Bao Huynh
CIS 320

Job and Employment Report in 2012

Introduction

Data include number of firms, number of establishments, employment, annual payroll and estimated receipts by employment and receipt size of the enterprise. The industry classification is based on 2012 North American Industry Classification System. In the year 2012, the unemployment rates was 8.1 percent according to Bureau of Labor Statistics.

Hypothesis

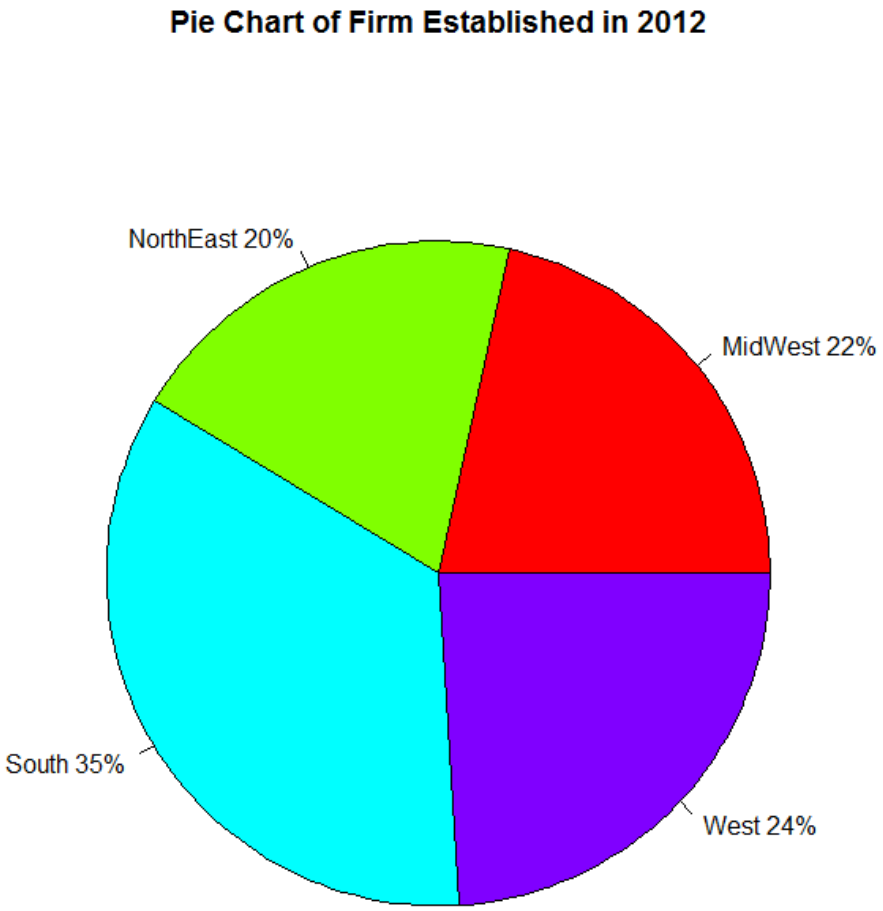
My goal is to find the relationship between the firms that established in each regions and number of jobs they can create for people across the America. Also, I want to look into what is the percentage of number of firms filed bankruptcy by the end of year 2012, which can cause people losing job.

Methods

The method I am going to using mostly in this report is linear regression, and coefficient to figure the relationship between number of firm established that create job for people in the US. I use 2 set of data, the first set of data include the information of 50 states and their number of firms, number of establishments, employment and annual payroll. I got the data from : http://www.census.gov/econ/susb/methodology.html. In this first set of data, I imported to excel to do some clean up data and format (*Figure 1*). The bar chart shows the distribution of firms in the USA. California, New York, Texas, Florida have the most employment in the US States. They have a lot of employees because of the location, capital of technology and financial industries. Also, States population which create more employees in the states. In the other hand, Alaska, Wyoming, Delaware doing poorly on the grows comparing to other states *(Figure 2)*. In

the pie chart, California, Texas, New York, and Florida are standing out, in previous bar chart, it shows that with the more firms have been created, mostly from technology and financial industries, it attracts more people to work in those industries. With new innovation, job opportunity are more in those states*(Figure 3)*. The next step I want to get descriptive Statistic, I loaded the dataset into SPSS(*Figure 4)*. The first descriptive statistics variable I choose to do is Employments, According to the descriptive statistics, the states that has lowest employees has about 408,317 employees, and California has the most employee which about 192,410,150 employees. The mean is 7,400,390.385. With this mean, whichever states has more than 7,400,390 employees is in good condition to grow, they have more job opportunity. However, whichever states below the mean, they are grow poorly, and require some investment to create job opportunity. Overall, the skewness of the data is 6.955 which is positive (*Figure 5)*. The average of firm establish is 369,592 firms across the country. The mean of firms are successful and keep up with the economy is 330,285. Which mean 11% of those firms established went bankruptcy because they could not keep up with the change and economy. The equation I used is 330285 / 369592 = 0.89, then convert to 89% of firms are doing good. Then I take 1- 89% to get 11% of firms that went bankruptcy(*Figure 6*). After using Excel and SPSS to do the analyze, I loaded my data into R. Before load data into R, I saved the file as .csv and use command in (*Figure 7*). After that, I use command summary(*Figure 8), and plot(Figure 9)* to see the distribution of my data in 50 states. However, the data too messy and it does not tell me anything. I chose to do grouping.

Instead of doing 50 states, I group 50 states into 4 different region by using pivot table. MidWest, NorthEast, South, and West. Then I loaded the table into R and call it PieChart(*Figure 10)*.




```
> # Pie Chart with Percentages of Firm Establlished
> slices <- c(4055735, 3724555, 6541381, 4527526)
> lbls <- c("MidWest", "NorthEast", "South", "West")
> pct <- round(slices/sum(slices)*100)
> lbls <- paste(lbls, pct) # add percents to labels
> lbls <- paste(lbls,"%",sep="") # ad % to labels
> pie(slices,labels = lbls, col=rainbow(length(lbls)),
+     main="Pie Chart of Firm Established in 2012")
```

According to the Pie Chart above, the South has the most firms established in 2012 with 35%, and the lowest grow region is NorthEast with 20%.

- Comparing correlation between 2 variables Successful Firm and Firm Established in 2012 by using command:

```
> cor(PieChart$Number.of.Firms,PieChart$Number.of.Firm.Established)
[1] 0.9990728
```

The correlation of two variables successful firms and firm Established is 0.9990728, they have a strong positive linear relationship.

