California State University, Los Angeles

CIS 5250 Visual Analytics

Instructor: Dr. Shilpa Balan

R Project

Forbes Global 2000 – 2019

Prepared by: Vishav Patel & Nikesh Beemola

INDEX

1. Introduction

1.1 Objective

Specifically, a dataset on Forbes Global from 2000 to 2019 is the main subject to examine and show. It attempts to research the top 2000 worldwide corporations' profits, revenues, industry types, and related areas. My ability to determine which industries are profitable and which are not will be aided by this data. which business generates the most and least money. Furthermore, based on this information, it is possible to identify which companies were performing well and poorly at any given period. In addition, we can identify which industry is controlling the marketplace.

1.2 Aim of Analysis

This project aims to find following analysis with this dataset.

- Market value of the company
- Which sectors are making huge profits
- Which sectors are making less profits
- Types of Industry of the top companies
- Companies with highest asset holdings
- Difference in revenue of top companies

After the data cleaning in the R this all analysis can be done in the R studio to generate the desired results and visualizations. Further, different charts also can be used to generate additional visualizations.[1]

[1] Krasnove, B. (2019, June 30). *Global 500 2000*. Fortune. Retrieved November 2, 2022, from https://fortune.com/global500/2000/

Wikimedia Foundation. (2022, October 30). *Forbes Global 2000*. Wikipedia. Retrieved November 2, 2022, from https://en.wikipedia.org/wiki/Forbes_Global_2000

2. Dataset

Data Set link: https://data.world/aroissues/forbes-global-2000-2008-2019/workspace/file?filename=Forbes+Global+2000+-+2019.csv

An established business publication called Forbes releases annual lists of the biggest corporations in the world. The largest private corporations, the largest publicly listed companies, and the companies with the greatest market values are often included on these lists in order of revenue size. Usually, these lists contain businesses from a range of industries. This project uses only one data set because it contains all the required column to analyze the global companies

	Company	Market Va	Revenue	Profits	Assets	Rank	Sector	Industry	Continent	Country	Headquar	State	CEO	Forbes Webpage	Profits as	Profits as %
2	ICBC	305.057	175.874	45.223	4034.482	1	Financials	Major Ban	Asia	China	China		Shu Gu	http://www.forbes.com/companies/i	0.011209	0.257133
3	JPMorgan	368.502	132.912	32.738	2737.188	2	Financials	Major Ban	North Am	United Sta	New York	New York	Jamie Dim	http://www.forbes.com/companies/j	0.01196	0.246313
4	China Con	224.988	150.313	38.841	3382.422	3	Financials	Major Ban	Asia	China	China		Wang Zuji	http://www.forbes.com/companies/c	0.011483	0.258401
5	agricultura	197.045	137.456	30.894	3293.105	4	Financials	Regional B	Asia	China	China		Huan Zhac	http://www.forbes.com/companies/a	0.009381	0.224756
6	Bank of ar	287.339	111.904	28.54	2377.164	5	Financials	Major Ban	North Am	United Sta	North Car	North Car	Brian Moy	http://www.forbes.com/companies/b	0.012006	0.25504
7	apple	961.257	261.705	59.431	373.719	6	Informatic	Computer	North Am	United Sta	California	California	Tim Cook	http://www.forbes.com/companies/a	0.159026	0.227092
8	Ping an In	220.197	151.788	16.3	1038.3	7	Financials	Diversifiec	Asia	China	China		Ma Mingz	http://www.forbes.com/companies/p	0.015699	0.107387
9	Bank of Ch	142.958	126.677	27.5	3097.612	8	Financials	Major Ban	Asia	China	China		Chen Siqir	http://www.forbes.com/companies/b	0.008878	0.217088
10	Royal Dut	264.939	382.626	23.329	399.194	9	Energy	Oil & Gas	Europe	Netherlan	Netherlands		Bernardus	http://www.forbes.com/companies/r	0.05844	0.060971
11	Wells Farg	214.676	101.456	23.117	1887.792	10	Financials	Major Ban	North Am	United Sta	California	California	Timothy S	http://www.forbes.com/companies/v	0.012246	0.227852
12	ExxonMob	343.431	279.209	20.84	346.196	11	Energy	Oil & Gas	North Am	United Sta	Texas	Texas	Darren W.	http://www.forbes.com/companies/e	0.060197	0.074639
13	aT&T	233.325	170.805	19.37	531.864	12	Telecomm	Telecomm	North Am	United Sta	Texas	Texas	Randall L.	http://www.forbes.com/companies/a	0.036419	0.113404
14	Samsung E	272.424	221.506	39.882	304.138	13	Informatic	Semicondi	Asia	South Kor	South Korea		Hyun-Suk	http://www.forbes.com/companies/s	0.131131	0.180049
15	Citigroup	161.112	99.974	17.928	1958.413	14	Financials	Major Ban	North Am	United Sta	New York	New York	Michael L.	http://www.forbes.com/companies/c	0.009154	0.179327

