, what about daily traded volume?



Defying the price trend, the traded volume has been decreasing. It might be obvious, because the price soared and the price now is over 2000 USD per share. Small investors might not be able to engage the trade.

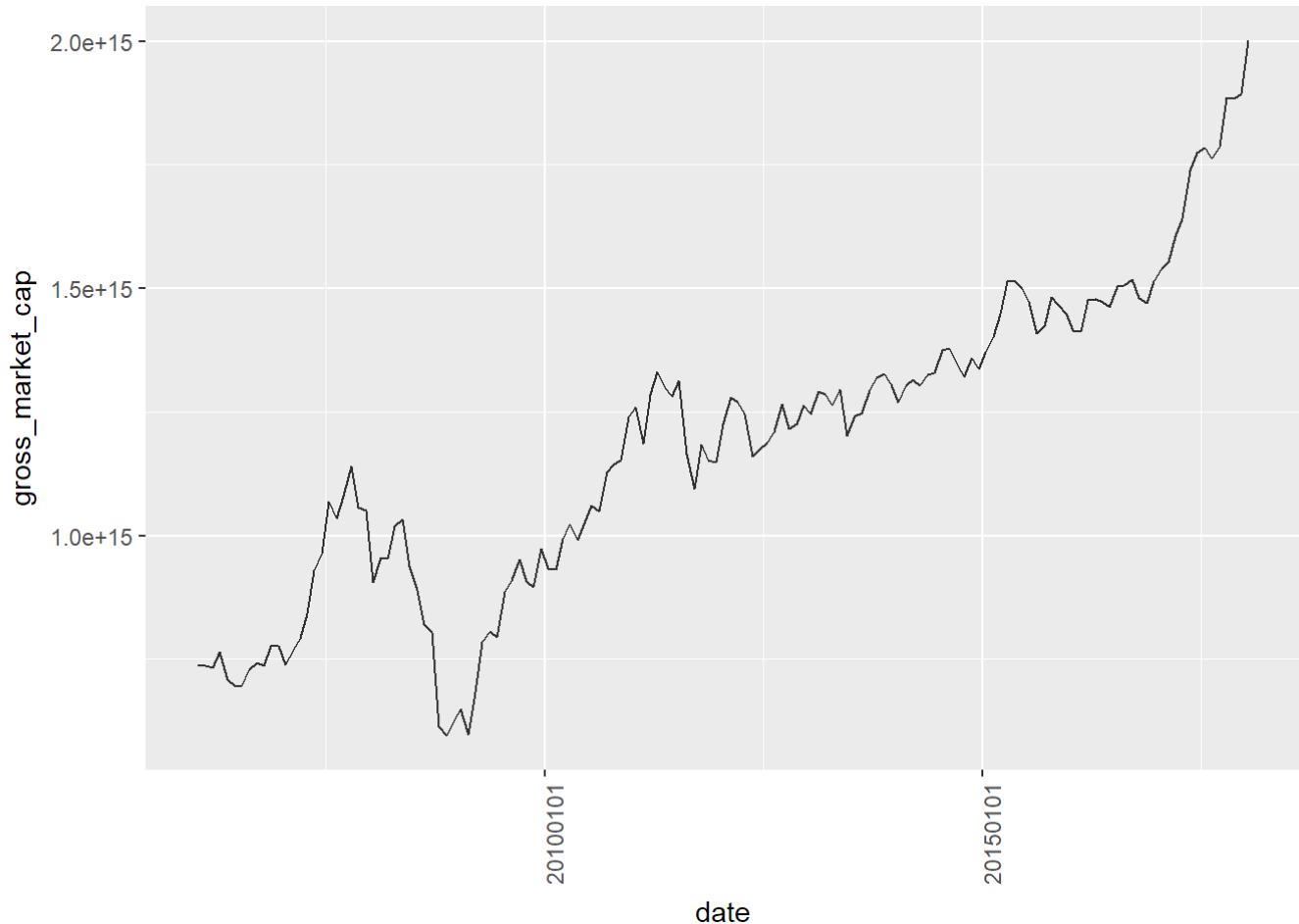
Notably, during the global crisis started in 2008, the volume increased. Samsung stock price at that time was increasing, which indicates that investors bought Samsung during the crisis. They must have assumed that crisis was a chance to buy Samsung. If the stock was bought in that period, the value would be x5 by now.

I had a check on the seasonality of the volume after standardizing the value.



I could not detect any pattern though I expected some boost around October, because we saw the pattern of rising price around that month in many years.

