**CIS 5270 PROJECT**

Analysis of the United States Drought Problem

By -:

Atinder Paul Singh

CIN – 305091298

1. **Data Set URL’s**

Dataset Retrieved from <https://www.kaggle.com/us-drought-monitor/united-states-droughts-by-county> .

The topic data includes two files i.e Us Droughts and County Information.

**File 1: US Drought.**

* **releaseDate**: when this data was released on the USDM website
* **FIPS**: the FIPS code for this county
* **county**: the county name
* **state**: the state the county is in
* **NONE**: percentage of the county that is *not in drought*
* **D0**: percentage of the county that is in *abnormally dry conditions*
* **D1**: percentage of the county that is in *moderate drought*
* **D2**: percentage of the county that is in *severe drought*
* **D3**: percentage of the county that is in *extreme drought*
* **D4**: percentage of the county that is in *exceptional drought*
* **validStart**: the starting date of the week that these observations represent
* **validEnd**: the ending date of the week that these observations represent
* **domStatisticFormatID**: seems to always be 1

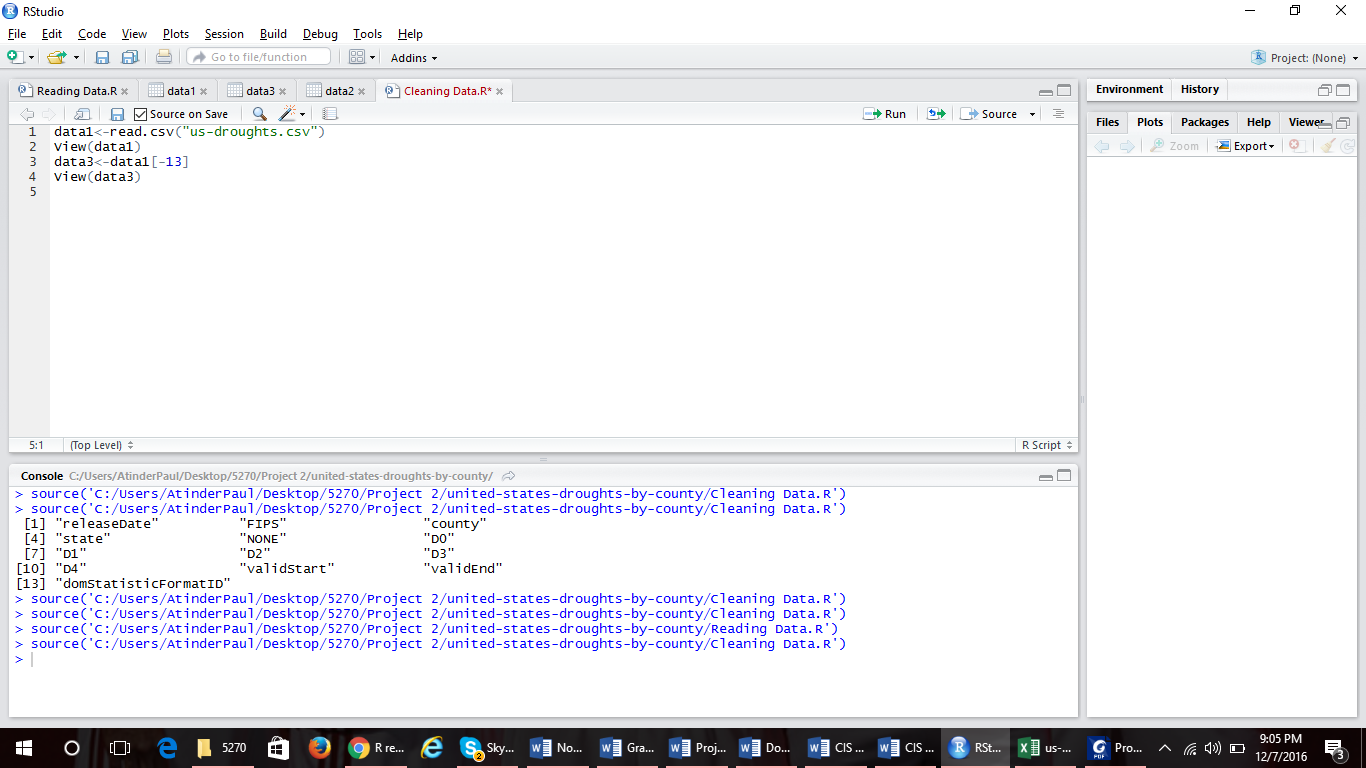
**File 2: County Info**

* **USPS**: United States Postal Service State Abbreviation
* **GEOID**: FIPS code
* **ANSICODE**: American National Standards Institute code
* **NAME**: Name
* **ALAND**: Land Area (square meters) - Created for statistical purposes only
* **AWATER**: Water Area (square meters) - Created for statistical purposes only
* **ALAND\_SQMI**: Land Area (square miles) - Created for statistical purposes only
* **AWATER\_SQMI**: Water Area (square miles) - Created for statistical purposes only
* **INTPTLAT**: Latitude (decimal degrees) First character is blank or "-" denoting North or South latitude respectively
* **INTPTLONG**: Longitude (decimal degrees) First character is blank or "-" denoting East or West longitude respectively.

