

CSE 3020 Data Visualisation

Lab Assessment 3

NITIN RANJAN, 18BCE0272

1. Create a Data Frame from the file

```
dat <- read.csv("C:/Users/lenovo/Desktop/Fortune1000.csv")
```

```
dat <- data.frame(dat) #18BCE0272
```

Screenshot:

```
> dat <- read.csv("C:/Users/lenovo/Desktop/Fortune1000.csv")
> dat <- data.frame(dat) #18BCE0272
> |
> dat      #18BCE0272
  Rank      Company      Sector
1     1      walmart      Retailing
2     2    Exxon Mobil      Energy
3     3       Apple      Technology
4     4 Berkshire Hathaway    Financials
5     5    McKesson      Health Care
6     6 UnitedHealth Group    Health Care
7     7    CVS Health    Food and Drug Stores
8     8  General Motors    Motor Vehicles & Parts
9     9    Ford Motor    Motor Vehicles & Parts
10    10      AT&T      Telecommunications
11    11  General Electric      Industrials
12    12 AmerisourceBergen    Health Care
13    13    verizon      Telecommunications
14    14    Chevron      Energy
15    15    Costco      Retailing
16    16  Fannie Mae      Financials
17    17    Kroger    Food and Drug Stores
18    18  Amazon.com      Technology
19    19 walgreens Boots Alliance    Food and Drug Stores
20    20      HP      Technology
21    21  Cardinal Health    Health Care
22    22 Express Scripts Holding    Health Care
23    23  J.P. Morgan Chase      Financials
24    24    Boeing    Aerospace & Defense
25    25  Microsoft      Technology
26    26 Bank of America Corp.    Financials
27    27  wells fargo      Financials
28    28  Home Depot      Retailing
29    29  Citigroup      Financials
30    30  Phillips 66      Energy
31    31      IBM      Technology
32    32  Valero Energy      Energy
33    33    Anthem      Health Care
34    34 Procter & Gamble    Household Products
```

18BCE0272
NITIN PRAMOD RANJAN

2. How many unique sectors are in the data set?

```
> #18BCE0272  
> length(unique(dat$Sector))  
[1] 21
```

So, there are 21 unique sectors in the dataset.

3. How many companies are in each sector?

```
> #Number of companies in each sector should be equivalent to the frequency of the occurrence of the sectors.  
>  
> as.data.frame(table(dat$Sector)) #18BCE0272, NITIN PRAMOD RANJAN
```

```
Var1 Freq  
1 Aerospace & Defense 20  
2 Apparel 15  
3 Business Services 51  
4 Chemicals 30  
5 Energy 122  
6 Engineering & Construction 26  
7 Financials 139  
8 Food and Drug Stores 15  
9 Food, Beverages & Tobacco 43  
10 Health Care 75  
11 Hotels, Restaurants & Leisure 25  
12 Household Products 28  
13 Industrials 46  
14 Materials 43  
15 Media 25  
16 Motor Vehicles & Parts 24  
17 Retailing 80  
18 Technology 102  
19 Telecommunications 15  
20 Transportation 36  
21 Wholesalers 40  
>  
> #18bce0272|
```

4. What are the top 5 sectors based on the total employee size.

Step 1:

Create a table of sectors vs employee size

```
> #18bce0272
>
> aggregate(dat$Employees, by=list(Category=dat$Sector), FUN=sum) #18BCE0272
```

| | Category | x |
|----|-------------------------------|---------|
| 1 | Aerospace & Defense | 968057 |
| 2 | Apparel | 346397 |
| 3 | Business Services | 1361050 |
| 4 | Chemicals | 463651 |
| 5 | Energy | 1188927 |
| 6 | Engineering & Construction | 406708 |
| 7 | Financials | 3359948 |
| 8 | Food and Drug Stores | 1395398 |
| 9 | Food, Beverages & Tobacco | 1211632 |
| 10 | Health Care | 2678289 |
| 11 | Hotels, Restaurants & Leisure | 2484245 |
| 12 | Household Products | 646038 |
| 13 | Industrials | 1545229 |
| 14 | Materials | 638123 |
| 15 | Media | 550314 |
| 16 | Motor Vehicles & Parts | 1082560 |
| 17 | Retailing | 6227629 |
| 18 | Technology | 3578949 |
| 19 | Telecommunications | 832468 |
| 20 | Transportation | 1536793 |
| 21 | wholesalers | 525597 |

```
.
```

Step 2:

Sort the table

```
> m <- m[order(m$x),]
> m
```

| | Category | x |
|----|-------------------------------|---------|
| 2 | Apparel | 346397 |
| 6 | Engineering & Construction | 406708 |
| 4 | Chemicals | 463651 |
| 21 | wholesalers | 525597 |
| 15 | Media | 550314 |
| 14 | Materials | 638123 |
| 12 | Household Products | 646038 |
| 19 | Telecommunications | 832468 |
| 1 | Aerospace & Defense | 968057 |
| 16 | Motor Vehicles & Parts | 1082560 |
| 5 | Energy | 1188927 |
| 9 | Food, Beverages & Tobacco | 1211632 |
| 3 | Business Services | 1361050 |
| 8 | Food and Drug Stores | 1395398 |
| 20 | Transportation | 1536793 |
| 13 | Industrials | 1545229 |
| 11 | Hotels, Restaurants & Leisure | 2484245 |
| 10 | Health Care | 2678289 |
| 7 | Financials | 3359948 |
| 18 | Technology | 3578949 |
| 17 | Retailing | 6227629 |

