|  |  |
| --- | --- |
| **THE FINANCIAL CRISIS OF 2008**  **PROJECT 2 – ANALYSIS USING RSTUDIO** | Abstract  Analyze the effects of the 2008 financial crisis on the American economy, especially its housing market, unemployment situation, the consumer price index, inflation levels, and the widely popular stock market crash of 2008.  Submitted to Dr. Shilpa Balan  Submitted by:  Alekhya Raidu Bojja Venkata  Meena Muthiah  CIS 5270 – BUSINESS INTELLIGENCE |

**INDEX**

1. Introduction ………………………………………………………………………...2
2. Data Set URL’s …………………………………………………………………….3
3. Data Cleaning ………………………………………………………………………4
4. Analysis & Visualizations …………………………………………………………10
5. Statistical Averages of data ………………………………………………………..16
6. Complete R Code from RStudio……………………………………………………17
7. References ………………………………………………………………………….22
8. **INTRODUCTION:**

Over the past few years, the American economy has experienced the most severe recession since the Great Depression of the 1930s. The 2008 Financial Crisis occurred despite aggressive efforts by the Federal Reserve and Treasury Department to prevent the U.S. banking system from collapsing. It led to the Great Recession.

There are several theories as to what caused this crisis. The most popular one being the sub-prime mortgage lending in the USA. Subprime mortgage is a type of loan given to individuals with poor credit rating who would otherwise not qualify for traditional loans. The unchecked expansion of this sector lead to booming real estate market triggering the housing bubble, with the housing price index rising to its all-time high of 228.2.

The first signs of distress came in June 2007. This led to US commercial and investment banks who were highly invested in these mortgages to suffer a liquidity crisis. The next biggest effect of the financial crisis was the stock market crash that occurred on September 29, 2008. The decline in overall economic activity was modest at first, but it steepened sharply in the fall of 2008 as stresses in financial markets reached their climax. From peak to trough, US gross domestic product fell by 4.3 percent, making this the deepest recession since World War II. It was also the longest, lasting eighteen months.

While it is still greatly debated, the recession was deemed to have ended by 2010 by several analysts, while its wide spanning effects could still be felt till the end of 2012. The US government responded with the Dodd-Frank Act, passed in 2010, implementing Basel 3, and requiring banks to follow stringent capital and liquidity standards. However, even 5 years after the crisis several companies are still suffering for liquidity, the stock market is just coming back to its pre-crisis levels, with even slower job market growth.

1. **DATA SET URL’S:**
   * + 1. Data Set for Housing Price Index:

URL: <https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx#qexe>

The Housing Price Index (HPI) is a broad measure of the movement of single-family house prices. It is a weighted, repeat-sales index, meaning that it measures average price changes in repeat sales or refinancing on the same properties. This information is obtained by reviewing repeat mortgage transactions on single-family properties whose mortgages have been purchased or securitized by Fannie Mae or Freddie Mac since January 1975. The HPI serves as a timely, accurate indicator of house price trends at various geographic levels. The above URL, taken from the Federal Housing Finance Agency database, provides various datasets for HPI in the United States. The 2 datasets chosen for this project are Quarterly data for all states in the US using Purchase Only Indexes, i.e. index estimated using sales price data and Monthly data for US as a whole.

* + - 1. Dataset for Unemployment:

URL: <https://www.bls.gov/webapps/legacy/cpsatab14.htm#a14.xlsx.f.1>

The dataset for unemployment in US is taken from the Bureau of Labor Statistics (BLS). Several BLS programs provide information about unemployment. The dataset used here is from the National Unemployment Rate (from the Current Population Survey) where a monthly household survey provides comprehensive information on the employment and unemployment of the population classified by age, sex, race, and other characteristics. The unemployment rate shows the number of unemployed persons over the total labor force.

* + - 1. Dataset for Failed Banks:

URL: <https://www5.fdic.gov/hsob/SelectRpt.asp?EntryTyp=30>

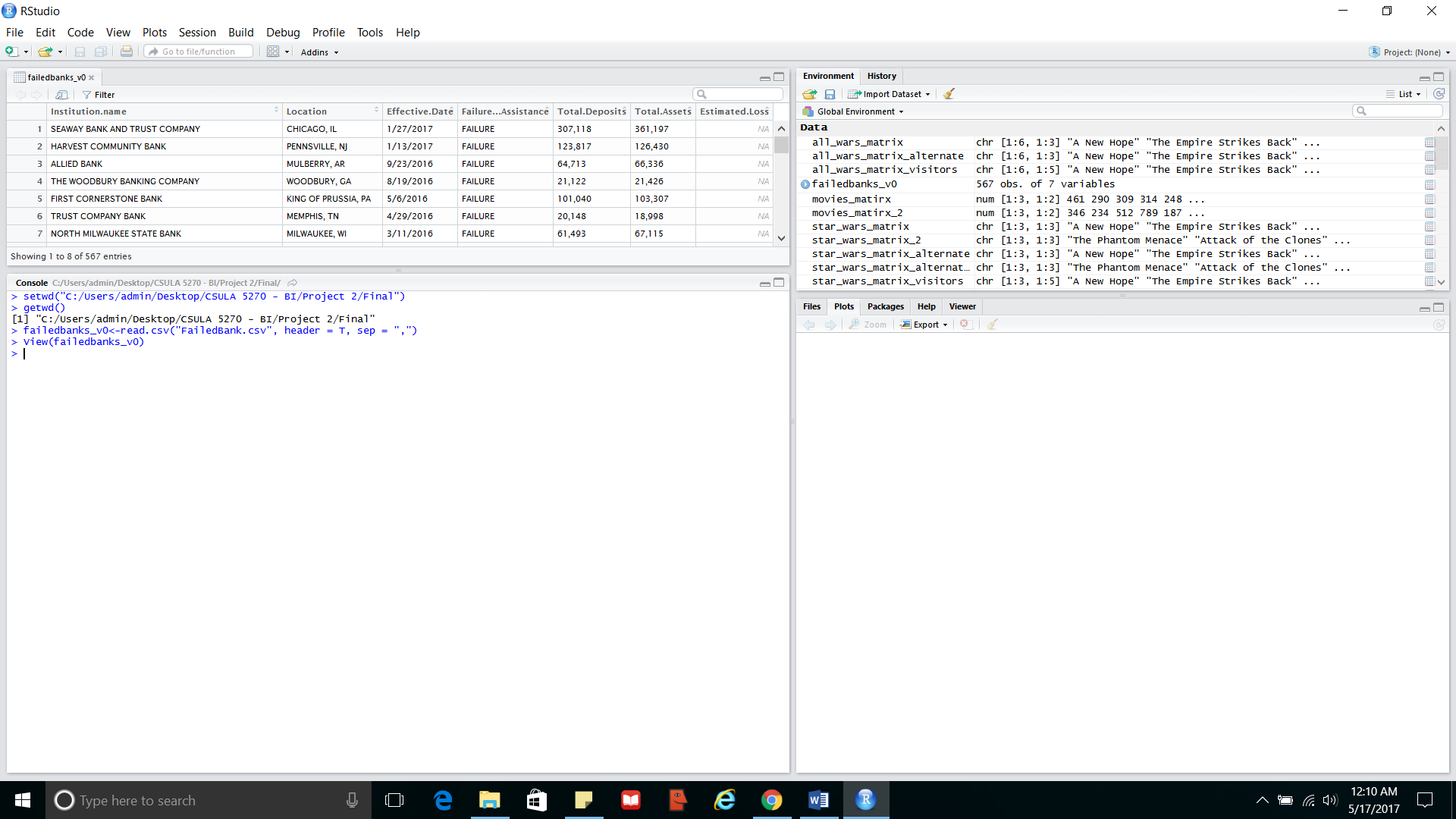
From the database of Federal Deposit Insurance Corporation (FDIC), an independent agency of the federal government that was created to promote confidence in the US banking system by insuring deposits in banks, we get the list of all banks that have failed between 2000 and 2017. The dataset provides the name of the bank with its location in US, along with information about its assets, liabilities and the estimated loss due to its failure among other details.

