UCLA Extension – Data Science and Visualization

COM SCI X 450.2

Name: Fengtian Zhang

Instructor: Ali El-Annan

Date: 2019-06-02

Influence of US-China Trade War on main stocks from Basic Material, Energy, Health Care and Utility

Contents

[1. Introduction: 3](#_Toc10969559)

[2. Data Collection 4](#_Toc10969560)

[2.1 US-China Trade War timeline and positive-negative coefficient 5](#_Toc10969561)

[2.2 Stock market historical data collection 5](#_Toc10969562)

[3. Data Analysis and Visualization: 5](#_Toc10969563)

[3.1 Data Import and Process 6](#_Toc10969564)

[3.2 Data Storage 10](#_Toc10969565)

[3.3 Data Visualization 11](#_Toc10969566)

[4. Conclusion 14](#_Toc10969567)

[5. Reference 14](#_Toc10969568)

# Introduction:

Since US implements the first China-specific tariffs increase on July 6th, 2018, the trade war is on. Even it went through a short period of truce, there is no deal made, and trade war is back on with the new tariff increase in May. No matter who should be blamed for the trade war, it has affected the stock market, especially tech giants like Apple and Google. There is no telling where the trend of trade war will go. Some say it will last for more than a decade, some say it will end with the Trump-Xi meeting on G20. Investors start to realize that stock market is sensitive to the trade war signal and try to pull the money out of tech giants to some “safer” stocks which could be more predictable than trade war. The purpose of this article is to answer the following question:

**which one of the four industries (Basic Material, Energy, Health Care and Utility) shows a weak influence by trade war? And furthermore, which company in that industry is least affected by trade war?**

The article starts from collecting trade war timeline and summarize a positive-negative coefficient based on the event. The top 10 stocks for each industry based on market capital will be collected as well. It will then find out the dates on which the Dow Jones Index(DJI) shows the same trend as the positive-negative coefficient. Based on the dates that picked, analyzing the average percentage change of stock price for each industry, and calculate the pearson correlation coefficient to determine the correlation between DJI and each industry on the picked dates. After finding out the least affected industry, further calculation will be performed on each top 10 stock inside that industry.

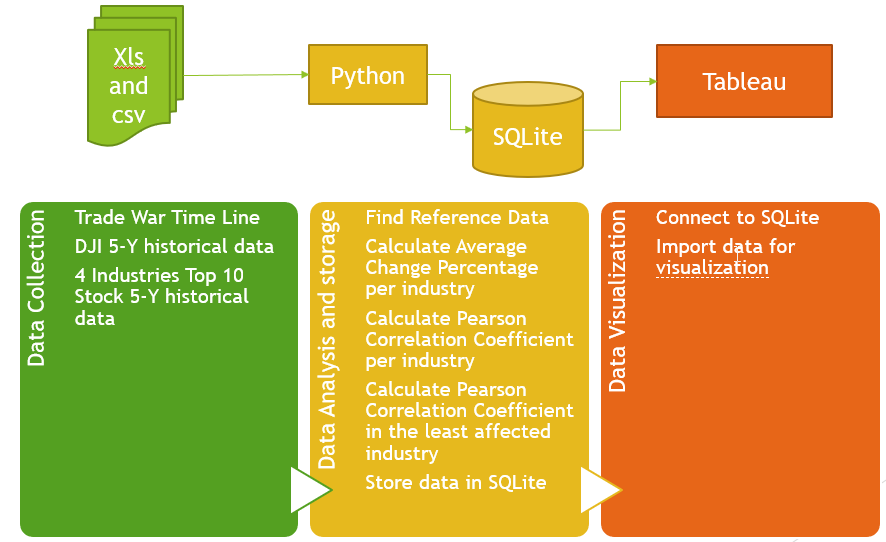


Figure Basic Process Flow

# Data Collection

US-China trade war is a complicated topic. It involves policies, tariff and retaliation. It is difficult to consider every possible factor to analyze the true effect on each industry. It is also hard to analyze every stock inside every industry, as the number of companies inside is large and each company’s relationship with China is unique. Thus, the following assumptions has been made in order to simplify the problem:

* Dow Jones Index is correctly representing the trending of the US stock market
* Dow Jones Index is sensitive to the US-China Trade War signal
* For any US-China Trade War event, there is only positive and negative effect, no neutral effect.
* Only investigate the stocks in Basic Material, Energy, Health Care and Utility
* The top 10 stocks in each industry based on market capital is good enough to represent its industry

The following date collection is done based on the assumptions above.