Next, the trend of total market capitalization in KOSPI

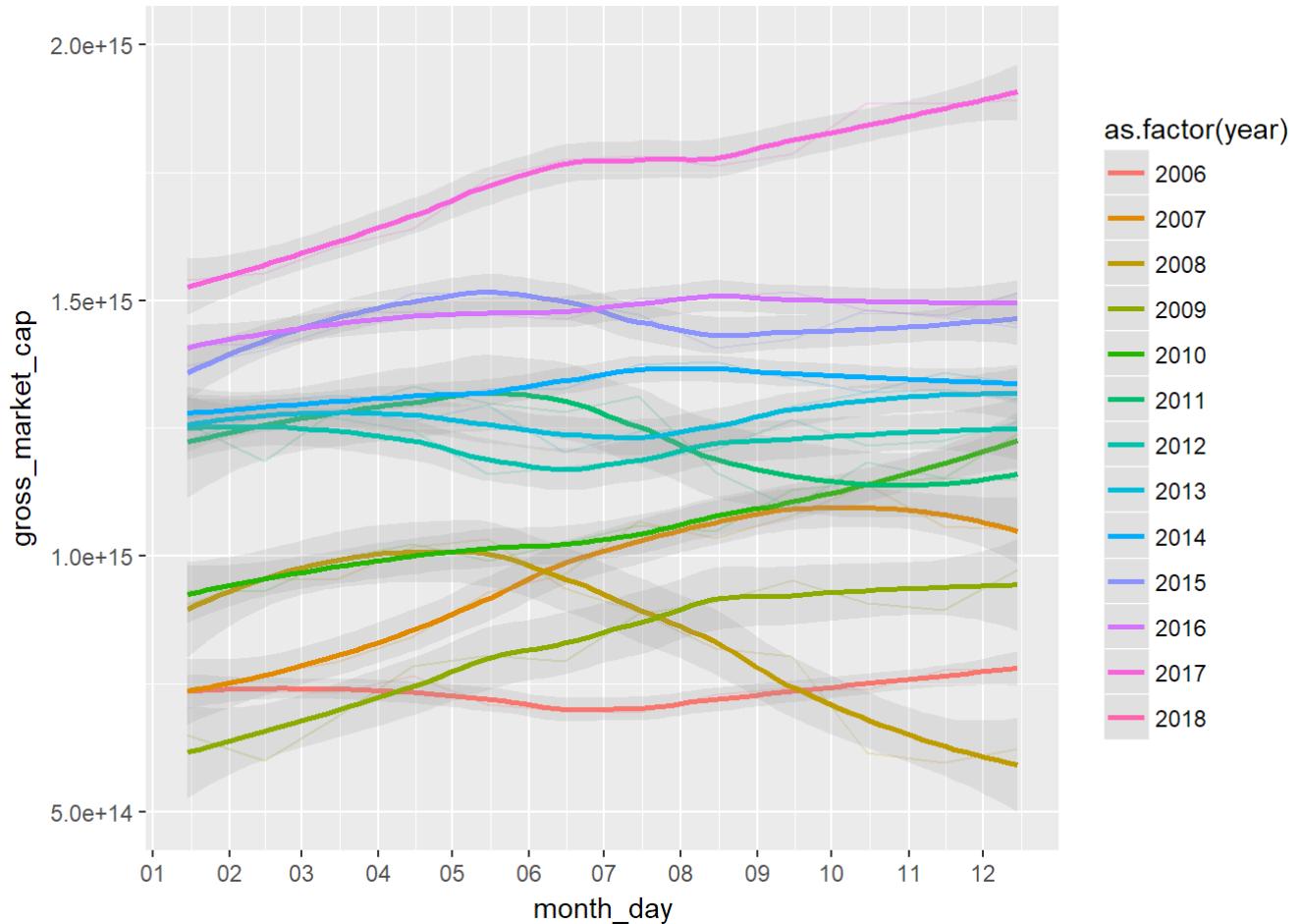


Acknowledgeably, the market capitalization of companies in Korea is escalating.

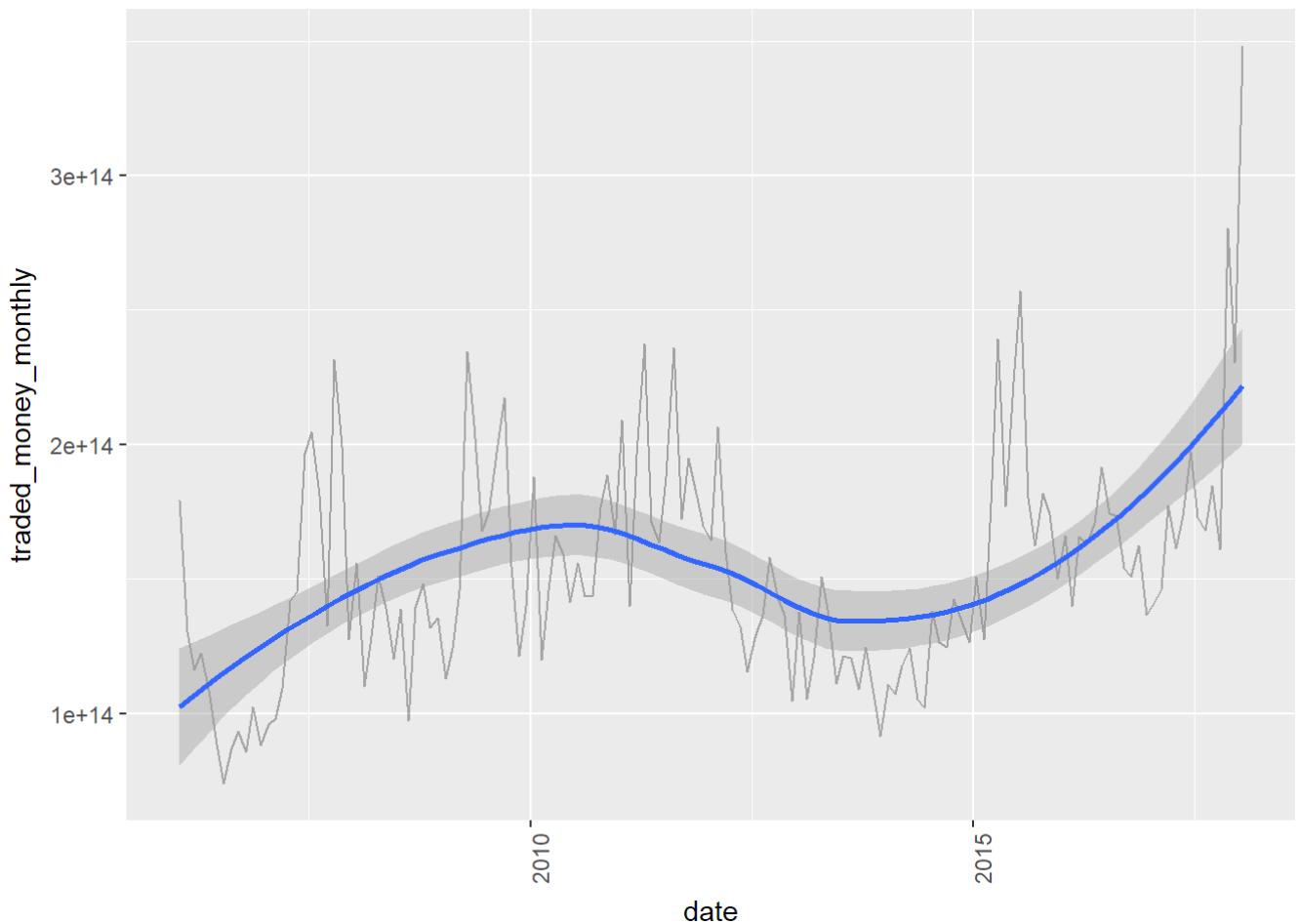
Sudden drop in 2008 when the global crisis from Lehman Brothers bankruptcy took place, is recognizable as well.

* Notice that it is not KOSPI, the Korea stock market index, but the total of market capitalization which we have interest in here.