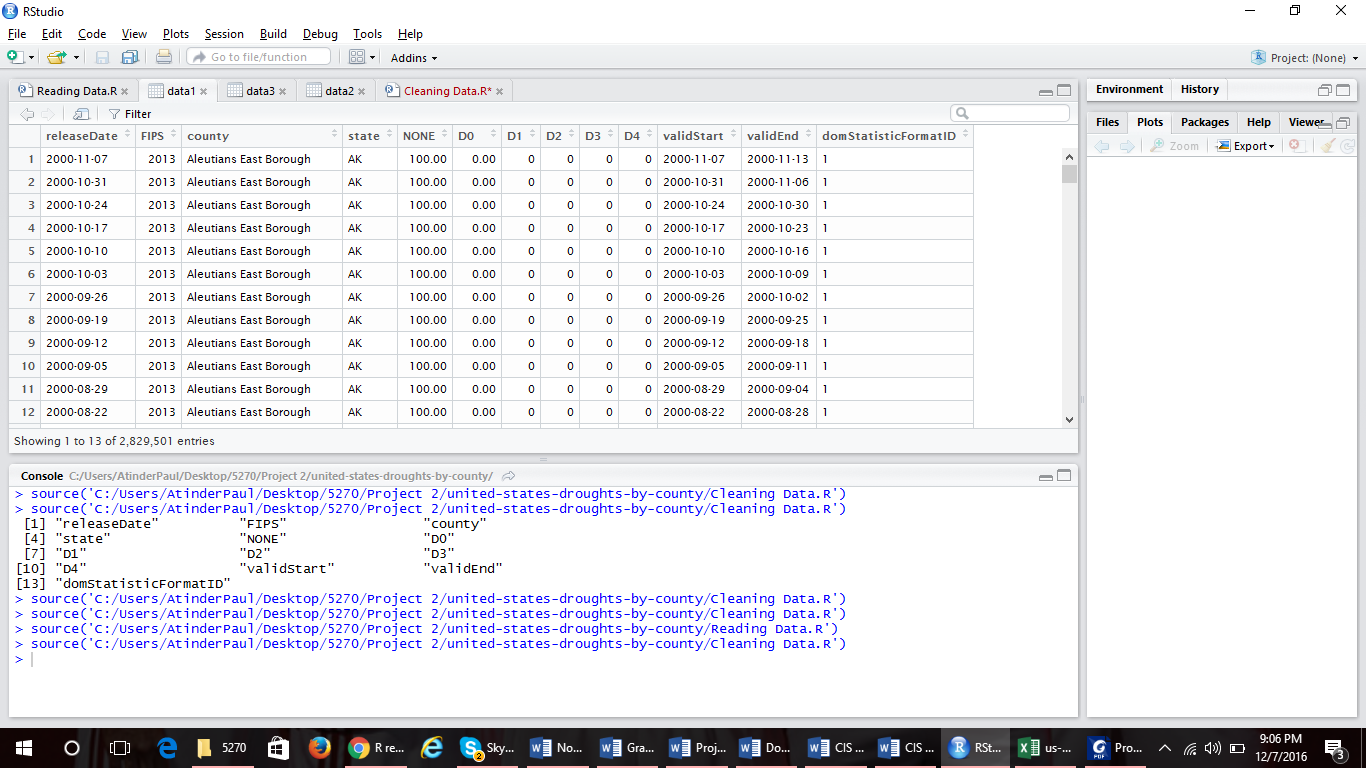
1. **Data Cleaning**