- Using LM command to apply Fitting Linear models.

```
> lm(formula = PieChart$Number.of.Firms~PieChart$Number.of.Firm.Established)

Call:
lm(formula = PieChart$Number.of.Firms ~ PieChart$Number.of.Firm.Established)

Coefficients:
                          (Intercept)  PieChart$Number.of.Firm.Established
                            1.630e+05                             8.591e-01
```

The relationship of these two variables is for every 8 million firms that established in 2012, there will be 1.6 million firms will survive and will not go bankruptcy

- Comparing correlation between 2 Variables, Firm Established and Number of Employment in 2012

```
> cor(PieChart$Sum.of.NUMBER.OF.ESTABLISHMENTS,PieChart$Sum.of.EMPLOYMENT)
[1] 0.9812507
```

The Correlation of two variables established firm and employee is 0.9812507, they have a strong positive linear relationship.

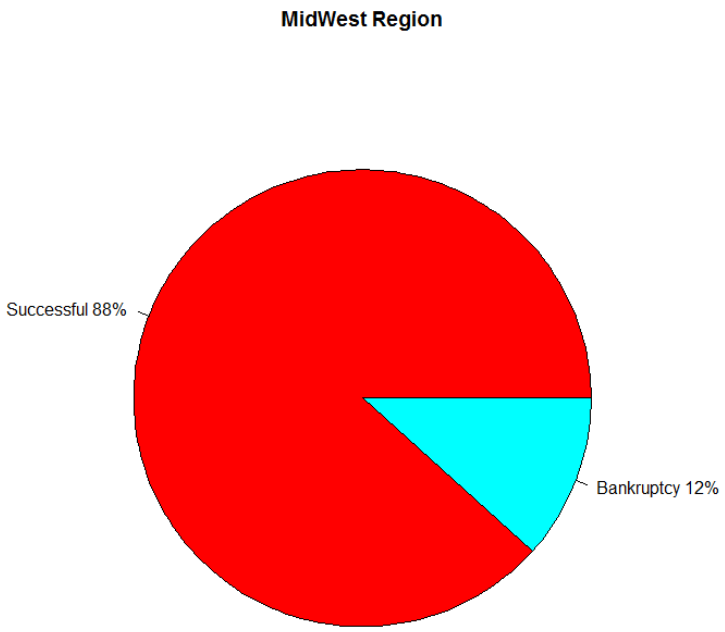
- Using LM Command to apply fitting Linear Models.

```
> lm(formula = PieChart$Sum.of.NUMBER.OF.ESTABLISHMENTS~PieChart$Sum.of.EMPLOYMENT)

Call:
lm(formula = PieChart$Sum.of.NUMBER.OF.ESTABLISHMENTS ~ PieChart$Sum.of.EMPLOYMENT)

Coefficients:
                (Intercept)  PieChart$Sum.of.EMPLOYMENT
                  2.726e+05                   9.230e-02

>
```

The relationship between Firm established and employee is for every 2 million firms established in 2012, they will create 9 million job for US citizen across the country.

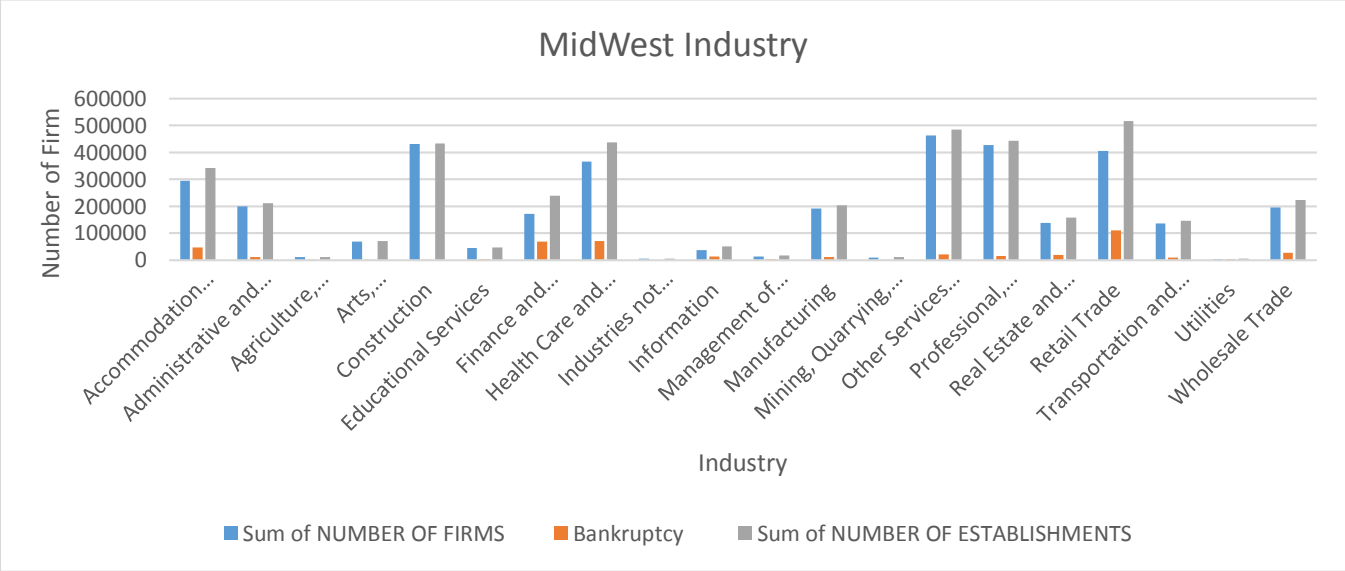
- Next I will explain the descriptive data of each regions, MidWest, NorthEast, South, and West.

1) MidWest:



```
>
> # Pie Chart with Percentages for Midwest
> slices <- c(3581565, 474170)
> lbls <- c("Successful", "Bankruptcy")
> pct <- round(slices/sum(slices)*100)
> lbls <- paste(lbls, pct) # add percents to labels
> lbls <- paste(lbls,"%",sep="") # ad % to labels
> pie(slices,labels = lbls, col=rainbow(length(lbls)),
+     main="Midwest Region")
>
```