3. Data Description

Company	Name of the company
Market Value	Market value of the company
Revenue	Revenue of the company
Profits	Profit made by the company
Assets	Assets of the company
Rank	Global rank of the company
Sector	Sector of the company
Industry	Focus of the Industry
Continent	Continent name
Country	Country located
Headquarters	Location of headquarters
State	Name of the state
CEO	Name of the CEO
Forbes Webpage	Source (URL)
Profits as % of Assets	Profit in terms of Assets
Profits as % of Revenue	Profit in terms of company revenue

4. Uploading Dataset

```
> setwd("C:/Users/COSMOS/Desktop/Final project")
> Forbes_Initial<-read.csv("Forbes Global 2000 - 2019.csv")
> View(Forbes_Initial)
> dim(Forbes_Initial)
[1] 2000   16
> head(Forbes_Initial)
                    Company Market.Value Revenue Profits   Assets Rank
1                      ICBC      305.057 175.874  45.223 4034.482    1
2              JPMorgan Chase      368.502 132.912  32.738 2737.188    2
3    China Construction Bank      224.988 150.313  38.841 3382.422    3
4 agricultural Bank of China      197.045 137.456  30.894 3293.105    4
5             Bank of america      287.339 111.904  28.540 2377.164    5
6                       apple      961.257 261.705  59.431  373.719    6
                Sector          Industry     Continent       Country   Headquarters
1           Financials       Major Banks          Asia         China          China
2           Financials       Major Banks North America United States       New York
3           Financials       Major Banks          Asia         China          China
4           Financials    Regional Banks          Asia         China          China
5           Financials       Major Banks North America United States North Carolina
6 Information Technology Computer Hardware North America United States     California
          State            CEO                                          Forbes.Webpage
1                      Shu Gu                        http://www.forbes.com/companies/icbc/
2       New York    Jamie Dimon             http://www.forbes.com/companies/jpmorgan-chase/
3                     Wang Zuji    http://www.forbes.com/companies/china-construction-bank/
4                     Huan Zhao http://www.forbes.com/companies/agricultural-bank-of-china/
5 North Carolina Brian Moynihan             http://www.forbes.com/companies/bank-of-america/
6     California       Tim Cook                  http://www.forbes.com/companies/apple/
  Profits.as...of.Assets Profits.as...of.Revenue
1             0.01120912               0.2571329
2             0.01196045               0.2463133
3             0.01148319               0.2584008
```

