



**STEVENS**  
INSTITUTE OF TECHNOLOGY  
THE INNOVATION UNIVERSITY®

# Which industries are more vulnerable under 2020 financial crisis in America?

*MIS 637 Final Project*

Instructor: Prof. M. Daneshmand  
Accomplished by: Ling Ao





# Content

- **Problem Statement (Business understanding phrase)**
- **Data Collection (Data understanding phrase)**
- **Data Processing (Data preparation phrase)**
- **Methodology (Modeling phrase)**
- **Software & Package**
- **Result**
- **Discussion (Evaluation phrase & Development phrase)**
- **Conclusion**
- **Reference**

# Problem Statement\_ Business Understanding

## Case Summary

Last updated at : 2020-05-07 23:02 EDT [Data source](#)

	Cases	Deaths	Recovered	Fatality
US	+29,699 1,291,496	+1,963 76,650 ⓘ	+3,804 174,888	5.9%
World	+92,650 3,755,341	+6,592 263,831	+39,328 1,284,741	7%

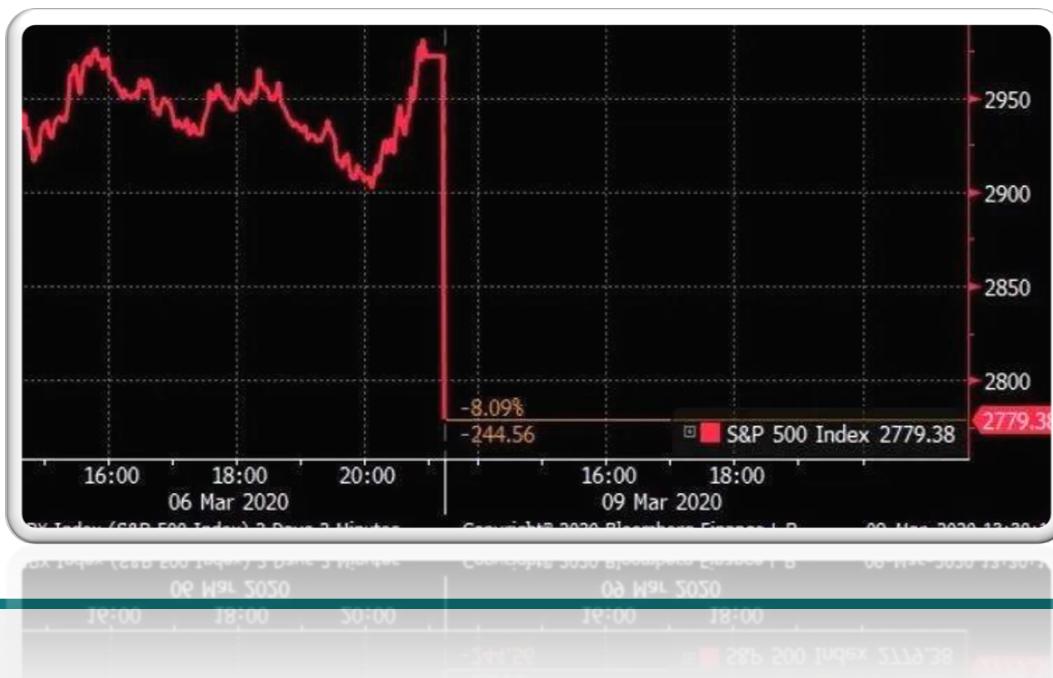
A new kind of virus, called COVID-19, is spreading around the world. Based on the information from 1point3Acres, there are around 1.29 million cases in America up to May 7<sup>th</sup>.

Most cities have imposed quarantine order during this special period and lots of companies let employees work from home.



# Problem Statement\_ Business Understanding

- In last three months, the economy of the United States has been greatly affected due to COVID-19
- In March 9<sup>th</sup>, the S&P 500 fell by 7% upon opening, triggering a circuit breaker that halted trading for 15 minutes for the first time since the 2008 financial crisis. In May 12<sup>th</sup> , 16<sup>th</sup>,18<sup>th</sup> S&P 500 triggered the circuit breaker again
- The financial crisis in America has came