1. **DATA CLEANING:**

Of the 3 datasets only the Failed Bank dataset, taken from the FDIC website, needs to be cleaned. The other 2 datasets, the Housing Price Index and the Unemployment datasets were already clean and nothing had to be modified in it. Using R, the Failed Bank dataset has been cleaned through the below steps.

* + - 1. Removal of NAs in the Failed Banks dataset:

The dataset taken for the banks that claimed bankruptcy during the Financial Crisis shows the estimated loss from each bank. However, a review of the dataset showed that in a few places the loss was not estimated (screenshot attached below).



Hence using the R code, the data has been cleaned to replace these NAs with the mean or the average loss to give a cleaner dataset (screenshot attached). During this process, the csv file was also modified a little to give both the string and the numeric values the same format.

*CODE:*

> setwd("C:/Users/admin/Desktop/CSULA 5270 - BI/Project 2/Final")

> getwd()

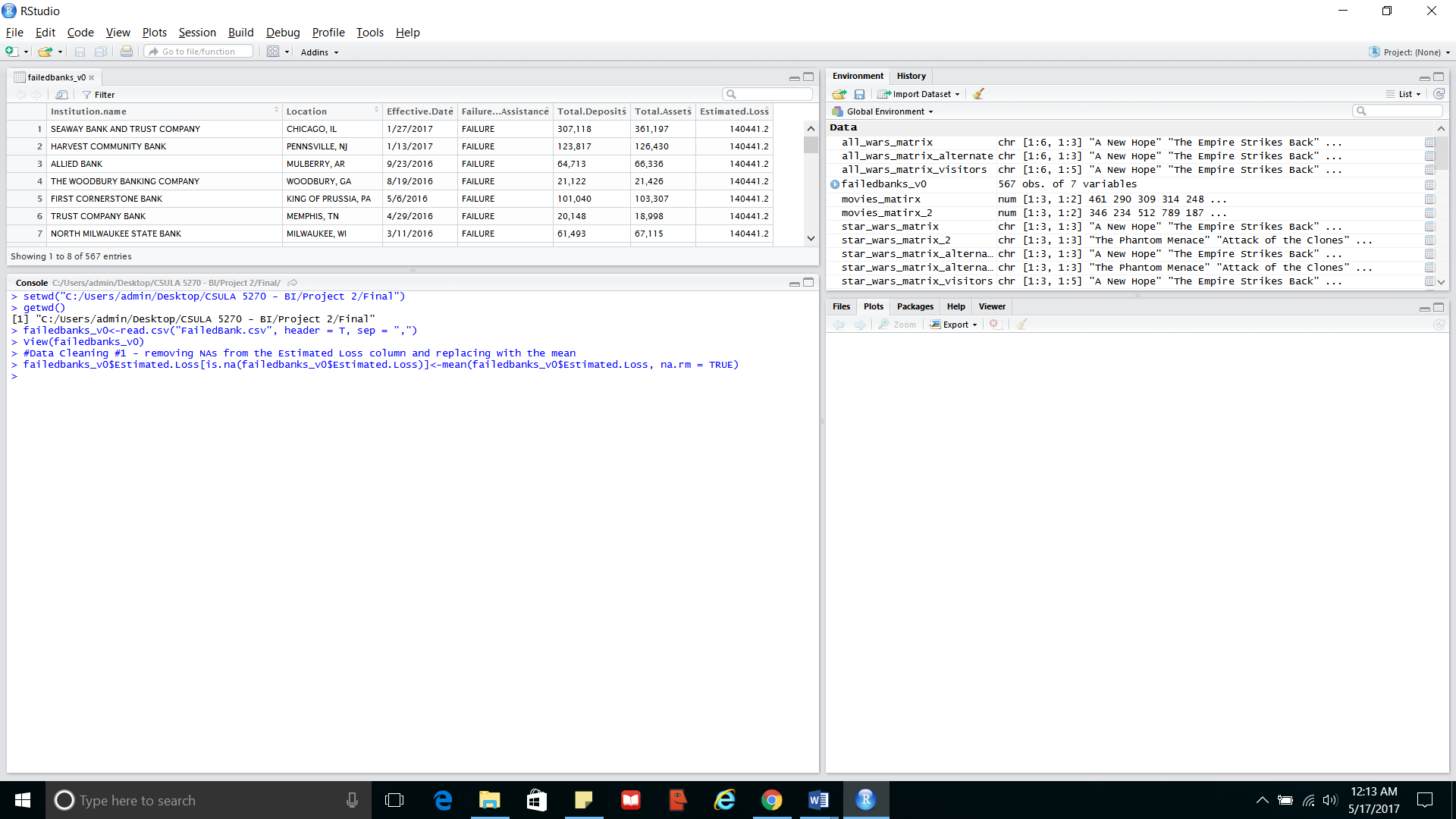
[1] "C:/Users/admin/Desktop/CSULA 5270 - BI/Project 2/Final"

> failedbanks\_v0<-read.csv("FailedBank.csv", header = T, sep = ",")

> View(failedbanks\_v0)