Step 3:

Find the last 5 elements of the table

```
> tail(m, n=5)
```

| | Category | x |
|----|-------------------------------|---------|
| 11 | Hotels, Restaurants & Leisure | 2484245 |
| 10 | Health Care | 2678289 |
| 7 | Financials | 3359948 |
| 18 | Technology | 3578949 |
| 17 | Retailing | 6227629 |

```
> #18BCE0272|
```

These 5 sectors employ most employees in reverse order, Retailing being the largest employer and Hotels, Restaurants and Leisure being 5th largest.

5. What are the top 10 industries based on the total revenue .

Step 1:

Generate a table of all company by revenue

```
> #18BCE0272  
> m <- aggregate(dat$Revenue, by=list(Category=dat$Company), FUN=sum) #18BCE0272  
> |
```

```
> m  


|    | Category                  | x     |
|----|---------------------------|-------|
| 1  | 3M                        | 30274 |
| 2  | 99 Cents Only Stores      | 1999  |
| 3  | A-Mark Precious Metals    | 6070  |
| 4  | A. Schulman               | 2392  |
| 5  | A.O. Smith                | 2537  |
| 6  | Aaron's                   | 3180  |
| 7  | Abbott Laboratories       | 20661 |
| 8  | AbbVie                    | 22859 |
| 9  | Abercrombie & Fitch       | 3519  |
| 10 | ABM Industries            | 5291  |
| 11 | Activision Blizzard       | 4664  |
| 12 | Acuity Brands             | 2707  |
| 13 | Adams Resources & Energy  | 1944  |
| 14 | Adobe Systems             | 4796  |
| 15 | ADP                       | 11477 |
| 16 | ADT                       | 3574  |
| 17 | Advance Auto Parts        | 9737  |
| 18 | Advanced Micro Devices    | 3991  |
| 19 | AECOM                     | 17990 |
| 20 | AES                       | 14963 |
| 21 | Aetna                     | 60337 |
| 22 | Affiliated Managers Group | 2485  |
| 23 | Aflac                     | 20872 |
| 24 | AGCO                      | 7467  |
| 25 | Agilent Technologies      | 4038  |
| 26 | AGL Resources             | 3941  |
| 27 | AIG                       | 58327 |
| 28 | Air Products & Chemicals  | 9895  |
| 29 | Airgas                    | 5305  |
| 30 | AK Steel Holding          | 6693  |
| 31 | Akamai Technologies       | 2197  |


```

Step 2:

Sort the table

```
> #18BCE0272  
> m <- m[order(m$x),]  
> |
```

Step 3:

Find the last 10 in the list

```
> tail(m, n=10)
      Category      x
98      AT&T 146801
355    Ford Motor 149558
373  General Motors 152356
257    CVS Health 153290
904 UnitedHealth Group 157107
566    McKesson 181241
126 Berkshire Hathaway 210821
83      Apple 233715
331    Exxon Mobil 246204
945    Walmart 482130
> #18BCE0272, Nitin Pramod Ranjan
> |
```

So, Walmart has maximum Revenue while AT&T has 10th largest revenue.

6. List the top 20 companies based on their profits, show their sectors and industries, and profits

Step 1:

Create a table of companies with their effective profit, sectors and industries

Syntax:

```
> #18BCE0272, Nitin Pramod Ranjan
```

```
> m <- aggregate(dat$Profits, by=list(dat$Sector, dat$Industry ,Category=dat$Company), FUN=sum)
#18BCE0272
```

```
> #18BCE0272, Nitin Pramod Ranjan
> m <- aggregate(dat$Profits, by=list(dat$Sector, dat$Industry ,Category=dat$Company), FUN=sum) #18BCE0272
> |
```

18BCE0272
NITIN PRAMOD RANJAN

Step 2:

Sort the table and print the last 20 elements

```
> #18BCE0272, Nitin Pramod Ranjan
> m <- aggregate(dat$Profits, by=list(dat$Sector, dat$Industry ,category=dat$Company), FUN=sum) #18BCE0272
> m <- m[order(m$x),]
> tail(m, n=20)
```