**Research Question:**  
**Which industries are more vulnerable under 2020 financial crisis in America?**

# Data Collection\_ Data understanding

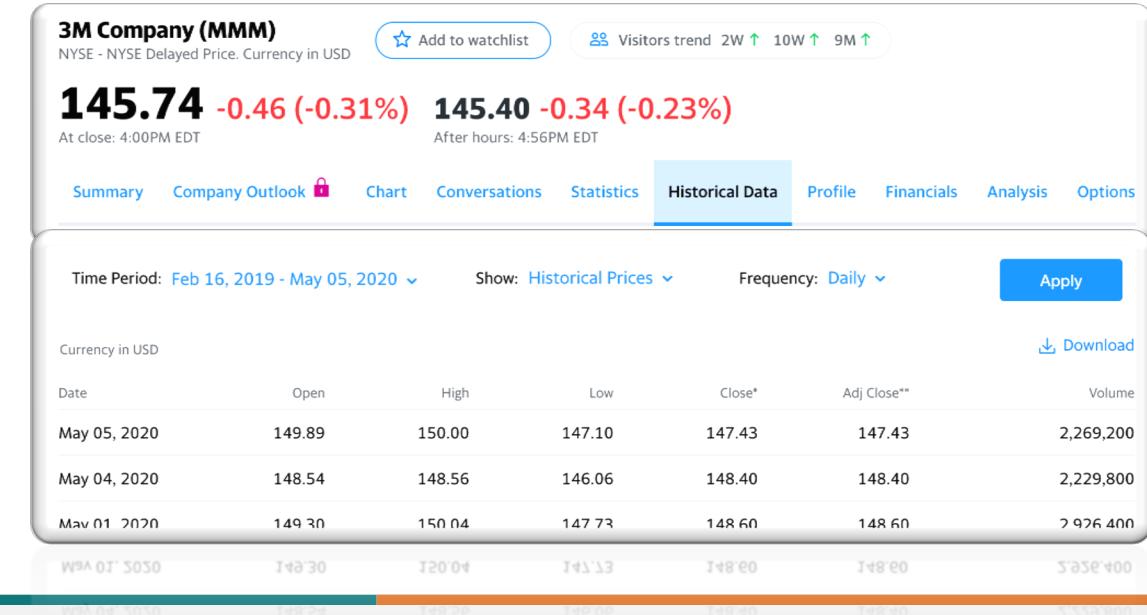
S&P 500 component stocks [edit]

Symbol	Security	SEC filings	GICS Sector	GICS Sub Industry	Headquarters Location	Date first added	CIK	Founded
MMM	3M Company	reports	Industrials	Industrial Conglomerates	St. Paul, Minnesota	1976-08-09	0000066740	1902
ABT	Abbott Laboratories	reports	Health Care	Health Care Equipment	North Chicago, Illinois	1964-03-31	0000001800	1888
ABBV	AbbVie Inc.	reports	Health Care	Pharmaceuticals	North Chicago, Illinois	2012-12-31	0001551152	2013 (1888)
ABMD	ABIOMED Inc	reports	Health Care	Health Care Equipment	Danvers, Massachusetts	2018-05-31	0000815094	1981
ACN	Accenture plc	reports	Information Technology	IT Consulting & Other Services	Dublin, Ireland	2011-07-06	0001467373	1989
ATVI	Activision Blizzard	reports	Communication Services	Interactive Home Entertainment	Santa Monica, California	2015-08-31	0000718877	2008

1) Information about all 505 companies contained in S&P500 index list in Wikipedia

