**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.

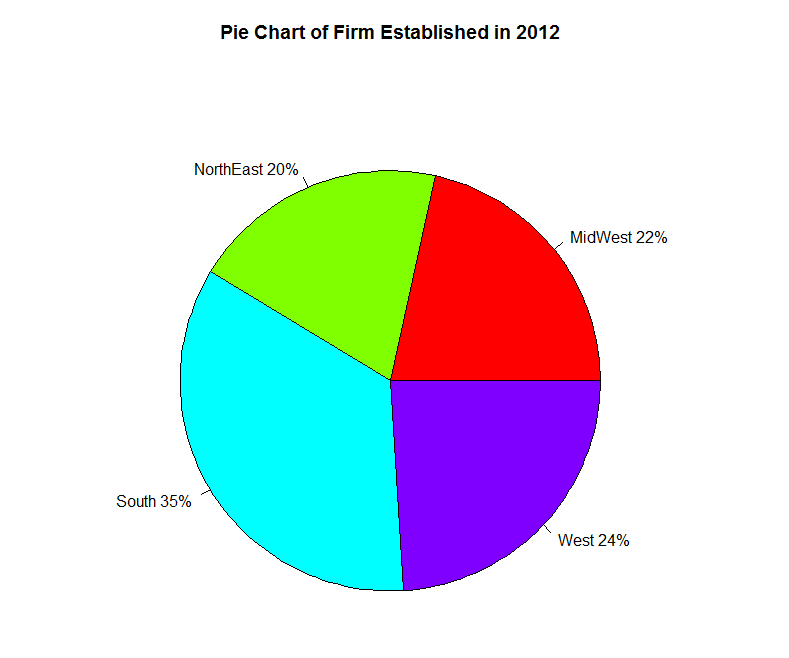
**Hypothesis**

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)),

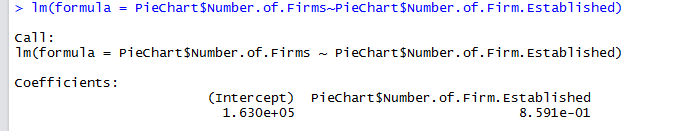
+ main="Pie Chart of Firm Established in 2012")

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: C:\Users\Bao Huynh\Desktop\Winter 2016\CIS 320\CIS 320\Project\Picture for R\Correlation.PNG

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.



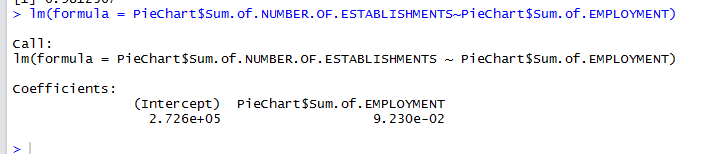
The relationship of these two variables is for every 8 million firms that established in 2012, there will be 1.6 million firms will survive and will not go bankruptcy

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

C:\Users\Bao Huynh\Desktop\Winter 2016\CIS 320\CIS 320\Project\Picture for R\Correlation of Employ vs Firm.PNG

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

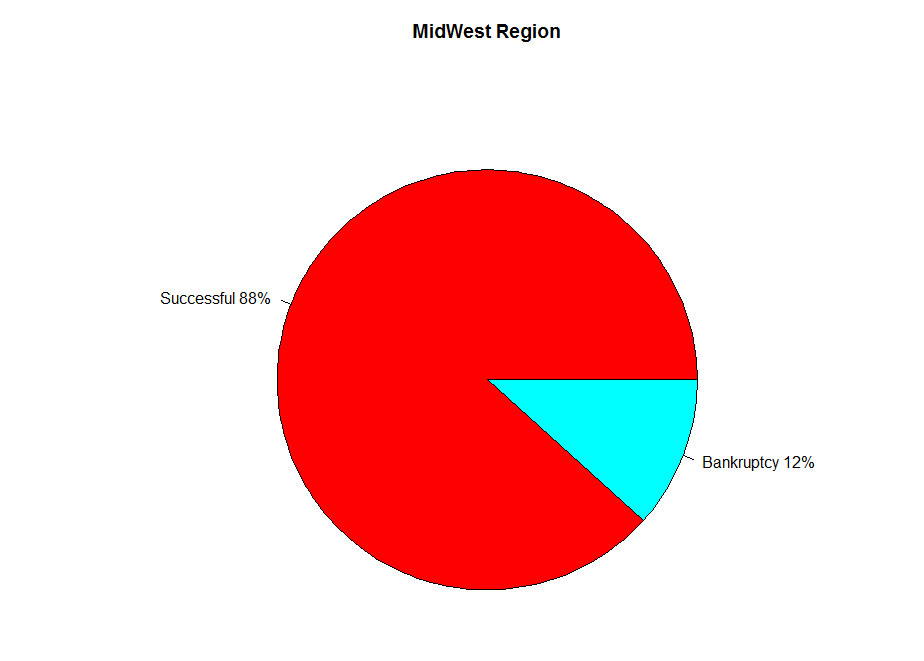
* Using LM Command to apply fitting Linear Models.