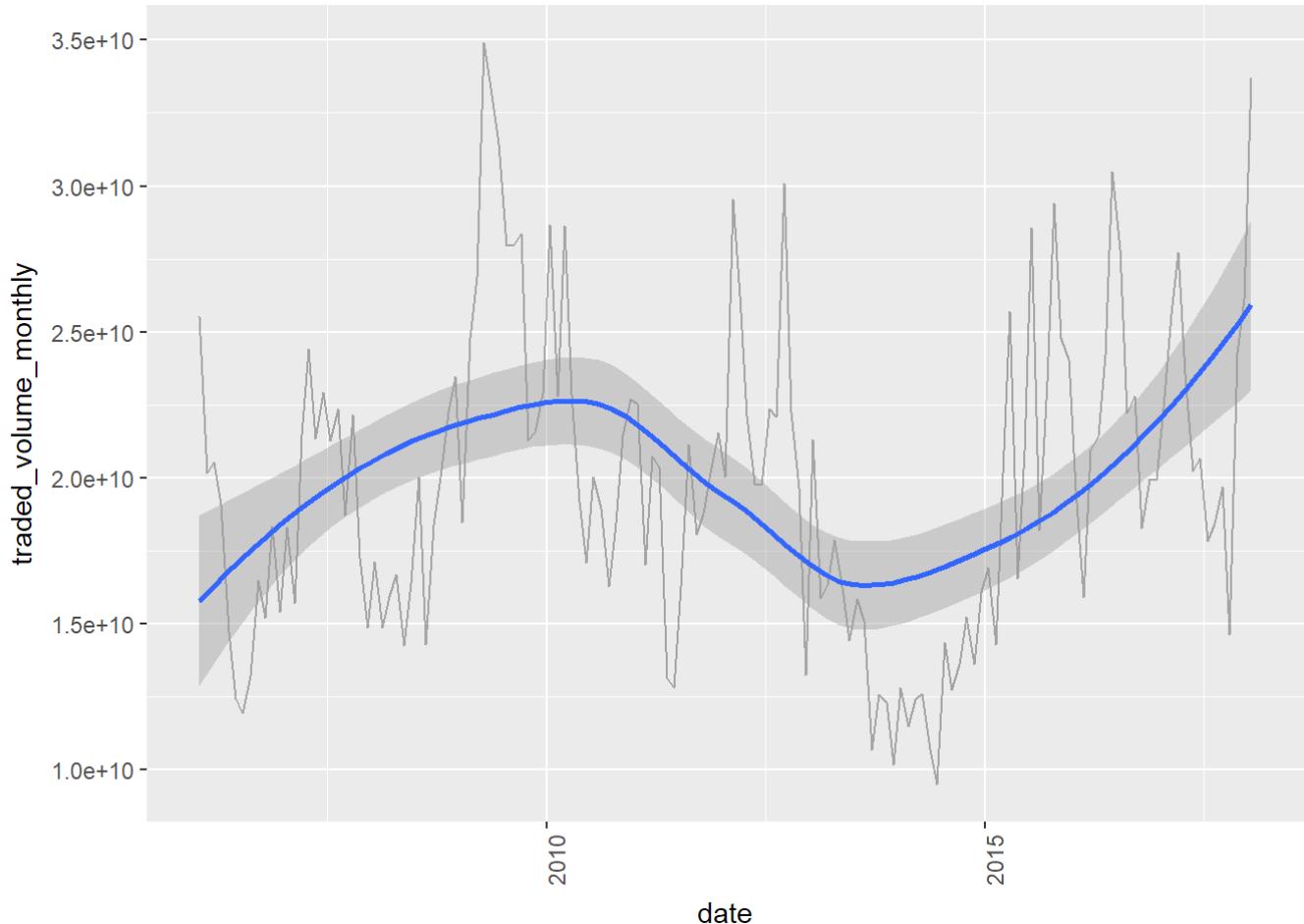
Now. Let's see whether there is seasonality in the records.



Not much, let's check on the traded money.

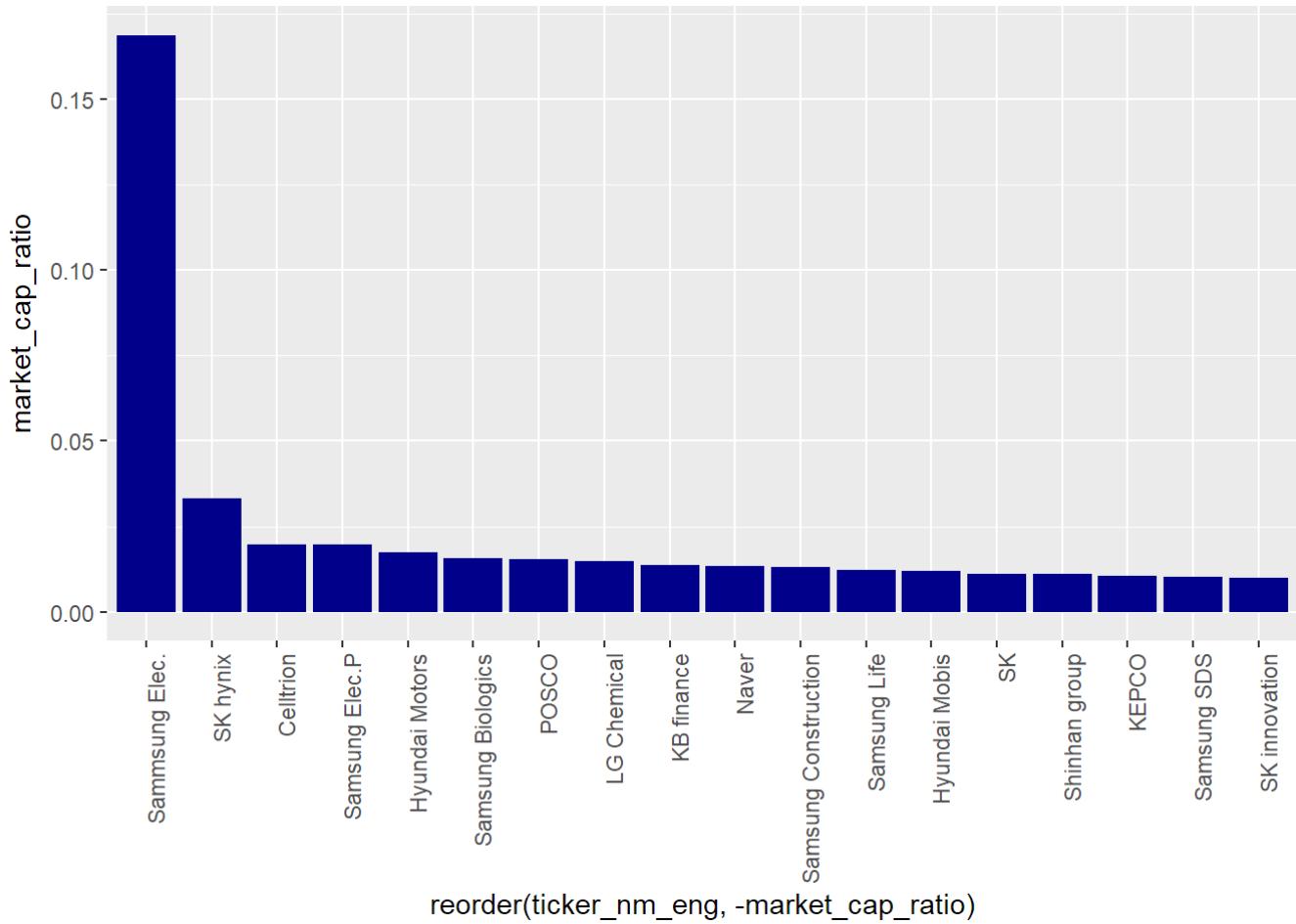


There is a big, 10-year-spanning wave. The market is approaching another peak now. This is not stock price but the traded money, but this surely implies the people's consideration of investing stock market. Accordingly we can guess the traded volume is increasing.