Code

```
Forbes_Initial<-read.csv("Forbes Global 2000 - 2019.csv")

View(Forbes_Initial)

dim(Forbes_Initial)

head(Forbes_Initial)

str(Forbes_Initial)
```

```
> str(Forbes_Initial)
'data.frame':   2000 obs. of  16 variables:
 $ Company               : chr  "ICBC" "JPMorgan Chase" "China Construction Bank" "agricultural Bank
 of China" ...
 $ Market.Value          : num  305 369 225 197 287 ...
 $ Revenue               : num  176 133 150 137 112 ...
 $ Profits               : num  45.2 32.7 38.8 30.9 28.5 ...
 $ Assets                : num  4034 2737 3382 3293 2377 ...
 $ Rank                  : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Sector                : chr  "Financials" "Financials" "Financials" "Financials" ...
 $ Industry              : chr  "Major Banks" "Major Banks" "Major Banks" "Regional Banks" ...
 $ Continent             : chr  "Asia" "North America" "Asia" "Asia" ...
 $ Country               : chr  "China" "United States" "China" "China" ...
 $ Headquarters          : chr  "China" "New York" "China" "China" ...
 $ State                 : chr  "" "New York" "" "" ...
 $ CEO                   : chr  "Shu Gu" "Jamie Dimon" "Wang Zuji" "Huan Zhao" ...
 $ Forbes.Webpage        : chr  "http://www.forbes.com/companies/icbc/" "http://www.forbes.com/compa
nies/jpmorgan-chase/" "http://www.forbes.com/companies/china-construction-bank/" "http://www.forbes.c
om/companies/agricultural-bank-of-china/" ...
```

5. Data Cleaning

Data science requires the clean-up of existing data. Working with contaminated data may be quite challenging. And today we'll talk about the same thing. Business might suffer from bad or inaccurate data since it can seriously impair choices that depend on it.

Although it may appear uninteresting and tedious, data cleansing is one of the most crucial activities a data science expert must perform. Having inaccurate or poor-quality data might hurt your procedures and analyses. In my case the initial data was in a good quality. However, it seemed little uninteresting and inconsistent. So, I did make sure that data is interesting and not distracting and for this I used two tools one of them is Microsoft Excel and the other is Open Refine which is a open source data cleaning tool

Further this project will describe the data cleaning with details in three categories

- Deleting Columns
- Replacing Missing values
- Merging & Splitting columns

5.1 Deleting Columns

Deleting table columns that aren't essential. This data cleaning approach shows how to eliminate the specified columns that are not necessary for the analysis and are not valuable. The table's "Forbes Webpage" columns were picked out and deleted. We eliminated these columns since we didn't need to provide a URL link for each organization and because users could find them confusing.

CEO	Forbes.Webpage	Profits.as...of.Assets
Shu Gu	http://www.forbes.com/companies/icbc/	0.011209122
Jamie Dimon	http://www.forbes.com/companies/jpmorgan-chase/	0.011960450
Wang Zuji	http://www.forbes.com/companies/china-construction-bank/	0.011483192
Huan Zhao	http://www.forbes.com/companies/agricultural-bank-of-china/	0.009381420
Brian Moynihan	http://www.forbes.com/companies/bank-of-america/	0.012005903
Tim Cook	http://www.forbes.com/companies/apple/	0.159025899
Ma Mingzhe	http://www.forbes.com/companies/ping-an-insurance/	0.015698738
Chen Siqing	http://www.forbes.com/companies/bank-of-china/	0.008877807
Bernardus Margriet van Beurden	http://www.forbes.com/companies/royal-dutch-shell/	0.058440257
Timothy Sloan	http://www.forbes.com/companies/wells-fargo/	0.012245523
Darren W. Woods	http://www.forbes.com/companies/exxon-mobil/	0.060197114
Randall L. Stephenson	http://www.forbes.com/companies/att/	0.036419085

Figure 1.1 - Pre-cleaning

ountry	Headquarters	State	CEO	Profits.as...of.Assets	Profits.as...of.Revenue
ina	China		Shu Gu	0.011209122	0.25713295
nited States	New York	New York	Jamie Dimon	0.011960450	0.24631335
ina	China		Wang Zuji	0.011483192	0.25840080
ina	China		Huan Zhao	0.009381420	0.22475556
nited States	North Carolina	North Carolina	Brian Moynihan	0.012005903	0.25504003
nited States	California	California	Tim Cook	0.159025899	0.22709157
ina	China		Ma Mingzhe	0.015698738	0.10738662
ina	China		Chen Siqing	0.008877807	0.21708755
etherlands	Netherlands		Bernardus Margriet van Beurden	0.058440257	0.06097077