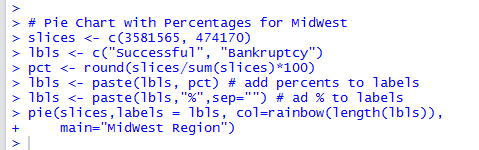


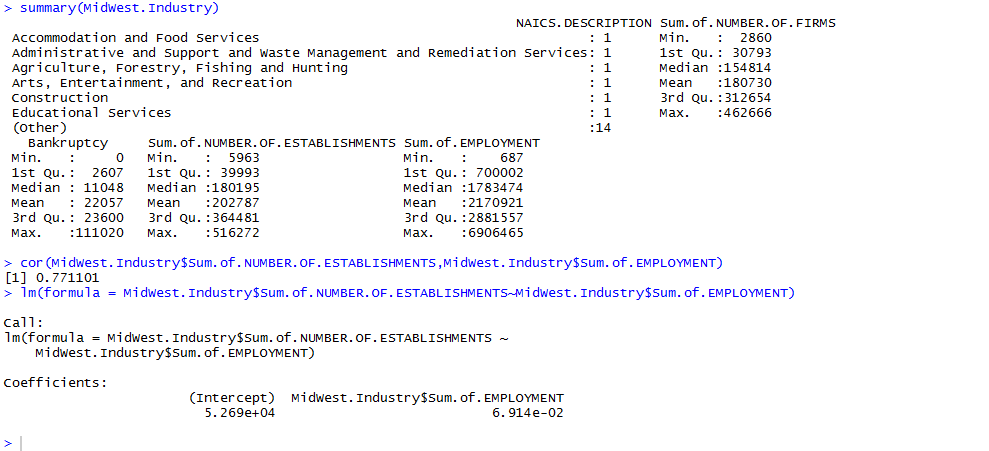
The relationship between Firm established and employee is for every 2 million firms established in 2012, they will create 9 million job for US citizen across the country.

