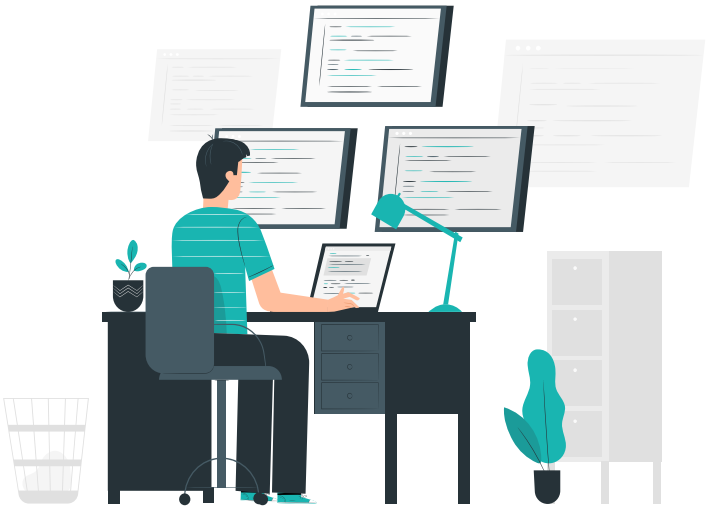
FINAL PROJECT

DBAS3017
User Experience Techniques
(Prof. George Campaies)

Presented by
Cheolsoon Im (W0439657)



Visualization and Analysis of Stock & Company Data

Focusing on Apple

TABLE OF CONTENTS

Data Sources 1A

Most Popular Real-Time Quotes

Symbol	Company Name
AMD	Advanced Micro Devices, Inc.
TSLA	Tesla, Inc.
AMZN	Amazon.com, Inc.
AAPL	Apple Inc.
ZNGA	Zynga Inc.
NVDA	NVIDIA Corporation
MSFT	Microsoft Corporation
JD	JD.com, Inc.
CSCO	Cisco Systems, Inc.
FB	Facebook, Inc.

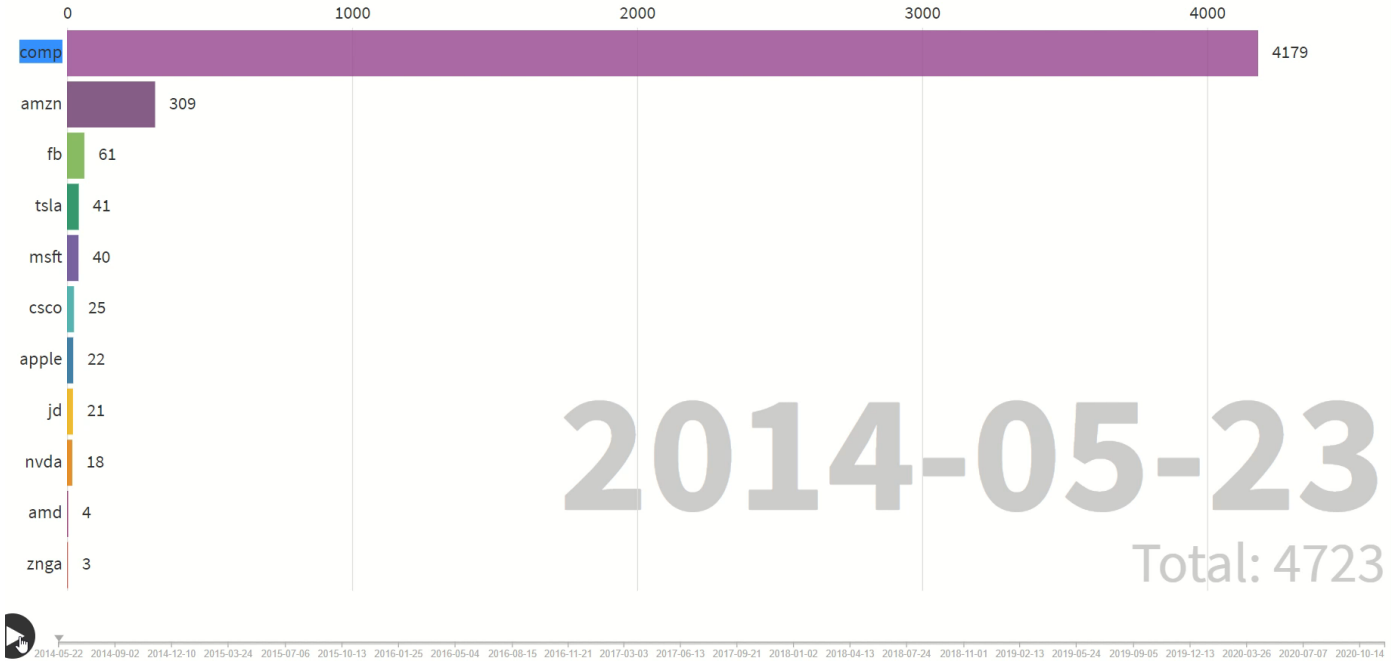
As of 2020-11-28
Source: Nasdaq (nasdaq.com, 2020)

Data Sources 1B

NASDAQ Composite Methodology

The NASDAQ Composite is calculated using a market capitalization weighting method. It means that the largest companies listed on this exchange exert the greatest impact on the final value of the index. The index is calculated by taking the total value of the share weights of all the stocks on the exchange, multiplied by each security's closing price. It is then divided by an index divisor to arrive at a more appropriate figure for reporting purposes.