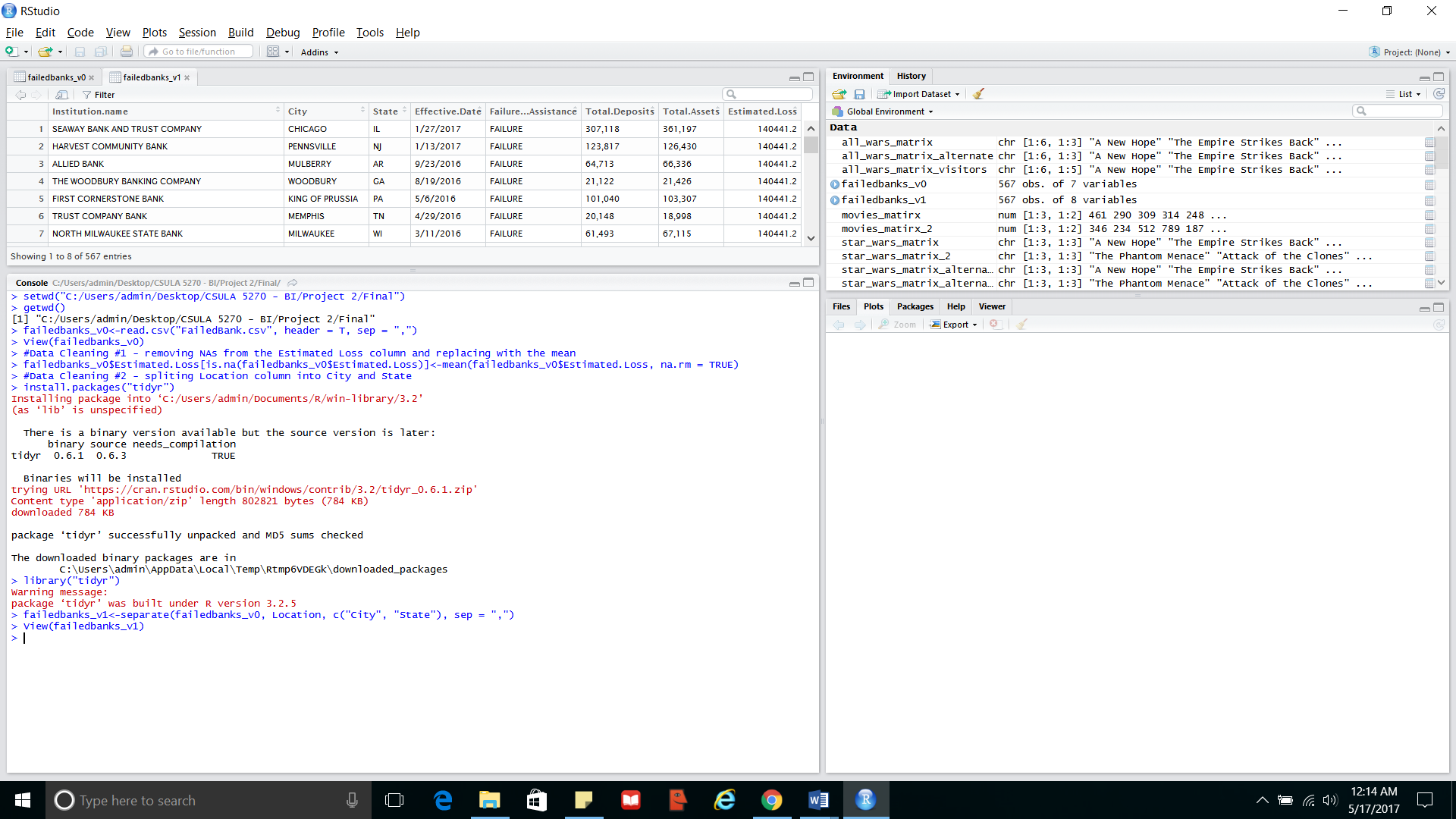
> #Data Cleaning #1 - removing NAs from the Estimated Loss column and replacing with the mean

> failedbanks\_v0$Estimated.Loss[is.na(failedbanks\_v0$Estimated.Loss)]<-mean(failedbanks\_v0$Estimated.Loss, na.rm = TRUE)



* + - 1. Splitting a column into 2 in the Failed Banks dataset

The screenshot above shows the Location column containing information about the City and the State of the bank in consideration separated by a comma. Using an R package “tidyr”, the separate function has been used to separate this column into City and State (screenshot attached).



*CODE:*

> #Data Cleaning #2 - spliting Location column into City and State

> install.packages("tidyr")

Installing package into ‘C:/Users/admin/Documents/R/win-library/3.2’

(as ‘lib’ is unspecified)

There is a binary version available but the source version is later:

binary source needs\_compilation

tidyr 0.6.1 0.6.3 TRUE

Binaries will be installed

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.2/tidyr\_0.6.1.zip'

Content type 'application/zip' length 802821 bytes (784 KB)

downloaded 784 KB

package ‘tidyr’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\admin\AppData\Local\Temp\Rtmp6VDEGk\downloaded\_packages

> library("tidyr")

Warning message:

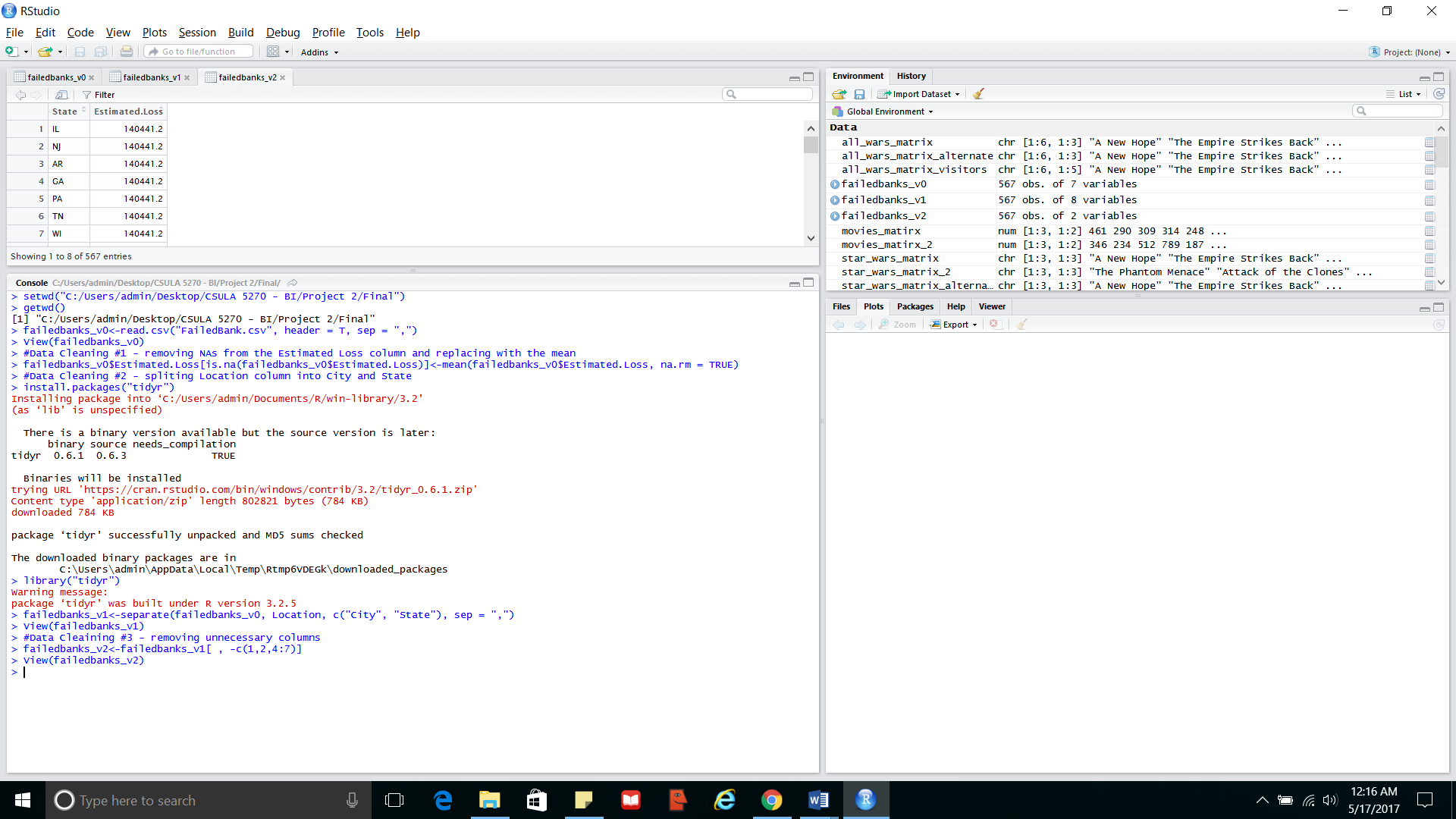
package ‘tidyr’ was built under R version 3.2.5

> failedbanks\_v1<-separate(failedbanks\_v0, Location, c("City", "State"), sep = ",")

> View(failedbanks\_v1)

* + - 1. Removal of unnecessary columns in the Failed Banks dataset

From the Failed Banks dataset only the columns State and Estimated Loss will be used for visualization purposes. Hence, the other columns have been removed (screenshot attached).