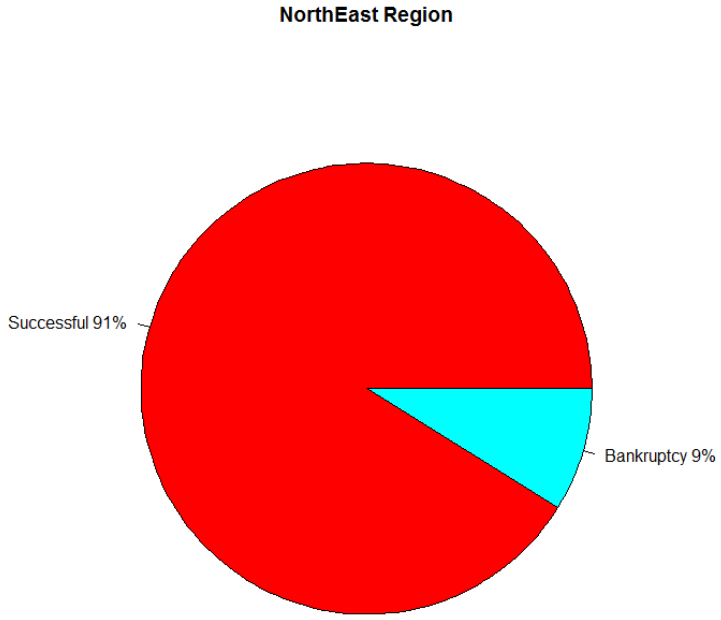
```
> summary(MidWest.Industry)
                                                                NAICS.DESCRIPTION Sum.of.NUMBER.OF.FIRMS
 Accommodation and Food Services                                              : 1      Min.    :  2860
 Administrative and Support and Waste Management and Remediation Services: 1      1st Qu.: 30793
 Agriculture, Forestry, Fishing and Hunting                                   : 1      Median :154814
 Arts, Entertainment, and Recreation                                          : 1      Mean   :180730
 Construction                                                                 : 1      3rd Qu.:312654
 Educational Services                                                         : 1      Max.   :462666
 (Other)                                                                      :14
   Bankruptcy      Sum.of.NUMBER.OF.ESTABLISHMENTS Sum.of.EMPLOYMENT
 Min.   :     0   Min.   :  5963                  Min.   :    687
 1st Qu.:  2607   1st Qu.: 39993                  1st Qu.: 700002
 Median : 11048   Median :180195                  Median :1783474
 Mean   : 22057   Mean   :202787                  Mean   :2170921
 3rd Qu.: 23600   3rd Qu.:364481                  3rd Qu.:2881557
 Max.   :111020   Max.   :516272                  Max.   :6906465

> cor(MidWest.Industry$Sum.of.NUMBER.OF.ESTABLISHMENTS,MidWest.Industry$Sum.of.EMPLOYMENT)
[1] 0.771101
> lm(formula = MidWest.Industry$Sum.of.NUMBER.OF.ESTABLISHMENTS~MidWest.Industry$Sum.of.EMPLOYMENT)

Call:
lm(formula = MidWest.Industry$Sum.of.NUMBER.OF.ESTABLISHMENTS ~
    MidWest.Industry$Sum.of.EMPLOYMENT)

Coefficients:
                       (Intercept)  MidWest.Industry$Sum.of.EMPLOYMENT
                         5.269e+04                            6.914e-02

>
```



- In the Pie Chart Firm Established in 2012, MidWest grows 22%, the second slowest region out of 4 regions. By the end of 2012, in those 22% new firms, there were 12% firms filed bankruptcy, only 88% firms be able to keep up with the economy. After load

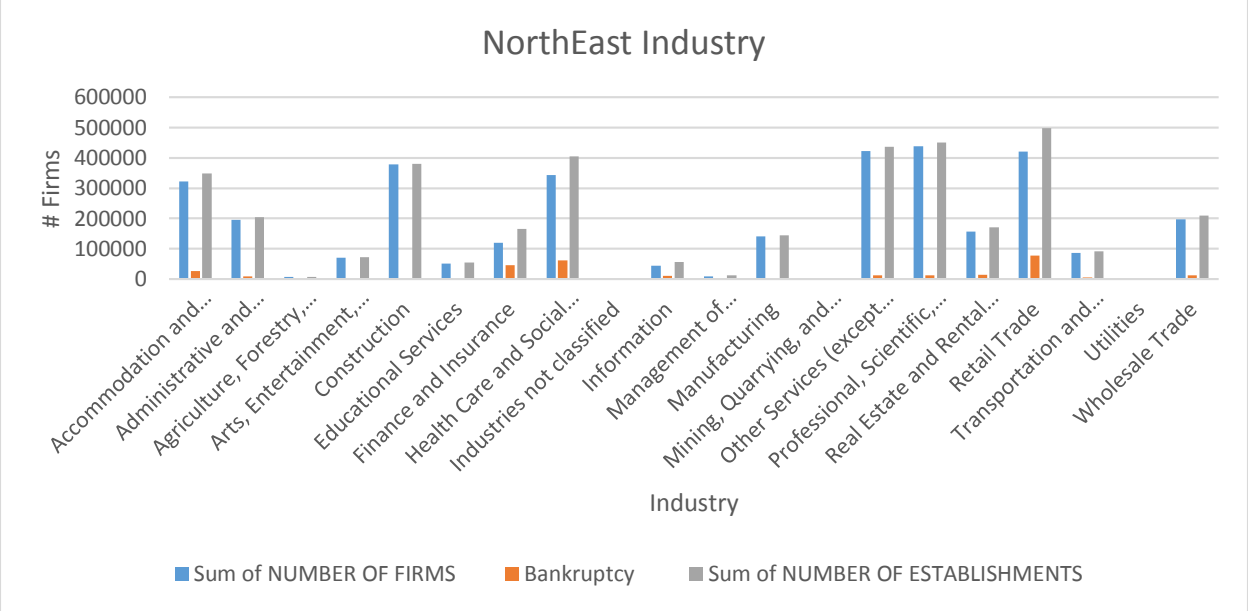
data into R, there are 4 variables Successful firm, bankruptcy, total firm establish and employment. According to mean, there were 202,787 total firms established, only 180,730 firms are successful in 2012 and 22,057 firms filed bankruptcy by the end of 2012. The correlation is 0.771101 between variables total firms establish and employment. It is a strong linear relationship. Also, from Coefficients, for every 5 thousand firms established, they created 6.9 million jobs for people in MidWest region. Also, according to the bar chart, industry Professional scientific, technology, retail trade have the most grow in this region.