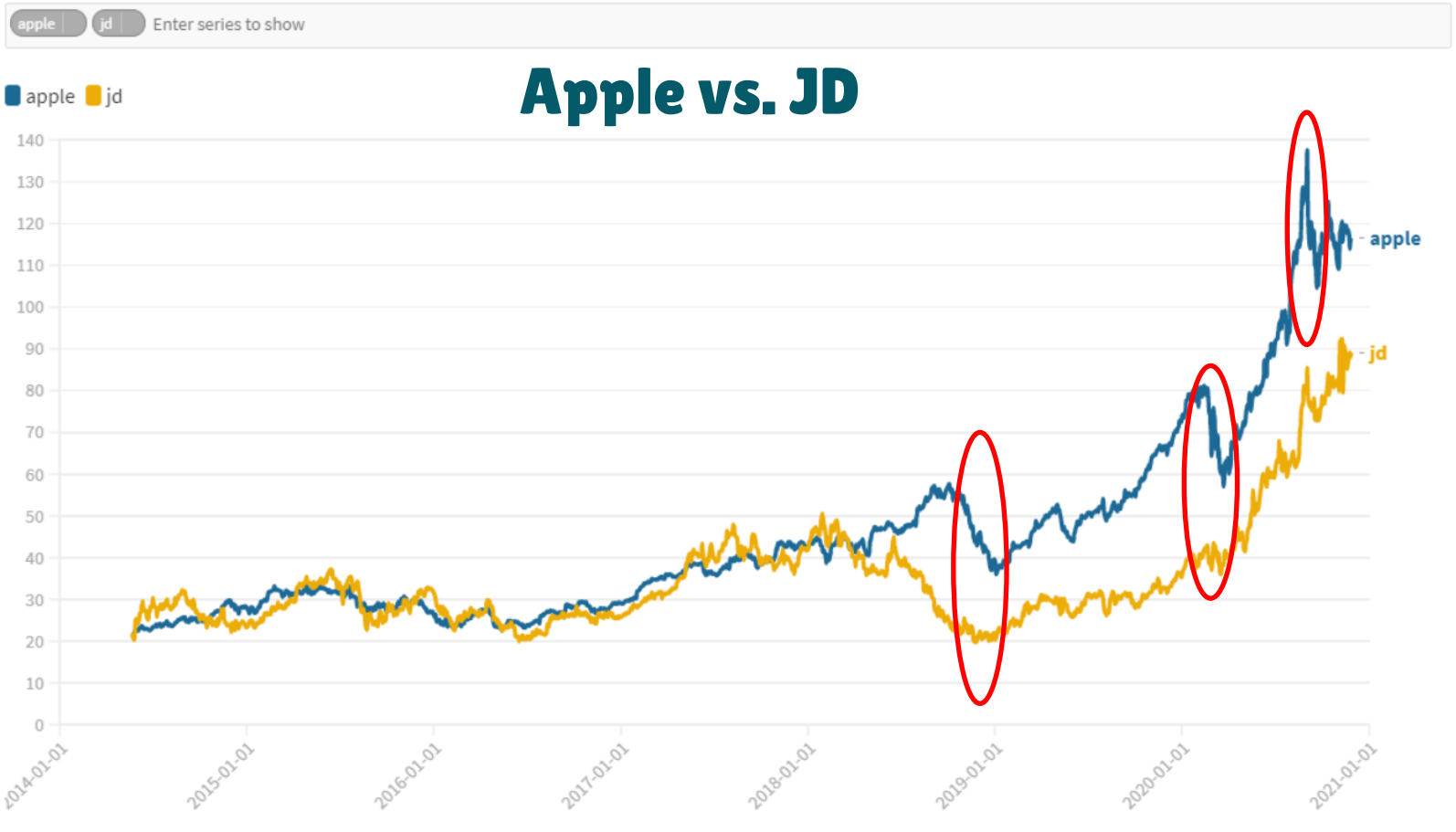
Top10 + Nasdaq comp




apple
jd
fb
amd
csco
nvda
msft
tsla
amzn
znga
Top 10 Only (without Comp)
4000
3500
3000
2500
2000
1500
1000
500
0
amzn
amzn
2017-06-06: 1012
tsla
nvda
fb
msft
apple
jd
amd
csco
znga
2014-01-01
2015-01-01
2016-01-01
2017-01-01
2018-01-01
2019-01-01
2020-01-01
2021-01-01


apple
jd
fb
amd
csco
nvda
msft
tsla
znga
Top 10 without Amazon
600
550
500
450
400
350
300
250
200
150
100
50
0
2014-01-01
2015-01-01
2016-01-01
2017-01-01
2018-01-01
2019-01-01
2020-01-01
2021-01-01
tsla
nvda
fb
msft
apple
jd
amd
csco
znga


apple
jd
amd
csco
znga
Top 10 without Big5
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
2014-01-01
2015-01-01
2016-01-01
2017-01-01
2018-01-01
2019-01-01
2020-01-01
2021-01-01
apple
jd
amd
csco
znga


apple
jd
Enter series to show
Apple vs. JD
apple
jd
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
2014-01-01
2015-01-01
2016-01-01
2017-01-01
2018-01-01
2019-01-01
2020-01-01
2021-01-01
apple
jd

Apple vs. JD

Three Critical Periods

A. Why up & down?
B. Why different?

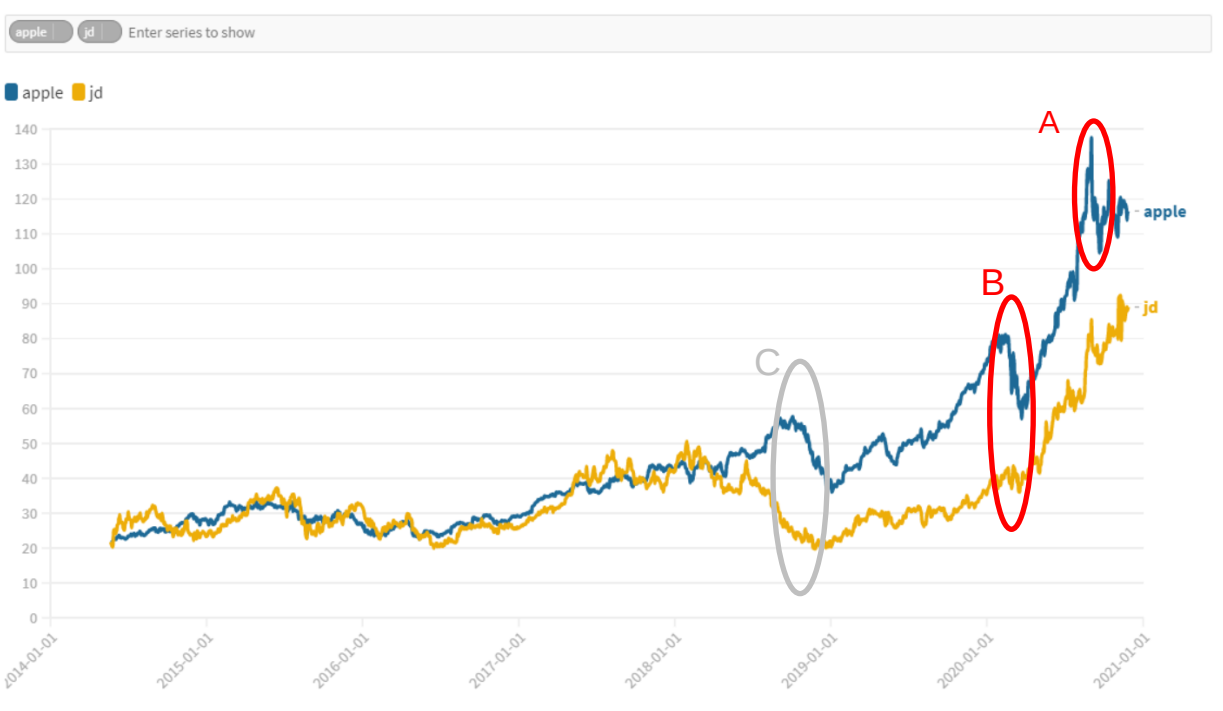
C. Correlation or Causation?

Apple vs. JD

Three Critical Periods

A. Why up & down?
B. Why different?

C. Correlation or Causation?



Stock Split

Q1. Do you know what it is? Yes or No.
Q2. Do you know why the companies do it? Yes or No.

In the chat, you can type, YY, YN, or NN.

One $200 AAPL stock

➔Stock Split

Four $50 AAPL stocks

AAPL's Stock Split

Apple's stock has split five times since the company went public. The stock split on a 4-for-1 basis on August 28, 2020, a 7-for-1 basis on June 9, 2014, and split on a 2-for-1 basis on February 28, 2005, June 21, 2000, and June 16, 1987.

- 4-for-1 basis on August 28, 2020
- 7-for-1 basis on June 9, 2014
- 2-for-1 basis on February 28, 2005
- 2-for-1 basis on June 21, 2000
- 2-for-1 basis on June 16, 1987

JD.com

JD.com, Inc., formerly called 360buy, is a **Chinese e-commerce company headquartered in Beijing, China.**

It has heavily invested in **high technology and Artificial Intelligence delivery** through drones, autonomous technology and robots, owing the largest drone delivery system, infrastructure and capability in the world.