## US-China Trade War timeline and positive-negative coefficient

The US-China Trade War timeline is based on an article reported by China Briefing[1]. It summarizes the trade war related big event started from July 6th, 2018. Even though there is data from Reuters, which includes the key dates and also S&P 500 change, I prefer to use the raw value in this article[6]. For each event, if it involves the tariff, retaliation and bad effect on market, it will be marked as -1 on that date. If it is about dropping tariff, trade war truce and positive signal for a deal, it will be marked as 1 on that date.

## Stock market historical data collection

The stock market data is collected from Yahoo finance[2]. For Basic Material, Energy, Health Care, and Utility, based on the market capital, the top 10 companies with the 5-year historical stock data are collected. If the 5-year historical data is not available, the stock has to have the data as early as July 6th, 2018, otherwise it will be dropped and replaced by the next stock with a high market capital. Dow Jones Index 5-year data is also collected from Yahoo finance.

# Data Analysis and Visualization:

The data is stored in excel and csv. Using python, import the data, calculate the average price change percentage based on reference dates, and calculate the pearson correlation coefficient. Once calculation is finalized, store the data in SQLite and visualize the results using Tablea. For details, one can refer to the python code and tablea file attached.

## Data Import and Process

The following steps are done to process the data and find out the correlation:

* Import Data

The positive-negative coefficient is stored in xls file “PositiveNegativeIndex.xlsx”. And the original stock data is stored in csv file. Using pandas.read\_xls and pandas.read\_csv to import data for further processing. The following shows an example of data imported:

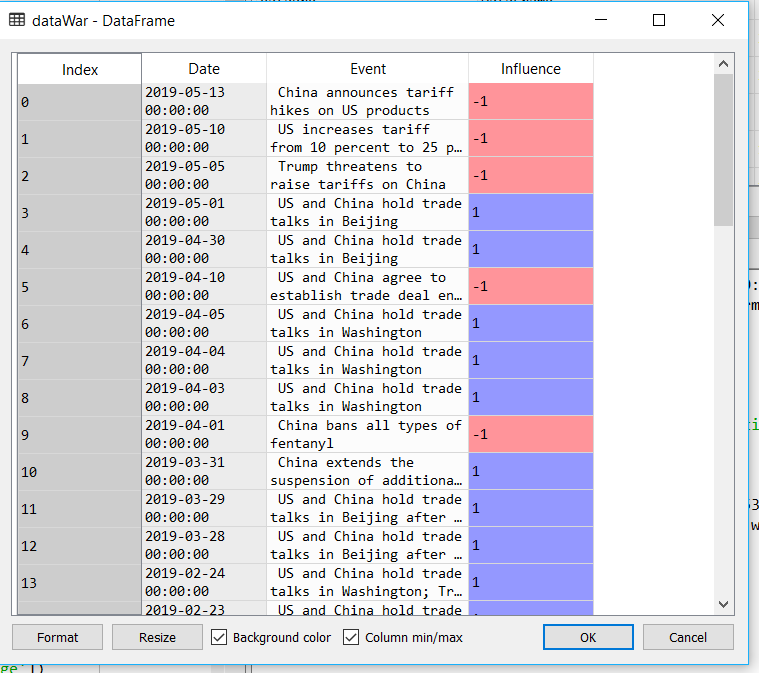


Figure 2 Timeline and influence data from excel

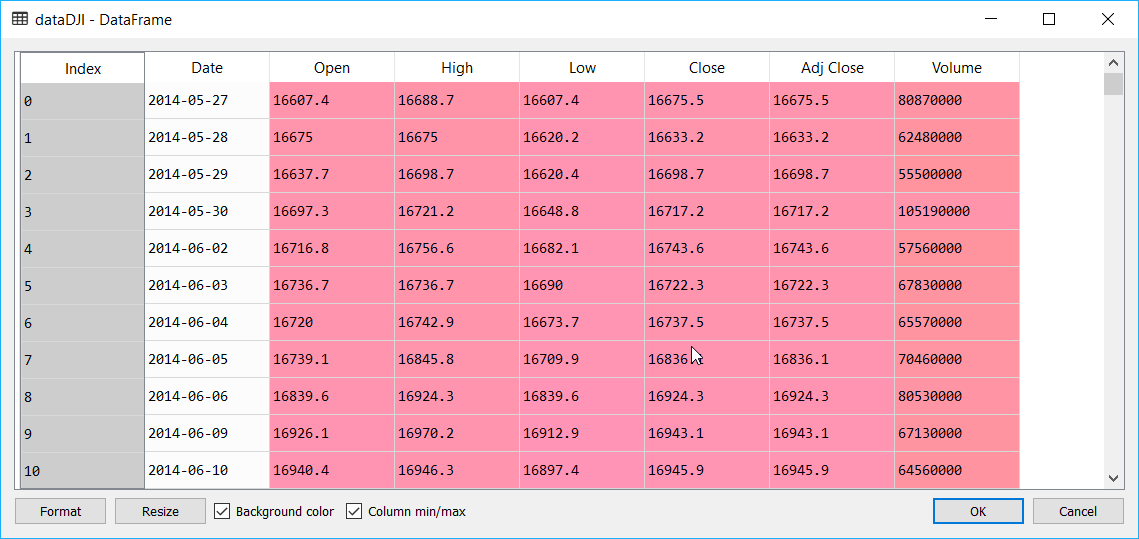


Figure 3 Raw Data for Dow Jones Industrial Index from csv

* Retrieve the reference dates

Calculate the change percentage of DJI on each day using the following equation:

Change=(Close Price-Open Price)/Open Price

Compare with the positive-negative coefficient and get a processed value of DJI on the dates which align with the change of trade war positive-negative affects. The following is an example of DJI data based on reference dates:

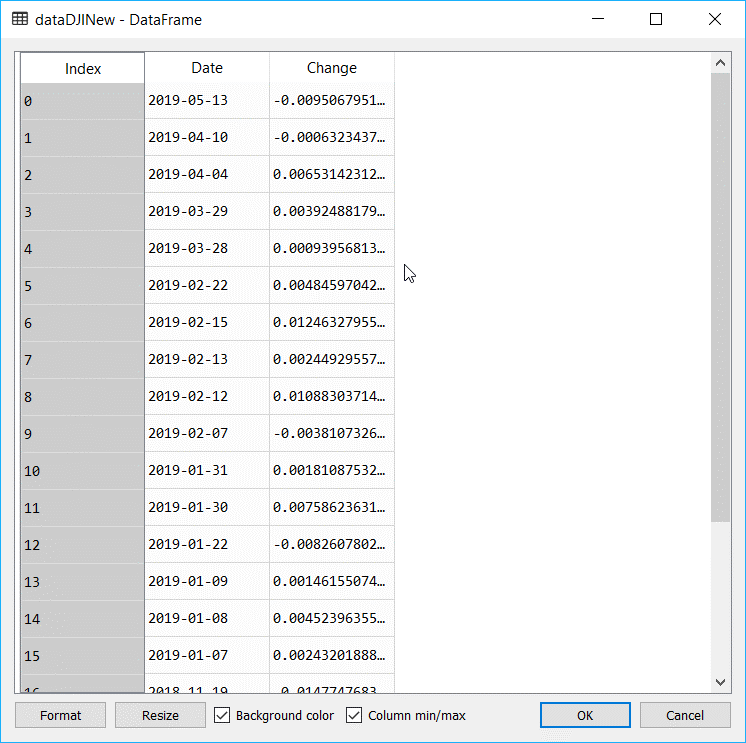


Figure 4 Process DJI data with percentage change on reference dates and influence

* Calculate the average percentage change per industry

After getting the dates of reference, calculate the average change per industry on those dates. The following is an example of the data calculated in one industry:

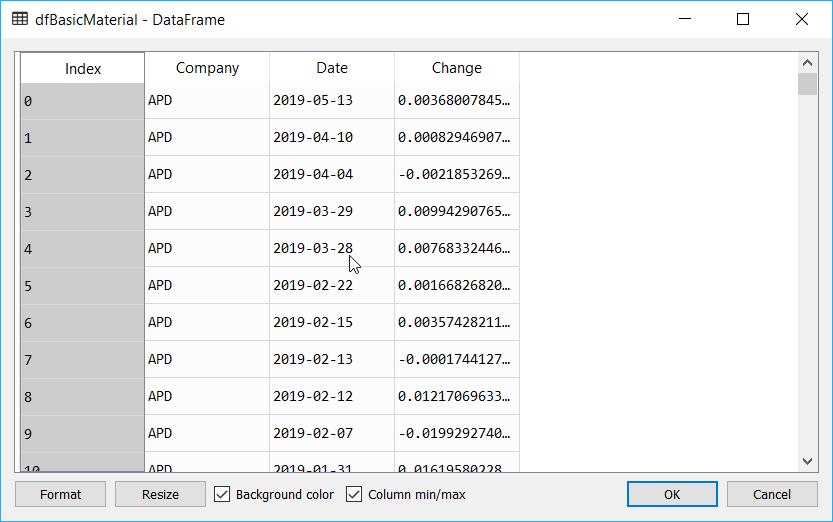


Figure 5 Basic Material processed data with company, date and percentage change

* Calculate the Pearson Correlation Coefficient

Use scipy.stats.pearsonr(x, y) to calculate the pearson coeffiecnt between DJI average change and average change per Industry.

Pearson coefficient is calculated using the formula below:

where mx is the mean of the vector x and my is the mean of the vector y.[3]

scipy.stats.pearsonr(x, y) gives a result of [r,p-value], where r is Pearson correlation coefficient. R will be from -1 to +1, where +1 shows extreme positive correlation between x and y and -1 shows extreme negative correlation. The smaller the absolute pearson correlation coefficient value is, the less correlation between x and y will be. P-value roughly indicates the probability of an uncorrelated system producing datasets that have a Pearson correlation at least as extreme as the one computed from these datasets. However, this p-value will not be accurate when the sample is less than 500. Since the there are only 21 reference dates, this p-value is not considered for analysis.[4] The following is an example of calculated person coefficient data in Utility per company:

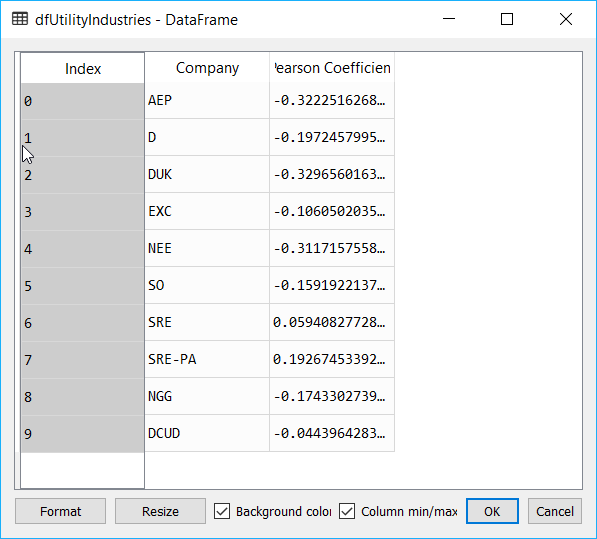


Figure 6 Pearson Correlation Coefficient Calculation for Utility per company

## Data Storage

Using python, the data required for visualization is then stored in SQLite database. It includes the positive-negative coefficient, average change for DJI and each industry, and the pearson correlation coefficient calculation between each industry and DJI. The pearson correlation coefficient calculation on top 10 company in Utility is also stored.

## Data Visualization

Using tabulea, the following plot and table is created:

* Positive-Negative Coefficient vs DJI price change:

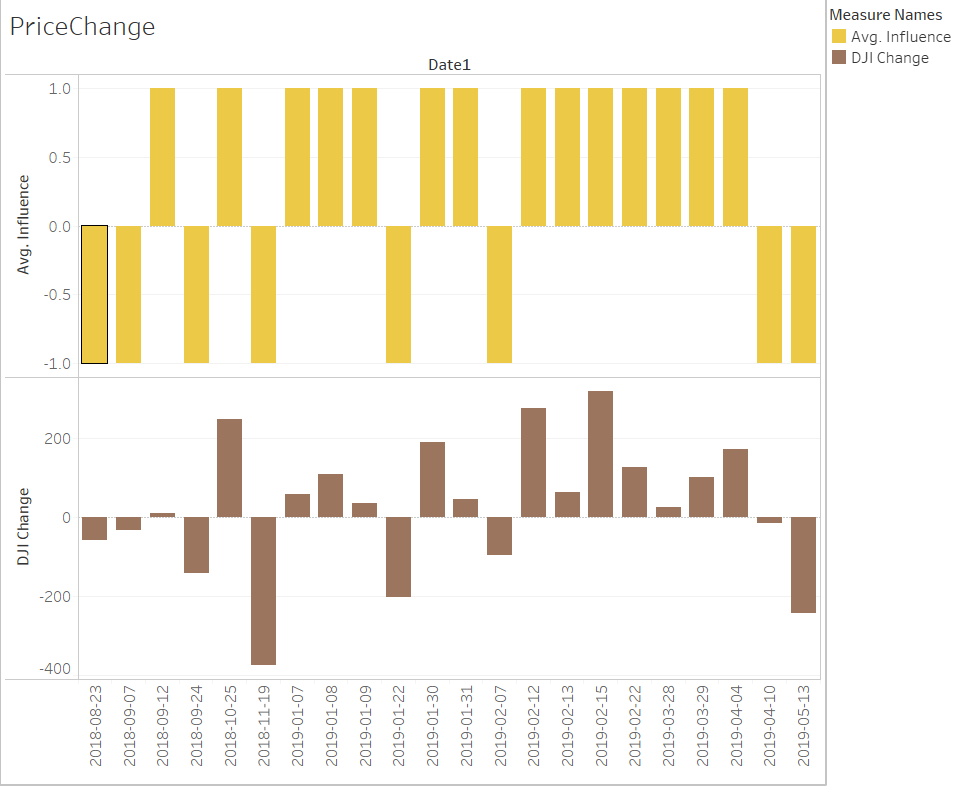


Figure Positive-Negative Coefficient vs DJI Price Change

* Change percentage per Industry vs DJI change percentage

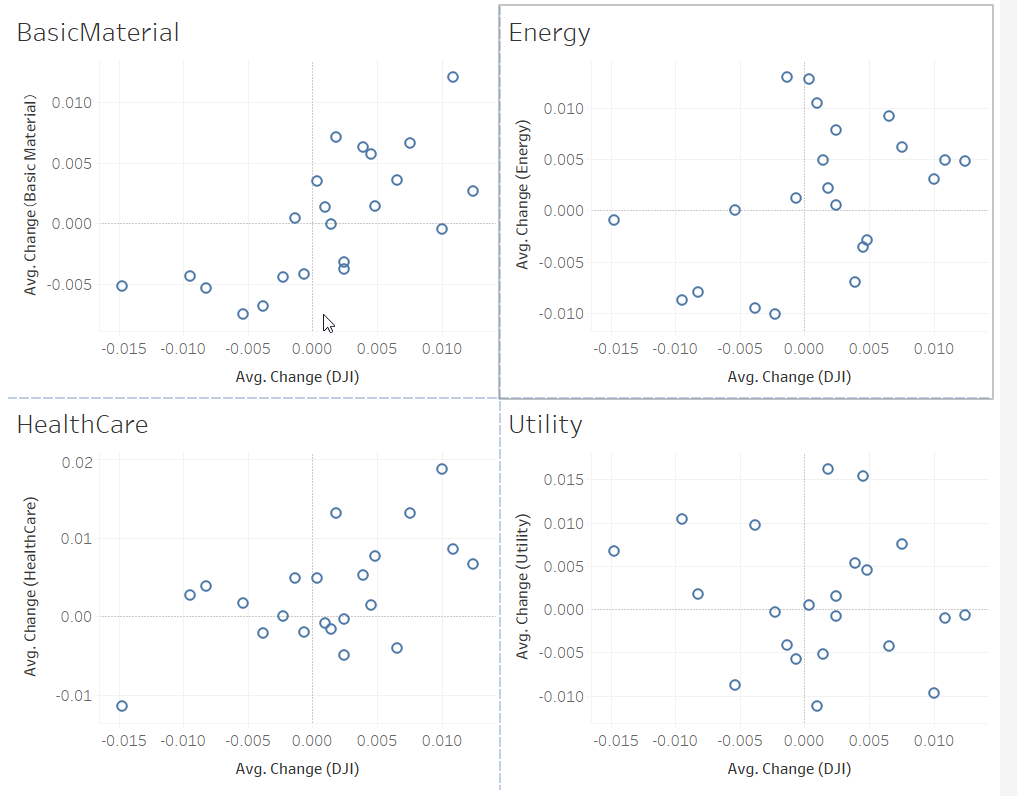


Figure Coorelation between industry and DJI

* Pearson Correlation Coefficient per industry vs DJI and Pearson Correlation Coefficient between top 10 companies and DJI

From the figure below, one can see that for industry, Utility shows a lowest absolute pearson correlation coefficient, and for utility, company DCUD shows the lowest absolute pearson correlation coefficient.