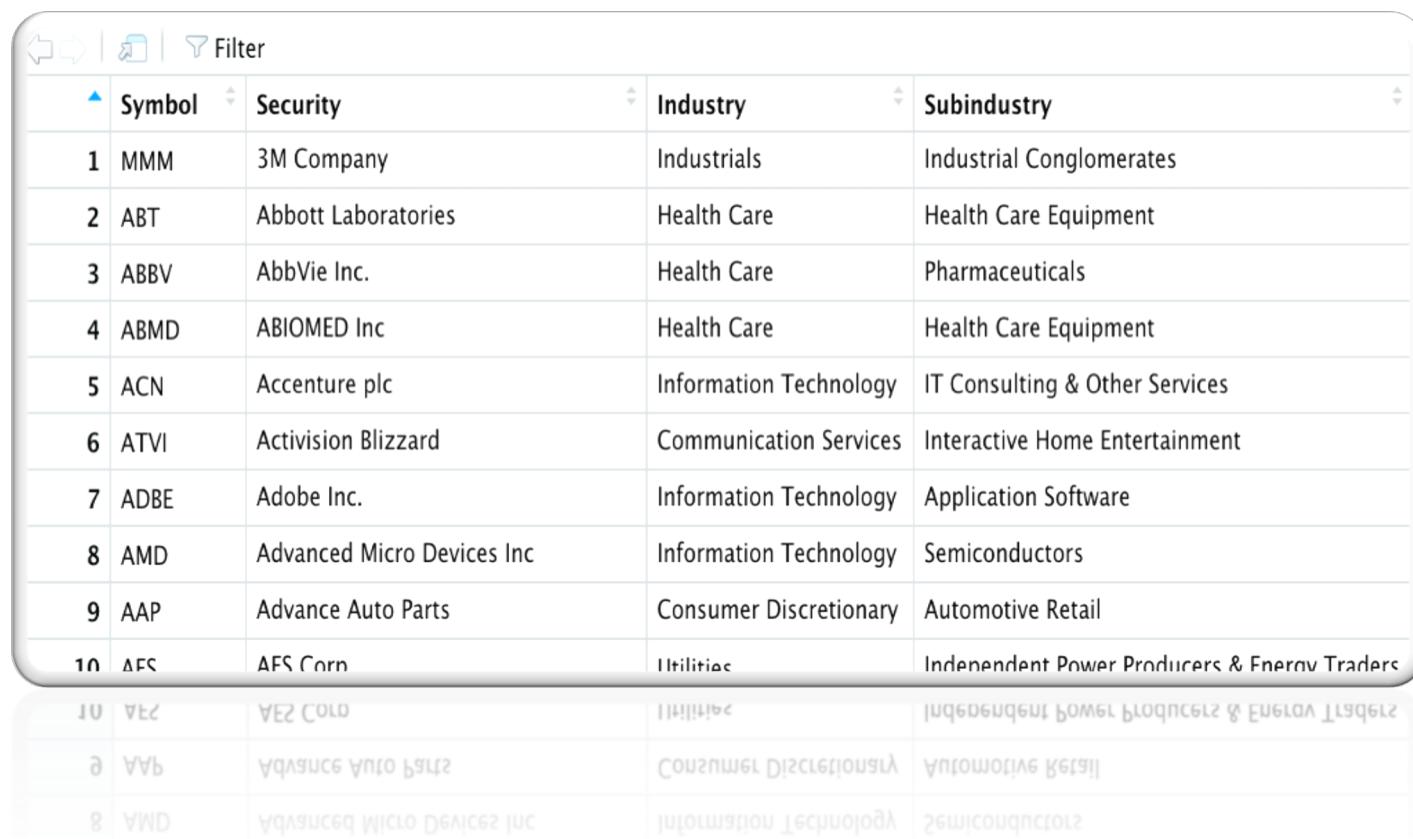
Contain: Stock Symbol, Security (company name), SEC filings, GICS Sector, GICS Sub Industry, Headquarters Location, Date first added, CIK and funded year.

2) Weekly adjusted close stock price of each share in S&P500 from February 17th until May 5th from Yahoo Finance website.



# Data Processing\_ Data Preparation

- 1) Imported “S&P 500 data. csv” into R
- 2) Deleted the possible N/A value contained in the dataset
- 3) Chose the first four columns of the dataset and named them as 'Symbol', 'Security', 'Industry', 'Subindustry' respectively
- 4) Used “unique” function to delete the possible repeat column in the dataset.



The screenshot shows a data visualization interface with a table of S&P 500 company data. The table has columns: Symbol, Security, Industry, and Subindustry. The data includes:

	Symbol	Security	Industry	Subindustry
1	MMM	3M Company	Industrials	Industrial Conglomerates
2	ABT	Abbott Laboratories	Health Care	Health Care Equipment
3	ABBV	AbbVie Inc.	Health Care	Pharmaceuticals
4	ABMD	ABIOMED Inc	Health Care	Health Care Equipment
5	ACN	Accenture plc	Information Technology	IT Consulting & Other Services
6	ATVI	Activision Blizzard	Communication Services	Interactive Home Entertainment
7	ADBE	Adobe Inc.	Information Technology	Application Software
8	AMD	Advanced Micro Devices Inc	Information Technology	Semiconductors
9	AAP	Advance Auto Parts	Consumer Discretionary	Automotive Retail
10	AES	AES Corp	Utilities	Independent Power Producers & Energy Traders
11	AEC	AECO INC	Construction	Independent Power Producers & Energy Traders
12	AAI	AAI CORP	Consumer Discretionary	Automotive Retail
13	WMA	WILCO MICRO DEVICES INC	Information Technology	Software

# Data Processing\_ Data Preparation

5) Installed and library the R package 'quantmod'

6) Used 'quantmod' package to gain all the daily adjusted closed price of each company from Yahoo Finance website; however, there were five companies didn't have complete weekly data, I deleted the five companies.