(Wikipedia, 2020)

Apple vs. JD

Three Critical Periods

A. Why up & down?
B. Why different?

C. Correlation or Causation?

What caused JD.com's fall in 2018?

In 2018, economic woes in China, falling profits, and its CEO issues has set this Chinese e-commerce stock plummeting.

- **The Chinese sell-off**
- **Slowing profit growth** ➔ Hope for the future
- **Liu's arrest**

(BowmanJeremy, 2018)

• Jan 2018: JD.com opens first of a chain of high-tech supermarket 7Fresh.
• Jan 2018: JD.com invest in Vietnam's online retail service tiki.vn for $50 million.
• Feb 2018: JD.com invest in France and the UK, and wants to be everywhere in Europe in a few years.
• Feb 2018: JD.com spins off JD Finance and raises $2.1 billion in a capital raise.

(Wikipedia, 2020)

What caused AAPL's fall in 2018?

1. Investors' concern about **weak demand** for the newest iPhone models (KimTae, 2018)
2. Investors' **lack of confidence** in Apple's services business.
3. Apple reflects broader market concerns – such as **US-China trade tensions**.

(ShermanNatalie, 2018)

Causes of the fall in 2018

Correlation vs Causation

- Slowing profit growth
- Liu's arrest

US –China Tension

- Concerns about iPhone sales.
- Apple's high prices
- Not confident about Apple's services business.

(BowmanJeremy, 2018)

(ShermanNatalie, 2018)

Apple vs. JD – Recap

Three Critical Points
A. Why up & down?
B. Correlation or Causation?
C. Why different?

Data Sources 2
AAPL & its Ties

QCOM & INTC (SHOBHIT SETHJULIE, 2020)

and Gold Price(GOLD HUB, 2020)

and Nasdaq Composite (nasdaq.com, 2020)

Can you match these with AAPL? (Pearson Correlation: –1.00 to +1.00)

a. Gold
b. Nasdaq composite
c. Qcom
d. Intel

1. 0.90
2. 0.96
3. 0.72

Your answer

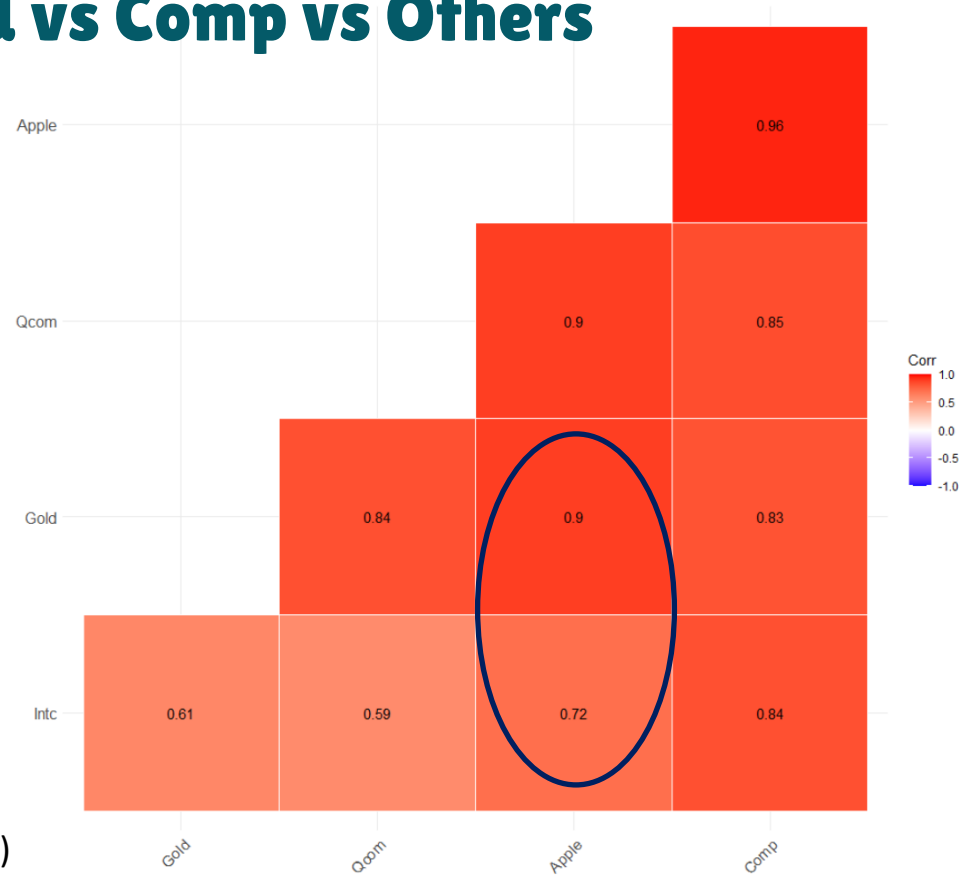
a. _____
b. _____
c. _____
d. _____

e.g.) 3 2 2 1
OR 0.72 0.96 0.96 0.90

Apple vs Gold vs Comp vs Others

```
> corr1
       Gold Qcom Apple Intc Comp
Gold   1.00 0.84  0.90 0.61 0.83
Qcom   0.84 1.00  0.90 0.59 0.85
Apple  0.90 0.90  1.00 0.72 0.96
Intc   0.61 0.59  0.72 1.00 0.84
Comp   0.83 0.85  0.96 0.84 1.00
```