Figure 1.2 - Post-cleaning

Code

```
> forbes_col14<-forbes[,-14]
> view(forbes_col14)
> |
```

forbes_col14<-forbes[,-14]

View(forbes_col14)

5.2 Merging & Splitting columns

To construct location, we combined the state and country columns. In order to prevent misunderstanding, avoid using two independent columns with the same meaning. But after that, we discovered that the state column is missing some data. To get rid of the missing data, we had to split the column once again.

	Industry	Continent	Location	Headquarters	CEO
s	Major Banks	Asia	China-	China	Shu Gu
s	Major Banks	North America	United States-New York	New York	Jamie Dimon
s	Major Banks	Asia	China-	China	Wang Zuji
s	Regional Banks	Asia	China-	China	Huan Zhao
s	Major Banks	North America	United States-North Carolina	North Carolina	Brian Moynihan
ion Technology	Computer Hardware	North America	United States-California	California	Tim Cook
s	Diversified Insurance	Asia	China-	China	Ma Mingzhe
s	Major Banks	Asia	China-	China	Chen Siqing
	Oil & Gas Operations	Europe	Netherlands-	Netherlands	Bernardus Margriet va
s	Major Banks	North America	United States-California	California	Timothy Sloan
	Oil & Gas Operations	North America	United States-Texas	Texas	Darren W. Woods
nunication Services	Telecommunications services	North America	United States-Texas	Texas	Randall L. Stephenson
ion Technology	Semiconductors	Asia	South Korea-	South Korea	Hyun-Suk Kim

Figure 3.1 – After Combining

	Industry	Continent	Country	Headquarters	State	CEO
	Major Banks	Asia	China	China		Shu Gu
	Major Banks	North America	United States	New York	New York	Jamie Dimon
	Major Banks	Asia	China	China		Wang Zuji
	Regional Banks	Asia	China	China		Huan Zhao
	Major Banks	North America	United States	North Carolina	North Carolina	Brian Moynihan
	Computer Hardware	North America	United States	California	California	Tim Cook
	Diversified Insurance	Asia	China	China		Ma Mingzhe
	Major Banks	Asia	China	China		Chen Siqing
	Oil & Gas Operations	Europe	Netherlands	Netherlands		Bernardus Margriet van Beurden
	Major Banks	North America	United States	California	California	Timothy Sloan
	Oil & Gas Operations	North America	United States	Texas	Texas	Darren W. Woods
ices	Telecommunications services	North America	United States	Texas	Texas	Randall L. Stephenson
	Semiconductors	Asia	South Korea	South Korea		Hyun-Suk Kim
	Major Banks	North America	United States	New York	New York	Michael L. Corbat

Figure 3.2 – After Splitting

Code

```
install.packages"
> library("tidyr")
```

install.packages("tidyr")

library(“tidyr”)

Combine Column

```
> Combine_location_column<-unite(forbes_col14,Location,Country,State,sep="-")
> View(Combine_location_column)
>
```

Combine_Location_column<-unite(forbes_col14,Location,Country,State,sep="-")

View(Combine_Location_column)

Split Column

```
> Combine_location_column_split<-separate(Combine_location_column,Location,c("Country","State"),sep="-")
> View(Combine_location_column_split)
>
```

Combine_Location_column_split-
<separate(Combine_Location_column,Location,c("Country","State"),sep="-")

View(Combine_Location_column_split)

5.3 Replacing Missing Values

We removed the missing values as the second step in our data cleaning process. in the state column, which. Since we lacked state-specific data to replace it. To keep the data consistent, we replaced any empty spaces to NA values. so that viewers may see the graphic well.