7) Used 'quantmod' package to calculate the weekly return of each company from February 17th until May 5<sup>th</sup>, totally 10 weekly returns

	Symbol	X2020.02.28	X2020.03.06	X2020.03.13	X2020.03.20	X2020.03.27	X2020.04.03	X2020.04.09	X2020.04.17	X2020.04.24	X2020.05.01
1	MMM	-0.0490026657937849	0.0295496438773237	-0.0779043375686692	-0.1185064570126	0.0668588843531017	0.00412779930472085	0.10456690882703	-0.00893214243424112	0.00358696554821279	0.0108843945578232
2	ABT	-0.119153783390067	0.0611449962500972	-0.00110100320775641	-0.167176995292664	0.0964705588235293	0.06553847522957285	0.082945302062126	0.115876346863362	-0.0203104255742023	-0.044758656338222
3	ABBV	-0.0974094365776057	0.036285159681311	-0.0388425692411618	-0.193862040745155	0.0559430107526884	0.00963265473049812	0.0869554772949512	0.0463949467084639	0.00167763936528376	-0.0089723655447955
4	ABMD	-0.0988365689445673	0.0453880422397193	-0.0907817342655751	-0.0861924898239222	0.112864995512413	-0.02643907234787	0.132107558646714	0.0371064588067658	0.0150584263327689	0.0888321899161597
5	ACN	-0.147718170502233	-0.0101888035924206	-0.0684755524475524	-0.0995135036788634	0.0863012059983834	-0.0658767845691067	0.169372352905557	-0.0159060365996631	0.00211326179937754	0.0263877169456628
6	ATVI	-0.0828337001761802	0.0316531905788202	-0.0155077536183468	-0.118394340813104	0.0943323745308815	0.053019681408351	0.00800265088362795	0.10618587671493	-0.00343892060880324	-0.0289572534446436
7	ADBE	-0.074621306085385	-0.0241945008141299	-0.00377108721525665	-0.119701949329359	0.0355183555971879	-0.0399568470046726	0.0854535890528383	0.0797300660283626	-2.89994491150525e-05	-0.0007556233521252
8	AMD	-0.146396380375307	0.0683817062445031	-0.0965218769294094	-0.0977221139989927	0.175965686039745	-0.0856591204096555	0.135947428973938	0.1699048538672	-0.00742045962616467	-0.112139533641865
9	AAP	-0.0542636010649058	-0.0381259749774695	-0.111171929914098	-0.340047517986674	0.274156994191084	-0.114539709509199	0.262374455153586	0.112296461610552	-0.0106848812831493	-0.0085041247896632
10	AES	-0.193734939759036	0.00537955768081289	-0.227110582639715	-0.086923076923077	0.133108677337826	-0.0899628252738104	0.202514379084957	-0.108695652173913	-0.0198170731707317	0.00622083981337473
11	AFL	-0.16844558364944	-0.0483080302594181	-0.149337914402597	-0.229460917539951	0.324354732510288	-0.0844633285613939	0.174945056029003	-0.0420168570369299	-0.00320830592105259	0.00909335880150584



# Methodology\_ Modeling

## ❑ K-Means clustering

### ❑ Goal

- ✓ Maximizing the homogeneity or similarity of records within a cluster and minimizing the similarity of different clusters
- ✓ Clustering all companies in S&P 500 based on the stock volatility

### ❑ Steps

- ✓ Randomly chose 3 points as cluster center, K=3
- ✓ Calculated the Euclidean Distance from each point to each center and chose the nearest center for each point
- ✓ Calculated the BCV and WCV values and calculated BCV/WCV
- ✓ Calculated the next iteration until find the max value of BCV/WCV