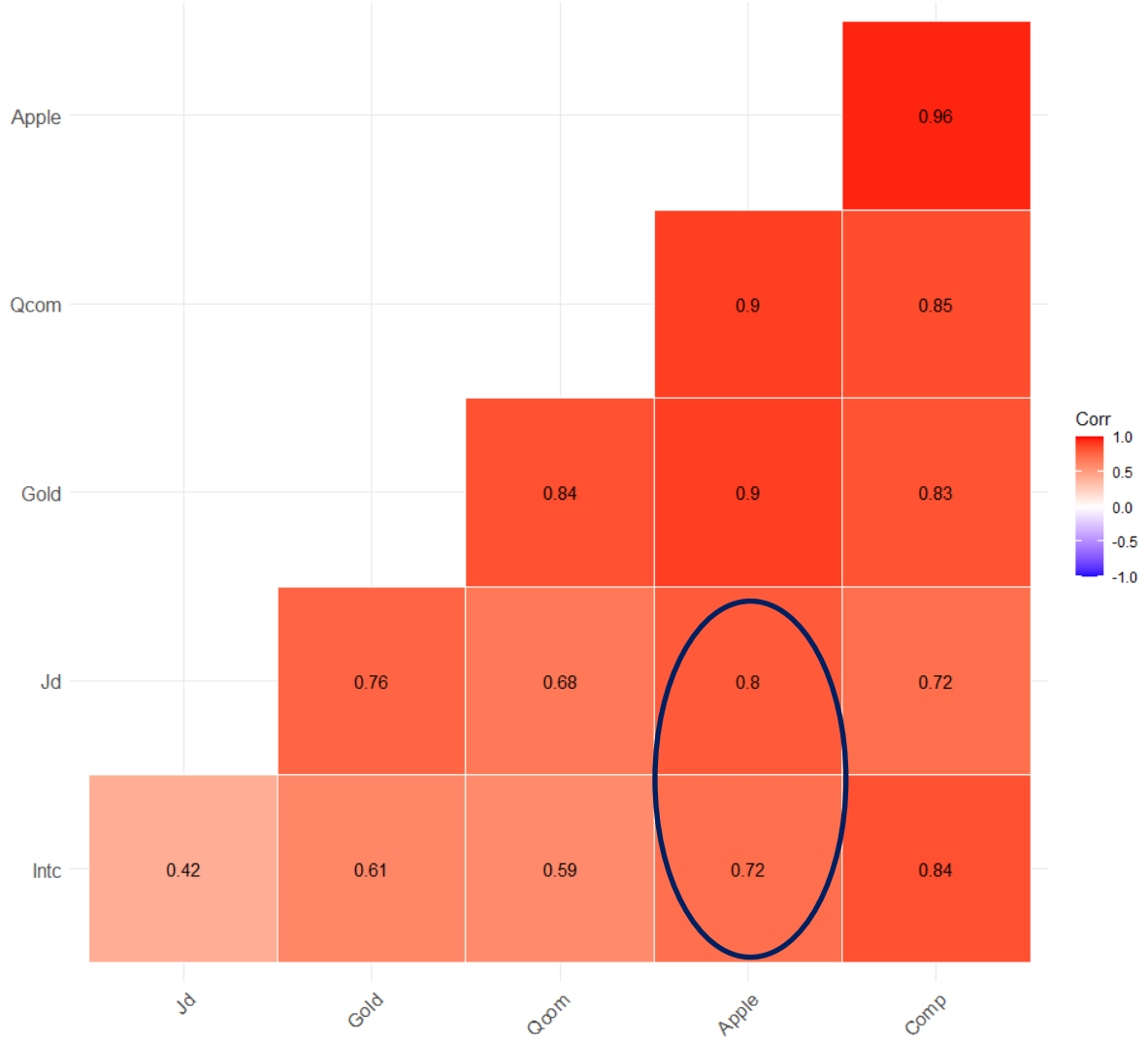
Apple has the most positive relationship with comp. The order of positive relationship is Comp, Gold = Qcom, and Intel.



(STHDA, no date)

AAPL vs JD revisit

```
> corr1
       Gold Qcom Apple Intc Comp   Jd
Gold   1.00 0.84  0.90 0.61 0.83 0.76
Qcom   0.84 1.00  0.90 0.59 0.85 0.68
Apple  0.90 0.90  1.00 0.72 0.96 0.80
Intc   0.61 0.59  0.72 1.00 0.84 0.42
Comp   0.83 0.85  0.96 0.84 1.00 0.72
Jd     0.76 0.68  0.80 0.42 0.72 1.00
```



Prediction Models

AR vs. MA vs. ARMA vs. ARIMA

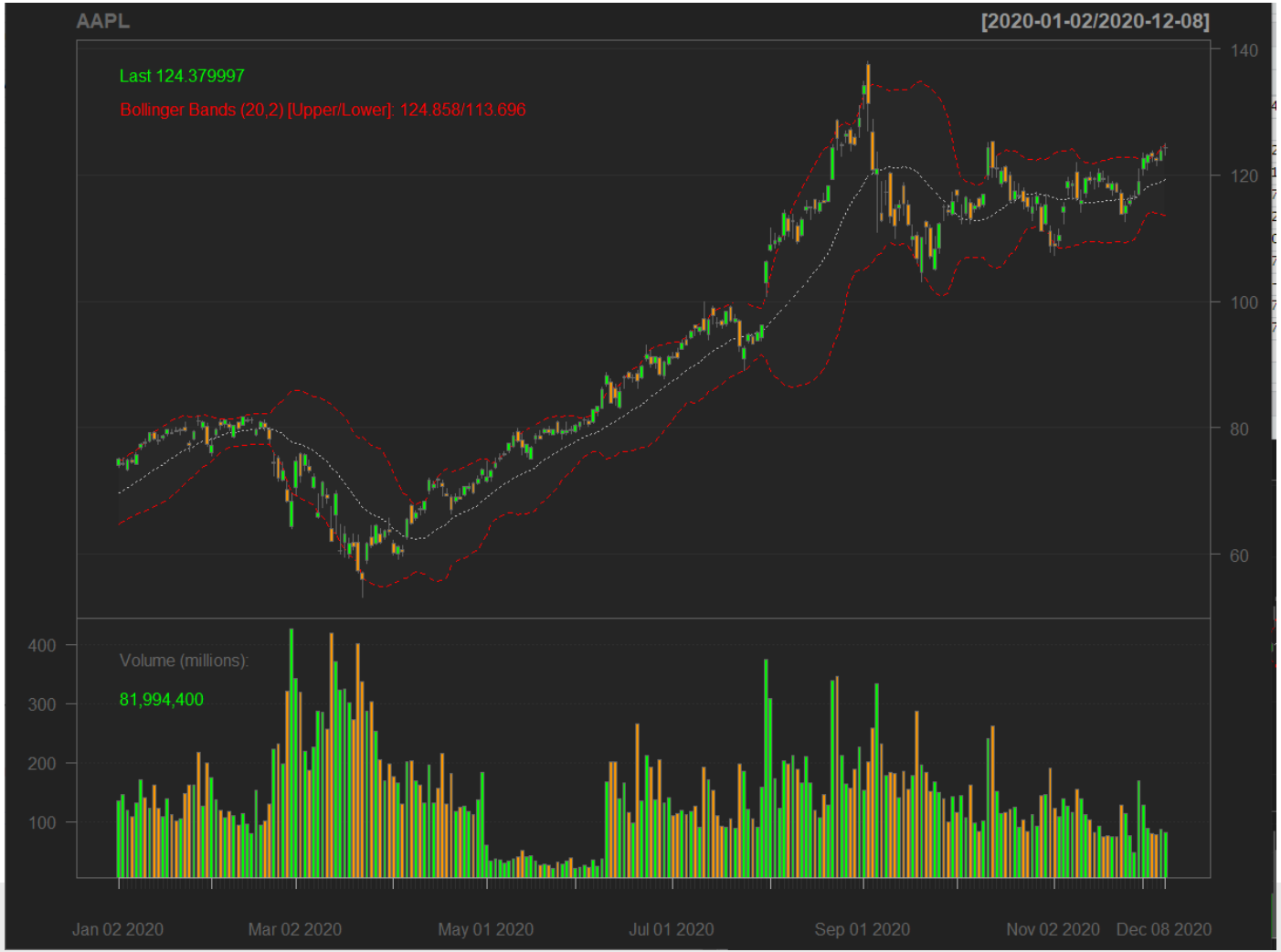
- AR: < Auto Regressive > means that the model uses the dependent relationship between an observation and some predefined number of lagged observations (also known as “time lag” or “lag”).
- I:< Integrated > means that the model employs differencing of raw observations (e.g. it subtracts an observation from an observation at the previous time step) in order to make the time-series stationary.MA:
- MA: < Moving Average > means that the model exploits the relationship between the residual error and the observations.