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

1. **MidWest:**

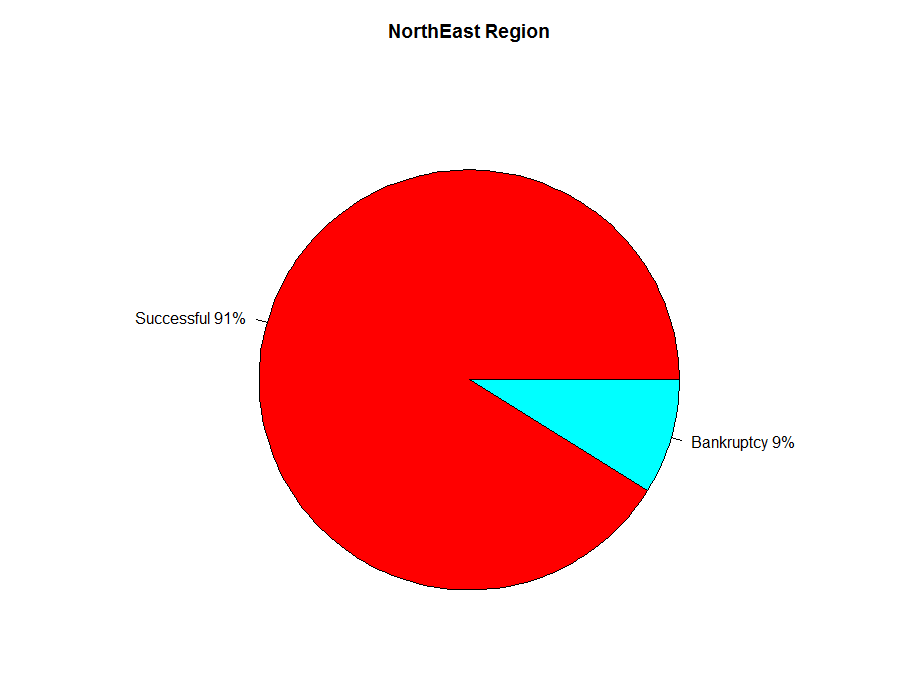


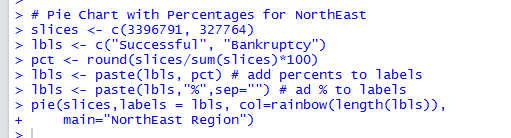


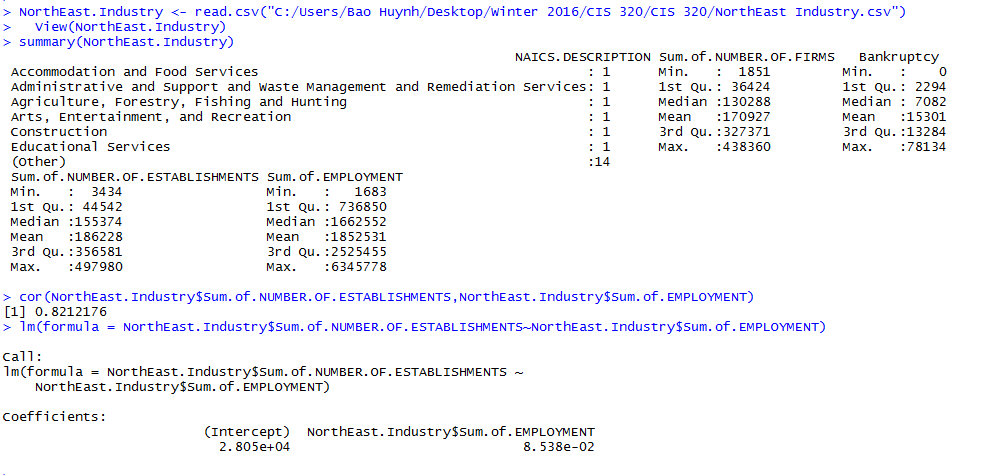


* In the Pie Chart Firm Established in 2012, MidWest grows 22%, the second slowest region out of 4 regions. By the end of 2012, in those 22% new firms, there were 12% firms filed bankruptcy, only 88% firms be able to keep up with the economy. After load data into R, there are 4 variables Successful firm,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