# Software & Package

## □ Software

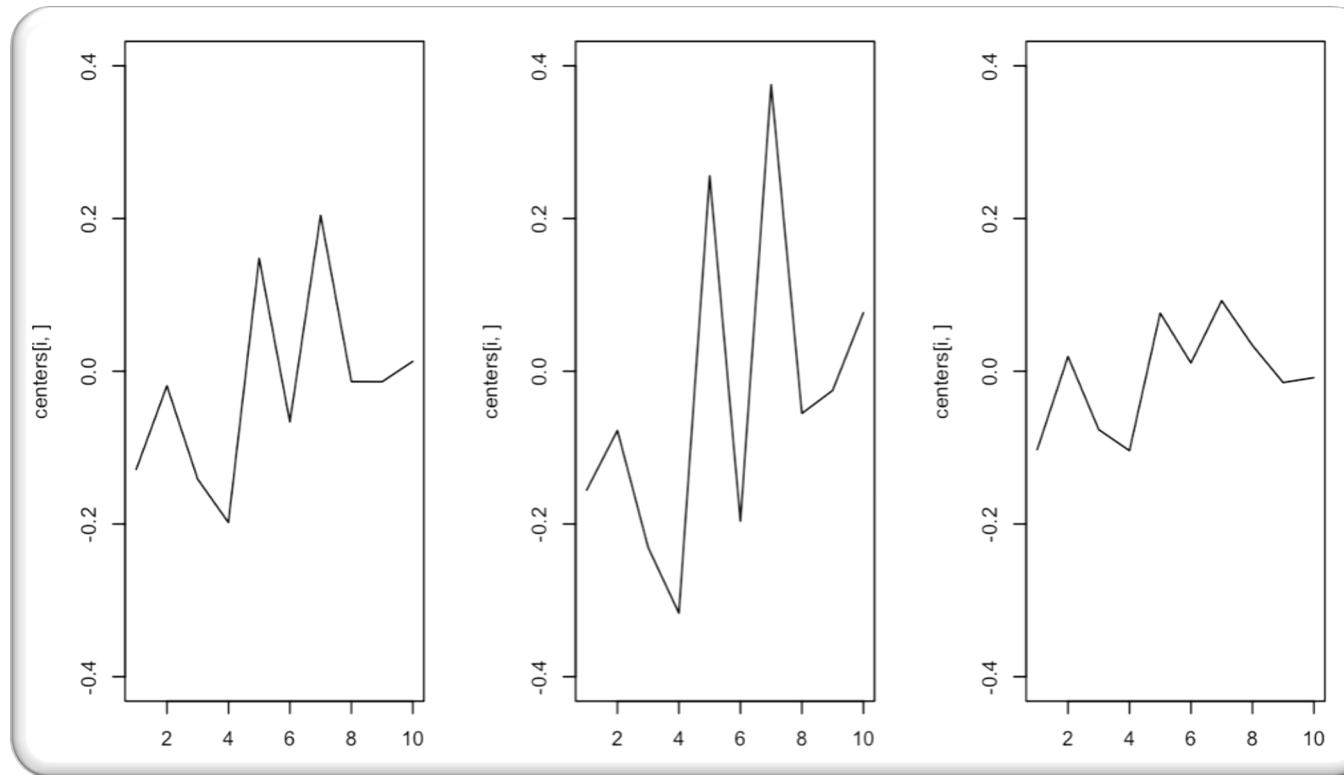


## □ Package: Quantmod

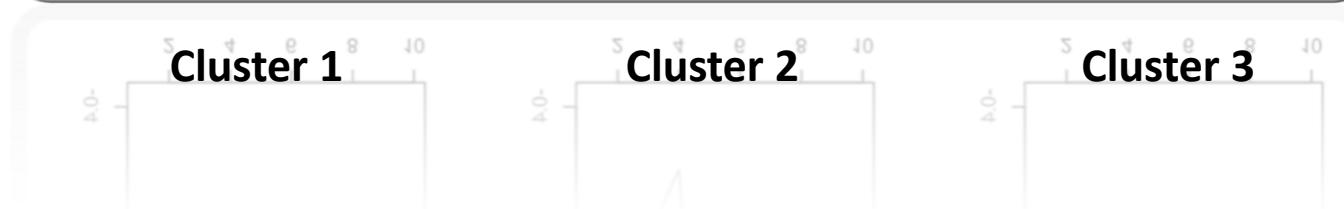
This package is a quantitative financial modeling and trading framework for R. It can gain the data from Yahoo Finance website and calculate the periodic return directly.



# Result



- ✓ These are the plots of the three clusters
- ✓ According to the graphs, we can know that the volatility of the three clusters is median, high, low separately
- ✓ Cluster1 contains 232 companies, cluster 2 contains 65 companies and cluster3 contains 203 companies.