Stock Data Summary: AAPL & JD

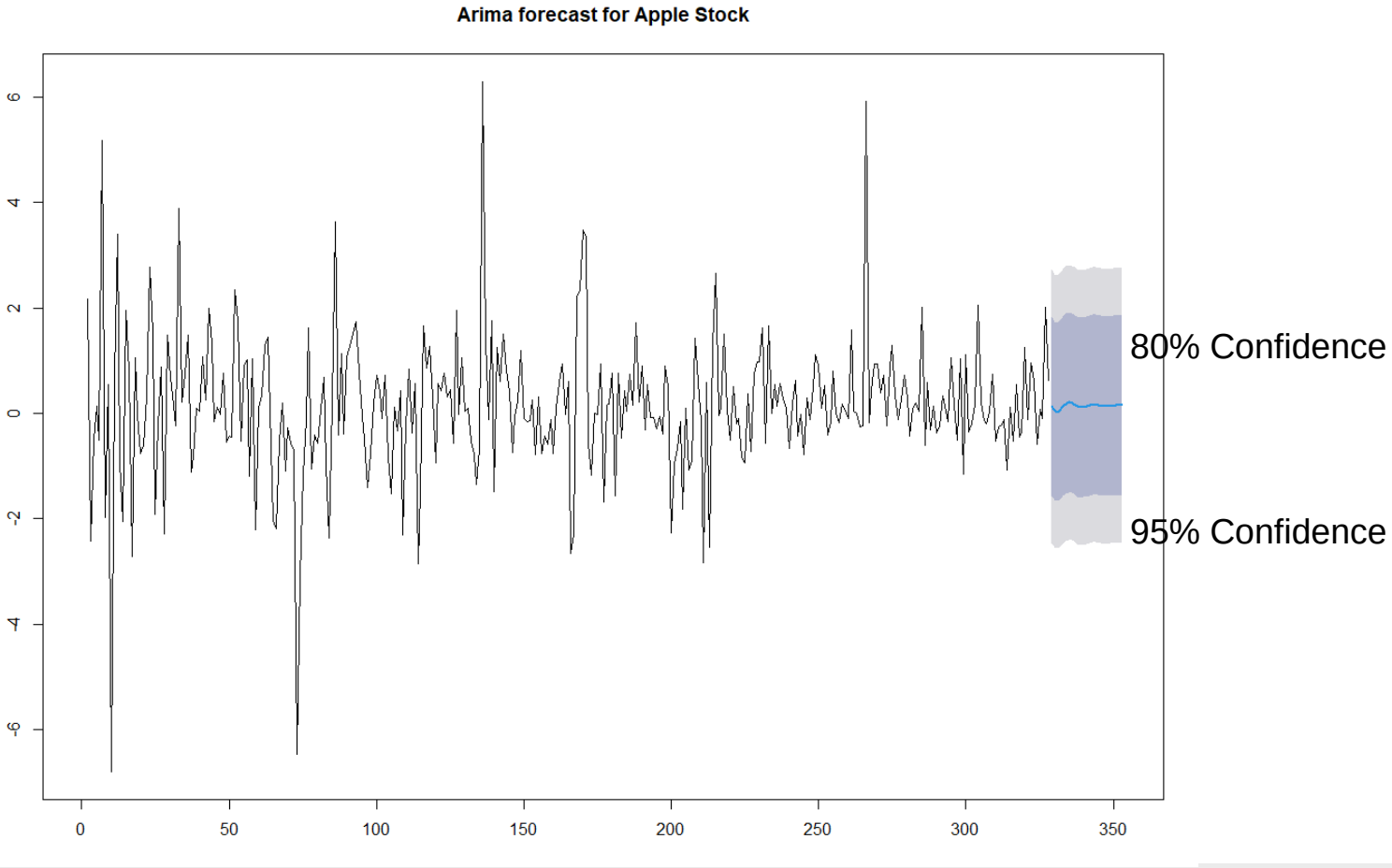
```
> summary(AAPL)
     Index              AAPL.Open        AAPL.High         AAPL.Low          AAPL.Close       AAPL.Volume
 Min.   :2007-01-03   Min.   :  2.835   Min.   :  2.929   Min.   :  2.793   Min.   :  2.793   Min.   :2.020e+07
 1st Qu.:2010-06-28   1st Qu.:  9.141   1st Qu.:  9.255   1st Qu.:  8.985   1st Qu.:  9.141   1st Qu.:1.367e+08
 Median :2013-12-19   Median : 21.365   Median : 21.552   Median : 21.107   Median : 21.389   Median :2.909e+08
 Mean   :2013-12-20   Mean   : 27.127   Mean   : 27.412   Mean   : 26.839   Mean   : 27.137   Mean   :4.176e+08
 3rd Qu.:2017-06-15   3rd Qu.: 37.075   3rd Qu.: 37.532   3rd Qu.: 36.690   3rd Qu.: 37.240   3rd Qu.:5.717e+08
 Max.   :2020-12-08   Max.   :137.590   Max.   :137.980   Max.   :130.530   Max.   :134.180   Max.   :3.373e+09
 AAPL.Adjusted
 Min.   :  2.390
 1st Qu.:  7.822
 Median : 18.517
 Mean   : 25.510
 3rd Qu.: 35.408
 Max.   :133.949
> summary(JD)
     Index              JD.Open          JD.High           JD.Low            JD.Close         JD.Volume          JD.Adjusted
 Min.   :2014-05-22   Min.   :19.76   Min.   :19.88   Min.   :19.21   Min.   :19.27   Min.   : 1217000   Min.   :19.27
 1st Qu.:2016-01-11   1st Qu.:26.32   1st Qu.:26.82   1st Qu.:25.84   1st Qu.:26.32   1st Qu.: 7489925   1st Qu.:26.32
 Median :2017-08-29   Median :30.63   Median :31.08   Median :30.19   Median :30.64   Median :10764450   Median :30.64
 Mean   :2017-08-30   Mean   :34.91   Mean   :35.48   Mean   :34.27   Mean   :34.88   Mean   :12460264   Mean   :34.88
 3rd Qu.:2019-04-22   3rd Qu.:39.57   3rd Qu.:40.12   3rd Qu.:38.90   3rd Qu.:39.38   3rd Qu.:15310200   3rd Qu.:39.38
 Max.   :2020-12-08   Max.   :92.40   Max.   :92.77   Max.   :90.91   Max.   :92.49   Max.   :95370800   Max.   :92.49
```