| | Group.1 | Group.2 | Category | x |
|-----|------------------------|--|-----------------------|-------|
| 208 | Technology | Network and Other Communications Equipment | Cisco Systems | 8981 |
| 373 | Motor Vehicles & Parts | Motor Vehicles and Parts | General Motors | 9687 |
| 656 | Technology | Computer Software | Oracle | 9938 |
| 334 | Financials | Diversified Financials | Fannie Mae | 10954 |
| 463 | Technology | Semiconductors and Other Electronic Components | Intel | 11420 |
| 583 | Technology | Computer Software | Microsoft | 12193 |
| 450 | Technology | Information Technology Services | IBM | 13190 |
| 98 | Telecommunications | Telecommunications | AT&T | 13345 |
| 945 | Retailing | General Merchandisers | Walmart | 14694 |
| 483 | Health Care | Pharmaceuticals | Johnson & Johnson | 15409 |
| 114 | Financials | Commercial Banks | Bank of America Corp. | 15888 |
| 331 | Energy | Petroleum Refining | Exxon Mobil | 16150 |
| 49 | Technology | Internet Services and Retailing | Alphabet | 16348 |
| 210 | Financials | Commercial Banks | Citigroup | 17242 |
| 927 | Telecommunications | Telecommunications | Verizon | 17879 |
| 380 | Health Care | Pharmaceuticals | Gilead Sciences | 18108 |
| 955 | Financials | Commercial Banks | Wells Fargo | 22894 |
| 126 | Financials | Insurance: Property and Casualty (Stock) | Berkshire Hathaway | 24083 |
| 478 | Financials | Commercial Banks | J.P. Morgan Chase | 24442 |
| 83 | Technology | Computers, Office Equipment | Apple | 53394 |

Output:

```
> #18BCE0272
```

| | Group.1 | Group.2 | Category | x |
|-----|------------------------|--|-----------------------|-------|
| 208 | Technology | Network and Other Communications Equipment | Cisco Systems | 8981 |
| 373 | Motor Vehicles & Parts | Motor Vehicles and Parts | General Motors | 9687 |
| 656 | Technology | Computer Software | Oracle | 9938 |
| 334 | Financials | Diversified Financials | Fannie Mae | 10954 |
| 463 | Technology | Semiconductors and Other Electronic Components | Intel | 11420 |
| 583 | Technology | Computer Software | Microsoft | 12193 |
| 450 | Technology | Information Technology Services | IBM | 13190 |
| 98 | Telecommunications | Telecommunications | AT&T | 13345 |
| 945 | Retailing | General Merchandisers | Walmart | 14694 |
| 483 | Health Care | Pharmaceuticals | Johnson & Johnson | 15409 |
| 114 | Financials | Commercial Banks | Bank of America Corp. | 15888 |
| 331 | Energy | Petroleum Refining | Exxon Mobil | 16150 |
| 49 | Technology | Internet Services and Retailing | Alphabet | 16348 |
| 210 | Financials | Commercial Banks | Citigroup | 17242 |
| 927 | Telecommunications | Telecommunications | Verizon | 17879 |
| 380 | Health Care | Pharmaceuticals | Gilead Sciences | 18108 |
| 955 | Financials | Commercial Banks | Wells Fargo | 22894 |
| 126 | Financials | Insurance: Property and Casualty (Stock) | Berkshire Hathaway | 24083 |
| 478 | Financials | Commercial Banks | J.P. Morgan Chase | 24442 |
| 83 | Technology | Computers, Office Equipment | Apple | 53394 |

So, While Computers, Office Equipment is the company with maximum profit, the 20th largest profit is made by Network and other communications Equipment.

7. List the average revenue, profit, and employee size for each industry.

```
> m <- aggregate(dat$Employees, by = list(Category=dat$Industry), FUN = mean)
> n <- aggregate(dat$Profits, by = list(Category=dat$Industry), FUN = mean) #18BCE0272
> o <- aggregate(dat$Revenue, by = list(Category=dat$Industry), FUN = mean) #18BCE0272
.
|
> library(data.table)
> m <- data.table(m, key= "Category") #18BCE0272, Nitin Pramod Ranjan
> n <- data.table(n, key= "Category")
> o <- data.table(o, key= "Category") #18BCE0272
> p <- merge(m,n)
> p <- merge(p,o)
.
|
```

OUTPUT:

x.x will represent Employee size, x.y will represent Profits and x will represent Revenue because the merge() function will create a join and x.x and x.y are extra columns in the final join, while Revenue is the new column being added and Category or the Industry is the key column for carrying out the join.

18BCE0272
NITIN PRAMOD RANJAN

```
> o <- data.table(o, key= "category") #18BCE0272
> p <- merge(m,n)
> p <- merge(p,o)
>
> p
```