*CODE:*

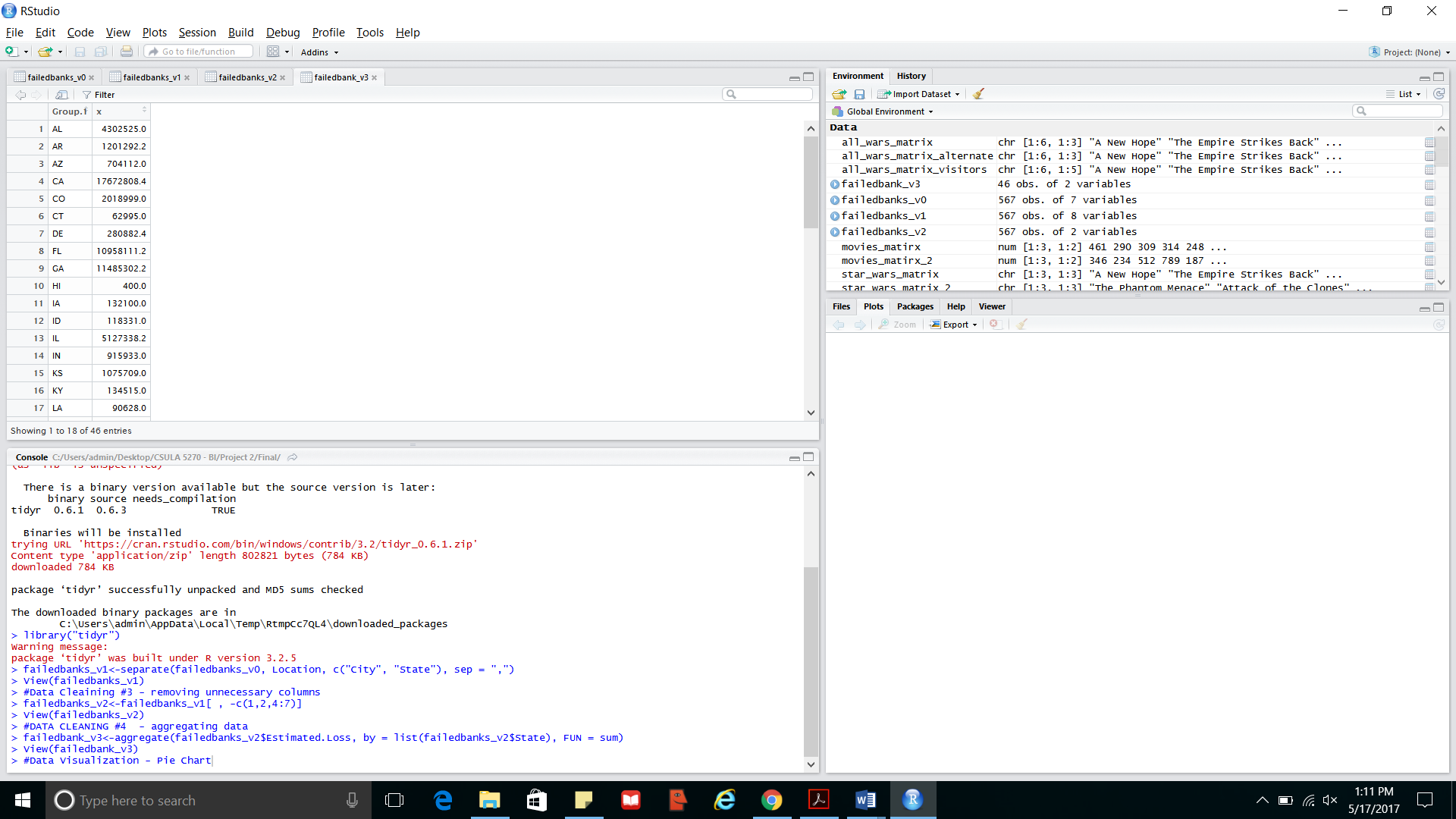
> #Data Cleaining #3 - removing unnecessary columns

> failedbanks\_v2<-failedbanks\_v1[ , -c(1,2,4:7)]

> View(failedbanks\_v2)

* + - 1. Aggregating values in the Failed Banks dataset

In the same dataset, after removal of columns, it was clearly seen that there are multiple values for the same state. This is due to the failure of several banks in the same area. Since the visualizations do not look for bank wise information, but rather a state wise information, the dataset has been aggregated to give a consolidated view (screenshot attached).