Continent	Country	Headquarters	State	CEO	Forbes.Webpage
Asia	China	China		Shu Gu	http://www.forbes.com/companie
North America	United States	New York	New York	Jamie Dimon	http://www.forbes.com/companie
Asia	China	China		Wang Zuji	http://www.forbes.com/companie
Asia	China	China		Huan Zhao	http://www.forbes.com/companie
North America	United States	North Carolina	North Carolina	Brian Moynihan	http://www.forbes.com/companie
North America	United States	California	California	Tim Cook	http://www.forbes.com/companie
Asia	China	China		Ma Mingzhe	http://www.forbes.com/companie
Asia	China	China		Chen Siqing	http://www.forbes.com/companie
Europe	Netherlands	Netherlands		Bernardus Margriet van Beurden	http://www.forbes.com/companie
North America	United States	California	California	Timothy Sloan	http://www.forbes.com/companie
North America	United States	Texas	Texas	Darren W. Woods	http://www.forbes.com/companie

Figure 2.1 – Pre cleaning

	Continent	Country	State	Headquarters	CEO	Profits.as...(
	Asia	China	NA	China	Shu Gu	
	North America	United States	New York	New York	Jamie Dimon	
	Asia	China	NA	China	Wang Zuji	
	Asia	China	NA	China	Huan Zhao	
	North America	United States	North Carolina	North Carolina	Brian Moynihan	
	North America	United States	California	California	Tim Cook	
	Asia	China	NA	China	Ma Mingzhe	

Figure 2.2 – Post cleaning

Code

```
> Forbes_dtb<-replace(Forbes_dt, Forbes_dt=='',NA )
```

Forbes_dtb<-replace(Forbes_dt, Forbes_dt='',NA)

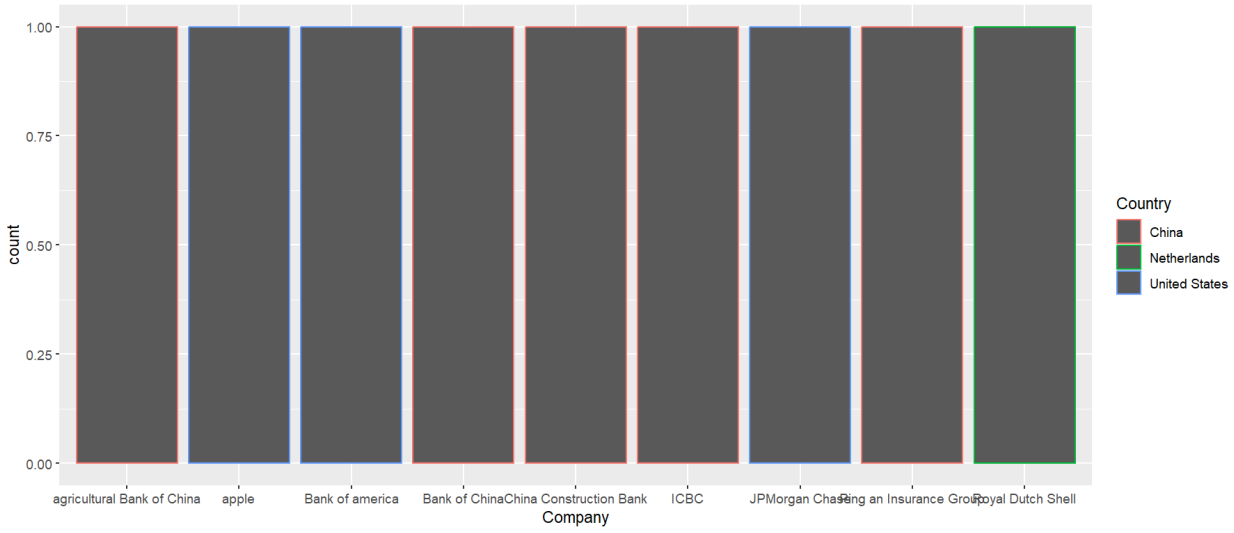
6. Analysis & Visualizations

The graphic display of information and data is known as data visualization. Data visualization tools offer an easy approach to observe and analyze trends, outliers, and patterns in data by utilizing visual components like charts, graphs, and maps. Additionally, it offers a great tool for staff members or business owners to clearly deliver data to non-technical audiences.