# Result

Industry Name	Communication.Services	Consumer.Discretionary	Consumer.Staples	Energy	Financials	Health.Care	Industrials	Information.Technology	Materials	Real.Estate	Utilities
Total companies belong to cluster 2(A)	1	24	2	4	13	0	7	2	0	11	1
Total companies in the industry(B)	26	63	33	27	66	59	68	71	28	31	28
Ratio=A/B	0.038461538	0.380952381	0.060606061	0.148148148	0.1969697	0	0.10294118	0.028169014	0	0.35483871	0.03571429

- ✓ Based on the above table, there were totally 11 industries: Communication. Service; Consumer. Discretionary; Consumer. Staples; Energy; Financial; Healthy. Care; Industrials; Information. Technology; Materials; Real. Estate; Utilities
- ✓ First row showed the companies belonged to cluster 2, which has the highest volatility
- ✓ Second row showed total number of the companies in the industry
- ✓ Third row showed the ratio of first two rows
- ✓ According to the results, Consumer. Discretionary industry had the highest volatility and Consumer. Staples had the lowest volatility

# Discussion \_ Evaluation

Industry Name	Communication.Services	Consumer.Discretionary	Consumer.Staples	Energy	Financials	Health.Care	Industrials	Information.Technology	Materials	Real.Estate	Utilities
Total companies belong to cluster 3(C)	15	9	29	6	17	41	29	39	8	7	3
Total companies in the industry(D)	26	63	33	27	66	59	68	71	28	31	28
Ratio=C/D	0.576923077	0.142857143	0.878787879	0.222222222	0.25757576	0.694915254	0.42647059	0.549295775	0.28571429	0.22580645	0.10714286

- ✓ In order to evaluate the results I got, I also calculated the companies belonged to cluster3, which had the lowest volatility
- ✓ After calculating, I found the ratio of Consumer. Discretionary industry had the highest ratio, which showed that this industry is most stable. Although Utilities industry had the lowest ratio, the ratio of Consumer. Discretionary is relatively low.
- ✓ The evaluation results were consistent with the results I calculated before, so I got the right answer



# Discussion \_ Evaluation

- ✓ Consumer. Discretionary industry contains four sub industry groups: Automobiles & Components, Consumer Durables & Apparel, customer service and retailing. Most of these products are non-necessary. Thus, the industry was not stable under financial crisis.
  
- ✓ Consumer. Staples industry contains Food & Staples Retailing; Beverages; Food Products; Household Products, Personal Products. These products are necessary product. Thus, the industry has been less affectable under financial crisis is reasonable.



# Discussion \_ Development

- ✓ Use other clustering methods

For example:

Hierarchical clustering with three different methods:

single, complete and average

- ✓ Consider more features as the metrics to do divide cluster

Such as: future, bond, fund



# Conclusion

- The economy of the United States has been greatly affected due to COVID-19 in the past three months.
- In order to find out which industries are more vulnerable under 2020 financial crisis in America, I used the stock price of S&P 500 index in Yahoo Finance to deal with the problem.
- K-means algorithm, Software R, package quantmod were used.
- The results showed that Consumer. Discretionary industry is most volatile due to the leverage, and Consumer. Discretionary industry is most stable.

# Reference

- Cai, J. (2007). What's in the news? Information content of S&P 500 additions. *Financial Management*, 36(3), 113-124.
- Oyelade, O. J., Oladipupo, O. O., & Obagbuwa, I. C. (2010). Application of k Means Clustering algorithm for prediction of Students Academic Performance. *arXiv preprint arXiv:1002.2425*.
- Ryan, J. A., Ulrich, J. M., Thielen, W., Teetor, P., Bronder, S., & Ulrich, M. J. M. (2020). Package 'quantmod'.
- Yuan, X., Xu, J., Hussain, S., Wang, H., Gao, N., & Zhang, L. (2020). Trends and Prediction in Daily New Cases and Deaths of COVID-19 in the United States: An Internet Search-Interest Based Model. *Exploratory Research and Hypothesis in Medicine*, 5(2), 1.
- [https://en.wikipedia.org/wiki/List\\_of\\_S%26P\\_500\\_companies](https://en.wikipedia.org/wiki/List_of_S%26P_500_companies)
- [https://en.wikipedia.org/wiki/Global\\_Industry\\_Classification\\_Standard](https://en.wikipedia.org/wiki/Global_Industry_Classification_Standard)
- <https://finance.yahoo.com/>