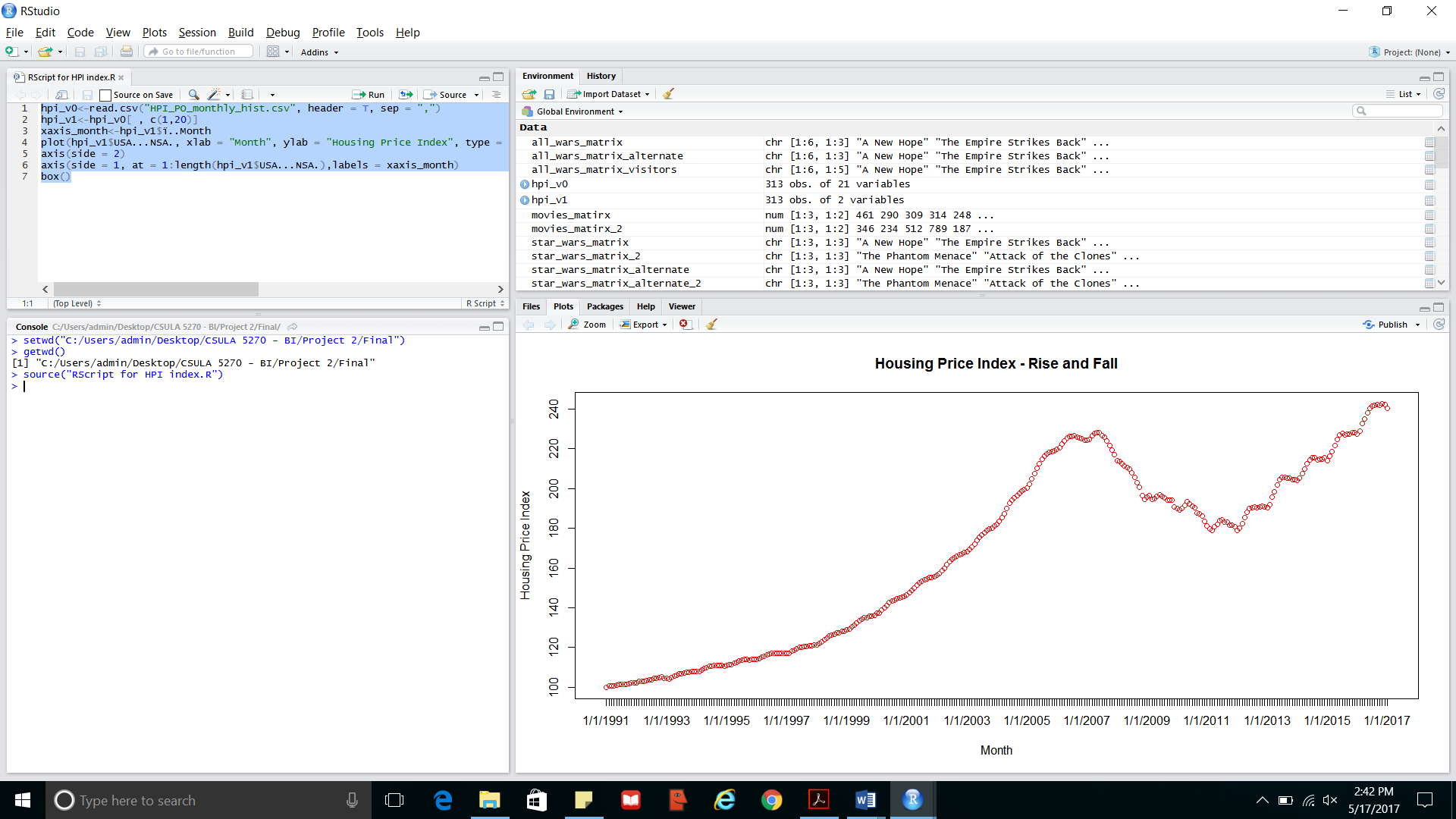
*CODE:*

> #DATA CLEANING #4 - aggregating data

> failedbank\_v3<-aggregate(failedbanks\_v2$Estimated.Loss, by = list(failedbanks\_v2$State), FUN = sum)

> View(failedbank\_v3)

1. **DATA VISUALIZATIONS:**
2. What was the main indicator of the housing bubble? And what was the effect of the financial crisis on the US housing market, as depicted by the housing price index?



From the line graph showing the trend of the housing price index, we can see that in the beginning of the 21st century the housing price index was at only around 140 points while within a very short period of 6 years there was an unnatural hike in housing prices leading to the index rising to its all time heights of around 230 points. This was one of the major indicators of the housing bubble, which even though several analysts had pointed out to the market, caution was not exercised. Like any other financial bubble through the history of time, this also reached it expected outcome. The bubble burst, housing prices started falling, investors lost their money, liquidity tightened in the market and the financial crisis, that decimated the world’s financial market, began.

Once the crisis hit the markets, the housing prices fell even faster than it rose after June 2007. It dropped to one of its lowest points in March 2011 to reach 178.9 points. After this the market showed signs of slight improvement, when the housing price index rose by around 5 points in July 2011. However, this seems have been a temporary and unsustainable increase. By January 2012, the index had hit bottom to stop at 179.1 points.

Techniques used:

The above visualization uses line graph with the period depicted on the x axis and the index on the y axis. This code was written in a script in RStudio. Both the script and the execution code are given below.

*RSCRIPT*

hpi\_v0<-read.csv("HPI\_PO\_monthly\_hist.csv", header = T, sep = ",")

hpi\_v1<-hpi\_v0[ , c(1,20)]

xaxis\_month<-hpi\_v1$ï..Month

plot(hpi\_v1$USA...NSA., xlab = "Month", ylab = "Housing Price Index", type = "b", axes = FALSE, col = "red", main = "Housing Price Index - Rise and Fall")

axis(side = 2)

axis(side = 1, at = 1:length(hpi\_v1$USA...NSA.),labels = xaxis\_month)

box()

*CODE*

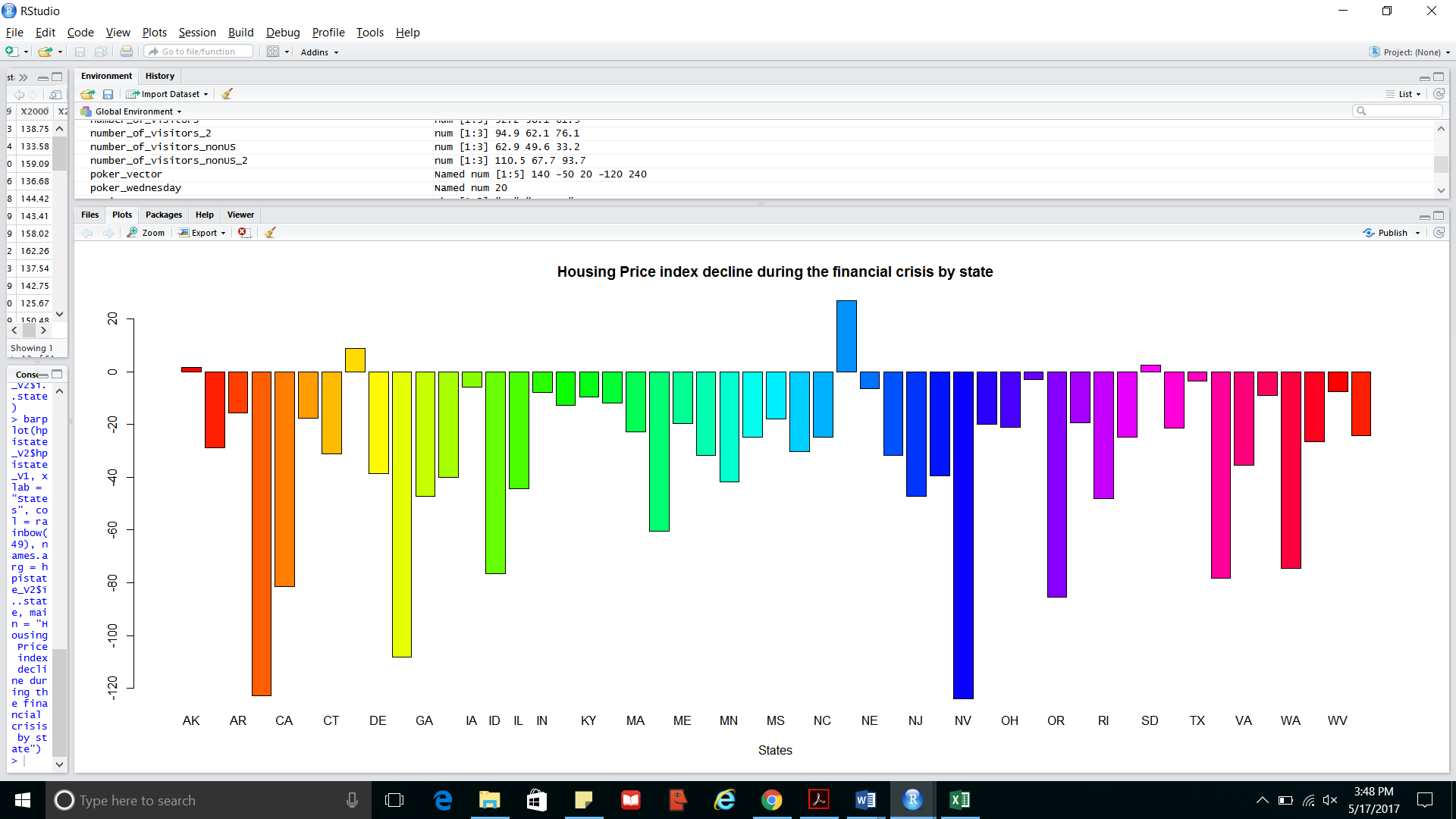
> setwd("C:/Users/admin/Desktop/CSULA 5270 - BI/Project 2/Final")

> getwd()

[1] "C:/Users/admin/Desktop/CSULA 5270 - BI/Project 2/Final"

> source("RScript for HPI index.R")

1. Which states had the worst fall in housing prices?



Almost all states in US were heavily affected by falling housing prices. As the bar graph above shows, the worst hit states were Nevada, Arizona and Florida, in that order. While other states might not have had such a severe effect as shown by the housing price index, the overall effect was still severe, affecting not only the home owners, but everyone connected to the financial market whose liquidity was affected by the fall in housing prices.