2) NorthEast:




```
> # Pie Chart with Percentages for NorthEast
> slices <- c(3396791, 327764)
> lbls <- c("Successful", "Bankruptcy")
> pct <- round(slices/sum(slices)*100)
> lbls <- paste(lbls, pct) # add percents to labels
> lbls <- paste(lbls,"%",sep="") # ad % to labels
> pie(slices,labels = lbls, col=rainbow(length(lbls)),
+     main="NorthEast Region")
>
```

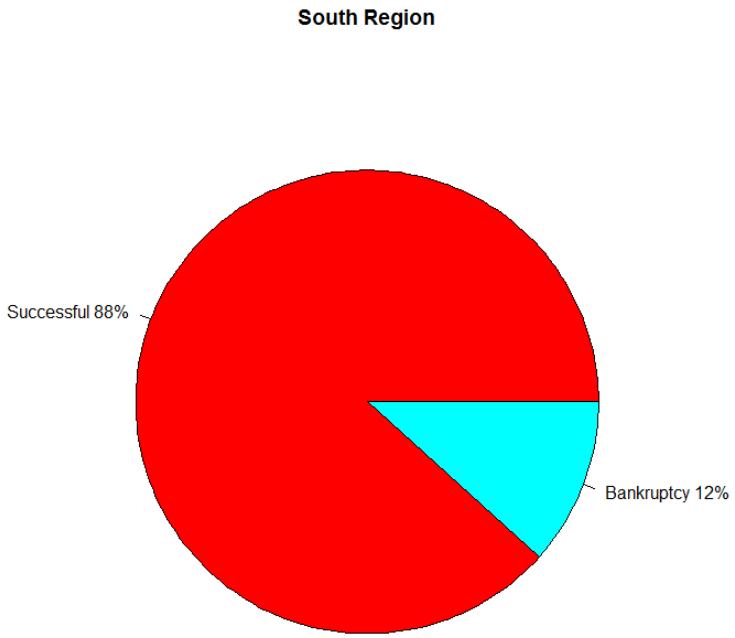
```
> NorthEast.Industry <- read.csv("C:/Users/Bao Huynh/Desktop/Winter 2016/CIS 320/CIS 320/NorthEast Industry.csv")
>   View(NorthEast.Industry)
> summary(NorthEast.Industry)
                                                                NAICS.DESCRIPTION Sum.of.NUMBER.OF.FIRMS   Bankruptcy
 Accommodation and Food Services                                              : 1      Min.   :  1851      Min.   :    0
 Administrative and Support and Waste Management and Remediation Services: 1      1st Qu.: 36424      1st Qu.: 2294
 Agriculture, Forestry, Fishing and Hunting                                   : 1      Median :130288      Median : 7082
 Arts, Entertainment, and Recreation                                          : 1      Mean   :170927      Mean   :15301
 Construction                                                                 : 1      3rd Qu.:327371      3rd Qu.:13284
 Educational Services                                                         : 1      Max.   :438360      Max.   :78134
 (Other)                                                                      :14
 Sum.of.NUMBER.OF.ESTABLISHMENTS Sum.of.EMPLOYMENT
 Min.   :  3434                  Min.   :   1683
 1st Qu.: 44542                  1st Qu.: 736850
 Median :155374                  Median :1662552
 Mean   :186228                  Mean   :1852531
 3rd Qu.:356581                  3rd Qu.:2525455
 Max.   :497980                  Max.   :6345778

> cor(NorthEast.Industry$Sum.of.NUMBER.OF.ESTABLISHMENTS,NorthEast.Industry$Sum.of.EMPLOYMENT)
[1] 0.8212176
> lm(formula = NorthEast.Industry$Sum.of.NUMBER.OF.ESTABLISHMENTS~NorthEast.Industry$Sum.of.EMPLOYMENT)

Call:
lm(formula = NorthEast.Industry$Sum.of.NUMBER.OF.ESTABLISHMENTS ~
    NorthEast.Industry$Sum.of.EMPLOYMENT)

Coefficients:
                        (Intercept)  NorthEast.Industry$Sum.of.EMPLOYMENT
                          2.805e+04                                8.538e-02
```



- NorthEast is one of the weakest region to grow, only 20% according to the pie chart firm established in 2012. By the end of 2012, in those 20% new firms, there were only 9% firms filed bankruptcy, and 91% firms be able to keep up with the economy. After load data into R, there are 4 variables Successful firm, bankruptcy, total firm establish and employment. According to mean, there were 186,228 total firms established, only 170,927 firms are successful in 2012 and 15,301 firms filed bankruptcy by the end of

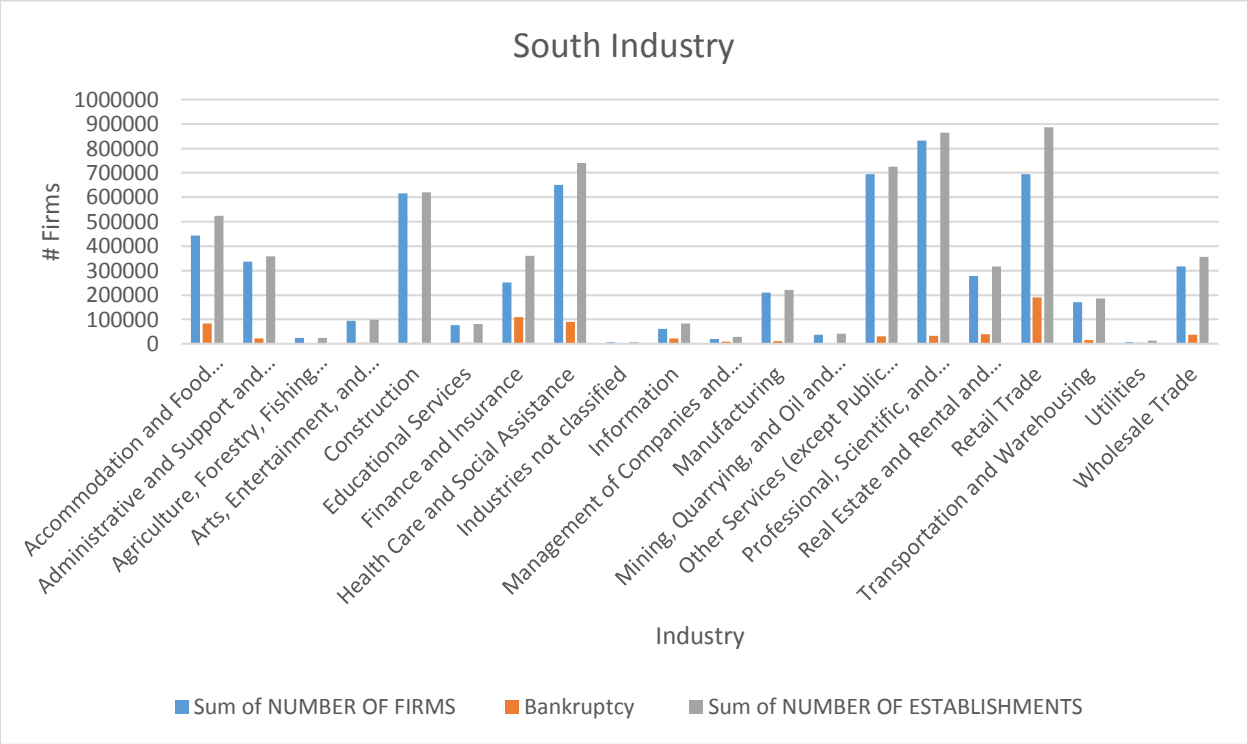
2012. The correlation is 0.8212176 between variables total firms establish and employment. It is a strong linear relationship. Also, from Coefficients, for every 2 thousand firms established, they created 8.5 million jobs for people in NorthEast region. Also, according to the bar chart, industry Professional scientific, technology, retail trade, Health Care and Social Assistance have the most grow in this region.