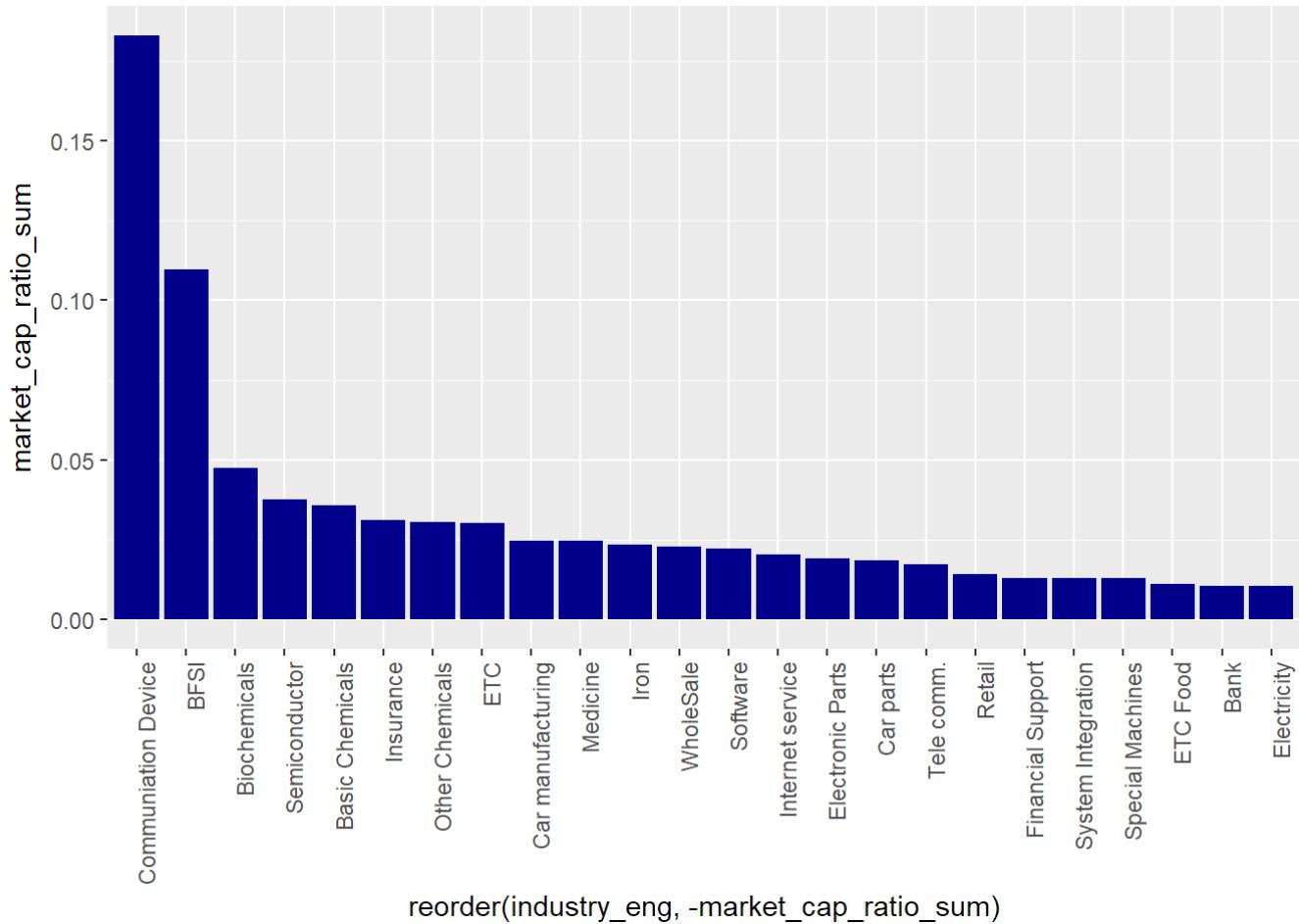
As expected.

Now, let's look into the final dataset, all the companies listed in Korea stock market. When I ordered the data by the market capitalization ratio, the plot is just as I heard of. Samsung is overwhelming.



There are many companies' name starting with 'Samsung'. They could be assumed a single company, Samsung. Then their market cap ratio might surpass 0.3 of the total Korea stock market. The ownership of that group companies are known to be very complicated and blurry at the edge. So, it would be next time to separate all the Samsung family companies and its market capitalization.

Now let's group by industry to see how diverse the indsutry and how we can characterize Korea industries.



reorder(industry_eng, -market_cap_ratio_sum)

1st item is, to simply put, cell phone manufacturing industry where Samsung belongs to. Samsung is core part of the industry according to the ratio. It is same as Samsung alone.

Other industries are much lower. So, when it comes to stability, the landscape of Korea industry is fragile or weak at risk in my view, because the whole economy is too much relying on Samsung and its export of non-diverse products.

I get that conventional manufacturing industry is decreasing. Rather the financial service like bank and insurance is taking much part, 10%. Software in IT is weak, compared to hardware like semiconductor, LED display.

Bivariate Analysis

Talk about some of the relationships you observed in this part of the investigation. How did the feature(s) of interest vary with other features in the dataset?

In the bivariate exploration, I could not directly compare the total market value and that of Samsung on the same time axis, but each plot gave me interesting view point. 1. Samsung stock price over time revealed there could be a cycle 2. Total Korea market revealed there is a big cycle which is approaching another peak. 3. Korea industry is unbalanced.

Did you observe any interesting relationships between the other features
(not the main feature(s) of interest)?

One thing I clearly found is that the price of stock and traded volume go the opposite. It would be obvious that expensive stock is not easy to trade, but I verified that it is true.

And there might be seasonality in the stock price. I noticed a Samsung stock flow repeating every year, not that assertive though. I need paper work to see what was behind such cycling.