Techniques used:

The above visualization uses a bar plot to show the fall in housing prices during the period of the financial crisis, i.e 2007 to 2011. Additionally a function has been used to calculate the fall in the housing price index between 2011 and 2007. Both the function (written in RScript) and the execution codes are given below.

*FUNCTION IN RSCRIPT*

function\_for\_HPI<-function(x)

{

y<-hpistate\_V0$X2011

z<-hpistate\_V0$X2007

y-z

}

*CODE*

> setwd("C:/Users/admin/Desktop/CSULA 5270 - BI/Project 2/Final")

> getwd()

[1] "C:/Users/admin/Desktop/CSULA 5270 - BI/Project 2/Final"

> hpistate\_V0<-read.csv("HPI\_PO\_state.csv", header = T, sep = ",")

> View(hpistate\_V0)

> source("Function\_for\_HPI.R")

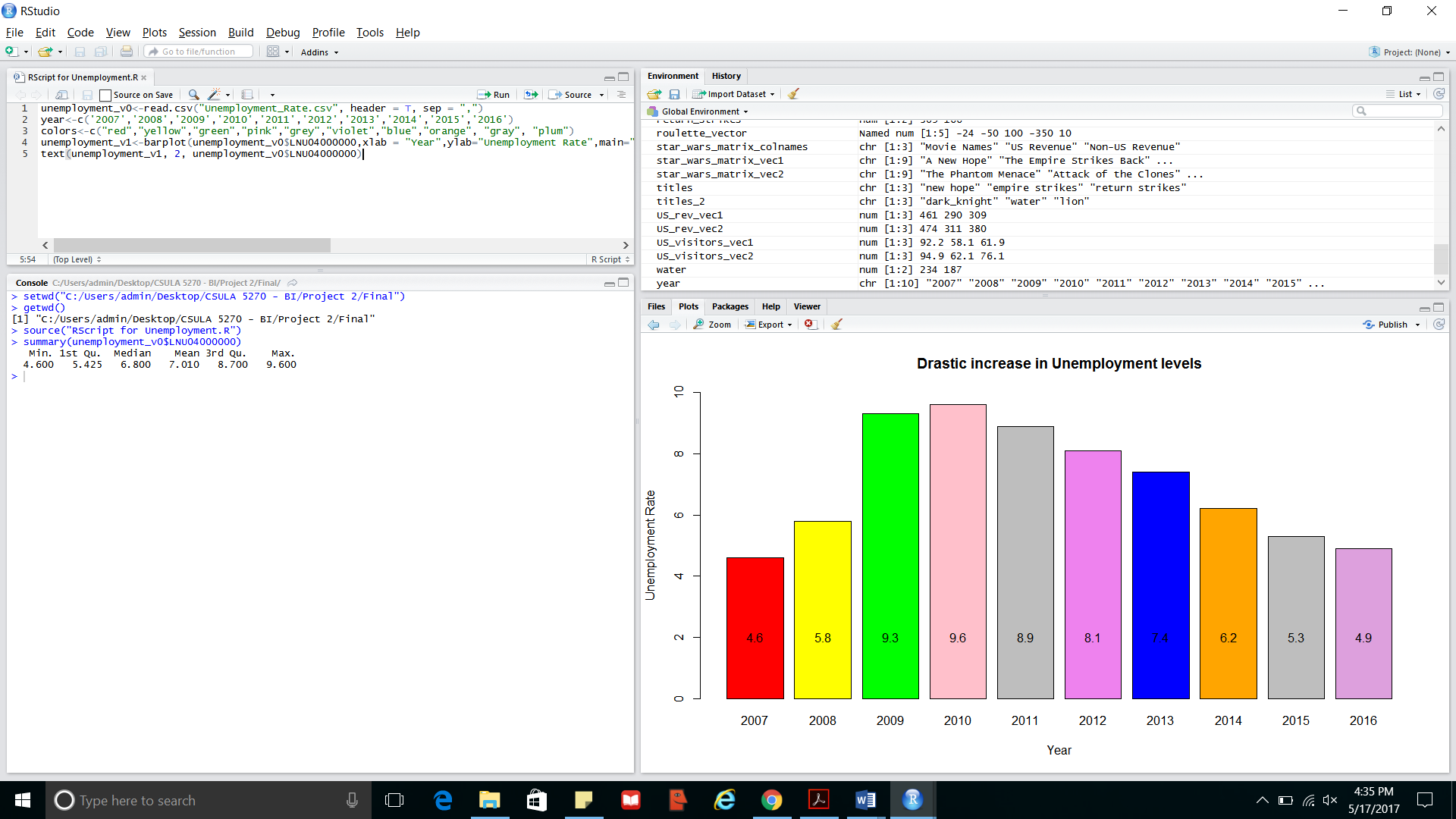
> hpistate\_V1<-function\_for\_HPI(hpistate\_V0$ï..state)

> hpistate\_V2<-data.frame(hpistate\_V0, hpistate\_V1)

> View(hpistate\_V2)

> states<-table(hpistate\_V2$ï..state)

> barplot(hpistate\_V2$hpistate\_V1, xlab = "States", col = rainbow(49), names.arg = hpistate\_V2$ï..state, main = "Housing Price index decline during the financial crisis by state")

1. How far the unemployment levels rose during the crisis period?

One of the other major effects of financial crisis was the drastic increase in unemployment. As seen from the above graph, unemployment rate rose to almost 10% in 2010. At the beginning of the crisis in 2007, the rate of unemployed people in the US was around 5%, and 6% in 2008. This drastically increased by more than 3% to rest at 9% in 2009 and reach its height, during the crisis, of 10% in 2010. As can be interpreted from the graph, while there has been a steady decline in unemployment after 2010, as of 2016 US has still not been able to reach its pre-crisis levels.

Techniques used:

The main tableau technique used in the above graphical representation is a bar chart, with unemployment rate being shown in the y-axis. The entire code has been written in RScript. Both the script and the execution code is given below.