1. **NorthEast:**

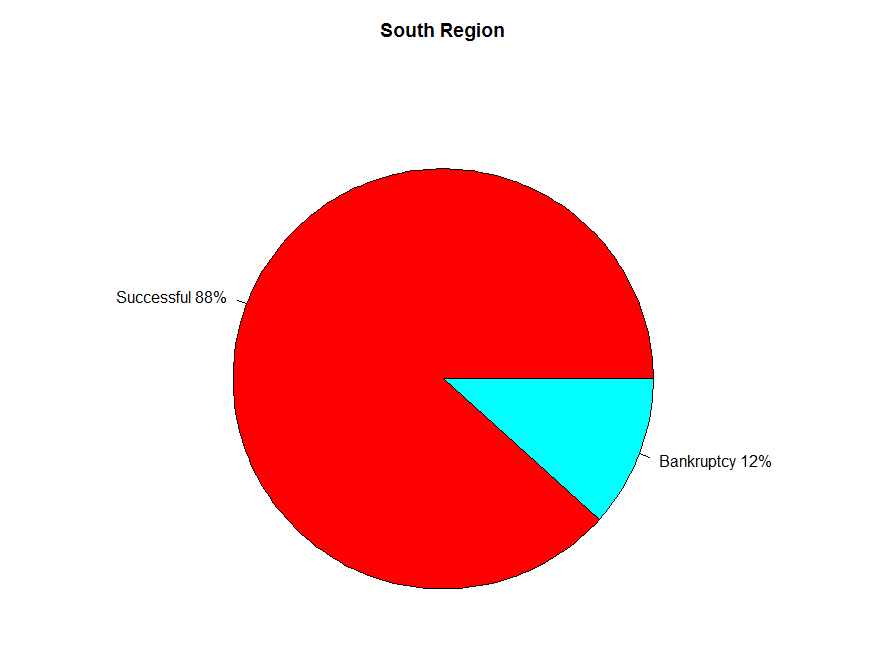
****

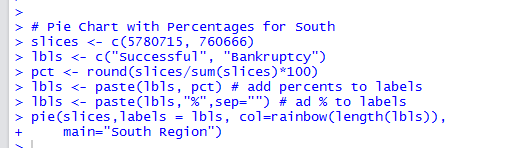
****

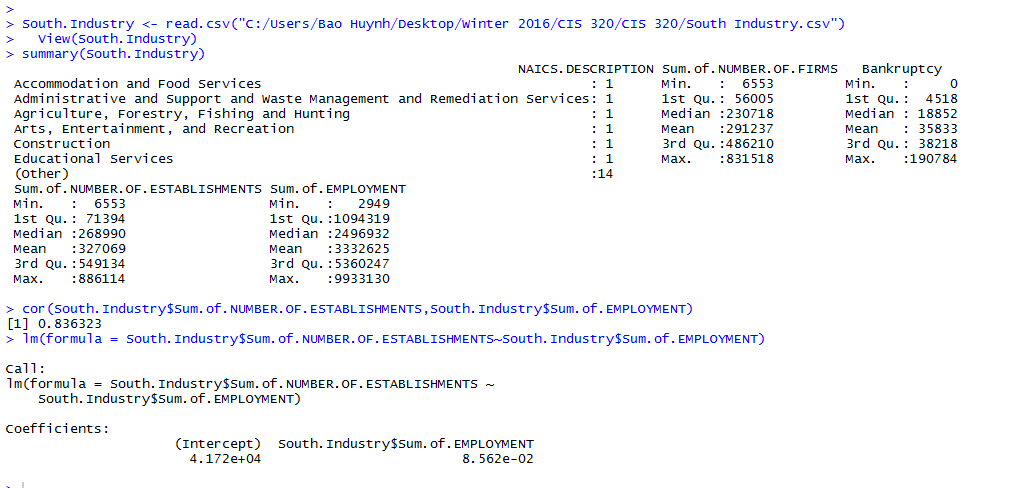
****

* 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.

1. **South:**

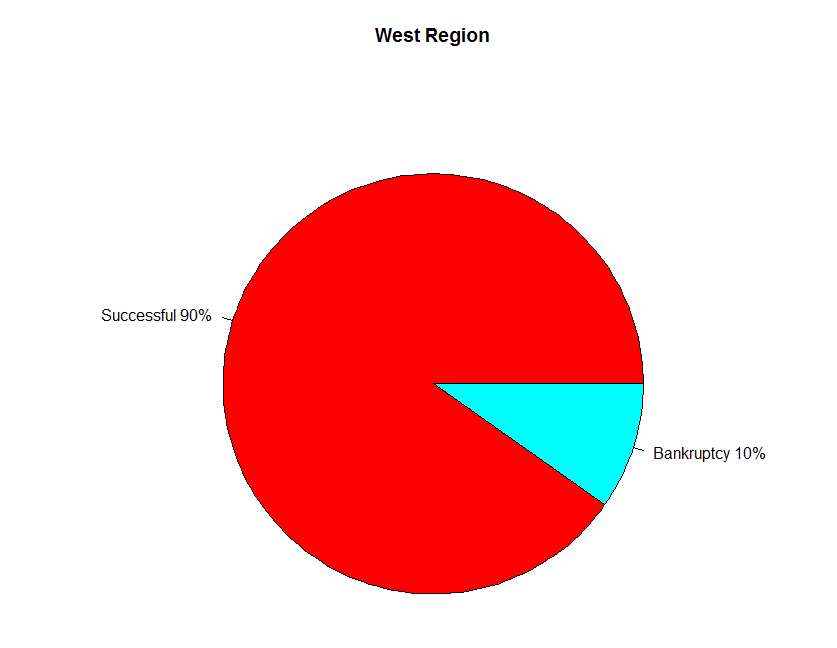
****

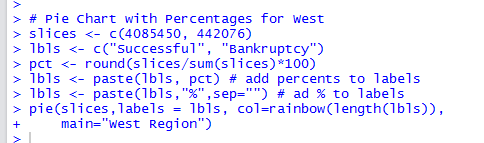
****

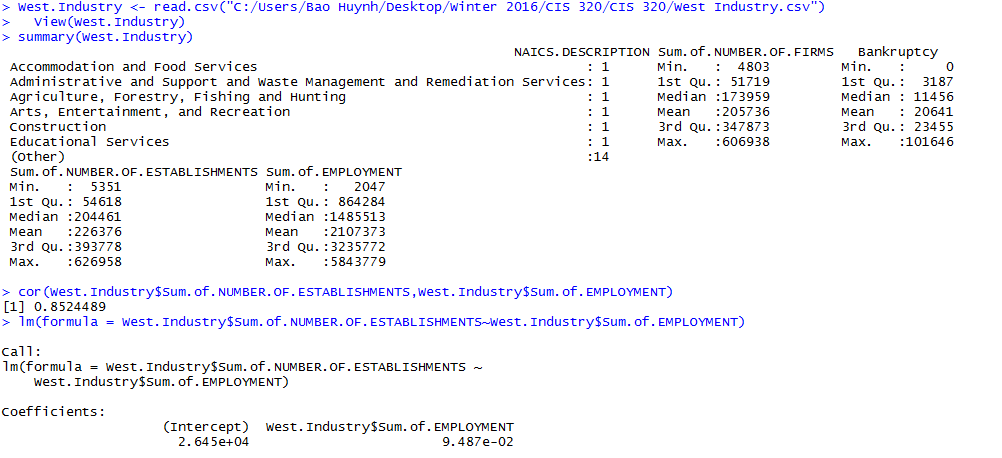
****

* 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.