| | Category | x.x | x.y | x |
|-----|--|------------|------------|-----------|
| 1: | Advertising, marketing | 62050.000 | 774.50000 | 11374.000 |
| 2: | Aerospace and Defense | 48402.850 | 1437.10000 | 17897.000 |
| 3: | Airlines | 43511.222 | 2644.44444 | 17660.889 |
| 4: | Apparel | 23093.133 | 549.06667 | 6397.867 |
| 5: | Automotive Retailing, Services | 17497.400 | 255.00000 | 11382.300 |
| 6: | Beverages | 22123.000 | 1299.77778 | 8821.778 |
| 7: | Building Materials, Glass | 10283.714 | 317.42857 | 3659.286 |
| 8: | Chemicals | 15455.033 | 754.26667 | 8129.900 |
| 9: | Commercial Banks | 54965.143 | 4634.28571 | 22273.893 |
| 10: | Computer Peripherals | 43814.750 | 993.75000 | 12237.500 |
| 11: | Computer Software | 27771.143 | 1889.78571 | 12793.714 |
| 12: | Computers, Office Equipment | 66869.286 | 8347.57143 | 50540.143 |
| 13: | Construction and Farm Machinery | 34107.667 | 553.44444 | 13401.333 |
| 14: | Diversified Financials | 21940.933 | 1903.20000 | 21805.000 |
| 15: | Diversified Outsourcing Services | 50595.000 | 307.50000 | 4630.643 |
| 16: | Education | 15585.000 | 23.00000 | 2495.000 |
| 17: | Electronics, Electrical Equip. | 37214.077 | 795.76923 | 9181.923 |
| 18: | Energy | 5005.143 | -931.28571 | 4839.214 |
| 19: | Engineering, Construction | 26293.143 | 126.07143 | 7126.714 |
| 20: | Entertainment | 21309.111 | 1427.00000 | 10348.667 |
| 21: | Financial Data Services | 13943.474 | 918.73684 | 5304.105 |
| 22: | Food Consumer Products | 33271.950 | 963.15000 | 11947.700 |
| 23: | Food Production | 24289.000 | 554.33333 | 19661.778 |
| 24: | Food Services | 142010.167 | 906.08333 | 7372.000 |
| 25: | Food and Drug Stores | 93026.533 | 1117.26667 | 32251.267 |
| 26: | Forest and Paper Products | 7816.000 | -1.00000 | 4314.000 |
| 27: | General Merchandisers | 340810.000 | 2013.70000 | 68432.000 |
| 28: | Health Care: Insurance and Managed Care | 41589.909 | 1365.36364 | 42124.818 |
| 29: | Health Care: Medical Facilities | 67323.357 | 213.07143 | 10649.857 |
| 30: | Health Care: Pharmacy and Other Services | 23728.077 | 434.15385 | 11653.923 |
| 31: | Home Equipment, Furnishings | 20950.833 | 282.33333 | 5252.583 |
| 32: | Homebuilders | 3217.000 | 294.91667 | 4517.417 |
| 33: | Hotels, Casinos, Resorts | 60000.462 | 755.60221 | 6227.077 |

18BCE0272

NITIN PRAMOD RANJAN

| | | | | |
|-----|--|------------|-------------|-----------|
| 37: | Insurance: Life, Health (Mutual) | 5880.100 | 560.10000 | 18133.600 |
| 38: | Insurance: Life, Health (stock) | 11235.167 | 1096.16667 | 13641.722 |
| 39: | Insurance: Property and Casualty (Mutual) | 19021.143 | 1147.28571 | 20526.000 |
| 40: | Insurance: Property and Casualty (stock) | 28361.536 | 1600.10714 | 19572.393 |
| 41: | Internet Services and Retailing | 28723.667 | 1482.13333 | 17275.733 |
| 42: | Mail, Package, and Freight Delivery | 332137.500 | 2947.00000 | 52908.000 |
| 43: | Medical Products and Equipment | 21249.250 | 677.12500 | 5828.312 |
| 44: | Metals | 13825.643 | -160.07143 | 7137.500 |
| 45: | Mining, Crude-oil Production | 5325.357 | -4448.39286 | 6301.250 |
| 46: | Miscellaneous | 17138.000 | 781.00000 | 6732.000 |
| 47: | Motor Vehicles and Parts | 45106.667 | 1079.08333 | 20105.833 |
| 48: | Network and Other Communications Equipment | 21379.286 | 1350.50000 | 8978.357 |
| 49: | Oil and Gas Equipment, Services | 25088.333 | -449.66667 | 9203.000 |
| 50: | Packaging, Containers | 20840.375 | 277.81250 | 6622.812 |
| 51: | Petroleum Refining | 14404.706 | 1875.76471 | 41498.353 |
| 52: | Pharmaceuticals | 31588.000 | 4551.53333 | 20864.667 |
| 53: | Pipelines | 6011.750 | 484.50000 | 11563.000 |
| 54: | Publishing, Printing | 23821.429 | -191.28571 | 4926.857 |
| 55: | Railroads | 24368.800 | 1801.00000 | 9710.800 |
| 56: | Real estate | 10705.267 | 632.00000 | 4183.200 |
| 57: | Scientific, Photographic and Control Equipment | 19291.750 | 778.83333 | 5486.000 |
| 58: | Securities | 7940.333 | 791.94444 | 4726.167 |
| 59: | Semiconductors and Other Electronic Components | 26065.391 | 1074.82609 | 7783.696 |
| 60: | Specialty Retailers: Apparel | 42807.944 | 254.61111 | 6831.944 |
| 61: | Specialty Retailers: Other | 44619.333 | 489.52381 | 12951.381 |
| 62: | Telecommunications | 55497.867 | 3242.46667 | 30788.933 |
| 63: | Temporary Help | 12004.000 | 200.00000 | 6943.200 |
| 64: | Tobacco | 25697.000 | 3093.40000 | 12132.200 |
| 65: | Transportation Equipment | 11945.200 | 487.60000 | 4771.000 |
| 66: | Transportation and Logistics | 24938.200 | 179.00000 | 6647.400 |
| 67: | Trucking, Truck Leasing | 18939.556 | 212.22222 | 3994.444 |
| 68: | Utilities: Gas and Electric | 10249.488 | 732.07317 | 8375.878 |
| 69: | Waste Management | 23839.800 | 343.60000 | 6090.800 |
| 70: | Wholesalers: Diversified | 9353.240 | 207.72000 | 7045.520 |
| 71: | Wholesalers: Electronics and Office Equipment | 20832.625 | 232.12500 | 18488.250 |
| 72: | Wholesalers: Food and Grocery | 19317.500 | 194.33333 | 18629.000 |
| 73: | Wholesalers: Health Care | 26000.000 | 560.16667 | 74085.333 |
| | Category | x.x | x.y | x |