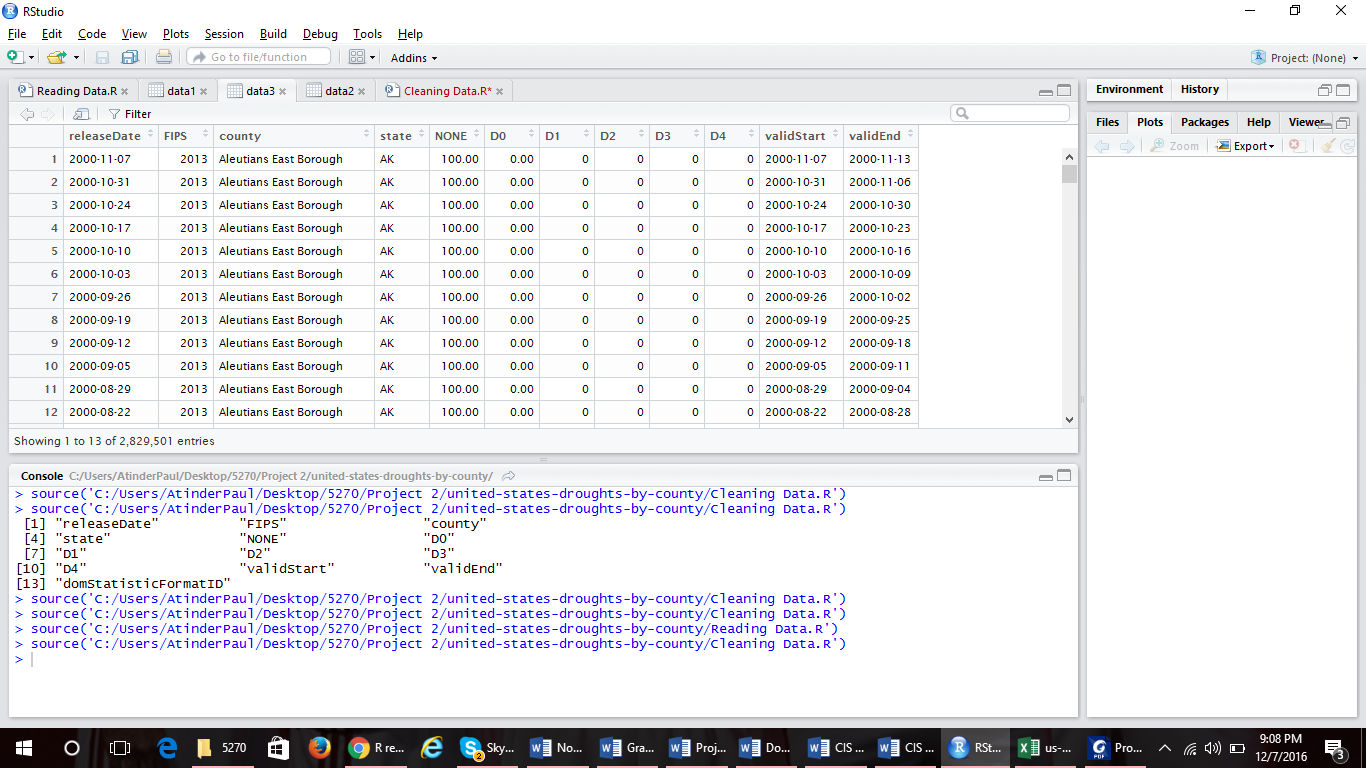
Category 1) Removing Irrelevant Columns.