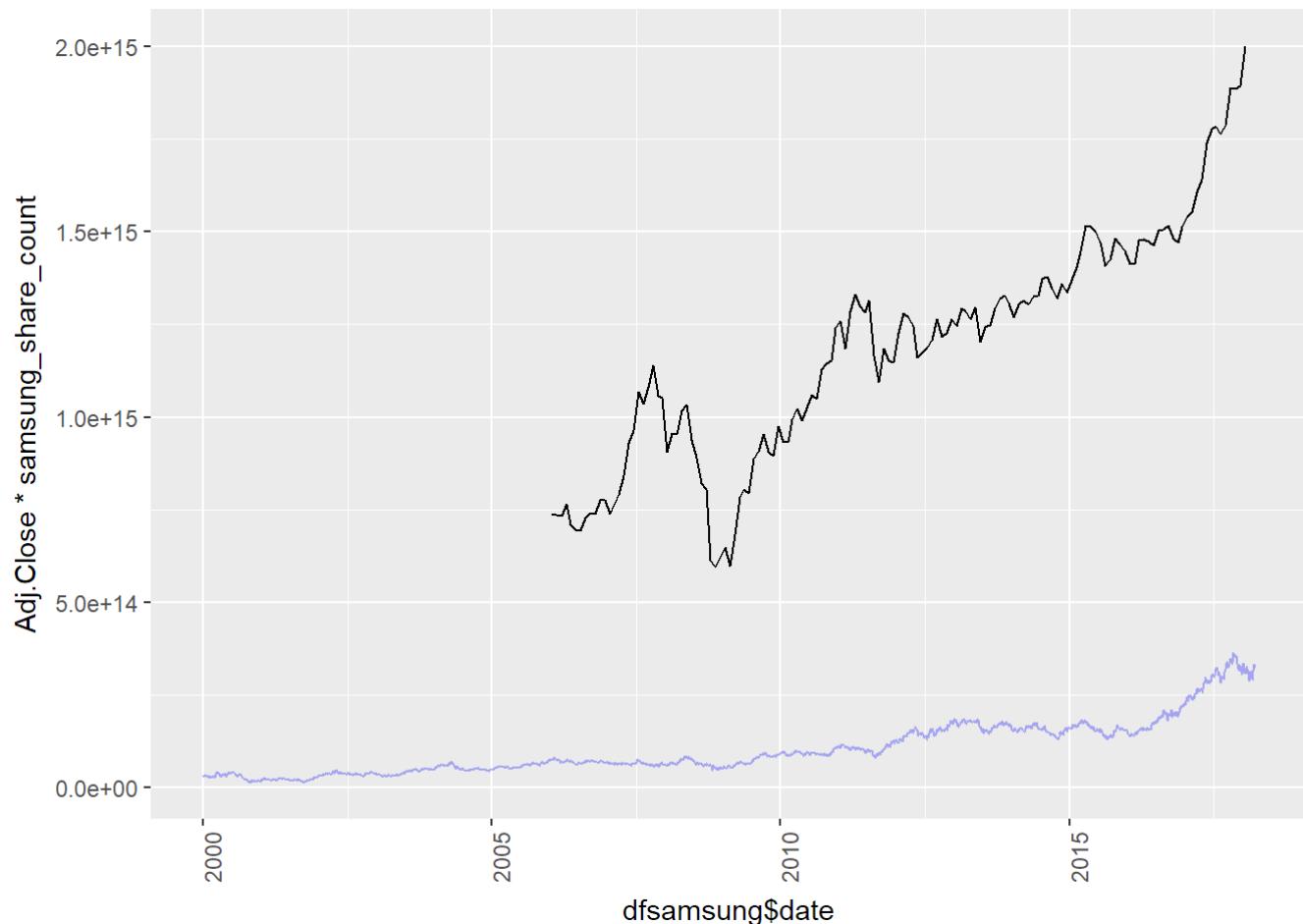
What was the strongest relationship you found?

Traded money and traded volume have much in common in the trend. It would be much clearer once we put them on the same plot in multivariate analysis.

And I got investment ideas using the finding that there seems to be a cycle in Samsung stock. Besides, there seems to be a 10-year-spanning cycle in total Korea stock market.

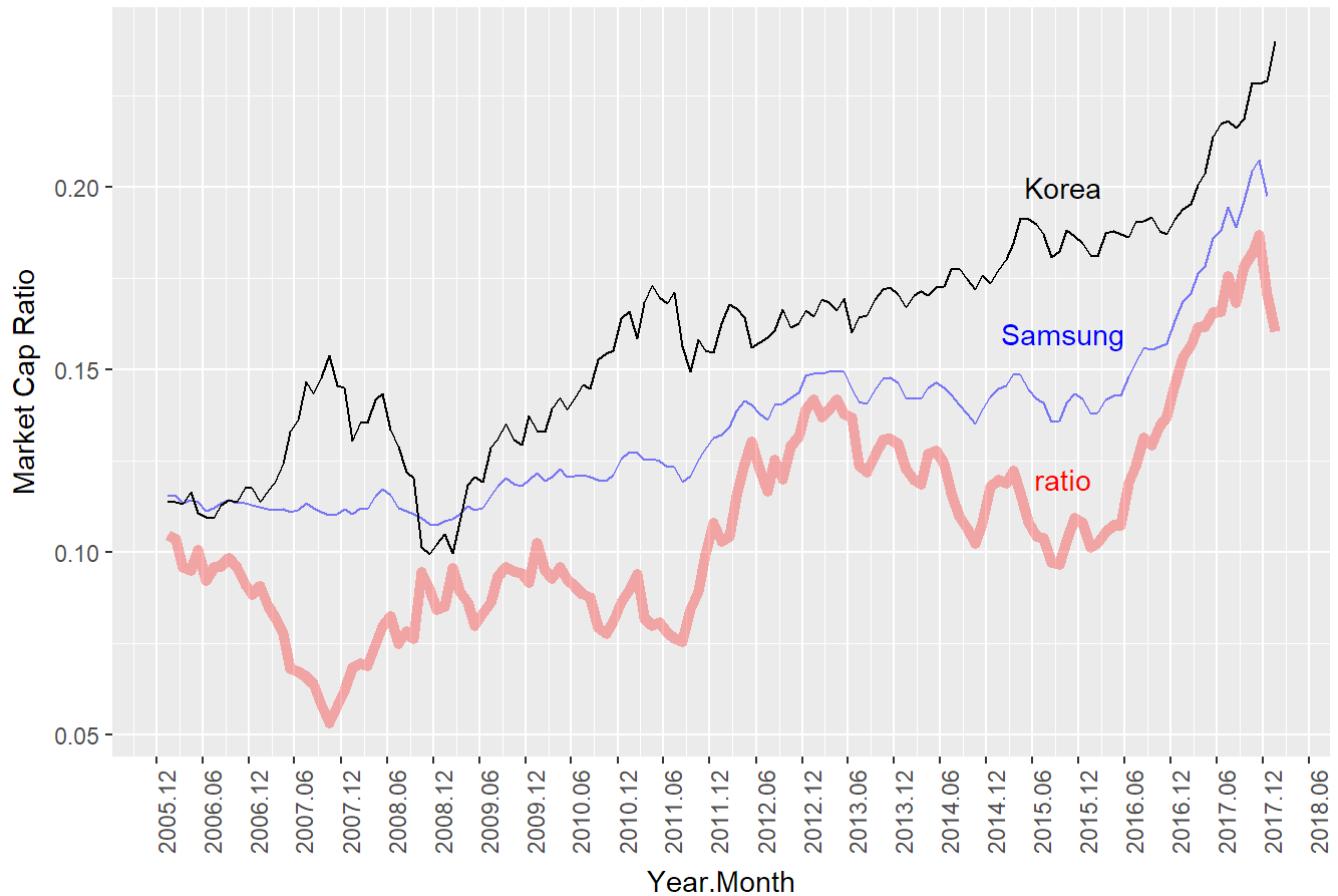
Multivariate Plots Section

Now, let's put Samsung and the whole Korea market together.



OK. we can see Samsung might have much portion against the whole stock market but not sure it is around 0.1 or 0.3 along with the time. Let's clarify that by calculating the ratio over the time.