3) South:




```
>
> # Pie Chart with Percentages for South
> slices <- c(5780715, 760666)
> lbls <- c("Successful", "Bankruptcy")
> pct <- round(slices/sum(slices)*100)
> lbls <- paste(lbls, pct) # add percents to labels
> lbls <- paste(lbls,"%",sep="") # ad % to labels
> pie(slices,labels = lbls, col=rainbow(length(lbls)),
+     main="South Region")
```

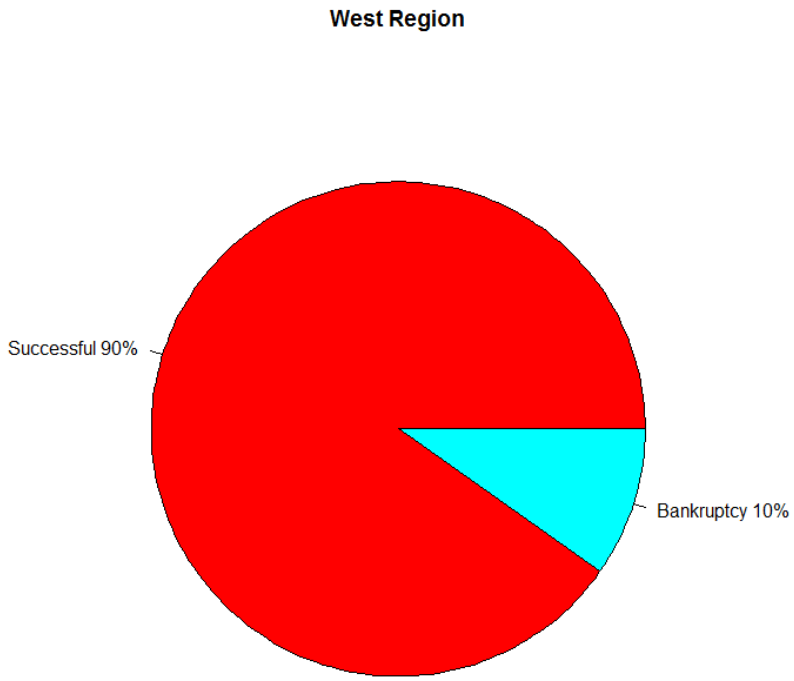
```
>
> South.Industry <- read.csv("C:/Users/Bao Huynh/Desktop/Winter 2016/CIS 320/CIS 320/South Industry.csv")
>   View(South.Industry)
> summary(South.Industry)
                                                                 NAICS.DESCRIPTION Sum.of.NUMBER.OF.FIRMS   Bankruptcy
 Accommodation and Food Services                                                : 1      Min.   :  6553         Min.   :     0
 Administrative and Support and Waste Management and Remediation Services: 1      1st Qu.: 56005         1st Qu.:  4518
 Agriculture, Forestry, Fishing and Hunting                                     : 1      Median :230718         Median : 18852
 Arts, Entertainment, and Recreation                                            : 1      Mean   :291237         Mean   : 35833
 Construction                                                                   : 1      3rd Qu.:486210         3rd Qu.: 38218
 Educational Services                                                           : 1      Max.   :831518         Max.   :190784
 (Other)                                                                        :14
 Sum.of.NUMBER.OF.ESTABLISHMENTS Sum.of.EMPLOYMENT
 Min.   :  6553                  Min.   :   2949
 1st Qu.: 71394                  1st Qu.:1094319
 Median :268990                  Median :2496932
 Mean   :327069                  Mean   :3332625
 3rd Qu.:549134                  3rd Qu.:5360247
 Max.   :886114                  Max.   :9933130

> cor(South.Industry$Sum.of.NUMBER.OF.ESTABLISHMENTS,South.Industry$Sum.of.EMPLOYMENT)
[1] 0.836323
> lm(formula = South.Industry$Sum.of.NUMBER.OF.ESTABLISHMENTS~South.Industry$Sum.of.EMPLOYMENT)

Call:
lm(formula = South.Industry$Sum.of.NUMBER.OF.ESTABLISHMENTS ~
    South.Industry$Sum.of.EMPLOYMENT)

Coefficients:
                      (Intercept)  South.Industry$Sum.of.EMPLOYMENT
                        4.172e+04                          8.562e-02
```



- The South has the strongest grow with 35% according to the pie chart firm established in 2012. By the end of 2012, in those 35% new firms, there were only 12% firms filed bankruptcy, and 88% firms be able to keep up with the economy. After load data into R, there are 4 variables Successful firm, bankruptcy, total firm establish and employment.