Stock Price Prediction by R

Stock Price Prediction by R



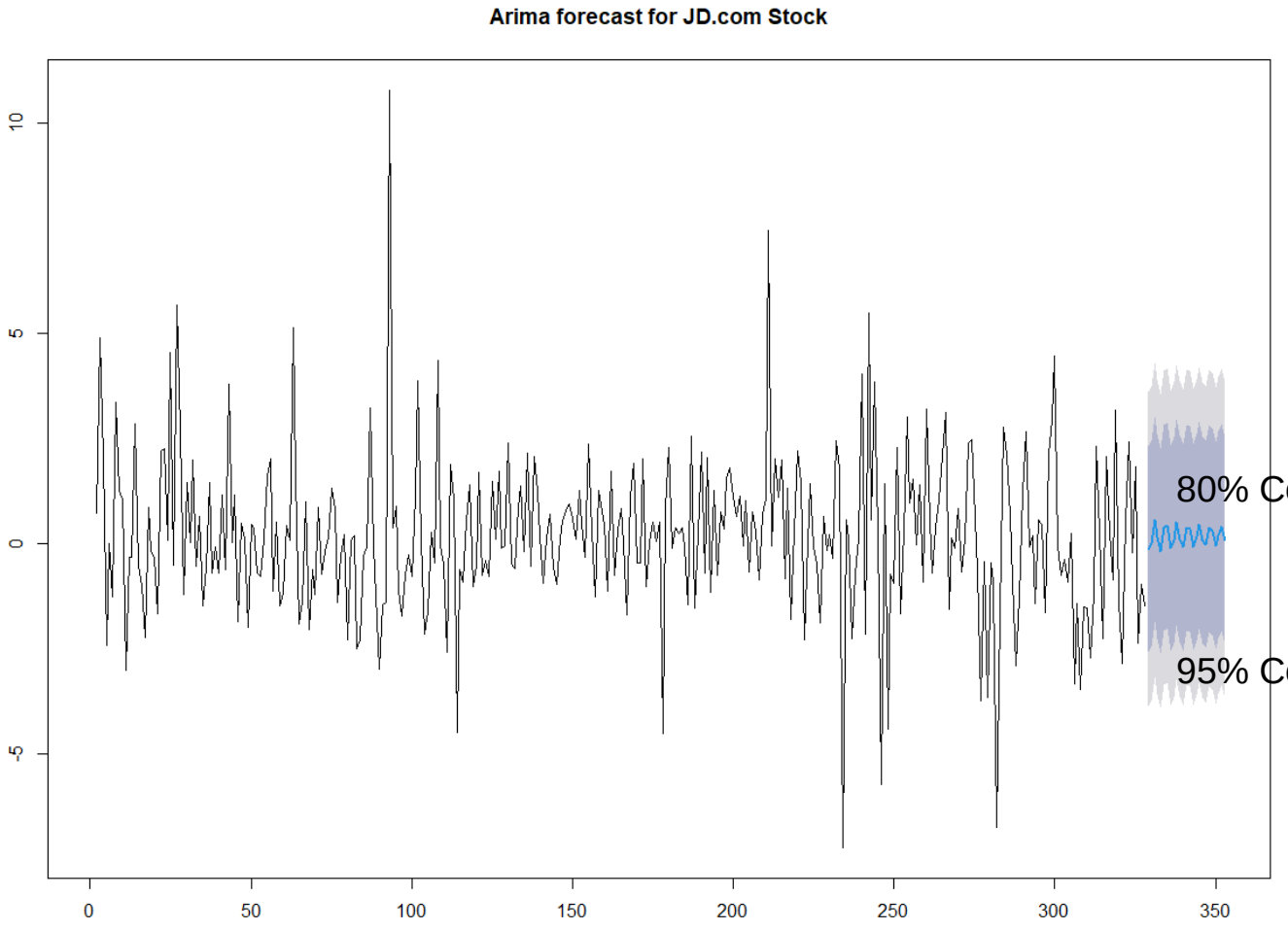
Stock Price Prediction by R



Stock Price Prediction by R

Stock Price Prediction by R

Stock Price Prediction by R



Stock Price Prediction by R

Accuracy Comparison:

RMSE (Root Mean Square Error)

(Stock Price Prediction Project, 2018)

```
> accuracy(preds, AAPL_ret_test)
                    ME     RMSE      MAE      MPE     MAPE
Test set -0.1634744 1.377053 0.9810064 102.1683 116.0075
```

```
> accuracy(preds, JD_ret_test)
                     ME     RMSE      MAE      MPE     MAPE
Test set -0.01437243 1.999257 1.434773 102.4694 104.7216
```

Future Study Directions

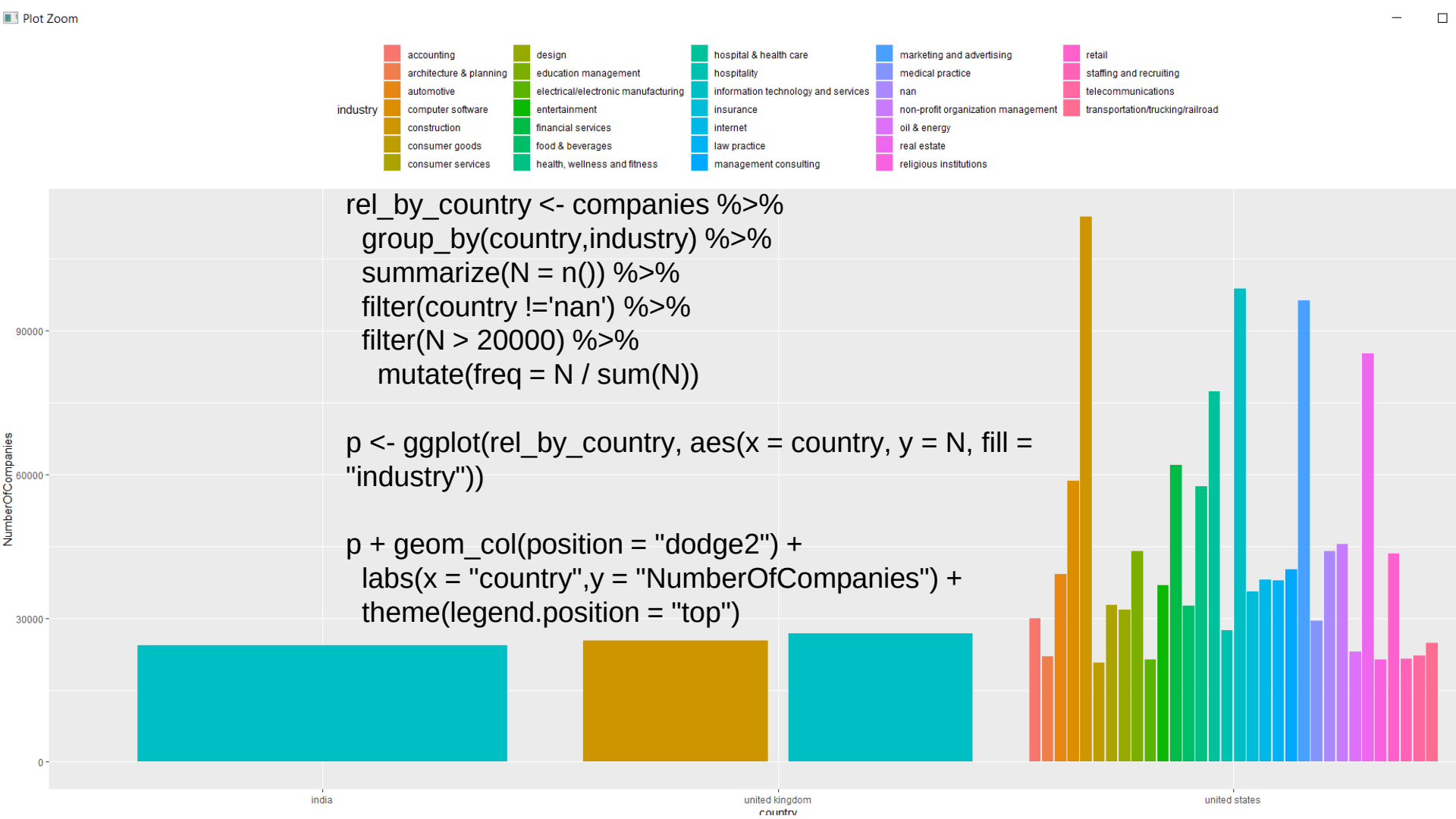
- Hypothesis Testing & Pearson Correlation on all the Nasdaq, Korean, and Chinese companies
- Comparison between economic superpowers and high-CO_2 emitting countries
- Time series analysis using other mathematical models to minimize prediction errors