The trend of Samsung ratio

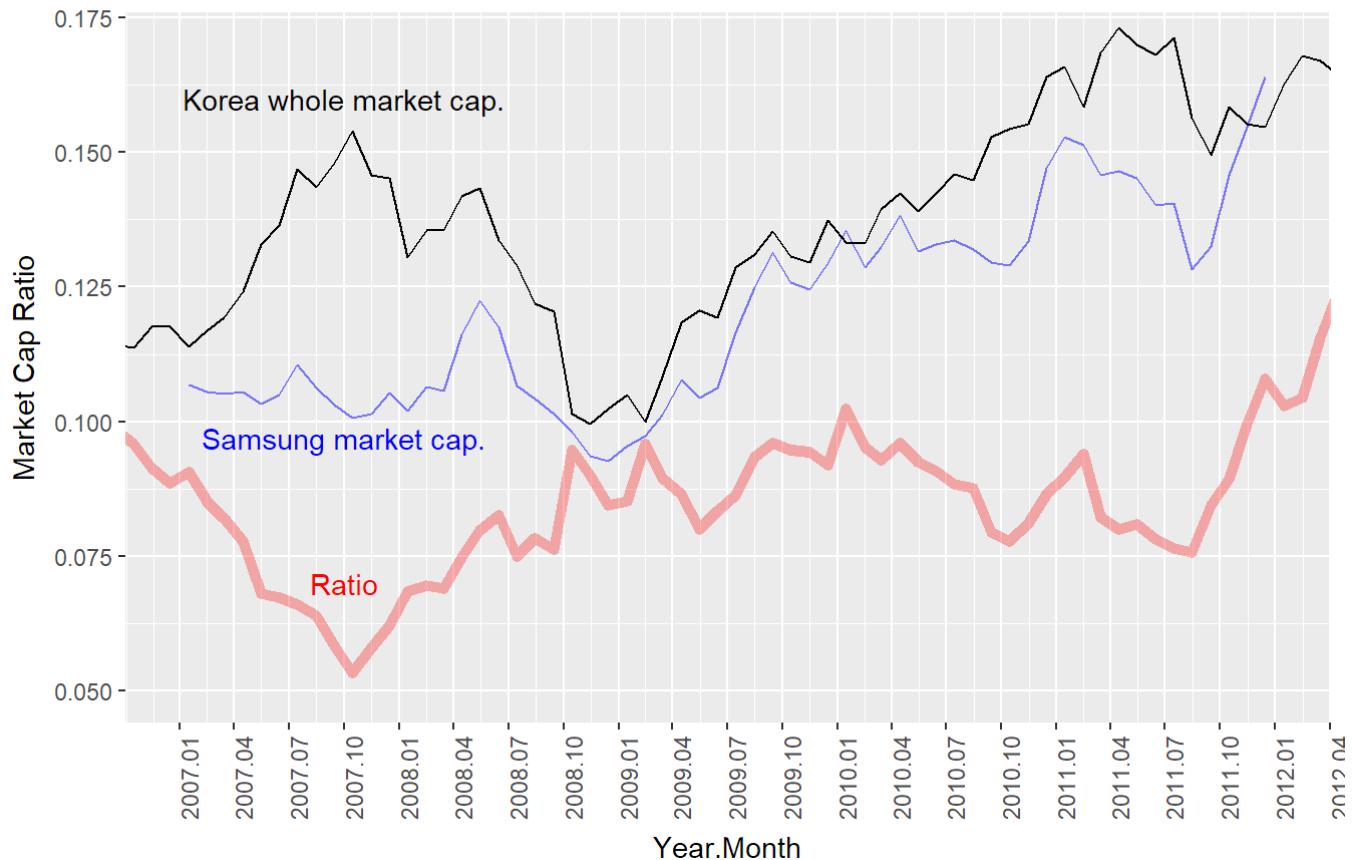


This plot depicts how the big ratio of Samsung Electronics against the whole Korea stock market has been established. As expected, the more recent the bigger ratio but there are several interesting parts.

1. There seems to be a cycle in the plot, spanning about 4 years. In other words, Samsung stock boosted every 4 years. I don't think the stock price solely can reveal this cycle.
2. There is a deep pit nearing zero in 2007~2008. I guess it overlaps the global crisis and the date is much interesting - the drop was already made 1 year ahead of the crisis. Let's zoom in that part, with other trends overlapped as reference.

Zooming in the Ratio of Samsung

Market Cap ratio of Samsung to the whole Korea stock market



The Lehman Brothers bankruptcy broke out in September 2008. The two other lines, market cap of Korea (black) and market cap of Samsung (blue), both dropped just after that period. However, the market cap ratio at that time already was climbing up after the drop 1 year ago.

Multivariate Analysis

Talk about some of the relationships you observed in this part of the investigation. Were there features that strengthened each other in terms of looking at your feature(s) of interest?

This is stock market data where people's collective mind is playing everywhere, there would not be variables which show strong relationship like physics.

Rather, I could see some cycle over time. 1) Samsung's money absorbing cycle every 4 years. 2) 10 years cycle of Korea stock market. These two might be useful guide for investment.

Were there any interesting or surprising interactions between features?