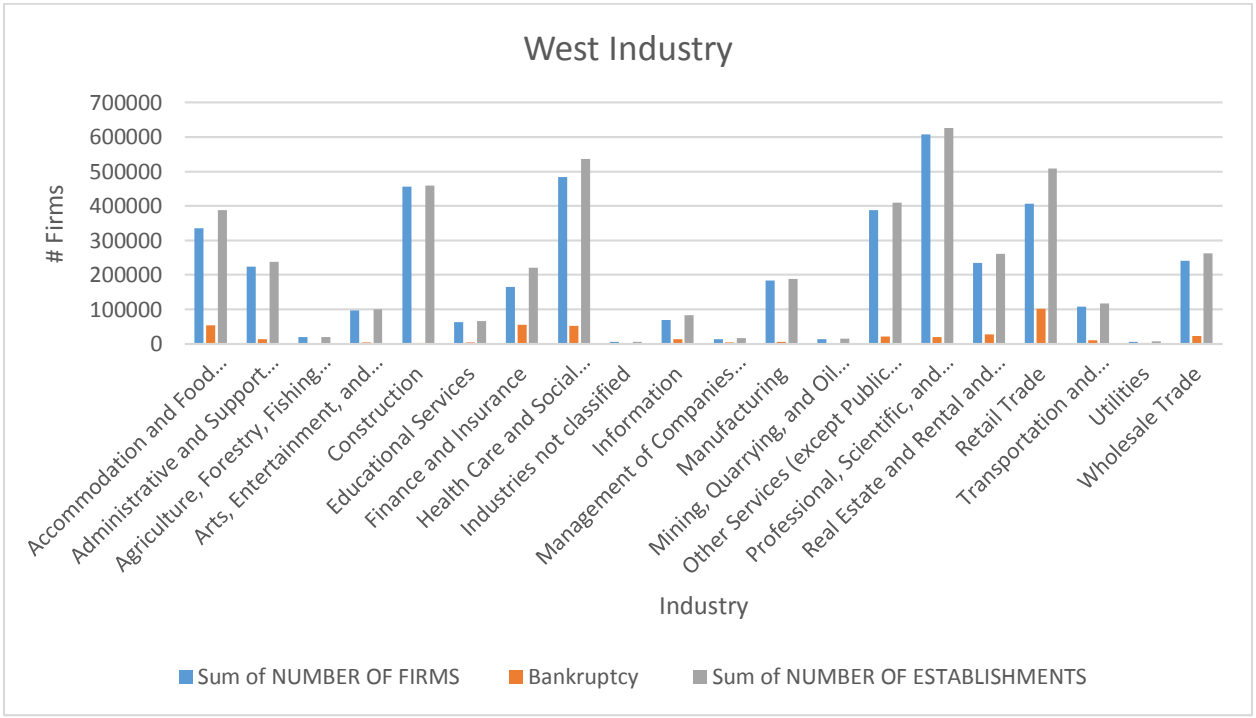
According to mean, there were 327,069 total firms established, only 291,237 firms are successful in 2012 and 35,833 firms filed bankruptcy by the end of 2012. The correlation is 0.836323 between variables total firms establish and employment. It is a strong linear relationship. Also, from Coefficients, for every 4 thousand firms established, they created 8.5 million jobs for people in South region. Also, according to the bar chart, industry Professional scientific, technology, retail trade, Health Care and Social Assistance have the most grow in this region.

4) West:



```
>
> # Pie Chart with Percentages for West
> slices <- c(4085450, 442076)
> lbls <- c("Successful", "Bankruptcy")
> pct <- round(slices/sum(slices)*100)
> lbls <- paste(lbls, pct) # add percents to labels
> lbls <- paste(lbls,"%",sep="") # ad % to labels
> pie(slices,labels = lbls, col=rainbow(length(lbls)),
+     main="West Region")
>
```

```
> West.Industry <- read.csv("C:/Users/Bao Huynh/Desktop/Winter 2016/CIS 320/CIS 320/West Industry.csv")
>   View(West.Industry)
> summary(West.Industry)
                                                                NAICS.DESCRIPTION Sum.of.NUMBER.OF.FIRMS   Bankruptcy
 Accommodation and Food Services                                              : 1      Min.   :  4803      Min.   :     0
 Administrative and Support and Waste Management and Remediation Services: 1      1st Qu.: 51719      1st Qu.:  3187
 Agriculture, Forestry, Fishing and Hunting                                   : 1      Median :173959      Median : 11456
 Arts, Entertainment, and Recreation                                          : 1      Mean   :205736      Mean   : 20641
 Construction                                                                 : 1      3rd Qu.:347873      3rd Qu.: 23455
 Educational Services                                                         : 1      Max.   :606938      Max.   :101646
 (Other)                                                                      :14
 Sum.of.NUMBER.OF.ESTABLISHMENTS Sum.of.EMPLOYMENT
 Min.   :  5351                  Min.   :   2047
 1st Qu.: 54618                  1st Qu.: 864284
 Median :204461                  Median :1485513
 Mean   :226376                  Mean   :2107373
 3rd Qu.:393778                  3rd Qu.:3235772
 Max.   :626958                  Max.   :5843779

> cor(West.Industry$Sum.of.NUMBER.OF.ESTABLISHMENTS,West.Industry$Sum.of.EMPLOYMENT)
[1] 0.8524489
> lm(formula = West.Industry$Sum.of.NUMBER.OF.ESTABLISHMENTS~West.Industry$Sum.of.EMPLOYMENT)

Call:
lm(formula = West.Industry$Sum.of.NUMBER.OF.ESTABLISHMENTS ~
    West.Industry$Sum.of.EMPLOYMENT)

Coefficients:
                  (Intercept)  West.Industry$Sum.of.EMPLOYMENT
                    2.645e+04                          9.487e-02
```



- The last region is the West with 24% grow according to the pie chart firm established in 2012. By the end of 2012, in those 24% new firms, there were only 10% firms filed bankruptcy, and 90% firms be able to keep up with the economy. After load data into R, there are 4 variables Successful firm, bankruptcy, total firm establish and employment. According to mean, there were 226,376 total firms established, only 205,736 firms are successful in 2012 and 20,641 firms filed bankruptcy by the end of 2012. The correlation is 0.8524489 between variables total firms establish and employment. It is a strong linear relationship. Also, from Coefficients, for every 2.6 thousand firms established, they created 9.4 million jobs for people in South region. Also, according to the bar chart, industry Professional scientific, technology, retail trade, Health Care and Social Assistance have the most grow in this region.

Conclusion

To answer the hypothesis about relationship between variables firm established and employment is for every 2 million firms established in 2012, they will create 9 million job for US citizen across the country. The industry that have growing the most are professional scientific and technology.