> #18BCE0272, Nitin Ranjan|

18BCE0272
NITIN PRAMOD RANJAN

8. Draw a bar chart to show the revenues of the top 20 companies

```
> #18BCE0272, Nitin Ranjan
> m <- aggregate(dat$Revenue, by = list(Category=dat$Company), FUN = sum) #18BCE0272

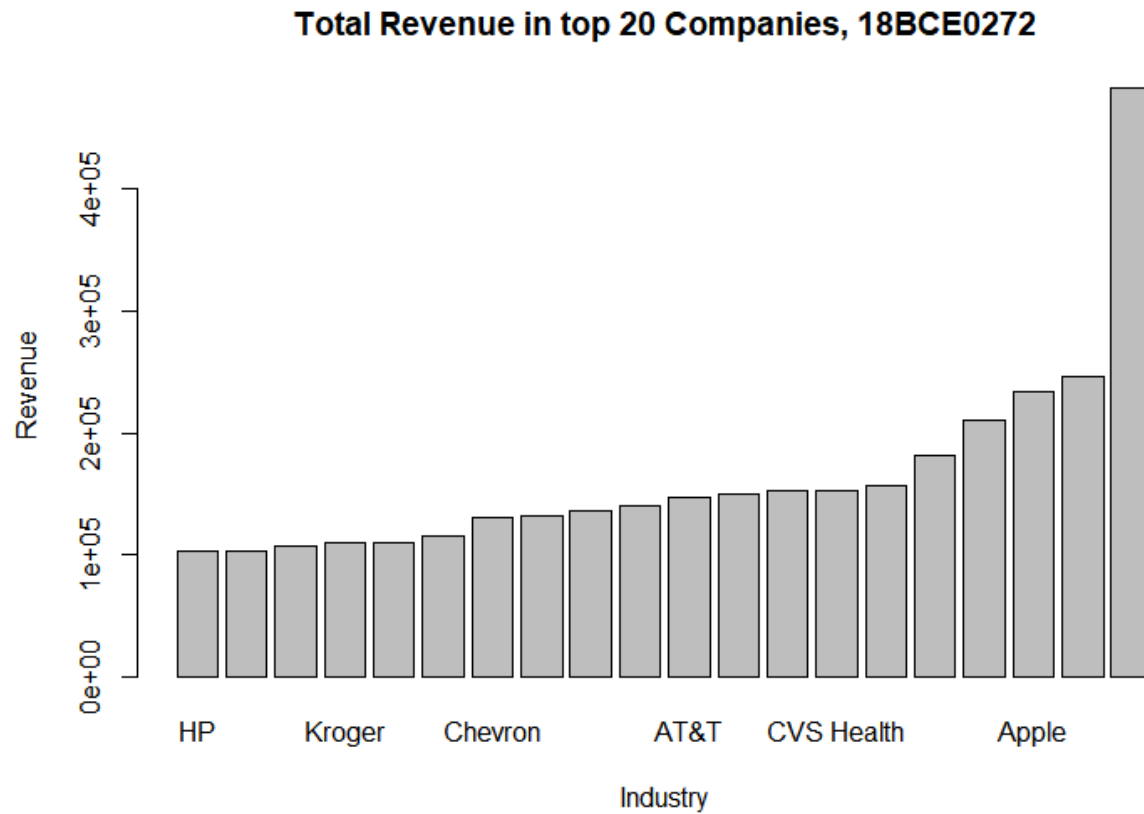
> m <- m[order(m$x),]
> m <- tail(m, n=20)
> #18BCE0272, Nitin PRamod Ranjan
.

> m #18BCE0272
```