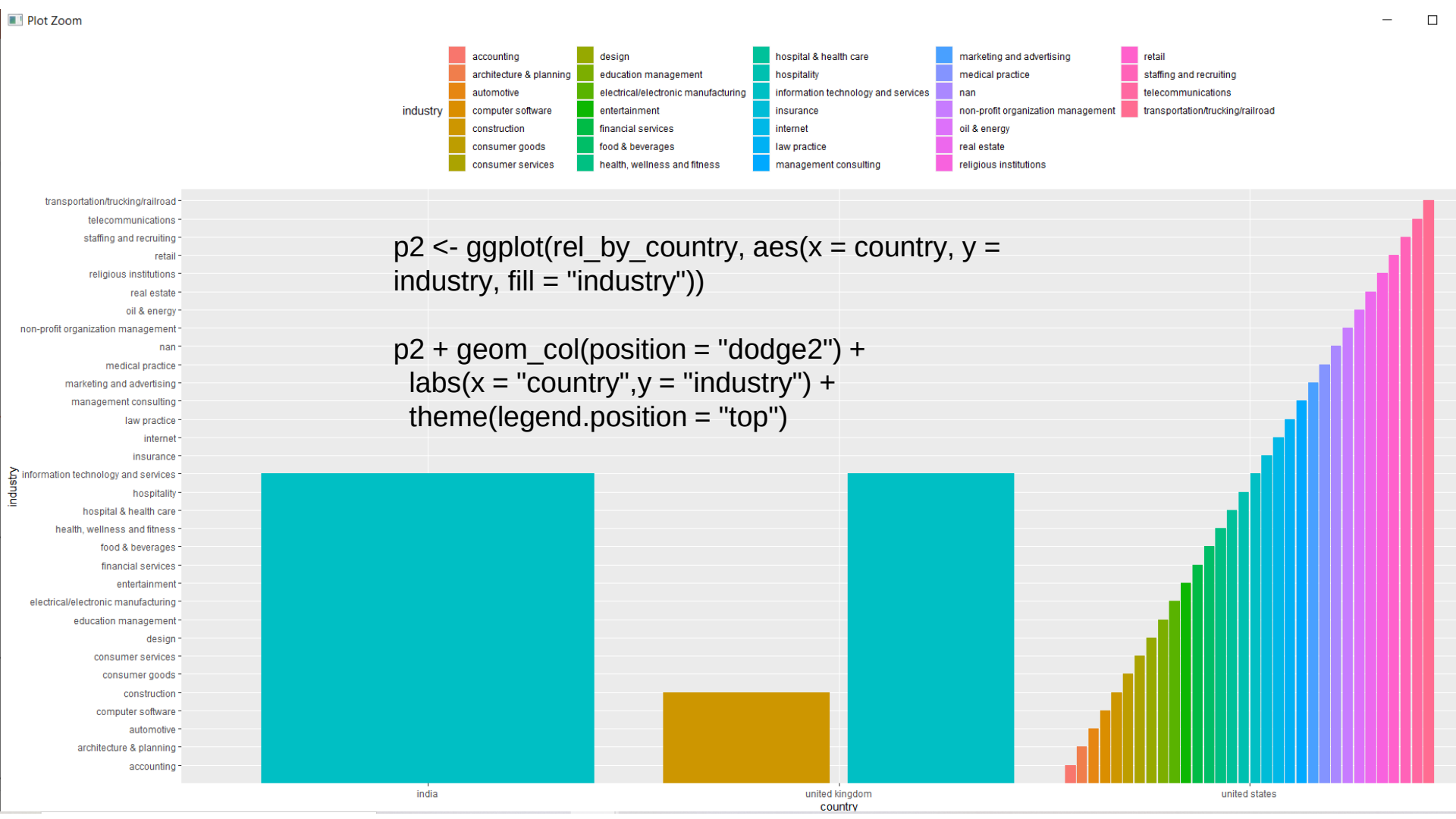
Data Sources 3 – exploratory study

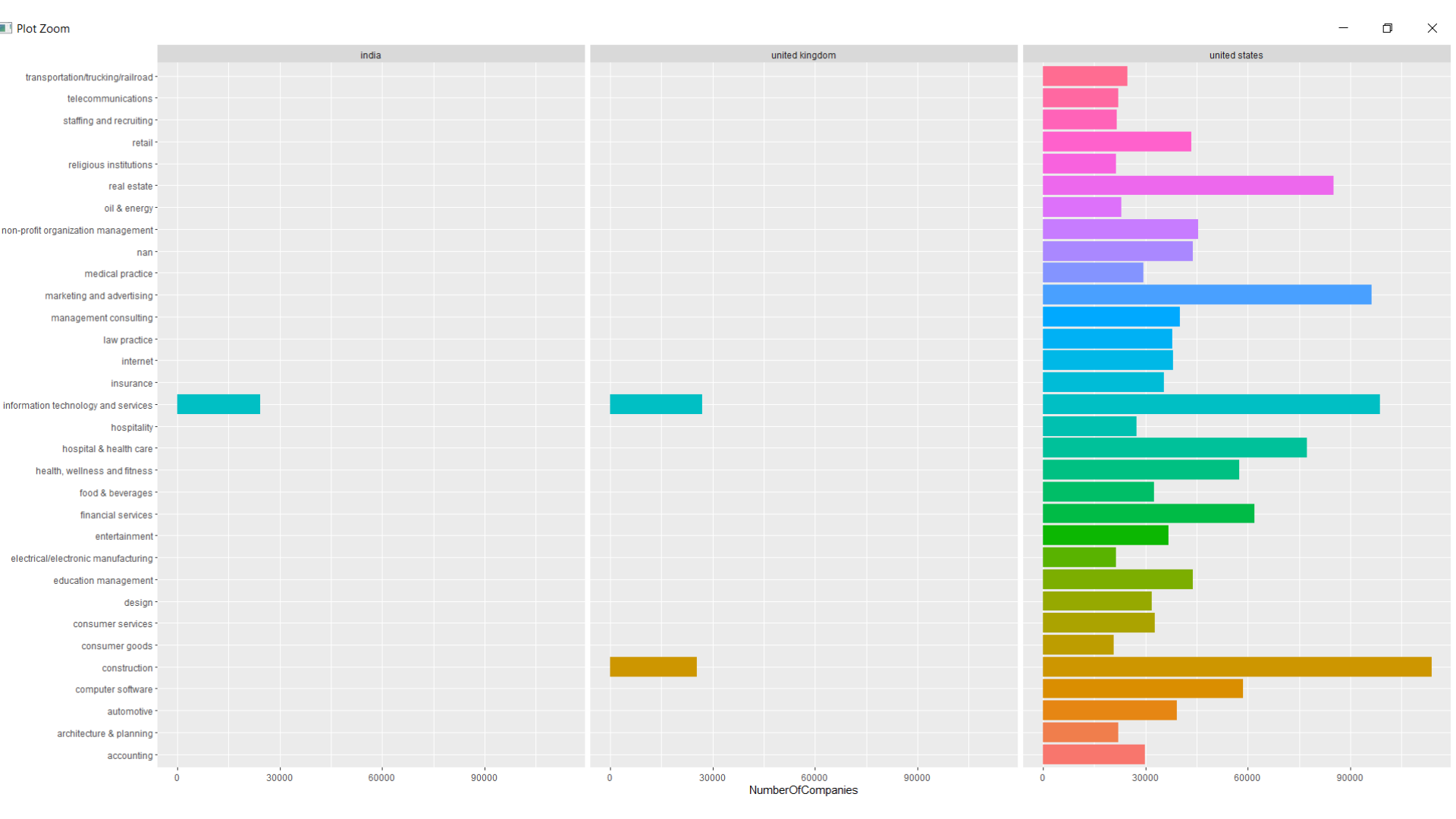
A collection of data on over **7 Million Global companies**.
Attributes: names, domains, size range, year funded, industry, locality, country, LinkedIn url, current employee estimate, and total employee estimate.