Bao Huynh
CIS 320

318 | Alaska

FIPS CODE	AREA DESCRIPTION	ENTERPRISE EMPLOYMENT SIZE	NUMBER OF FIRMS	NUMBER OF ESTABLISHMENTS	EMPLOYMENT	ANNUAL PAYROLL ($1,000)	ESTIMATED RECEIPTS ($1,000)
01	Alabama	02: 0-4	39,993	40,032	74,303	2,376,948	16,168,483
01	Alabama	03: 5-9	13,803	13,953	90,612	2,726,168	14,590,949
01	Alabama	04: 10-19	8,277	8,777	110,083	3,612,533	18,376,950
01	Alabama	05: <20	62,073	62,762	274,998	8,715,649	49,136,382
01	Alabama	06: 20-99	7,280	10,250	265,844	9,568,508	49,512,150
01	Alabama	07: 100-499	1,926	5,428	223,365	8,415,964	48,317,515
01	Alabama	08: <500	71,279	78,440	764,207	26,700,121	146,966,047
01	Alabama	09: 500+	2,382	19,498	821,554	34,584,810	264,677,916
02	Alaska	02: 0-4	9,979	10,005	15,036	809,344	4,095,233
02	Alaska	03: 5-9	2,826	2,872	18,657	791,222	3,281,028
02	Alaska	04: 10-19	1,732	1,872	23,049	1,049,710	4,226,278
02	Alaska	05: <20	14,537	14,749	56,742	2,650,276	11,602,539
02	Alaska	06: 20-99	1,192	1,752	41,453	1,976,375	8,217,375
02	Alaska	07: 100-499	334	1,109	40,183	2,011,768	10,136,464
02	Alaska	08: <500	16,063	17,610	138,378	6,636,419	29,956,378
02	Alaska	09: 500+	593	2,817	119,841	7,347,964	61,405,030
04	Arizona	02: 0-4	60,802	60,900	96,077	3,935,712	20,771,995
04	Arizona	03: 5-9	16,360	16,543	107,797	3,535,204	16,281,014
04	Arizona	04: 10-19	9,669	10,153	128,201	4,544,676	18,874,168
04	Arizona	05: <20	86,831	87,596	332,075	12,015,592	55,927,177
04	Arizona	06: 20-99	8,755	11,300	320,841	11,348,611	49,309,499
04	Arizona	07: 100-499	2,671	6,915	302,278	11,988,940	54,294,336
04	Arizona	08: <500	98,257	105,811	955,194	35,353,143	159,531,012
04	Arizona	09: 500+	2,948	25,564	1,179,058	55,485,473	322,415,405