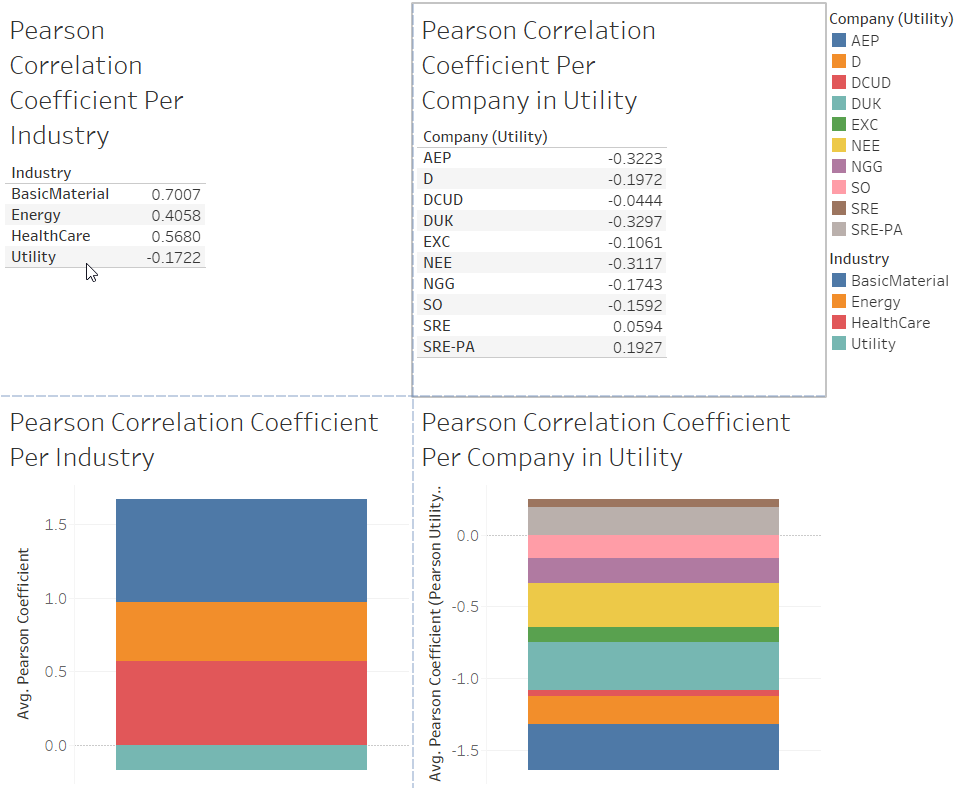


Figure Pearson Coorelation Coefficient Calculation

# Conclusion

As shown in section 3, Utility is less affected by US-China Trade war, and in its top 10 company, Dominion Energy, Inc. UT SER A 081519 (DCUD) is least affected based on the data collected.

However, as stated in section 2, the analysis done in this article relies on the assumption that DJI is sensitive to trade war, which means the good and bad effect will happen immediately after the news, and sometimes which is not so accurate. On the other hand, since I inserted the positive-negative index based on my own view, different opinions on this item may also affect the reference data, which could affect the final results. The work could be improved by the following:

* Revise the reference data. In this article, only DJI is used. There are other data could be used as reference, like NASDAQ and S&P 500.
* Research and find out the relationship between the China-US tariff list & entity list and the US companies. It could be quantized by using the dollar spent previously on the imports related to those lists per industry before and after tariff.
* Find out the industries that has least China import before the tariff, and investigate the relationship between reference input and their stock price change on the reference dates.
* Also analyze the weekly price change and compare against the daily price change to rule out some other factors.

# Reference

[1] https://www.china-briefing.com/news/the-us-china-trade-war-a-timeline/

[2] https://finance.yahoo.com/

[3] <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.pearsonr.html> scipy.stats.pearsonr

[4] <https://blog.csdn.net/zhouwenyuan1015/article/details/65938847> python learning - pearsonr(x,y) pearson correlation coefficient calculation

[5] <https://www.zhihu.com/question/19734616> How to understand pearson correlation coefficient calculation

[6]<https://www.reuters.com/article/us-usa-trade-china-timeline/timeline-key-dates-in-the-us-china-trade-war-idUSKCN1SE2OZ>