*RSCRIPT*

unemployment\_v0<-read.csv("Unemployment\_Rate.csv", header = T, sep = ",")

year<-c('2007','2008','2009','2010','2011','2012','2013','2014','2015','2016')

colors<-c("red","yellow","green","pink","grey","violet","blue","orange", "gray", "plum")

unemployment\_v1<-barplot(unemployment\_v0$LNU04000000,xlab = "Year",ylab="Unemployment Rate",main="Drastic increase in Unemployment levels",col = colors, names.arg = year, ylim = c(0,10))

text(unemployment\_v1, 2, unemployment\_v0$LNU04000000)

*CODE*

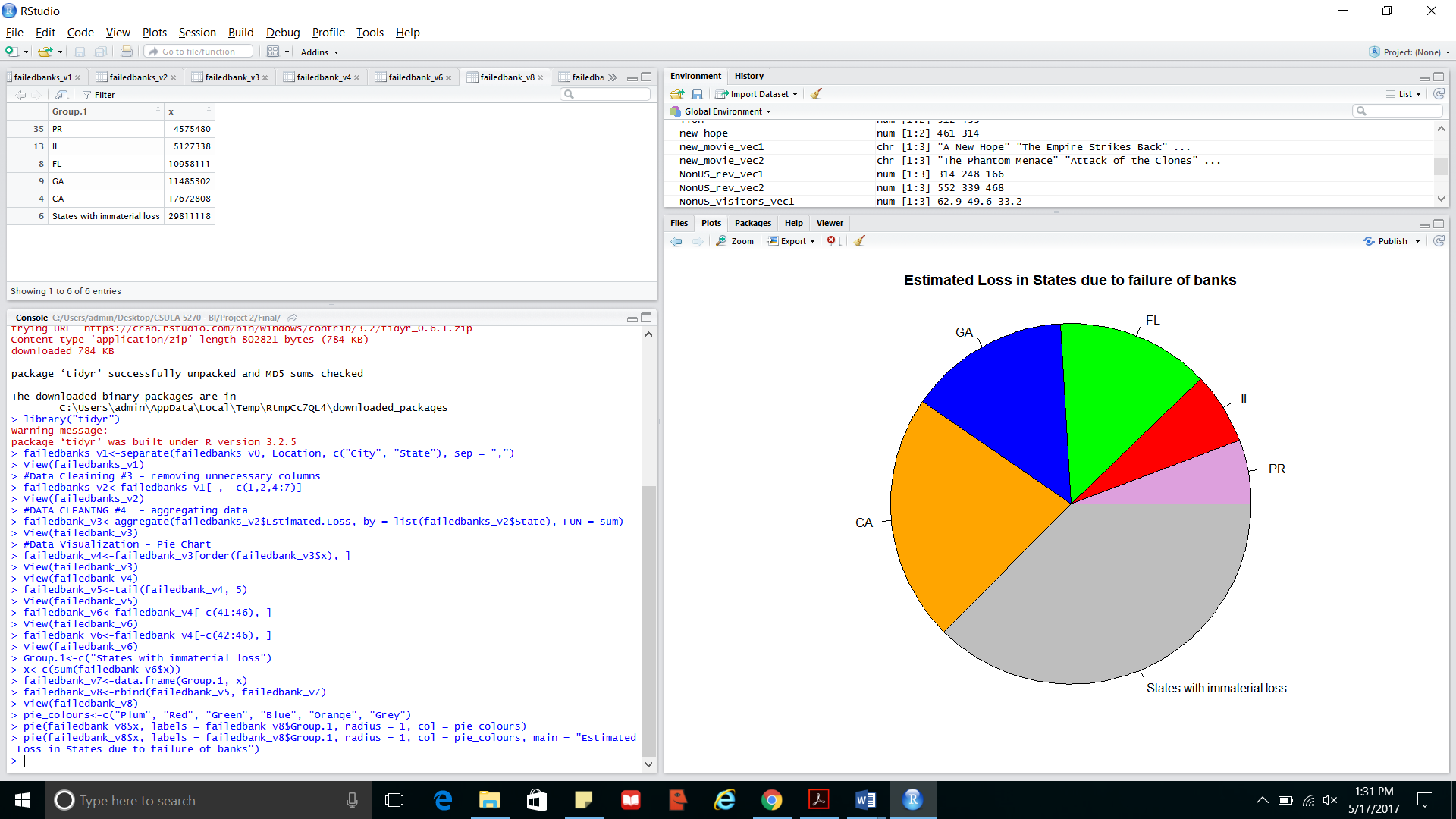
> setwd("C:/Users/admin/Desktop/CSULA 5270 - BI/Project 2/Final")

> getwd()

[1] "C:/Users/admin/Desktop/CSULA 5270 - BI/Project 2/Final"

> source("RScript for Unemployment.R")

1. What was the estimated loss in each state of US due to the failure of banks as shown by FDIC? Which states were worst hit?



Using the data released by the Federal Deposit Insurance Corporation (FDIC) about failed banks in the US, the above geographical representation shows each state in US differentiated by the extent of the estimated financial loss by banks. From the above, it shows that the worst hit state was California with an estimated loss of $17 thousand million, followed by Georgia and Florida with loss of almost $11 thousand million each and 4th comes Illinois with a loss of $5 thousand million.

Techniques used:

The above chart uses a pie chart to show the top 5 states that were worst hit by the financial crisis with regards to bank failures. All other states have been clubbed together as “States with Immaterial Loss”. To do this functions, sum, order, tail and data frames have been used. The code given below follows after the code for data cleaning.

*CODE:*

> #Data Visualization - Pie Chart

> failedbank\_v4<-failedbank\_v3[order(failedbank\_v3$x), ]

> View(failedbank\_v3)

> View(failedbank\_v4)

> failedbank\_v5<-tail(failedbank\_v4, 5)

> View(failedbank\_v5)

> failedbank\_v6<-failedbank\_v4[-c(41:46), ]

> View(failedbank\_v6)

> failedbank\_v6<-failedbank\_v4[-c(42:46), ]

> View(failedbank\_v6)

> Group.1<-c("States with immaterial loss")

> x<-c(sum(failedbank\_v6$x))

> failedbank\_v7<-data.frame(Group.1, x)

> failedbank\_v8<-rbind(failedbank\_v5, failedbank\_v7)

> View(failedbank\_v8)

> pie\_colours<-c("Plum", "Red", "Green", "Blue", "Orange", "Grey")

> pie(failedbank\_v8$x, labels = failedbank\_v8$Group.1, radius = 1, col = pie\_colours)

> pie(failedbank\_v8$x, labels = failedbank\_v8$Group.1, radius = 1, col = pie\_colours, main = "Estimated Loss in States due to failure of banks")

1. **STATISTICAL AVERAGES OF DATA**

The codes below give the summary of the data used to create the visualizations using the 4 datasets.

1. Summary of the Housing Price Index dataset:

> summary(hpi\_v1$USA...NSA.)

Min. 1st Qu. Median Mean 3rd Qu. Max.

100.0 120.3 179.7 167.5 205.3 242.6

1. Summary of the Housing Price Index dataset by states:

> summary(hpistate\_V2$hpistate\_V1)

Min. 1st Qu. Median Mean 3rd Qu. Max.

-123.80 -43.02 -24.81 -33.04 -12.27 26.86

1. Summary of the Unemployment dataset:

> summary(unemployment\_v0$LNU04000000)

Min. 1st Qu. Median Mean 3rd Qu. Max.

4.600 5.425 6.800 7.010 8.700 9.600

1. Summary of the Failed Bank dataset:

> summary(failedbank\_v8$x)

Min. 1st Qu. Median Mean 3rd Qu. Max.