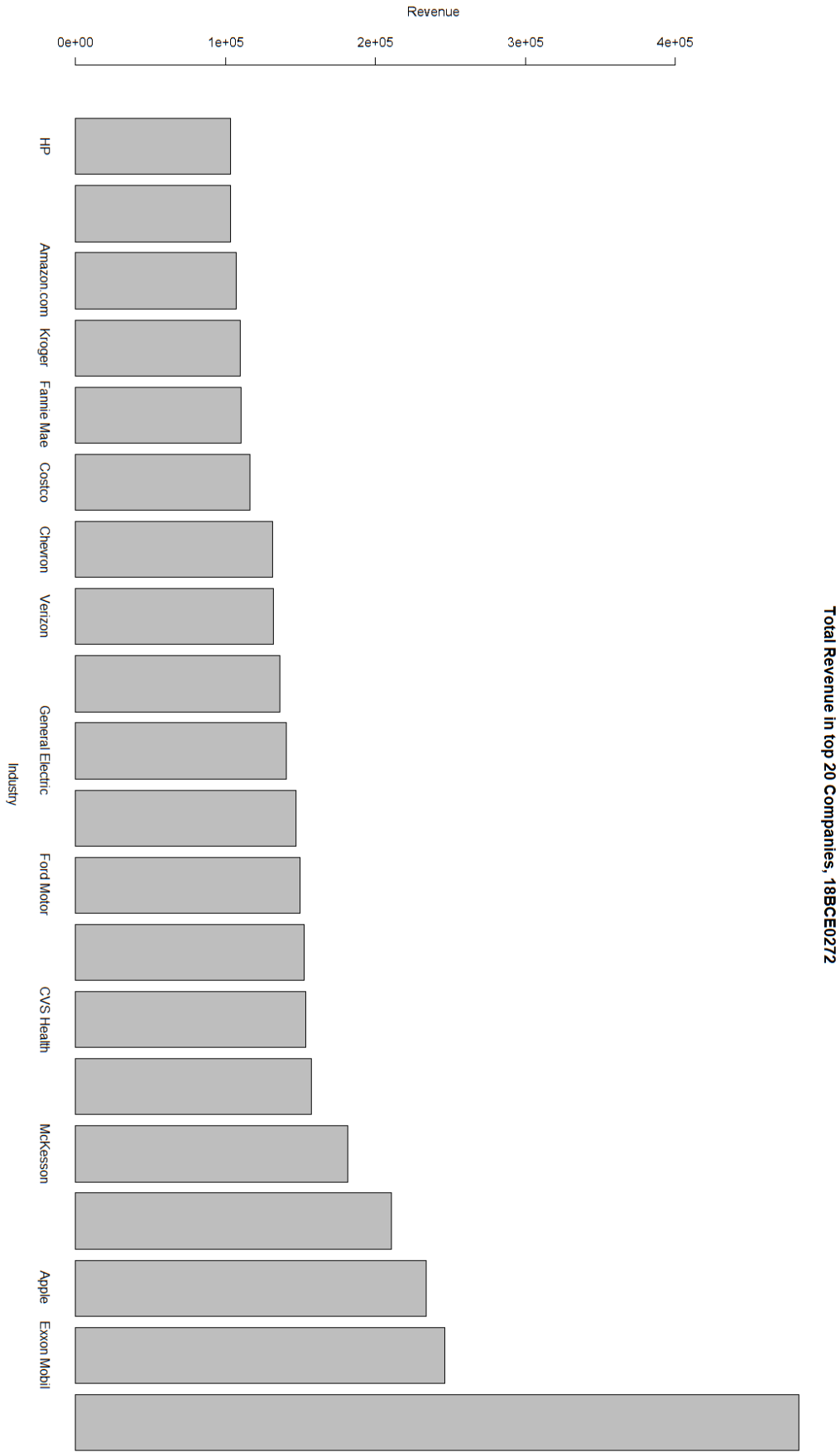
| | Category | x |
|-----|--------------------------|--------|
| 437 | HP | 103355 |
| 944 | walgreens Boots Alliance | 103444 |
| 51 | Amazon.com | 107006 |
| 510 | Kroger | 109830 |
| 334 | Fannie Mae | 110359 |
| 244 | Costco | 116199 |
| 198 | Chevron | 131118 |
| 927 | Verizon | 131620 |
| 67 | AmerisourceBergen | 135962 |
| 370 | General Electric | 140389 |
| 98 | AT&T | 146801 |
| 355 | Ford Motor | 149558 |
| 373 | General Motors | 152356 |
| 257 | CVS Health | 153290 |
| 904 | UnitedHealth Group | 157107 |
| 566 | McKesson | 181241 |
| 126 | Berkshire Hathaway | 210821 |
| 83 | Apple | 233715 |
| 331 | Exxon Mobil | 246204 |
| 945 | walmart | 482130 |

```
> barplot(m$x, names.arg = m$Category, xlab= 'Industry', ylab = 'Revenue',
  main = 'Total Revenue in top 20 Companies, 18BCE0272')
> #18BCE0272
> |
```

OUTPUT:



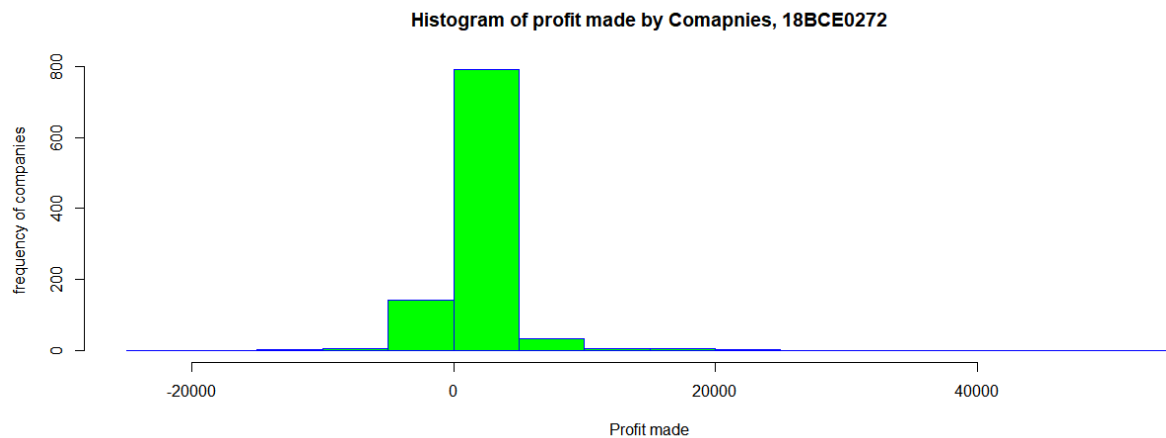
The names of Companies appear if the viewer space in R is expanded. The following image is the limit of my desktop screen.



9. Draw a histogram of the companies based on their profits

```
> #18BCE0272  
> m <- aggregate(dat$Profits, by = list(Category=dat$Company),  
  FUN = sum) #18BCE0272  
  
> hist(m$x, xlab= "Profit made", ylab = "frequency of companies",  
  main = "Histogram of profit made by Comapnies, 18BCE0272", border  
  = "blue", col = "green")  
> #18BCE0272  
> |
```

OUTPUT:



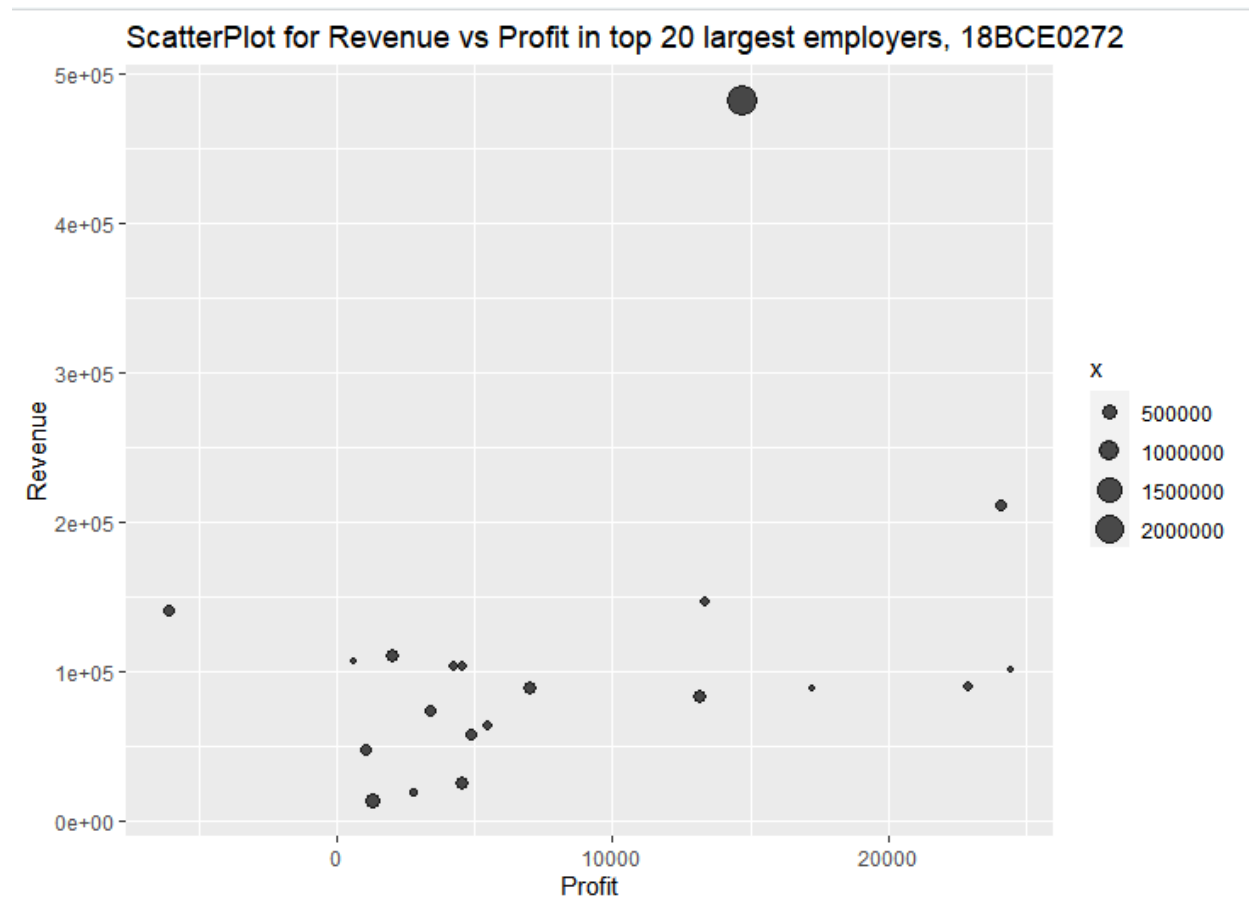
18BCE0272
NITIN PRAMOD RANJAN

10. Draw a scatter plot based on revenue and profit of the top 20 companies based on the employee size, use the employee number as the size of the circle

```
>
> #18BCE0272, Nitin Pramod Ranjan
> m <- aggregate(dat$Employees, by = list(dat$Profits, dat$Revenue, Category=dat$Company), FUN = sum) #18BCE0272
> m <- m[order(m$x),]
> m <- tail(m, n=20) #18BCE0272
> |

.
> dev.off()
null device
      1
> library(ggplot2)
> library(dplyr)
>
> #18BCE0272, Nitin Pramod Ranjana
> #18BCE0272, Nitin Pramod Ranjan

-----
>
> ggplot(m, aes(x= Group.1, y=Group.2, size = x))+geom_point(alpha = 0.7)
+ labs(title = "ScatterPlot for Revenue vs Profit in top 20 largest employers, 18BCE0272" , x= "Profit" , y = "Revenue" ) #18BCE0272
> #Nitin Pramod Ranjan
> |
```

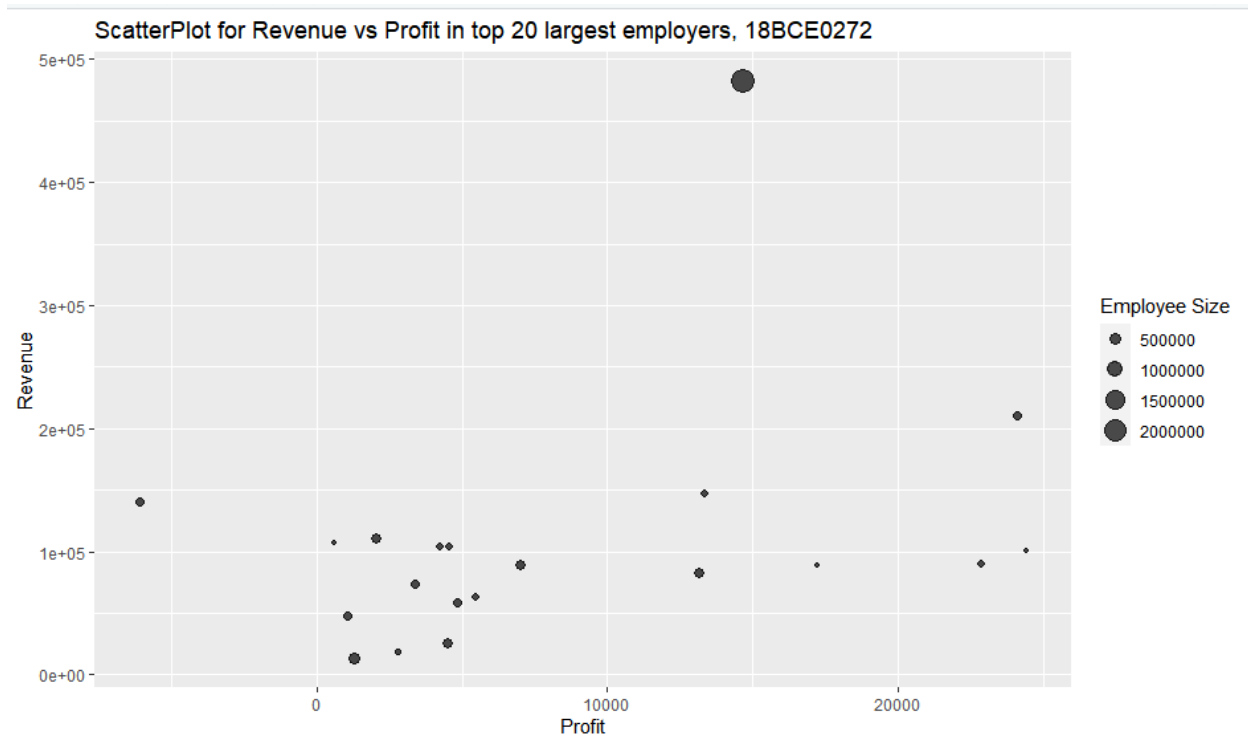


RENAMING THE LEGEND

```
> ggplot(m, aes(x= Group.1, y=Group.2, size = x))+geom_point  
(alpha = 0.7) + labs(title = "ScatterPlot for Revenue vs Prof  
it in top 20 largest employers, 18BCE0272" , x= "Profit" , y  
= "Revenue", size="Employee Size")  
> #18BCE0272  
> #Nitin Pramod Ranjan  
> |
```

FINAL OUTPUT:

18BCE0272
NITIN PRAMOD RANJAN



Submitted by:

18BCE0272,

Nitin Pramod Ranjan

For Data Visualisation Lab Assessment 3

On 15th April 2020