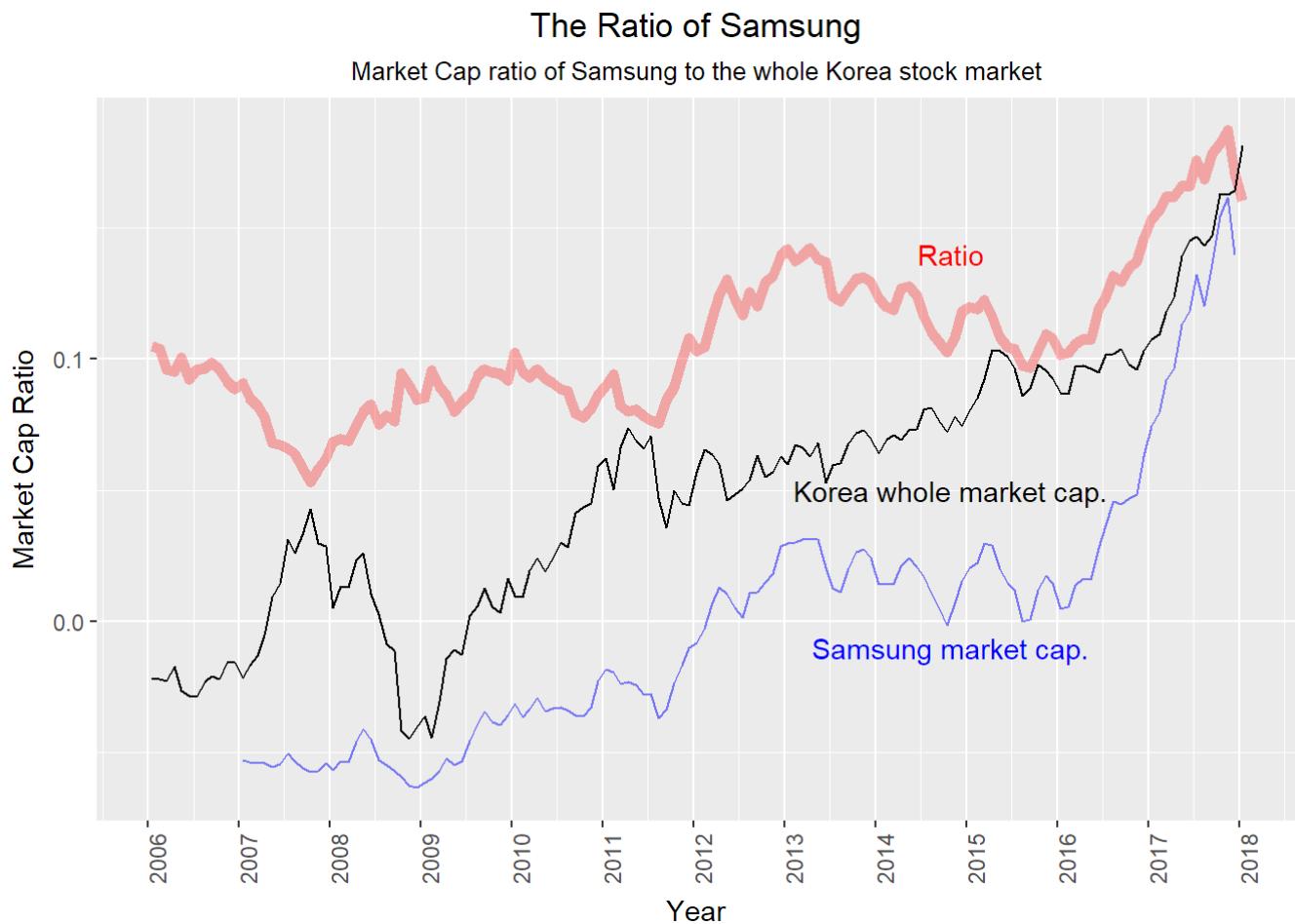
Most surprising was mismatched timing of the trends of price and market cap ratio. Seemingly price is a superficial metric and more essential part might be market cap ratio which already made a motion and direction before certain price forms. Of course, more data and research would be required.

OPTIONAL: Did you create any models with your dataset?
 Discuss the
 strengths and limitations of your model.

The 2 cycles, 4 years of Samsung and 10 years of Korea, could be the candidates of the model. We would need more data to verify that.

Final Plots and Summary

Plot One



Description One

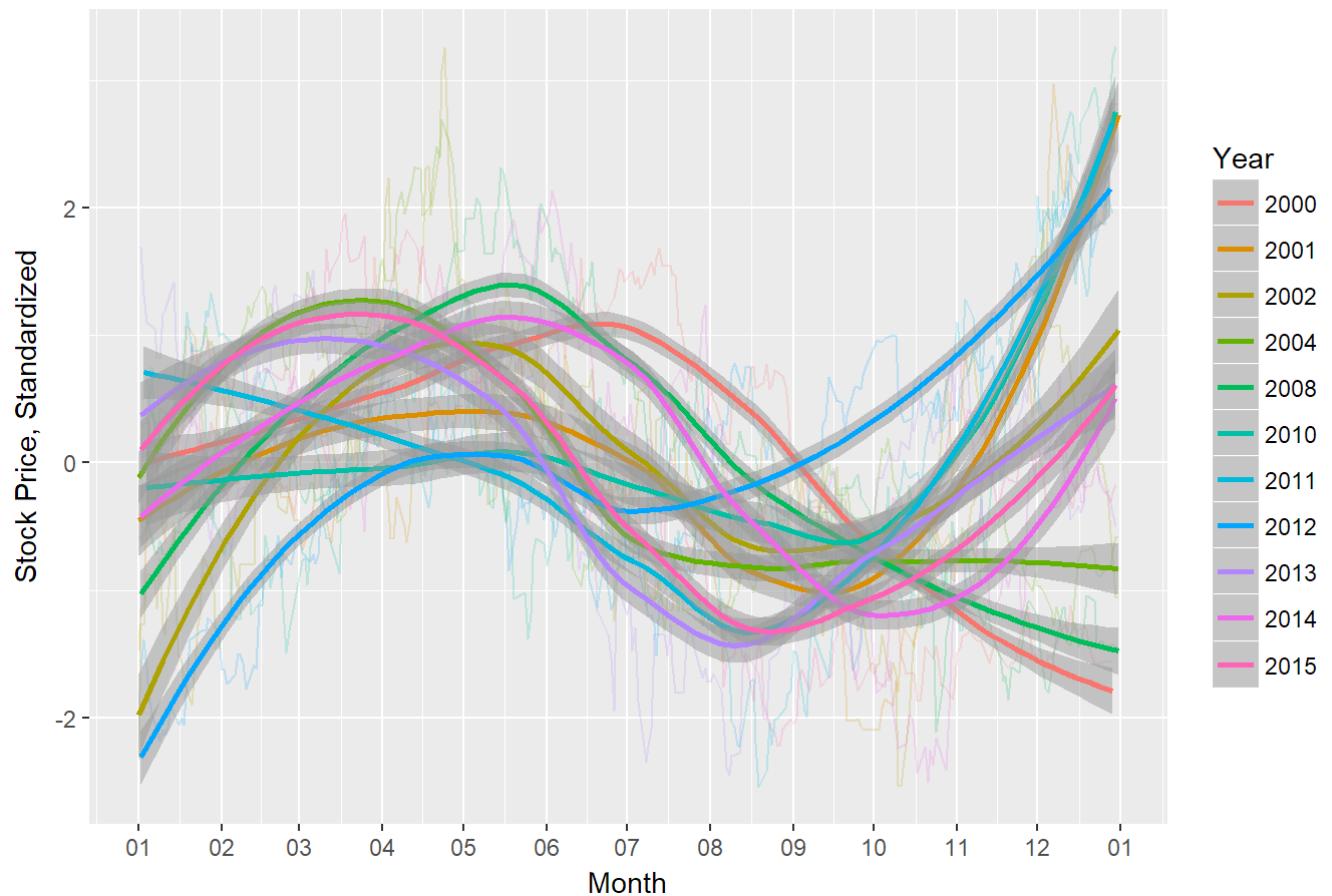
The current ratio of Samsung Electronics' market value against the entire Korea stock market is 17%. It is a huge portion. Then what about the ratio in the past? In this project I tracked the trend of the ratio over time. In the plot, red line is the ratio ranging 0.05 ~ 0.20. Black line indicates the entire market capitalization of Korea stock market and blue line corresponds to Samsung Electronics' market capitalization. Later two lines are just showing overall trends, irrelevant to the y-axis unit.