Further we will see the following data visualizations using various graphs

5.1 Top companies with countries

We can see that there are top companies with count and country as color in a bar graph.



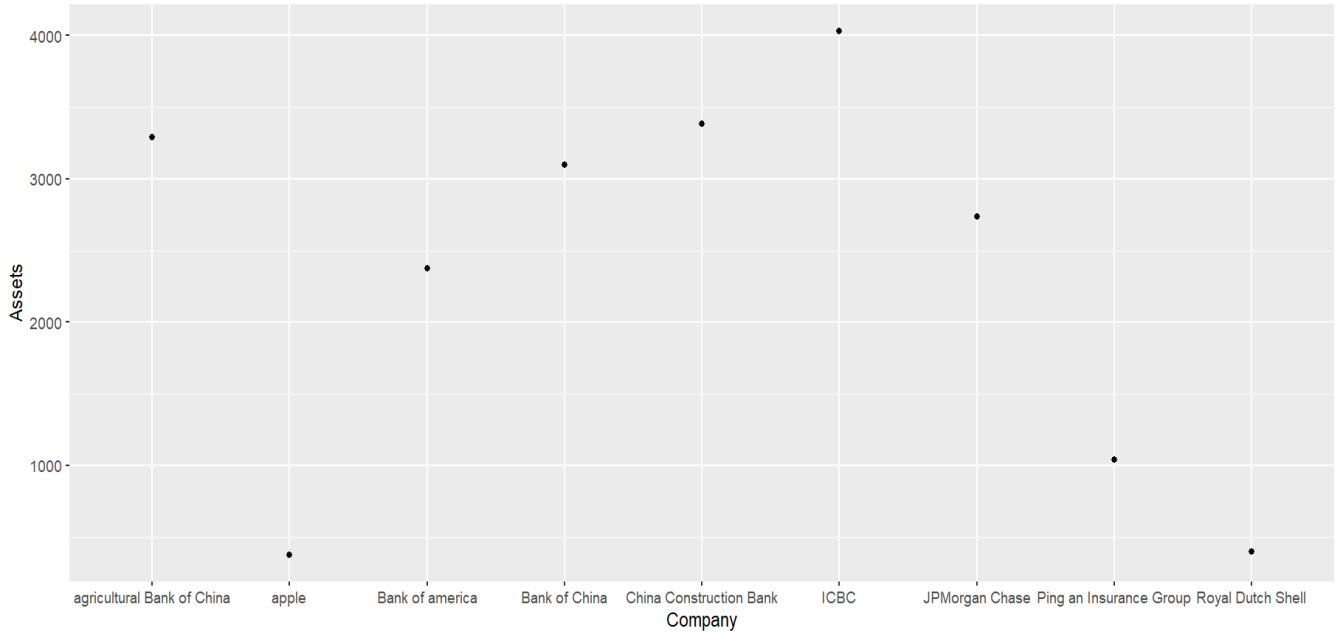

```
ggplot(data=Forbes) + geom_bar(mapping=aes(x=Company, color=Country))
```

Code

ggplot(data=Forbes) + geom_bar(mapping=aes(x=Company, color=Country))

5.2 Top companies with assets

A scatter plot visualizes the top company with its assets



Code

library(ggplot2)

ggplot(Forbes, aes(x=Company, y=Assets))

ggplot(Forbes, aes(x=Company, y=Assets)) + geom_point()