4575000 6585000 11220000 13270000 16130000 29810000

1. **COMPLETE R CODE FROM RSTUDIO**
2. Complete code for the Housing Price Index dataset in Question 1:

*SCRIPT*

hpi\_v0<-read.csv("HPI\_PO\_monthly\_hist.csv", header = T, sep = ",")

hpi\_v1<-hpi\_v0[ , c(1,20)]

xaxis\_month<-hpi\_v1$ï..Month

plot(hpi\_v1$USA...NSA., xlab = "Month", ylab = "Housing Price Index", type = "b", axes = FALSE, col = "red", main = "Housing Price Index - Rise and Fall")

axis(side = 2)

axis(side = 1, at = 1:length(hpi\_v1$USA...NSA.),labels = xaxis\_month)

box()

*CODE*

> setwd("C:/Users/admin/Desktop/CSULA 5270 - BI/Project 2/Final")

> getwd()

[1] "C:/Users/admin/Desktop/CSULA 5270 - BI/Project 2/Final"

> source("RScript for HPI index.R")

> summary(hpi\_v1$USA...NSA.)

Min. 1st Qu. Median Mean 3rd Qu. Max.

100.0 120.3 179.7 167.5 205.3 242.6

1. Complete code for the Housing Price Index dataset in Question 2:

*FUNCTION IN RSCRIPT*

function\_for\_HPI<-function(x)

{

y<-hpistate\_V0$X2011

z<-hpistate\_V0$X2007

y-z

}

*CODE*

> setwd("C:/Users/admin/Desktop/CSULA 5270 - BI/Project 2/Final")

> getwd()

[1] "C:/Users/admin/Desktop/CSULA 5270 - BI/Project 2/Final"

> hpistate\_V0<-read.csv("HPI\_PO\_state.csv", header = T, sep = ",")

> View(hpistate\_V0)

> source("Function\_for\_HPI.R")

> hpistate\_V1<-function\_for\_HPI(hpistate\_V0$ï..state)

> hpistate\_V2<-data.frame(hpistate\_V0, hpistate\_V1)

> View(hpistate\_V2)

> states<-table(hpistate\_V2$ï..state)

> barplot(hpistate\_V2$hpistate\_V1, xlab = "States", col = rainbow(49), names.arg = hpistate\_V2$ï..state, main = "Housing Price index decline during the financial crisis by state")

> summary(hpistate\_V2$hpistate\_V1)

Min. 1st Qu. Median Mean 3rd Qu. Max.

-123.80 -43.02 -24.81 -33.04 -12.27 26.86

1. Complete code for the Unemployment dataset:

*RSCRIPT*

unemployment\_v0<-read.csv("Unemployment\_Rate.csv", header = T, sep = ",")

year<-c('2007','2008','2009','2010','2011','2012','2013','2014','2015','2016')

colors<-c("red","yellow","green","pink","grey","violet","blue","orange", "gray", "plum")

unemployment\_v1<-barplot(unemployment\_v0$LNU04000000,xlab = "Year",ylab="Unemployment Rate",main="Drastic increase in Unemployment levels",col = colors, names.arg = year, ylim = c(0,10))

text(unemployment\_v1, 2, unemployment\_v0$LNU04000000)

*CODE*

> setwd("C:/Users/admin/Desktop/CSULA 5270 - BI/Project 2/Final")

> getwd()

[1] "C:/Users/admin/Desktop/CSULA 5270 - BI/Project 2/Final"

> source("RScript for Unemployment.R")

> summary(unemployment\_v0$LNU04000000)

Min. 1st Qu. Median Mean 3rd Qu. Max.

4.600 5.425 6.800 7.010 8.700 9.600

1. Complete code for the Failed Bank dataset in Question 4:

> setwd("C:/Users/admin/Desktop/CSULA 5270 - BI/Project 2/Final")

> getwd()

[1] "C:/Users/admin/Desktop/CSULA 5270 - BI/Project 2/Final"

> failedbanks\_v0<-read.csv("FailedBank.csv", header = T, sep = ",")

> View(failedbanks\_v0)

> #Data Cleaning #1 - removing NAs from the Estimated Loss column and replacing with the mean

> failedbanks\_v0$Estimated.Loss[is.na(failedbanks\_v0$Estimated.Loss)]<-mean(failedbanks\_v0$Estimated.Loss, na.rm = TRUE)

> #Data Cleaning #2 - spliting Location column into City and State

> install.packages("tidyr")

Installing package into ‘C:/Users/admin/Documents/R/win-library/3.2’

(as ‘lib’ is unspecified)

There is a binary version available but the source version is later:

binary source needs\_compilation

tidyr 0.6.1 0.6.3 TRUE

Binaries will be installed

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.2/tidyr\_0.6.1.zip'

Content type 'application/zip' length 802821 bytes (784 KB)

downloaded 784 KB

package ‘tidyr’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\admin\AppData\Local\Temp\RtmpCc7QL4\downloaded\_packages

> library("tidyr")

Warning message:

package ‘tidyr’ was built under R version 3.2.5

> failedbanks\_v1<-separate(failedbanks\_v0, Location, c("City", "State"), sep = ",")

> View(failedbanks\_v1)

> #Data Cleaining #3 - removing unnecessary columns

> failedbanks\_v2<-failedbanks\_v1[ , -c(1,2,4:7)]

> View(failedbanks\_v2)

> #DATA CLEANING #4 - aggregating data

> failedbank\_v3<-aggregate(failedbanks\_v2$Estimated.Loss, by = list(failedbanks\_v2$State), FUN = sum)

> View(failedbank\_v3)

> #Data Visualization - Pie Chart

> failedbank\_v4<-failedbank\_v3[order(failedbank\_v3$x), ]

> View(failedbank\_v3)

> View(failedbank\_v4)

> failedbank\_v5<-tail(failedbank\_v4, 5)

> View(failedbank\_v5)

> failedbank\_v6<-failedbank\_v4[-c(41:46), ]

> View(failedbank\_v6)

> failedbank\_v6<-failedbank\_v4[-c(42:46), ]

> View(failedbank\_v6)

> Group.1<-c("States with immaterial loss")

> x<-c(sum(failedbank\_v6$x))

> failedbank\_v7<-data.frame(Group.1, x)

> failedbank\_v8<-rbind(failedbank\_v5, failedbank\_v7)

> View(failedbank\_v8)

> pie\_colours<-c("Plum", "Red", "Green", "Blue", "Orange", "Grey")

> pie(failedbank\_v8$x, labels = failedbank\_v8$Group.1, radius = 1, col = pie\_colours)

> pie(failedbank\_v8$x, labels = failedbank\_v8$Group.1, radius = 1, col = pie\_colours, main = "Estimated Loss in States due to failure of banks")

> summary(failedbank\_v8$x)

Min. 1st Qu. Median Mean 3rd Qu. Max.

4575000 6585000 11220000 13270000 16130000 29810000

1. **References:**
2. "Origins of Financial Crisis Crash course." The Economist. The Economist Newspaper, 07 Sept. 2013. Web. 18 May 2017.
3. Amadeo, Kimberly. "What Caused the 2008 Financial Crisis and Could It Happen Again?" The Balance. N.p., n.d. Web. 17 May 2017.
4. Writer, Jake Grovum Stateline Staff. "2008 financial crisis impact still hurting states." USA Today. Gannett Satellite Information Network, 15 Sept. 2013. Web. 18 May 2017.
5. "Unemployment." U.S. Bureau of Labor Statistics. U.S. Bureau of Labor Statistics, n.d. Web. 18 May 2017.
6. Amadeo, Kimberly. "When and Why Did the Stock Market Crash in 2008?" The Balance. N.p., n.d. Web. 17 May 2017.