The red line shows a cycling pattern of the ratio, repeating every 4 years. When we see black and blue lines, the moment that the ratio hit the lowest points, the gap between black and blue is widest. In my interpretation, Samsung's value boosts up 1 year later of a big influx of money into the Korea stock market. It looks like pumping water into a lake which in fact is mostly shallow but has big depth area called Samsung. This plot is impressive to me because that implies not just the weight of Samsung in Korea but also how they behave each other over time.

The initial money influx was not targeting Samsung according to the plot, then what company would that be? further research is required.

Plot Two

Seasonality in Samsung Stock Price

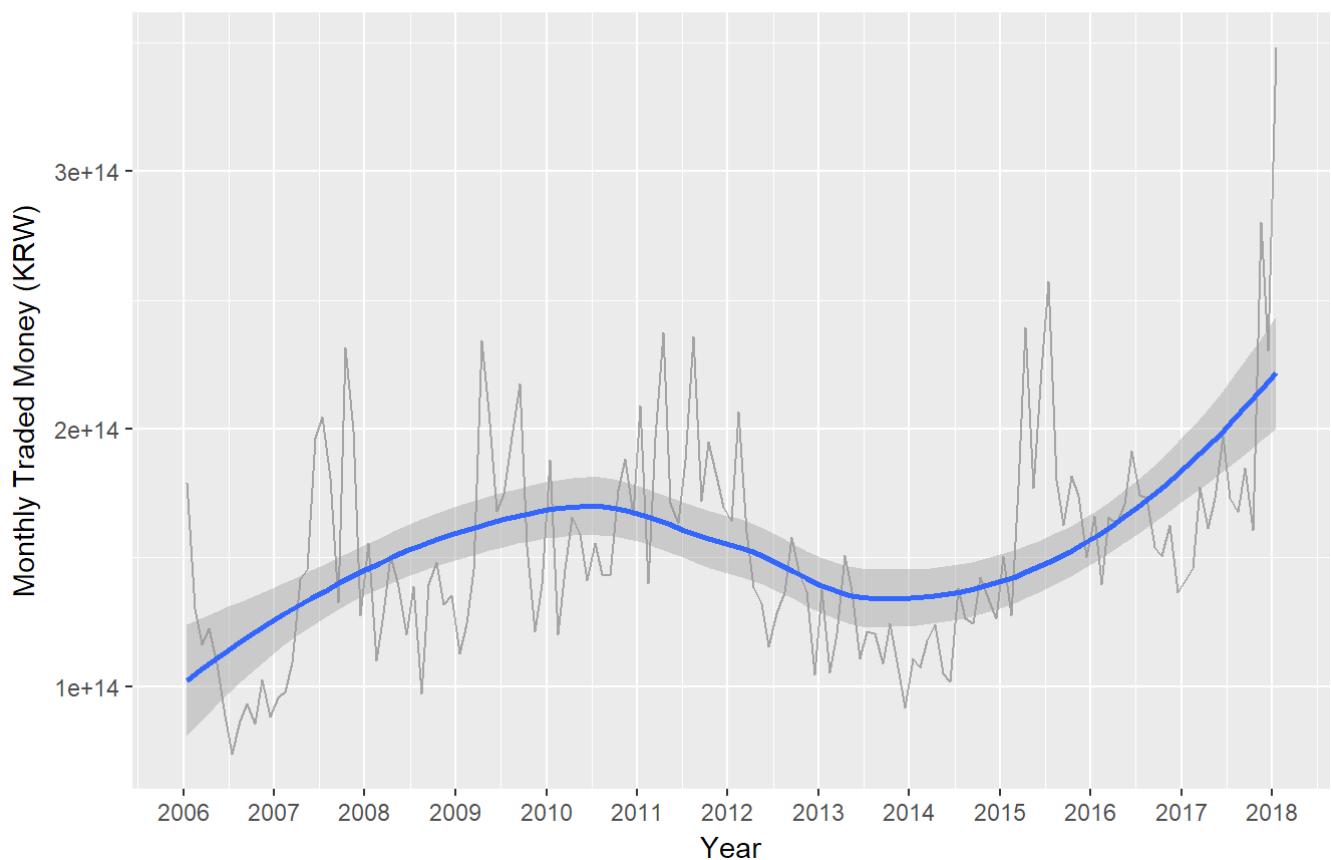


Description Two

The reason I am interested in this plot is that this might be a clue to invest Samsung better. Of course we already checked other pattern exist but the above looks prominent and frequent. There might be some reason behind, for example rolling out new product aiming at certain season.

Plot Three

Money Tide over 10 years
 Cycling pattern of money flow in Korea stock market



Description Three

Together with the traded volume plot which has the same pattern, this plot might be able to guide investors how the landscape is changing. It totally feels like tide. Presently people are shifting their wealth over to stock market, then from where? real estate maybe? Another research would be required.

Reflection

Main focus here is the ratio of Samsung's value against total value of Korea stock market over time. During the exploration, there were other interesting findings not just the shape of the ratio plot. One thing most impressive is that the ratio could become a guide for investors because certain cycle was detected and the ratio pattern preceded ahead of the price. Validation is necessary of course.

About 60% of the time was spent to prepare the data, 3 files. Adapting to Yahoo Finance API was most tedious part because of my limited knowledge on stock market and weak documentation of Yahoo Finance. I heavily relied on many blogs.

My exploration and analysis are lack of generalization. The pattern and cycle I found have to be validated with other data such as other company data like Hyundai, LG, and other global companies in each stock market. That will be the future items for research.

During the exploration, again I realized that R is easy, quick and intuitive. Because Python has been in my comfort zone, I searched Python equivalent to ggplot library and found “plotnine” (<http://pltn.ca/plotnine-superior-python-ggplot/>). The feature and grammar is the same and I will use that library in my other projects at early stage of data exploration .

References

- Samsung Electronics on Yahoo Finance : <https://finance.yahoo.com/quote/005935.KS?p=005935.KS>

- Yahoo Finance and python, guide : <https://ntguardian.wordpress.com/2016/09/19/introduction-stock-market-data-python-1/> (<https://ntguardian.wordpress.com/2016/09/19/introduction-stock-market-data-python-1/>)
- Korea stock index, KOSPI historical data download : http://kosis.kr/statHtml/statHtml.do?orgId=343&tblId=DT_343_2010_S0001# (http://kosis.kr/statHtml/statHtml.do?orgId=343&tblId=DT_343_2010_S0001#)
- Company information in KOSPI, KOSDAQ : www.krx.co.kr
- Plotting seasonality : <https://stackoverflow.com/questions/13520639/ggplot-year-by-year-comparison> (<https://stackoverflow.com/questions/13520639/ggplot-year-by-year-comparison>)
- Value standardization : <https://stackoverflow.com/questions/35775696/trying-to-use-dplyr-to-group-by-and-apply-scale> (<https://stackoverflow.com/questions/35775696/trying-to-use-dplyr-to-group-by-and-apply-scale>)