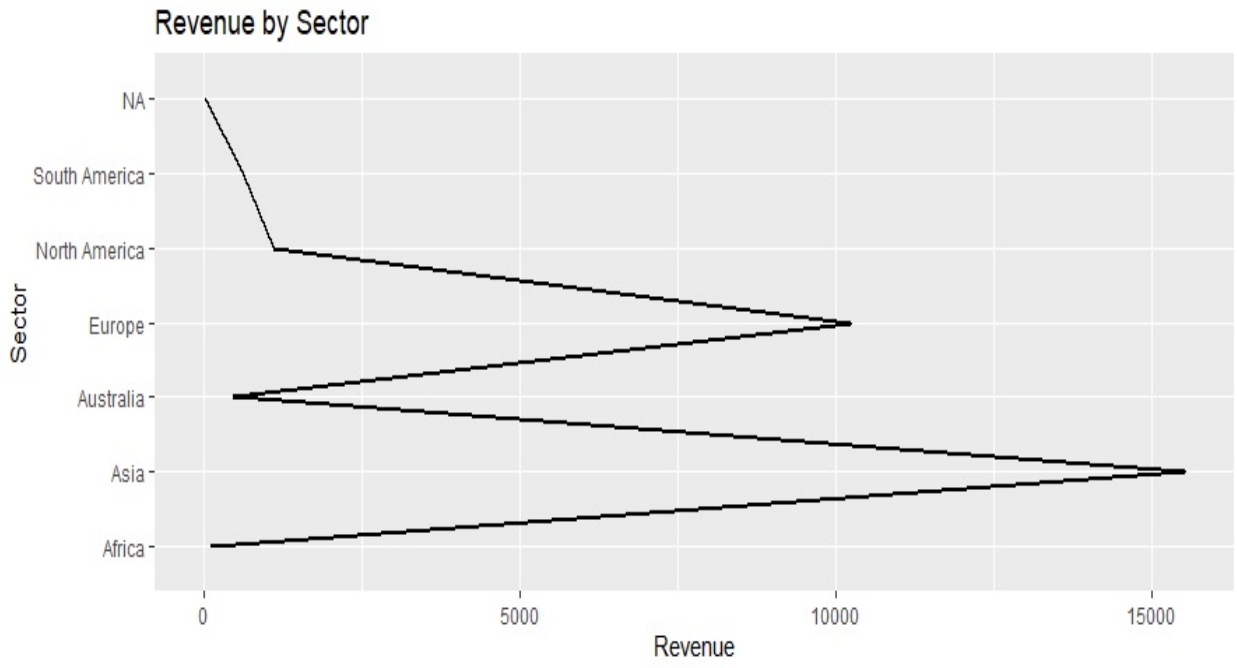
p <-ggplot(Forbes, aes(x=Company, y=Assets))

p + geom_point()

```
ggplot(Forbes, aes(x=Company, y=Assets))
ggplot(Forbes, aes(x=Company, y=Assets)) + geom_point()
p <-ggplot(Forbes, aes(x=Company, y=Assets))
p + geom_point()
```

5.3 Line Chart

This line chart describes sectors with their revenue with respect to countries




```
ggplot(data=Forbes_dtb,mapping=aes(x= Revenue, y= Continent)) +
  stat_summary(fun=sum,na.rm=TRUE,geom='line',aes(group=State),size=1)+
  labs(x="Revenue", y="Sector", title = "Revenue by Sector")
```

Code

ggplot(data=Forbes_dtb,mapping=aes(x= Revenue, y= Continent)) +

stat_summary(fun=sum,na.rm=TRUE,geom='line',aes(group=State),size=1)+

labs(x="Revenue", y="Sector", title = "Revenue by Sector")