1. **West:**

****

****

****

* 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.

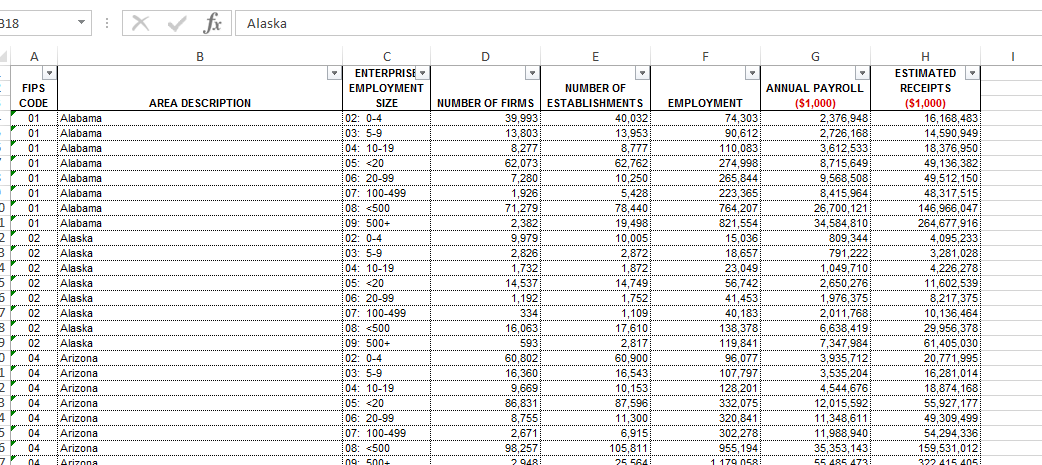


Figure 1: Data set 1

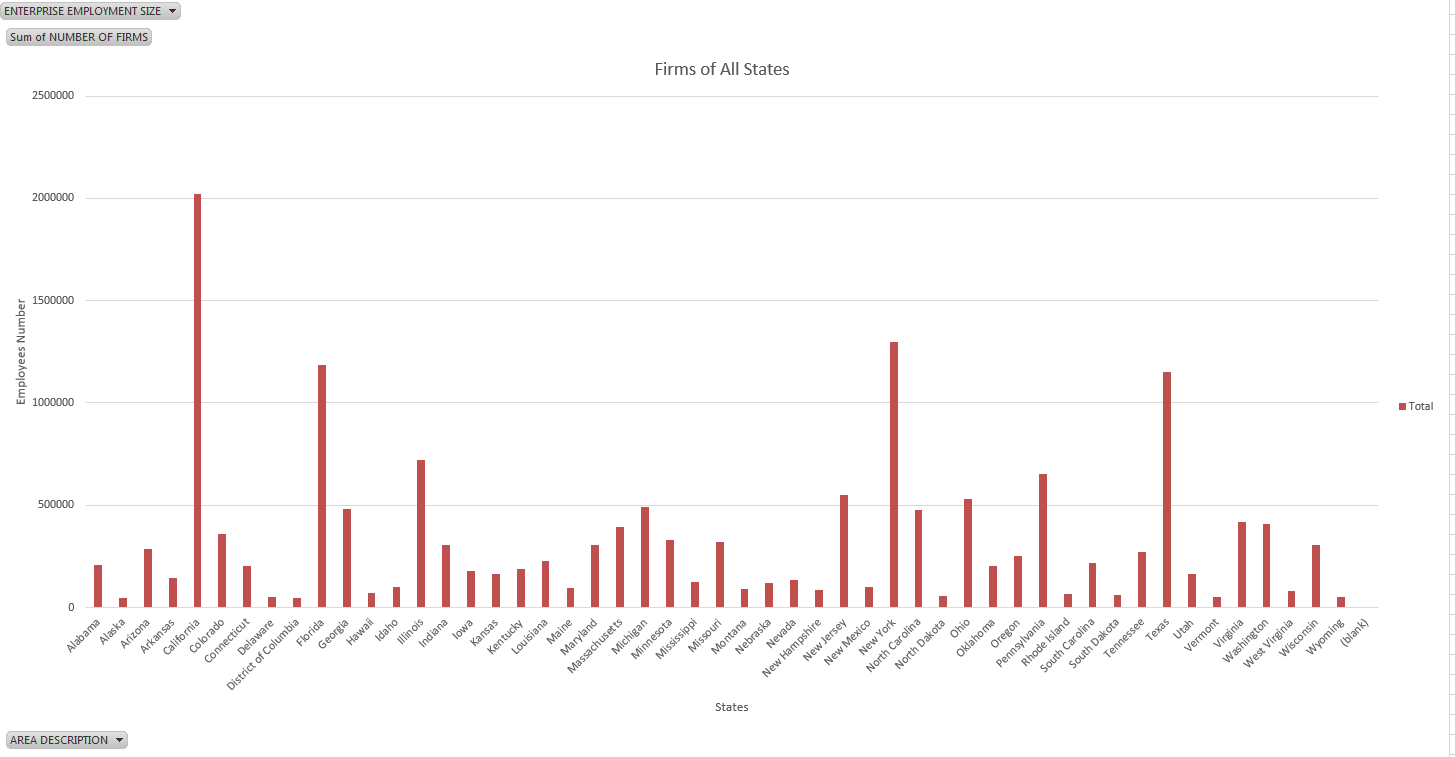
****

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