NAME	DOMAIN	YEAR_FOUNDED	INDUSTRY	SIZE_RANGE	LOCALITY	COUNTRY	LINKEDIN URL	CURRENT EMPLOYEE ESTIMATE	TOTAL EMPLOYEE ESTIMATE
7,004,634 Unique	5,474,766 Unique	236 Unique	148 Unique	8 Unique Ranges	city, state, country	country headquarters	7,173,427 Unique	Estimate for current employee count	Estimate for total historic employee count.
Never Null	35% Null	50% Null	4% Null		35% Null	33% Null	Never Null	Never Null	Never Null
Not Unique	Not Unique	Not Unique							
	48,041 Repeated								

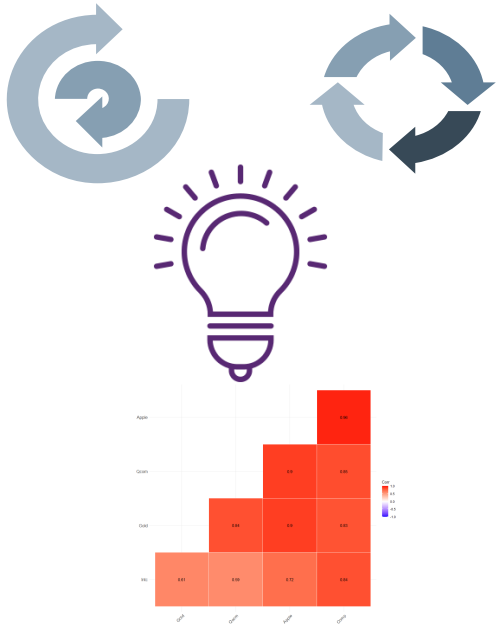
Plot Zoom


industry
accounting
architecture & planning
automotive
computer software
construction
consumer goods
consumer services
design
education management
electrical/electronic manufacturing
entertainment
financial services
food & beverages
health, wellness and fitness
hospital & health care
hospitality
information technology and services
insurance
internet
law practice
management consulting
marketing and advertising
medical practice
nan
non-profit organization management
oil & energy
real estate
religious institutions
retail
staffing and recruiting
telecommunications
transportation/trucking/railroad
rel_by_country <- companies %>%
group_by(country,industry) %>%
summarize(N = n()) %>%
filter(country !='nan') %>%
filter(N > 20000) %>%
mutate(freq = N / sum(N))
p <- ggplot(rel_by_country, aes(x = country, y = N, fill = "industry"))
p + geom_col(position = "dodge2") +
labs(x = "country",y = "NumberOfCompanies") +
theme(legend.position = "top")
90000
60000
30000
0
NumberOfCompanies
india
united kingdom
united states
country


Plot Zoom
industry
accounting
architecture & planning
automotive
computer software
construction
consumer goods
consumer services
design
education management
electrical/electronic manufacturing
entertainment
financial services
food & beverages
health, wellness and fitness
hospital & health care
hospitality
information technology and services
insurance
internet
law practice
management consulting
marketing and advertising
medical practice
nan
non-profit organization management
oil & energy
real estate
religious institutions
retail
staffing and recruiting
telecommunications
transportation/trucking/railroad
p2 <- ggplot(rel_by_country, aes(x = country, y = industry, fill = "industry"))
p2 + geom_col(position = "dodge2") +
labs(x = "country",y = "industry") +
theme(legend.position = "top")
industry
country
india
united kingdom
united states


Plot Zoom
india
united kingdom
united states
transportation/trucking/railroad
telecommunications
staffing and recruiting
retail
religious institutions
real estate
oil & energy
non-profit organization management
nan
medical practice
marketing and advertising
management consulting
law practice
internet
insurance
information technology and services
hospitality
hospital & health care
health, wellness and fitness
food & beverages
financial services
entertainment
electrical/electronic manufacturing
education management
design
consumer services
consumer goods
construction
computer software
automotive
architecture & planning
accounting
0
30000
60000
90000
NumberOfCompanies

Key Takeaways!



References

Bowman, J. (2018, November 24). Why JD.com Stock Is Down 53% This Year. Retrieved from The Motley Fool: https://www.fool.com/investing/2018/11/24/why-jdcom-stock-is-down-53-this-year.aspx

GOLD HUB. (2020, November 30). Gold prices. Retrieved from GOLD HUB: https://www.gold.org/goldhub/data/gold-prices?utm_source=google&utm_medium=cpc&utm_campaign=rwm-goldhub&utm_content=466092663098&utm_term=gold%20price%20time%20series&gclid=CjwKCAiAn7L-BRBbEiwAl9UtkATSeAolawteF5UkPDw1LpKWeyf7ZgNGWw-7JdHt4RPUUHt93RrOOhoCDD

Kim, T. (2018, December 27). What Happened With Apple Stock in 2018 and What's Next. Retrieved from BARRON'S: https://www.barrons.com/articles/apple-stock-2018-year-in-review-51545862266

nasdaq.com. (2020). AAPL Historical Data. Retrieved from Nasdaq: https://www.nasdaq.com/market-activity/stocks/aapl/historical

Sherman, N. (2018, November 21). Four reasons that Apple shares have been falling. Retrieved from BBC News: https://www.bbc.com/news/business-46281768

SHOBHIT SETH, J. Y. (2020, October 4). 9 Major Companies Tied to the Apple Supply Chain. Retrieved from Investopedia: https://www.investopedia.com/articles/investing/090315/10-major-companies-tied-apple-supply-chain.asp

STHDA. (no date). ggcorrplot: Visualization of a correlation matrix using ggplot2. Retrieved from Statistical tools for high-throughput data analysis: http://www.sthda.com/english/wiki/ggcorrplot-visualization-of-a-correlation-matrix-using-ggplot2

Stock Price Prediction Project. (2018). Stock Market Prediction with R. Retrieved from Stock Price Prediction Project: https://shakil-stat-bsmrstu.gitbook.io/stockproject/stock-market-prediction-with-r

wikipedia. (2020, December 6). JD.com. Retrieved from wikipedia: https://en.wikipedia.org/wiki/JD.com

Special THANKS

to George & all of my ITDA classmates!

Do you have any questions?