7. Statistical Summary

```
> View(Forbes_dtb)
> summary(Forbes_dtb)
   Company           Market.Value        Revenue
 Length:2000        Min.    :  0.009   Min.    : -9.140
 Class :character   1st Qu.:  6.857   1st Qu.:  4.748
 Mode  :character   Median : 13.367   Median : 10.255
                    Mean   : 28.408   Mean   : 20.599
                    3rd Qu.: 27.474   3rd Qu.: 20.588
                    Max.   :961.257   Max.   :514.405
                                      NA's   :2
    Profits             Assets             Rank
 Min.   :-22.3670   Min.   :   1.537   Min.   :   1.0
 1st Qu.:  0.4205   1st Qu.:  12.042   1st Qu.: 500.8
 Median :  0.7760   Median :  26.254   Median :1000.5
 Mean   :  1.7048   Mean   :  93.582   Mean   :1000.3
 3rd Qu.:  1.6840   3rd Qu.:  61.000   3rd Qu.:1500.2
 Max.   : 59.4310   Max.   :4034.482   Max.   :1999.0
 NA's   :1          NA's   :5
    Sector            Industry          Continent
 Length:2000        Length:2000        Length:2000
 Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character



  Country            State           Headquarters
 Length:2000        Length:2000        Length:2000
 Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character



    CEO             Profits.as...of.Assets Profits.as...of.Revenue
 Length:2000        Min.    :-0.97605      Min.    : -1.74987
 Class :character   1st Qu.: 0.01027       1st Qu.:  0.03979
 Mode  :character   Median : 0.03287       Median :  0.09559
```

We used R Studio's summary () function to condense the entirety of the Forbes Global 2000 - 2019.csv dataset. After summarizing each column, it delivers the values for the respective summary columns. We can examine the Forbes Global

2000 - 2019.csv dataset's summary values for the min, median, mean, and max for each column.

8. Individual Statistical Summary

8.1 Market Value

```
> summary(Forbes_dtb$Market.value)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.009   6.857  13.367  28.408  27.474 961.257
> min(Forbes_dtb$Market.value)
[1] 0.009
> max(Forbes_dtb$Market.value)
[1] 961.257
> mean(Forbes_dtb$Market.value)
[1] 28.40805
> median(Forbes_dtb$Market.value)
[1] 13.3665
> sd(Forbes_dtb$Market.value)
[1] 60.02556
> Forbes_dtb_Market<-filter(
+ Forbes_dtb,
+ Market.value > 28
+ )
> dim(Forbes_dtb_Market)
[1] 489  15
> 
```

Code

```
summary(Forbes_dtb$Market.Value)
min(Forbes_dtb$Market.Value)
max(Forbes_dtb$Market.Value)
mean(Forbes_dtb$Market.Value)
median(Forbes_dtb$Market.Value)
sd(Forbes_dtb$Market.Value)

Forbes_dtb_Market<-filter(
+ Forbes_dtb,
```

+ Market.Value > 28

+)

dim(Forbes_dtb_Market)

8.2 Rank

```
> summary(Forbes_dtb$Rank)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   500.8  1000.5  1000.3  1500.2  1999.0
> min(Forbes_dtb$Rank)
[1] 1
> max(Forbes_dtb$Rank)
[1] 1999
> mean(Forbes_dtb$Rank)
[1] 1000.317
> median(Forbes_dtb$Rank)
[1] 1000.5
> sd(Forbes_dtb$Rank)
[1] 577.3995
>
```

Code

summary(Forbes_dtb$Rank)
min(Forbes_dtb$Rank)

max(Forbes_dtb$Rank)

mean(Forbes_dtb$Rank)

median(Forbes_dtb$Rank)

sd(Forbes_dtb$Rank)

8. References

i. "andrea Murphy", " I. C. (n.d.). *The global 2000 2022*. Forbes. Retrieved November 2, 2022, from https://www.forbes.com/lists/global2000/?sh=11aaea0e5ac0

ii. Ponciano, J. (2019, May 19). The world's Largest Public Companies 2019: Global 2000 by the numbers. Forbes. Retrieved November 2, 2022, from https://www.forbes.com/sites/jonathanponciano/2019/05/15/worlds-largest-companies-2019-global-2000/?sh=7cdd7ed64ada

iii. M, S. (2022, October 21). Top 8 excel data cleaning techniques to know in 2022: Simplilearn. Simplilearn.com. Retrieved November 2, 2022, from https://www.simplilearn.com/tutorials/excel-tutorial/excel-data-cleaning

iv. Johnson, D. (2022, September 17). SAS tutorial for beginners: What is & programming example. Guru99. Retrieved November 2, 2022, from https://www.guru99.com/sas-tutorial.html