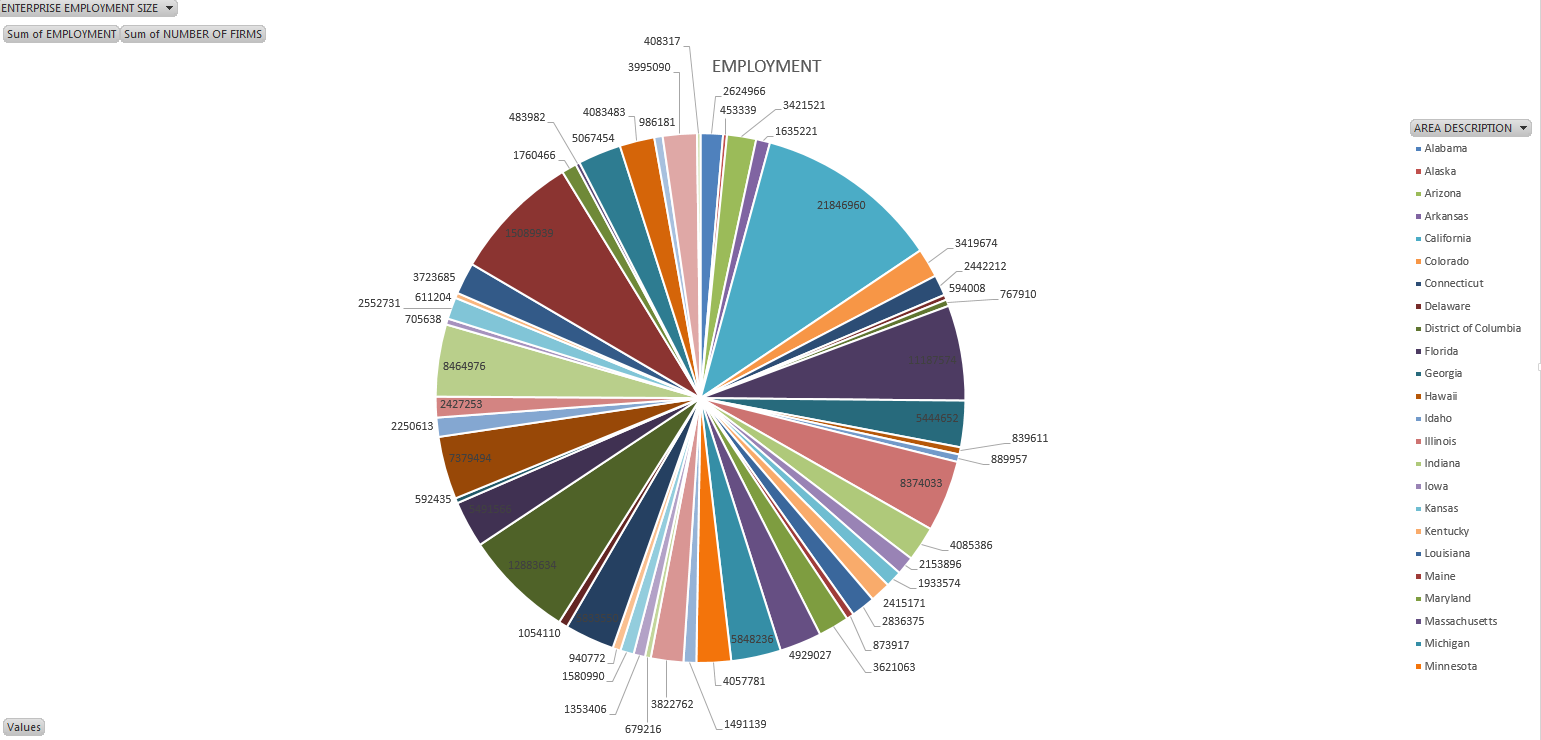
****

Figure 3: Pie Chart of Employment

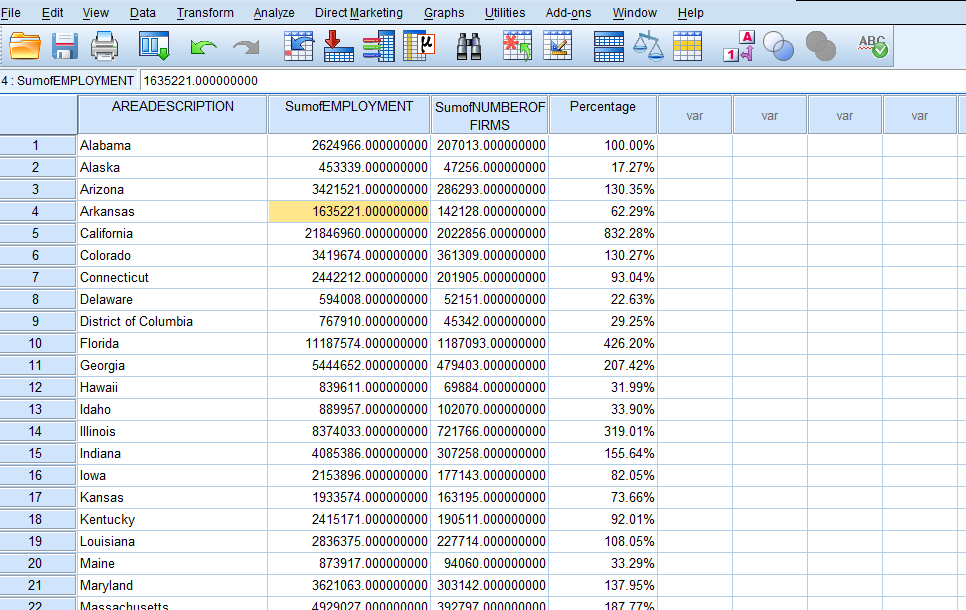
****

Figure 4: Load data into SPSS

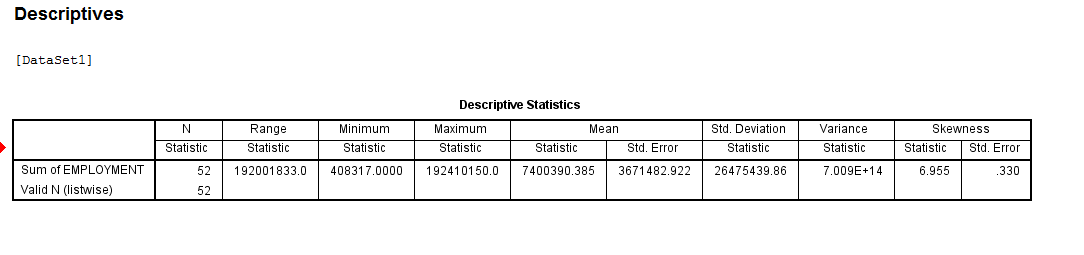
****

Figure 5: Descriptive Statistics of Employment.

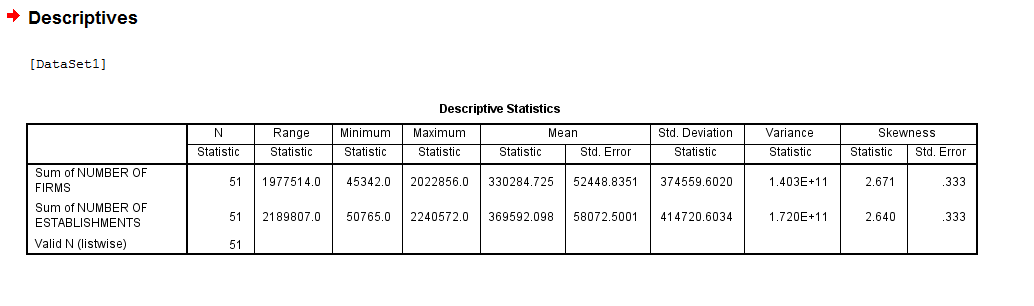
****

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

C:\Users\Bao Huynh\Desktop\Winter 2016\CIS 320\CIS 320\Project\Picture for R\Load Data into R.PNG