Figure 1: Data set 1

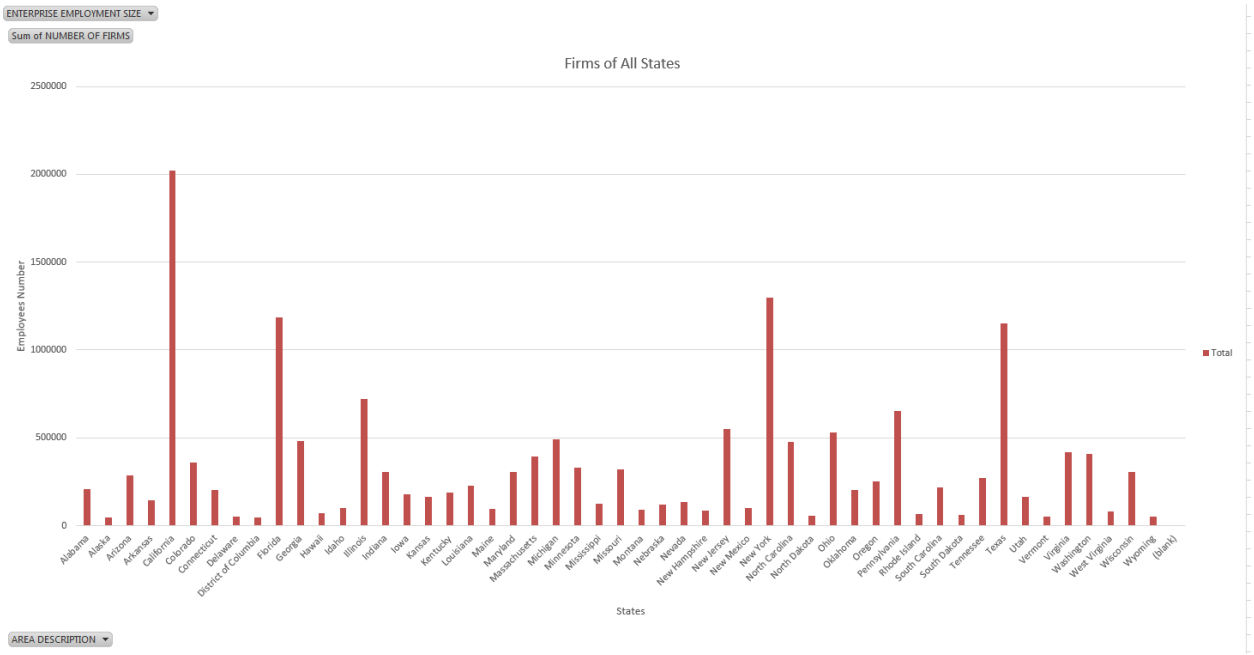


Figure 2: Bar chart of distribution of firms in the USA

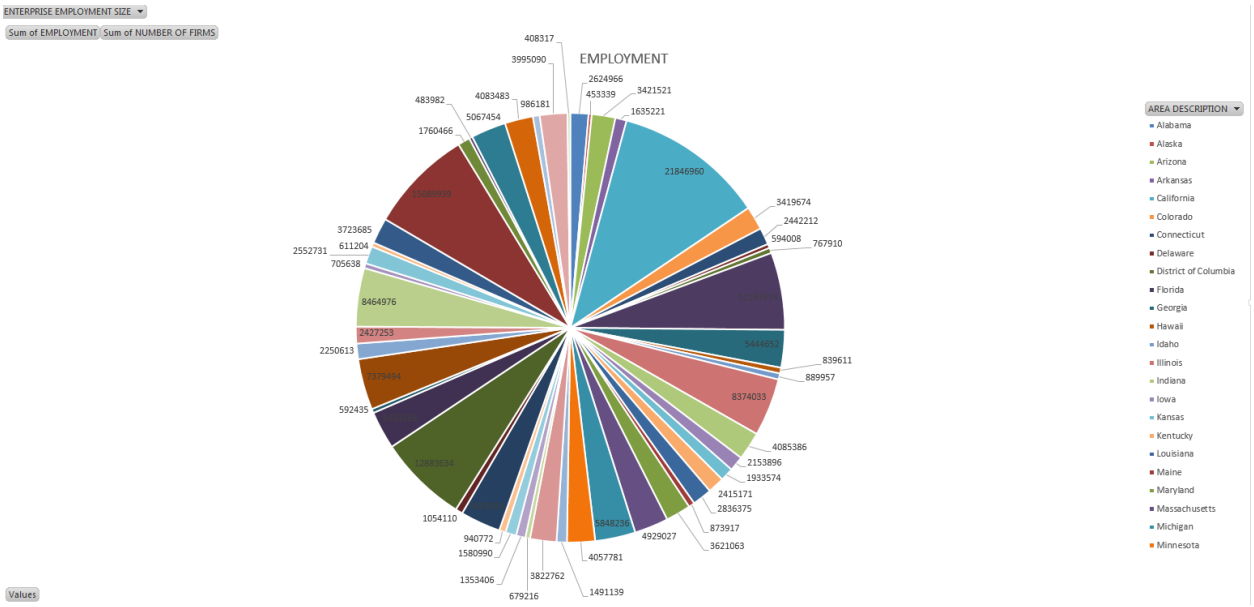


Figure 3: Pie Chart of Employment

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

4 : SumofEMPLOYMENT 1635221.000000000

	AREADESCRIPTION	SumofEMPLOYMENT	SumofNUMBEROFFIRMS	Percentage	var	var	var	var
1	Alabama	2624966.000000000	207013.000000000	100.00%				
2	Alaska	453339.000000000	47256.000000000	17.27%				
3	Arizona	3421521.000000000	286293.000000000	130.35%				
4	Arkansas	1635221.000000000	142128.000000000	62.29%				
5	California	21846960.000000000	2022856.00000000	832.28%				
6	Colorado	3419674.000000000	361309.000000000	130.27%				
7	Connecticut	2442212.000000000	201905.000000000	93.04%				
8	Delaware	594008.000000000	52151.000000000	22.63%				
9	District of Columbia	767910.000000000	45342.000000000	29.25%				
10	Florida	11187574.000000000	1187093.00000000	426.20%				
11	Georgia	5444652.000000000	479403.000000000	207.42%				
12	Hawaii	839611.000000000	69884.000000000	31.99%				
13	Idaho	889957.000000000	102070.000000000	33.90%				
14	Illinois	8374033.000000000	721766.000000000	319.01%				
15	Indiana	4085386.000000000	307258.000000000	155.64%				
16	Iowa	2153896.000000000	177143.000000000	82.05%				
17	Kansas	1933574.000000000	163195.000000000	73.66%				
18	Kentucky	2415171.000000000	190511.000000000	92.01%				
19	Louisiana	2836375.000000000	227714.000000000	108.05%				
20	Maine	873917.000000000	94060.000000000	33.29%				
21	Maryland	3621063.000000000	303142.000000000	137.95%				
22	Massachusetts	4929027.000000000	392797.000000000	187.77%				

Figure 4: Load data into SPSS

Descriptives

[DataSet1]

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error
Sum of EMPLOYMENT	52	192001833.0	408317.0000	192410150.0	7400390.385	3671482.922	26475439.86	7.009E+14	6.955	.330
Valid N (listwise)	52									

Figure 5: Descriptive Statistics of Employment.

Descriptives

[DataSet1]

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error
Sum of NUMBER OF FIRMS	51	1977514.0	45342.0	2022856.0	330284.725	52448.8351	374559.6020	1.403E+11	2.671	.333
Sum of NUMBER OF ESTABLISHMENTS	51	2189807.0	50765.0	2240572.0	369592.098	58072.5001	414720.6034	1.720E+11	2.640	.333
Valid N (listwise)	51									

Figure 6: Descriptive Statistics of Number of firms and Number of establishments.

```
> firmdata <- read.csv(file.choose(), header = TRUE, sep = "")
```

Figure 7: Command load data into R

```
> summary(firmdata)
     Number              AREA.DESCRIPTION Sum.of.NUMBER.OF.FIRMS Sum.of.NUMBER.OF.ESTABLISHMENTS
 Min.   : 1.0   Alabama    : 1   Min.   :  45342   Min.   :  50765
 1st Qu.:13.5   Alaska     : 1   1st Qu.:  95935   1st Qu.: 107093
 Median :26.0   Arizona    : 1   Median : 207013   Median : 239140
 Mean   :26.0   Arkansas   : 1   Mean   : 330285   Mean   : 369592
 3rd Qu.:38.5   California: 1   3rd Qu.: 401213   3rd Qu.: 444887
 Max.   :51.0   Colorado   : 1   Max.   :2022856   Max.   :2240572
                (Other)    :45
```

Figure 8: Descriptive statistic of firmdata

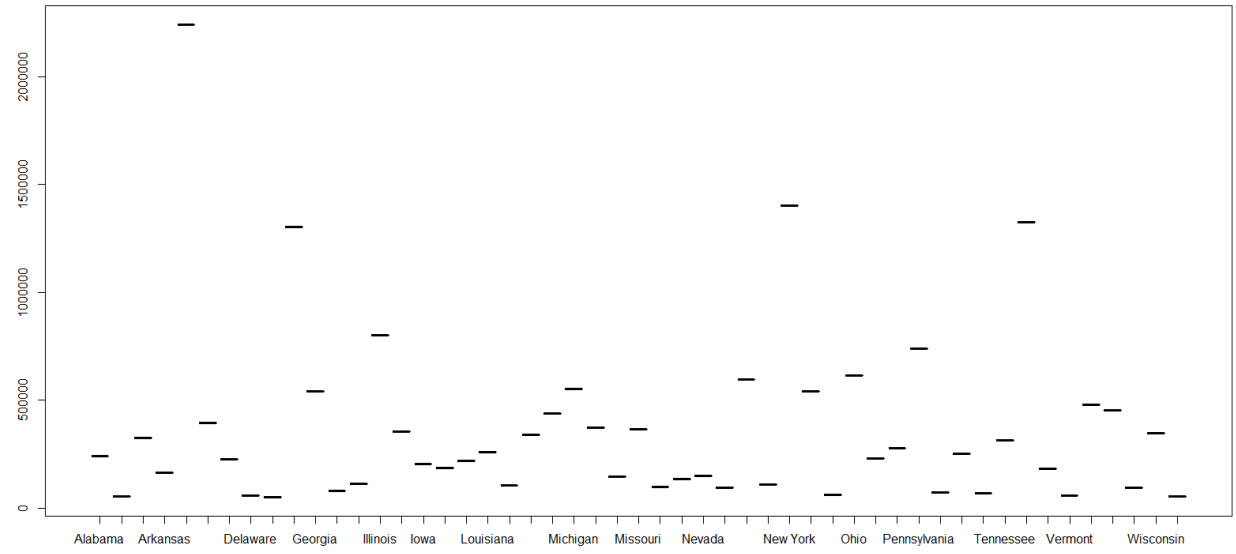


Figure 9: Plot command:

plot(firmdata$AREA.DESCRIPTION,firmdata$Sum.of.NUMBER.OF.ESTABLISHMENTS)

	Region	Number.of.Firms	Number.of.Bankruptcy	Number.of.Firm.Established
1	MidWest	3581565	474170	4055735
2	NorthEast	3396791	327764	3724555
3	South	5780715	760666	6541381
4	West	4085450	442076	4527526

Figure 10: PieChart <- read.csv(file.choose(), header = TRUE, sep = "")