Figure 7: Command load data into R

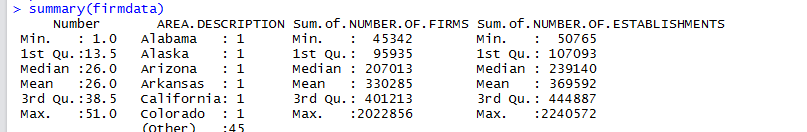


Figure 8: Descriptive statistic of firmdata

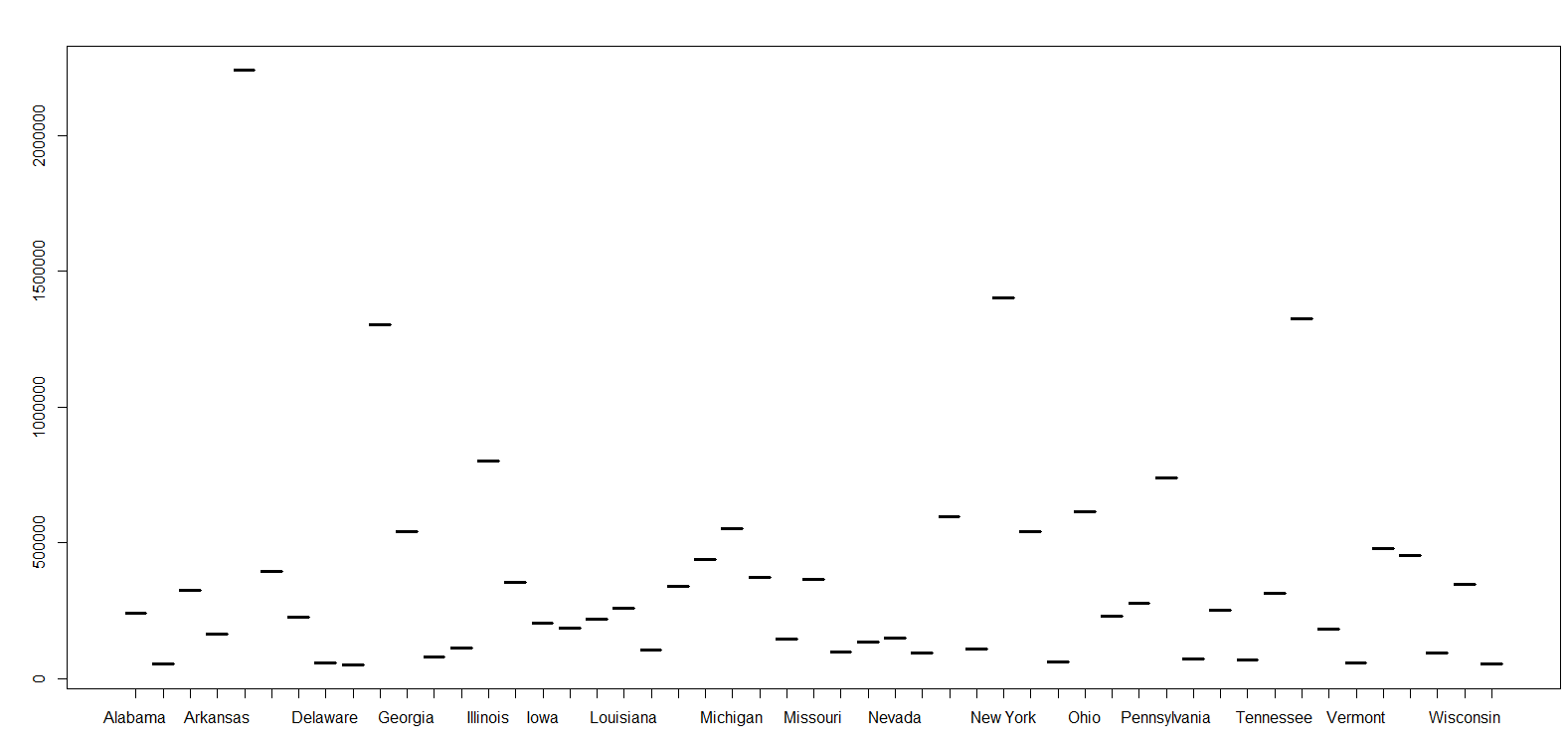


Figure 9: Plot command: plot(firmdata$AREA.DESCRIPTION,firmdata$Sum.of.NUMBER.OF.ESTABLISHMENTS)

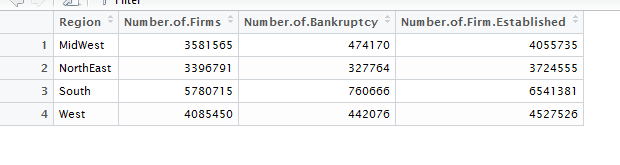


Figure 10: PieChart <- read.csv(file.choose(), header = TRUE, sep = "")