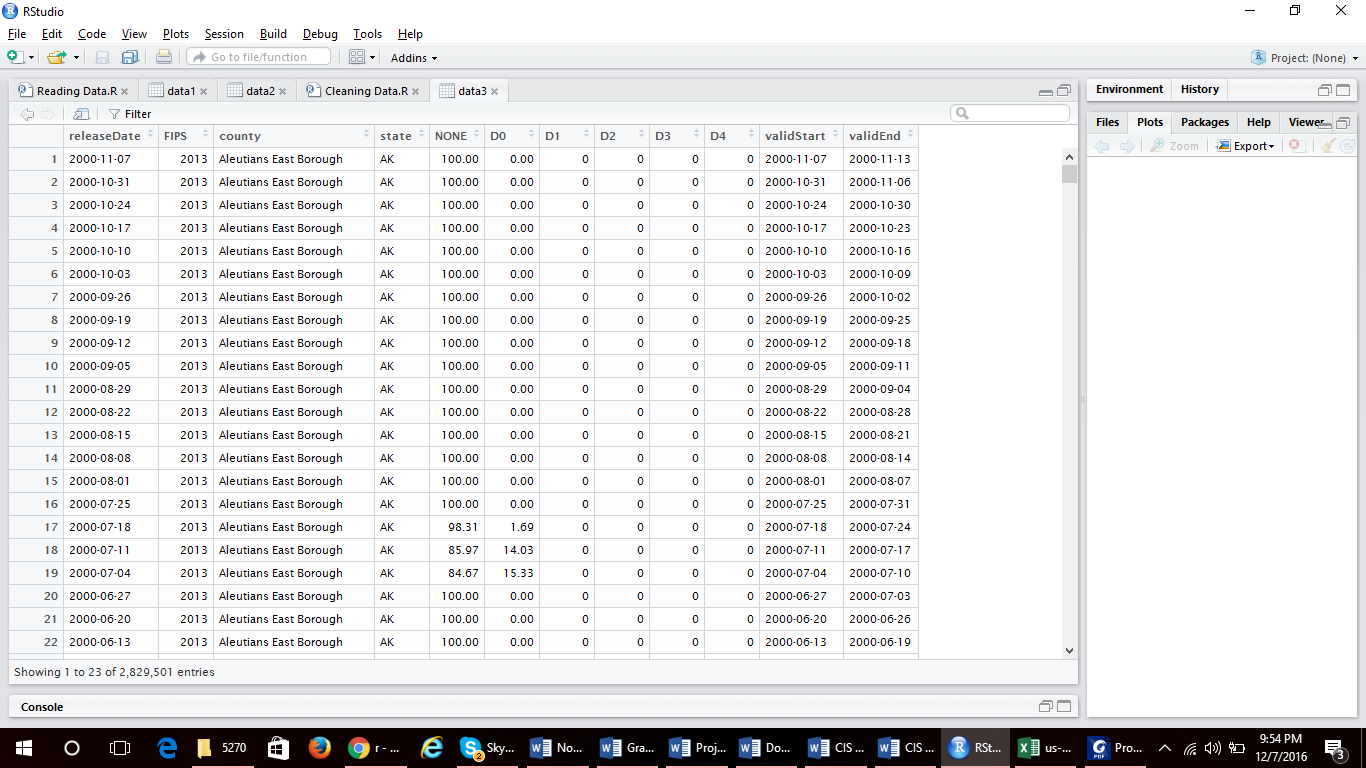
Code to Remove the **domStatisticFormatID** Column.

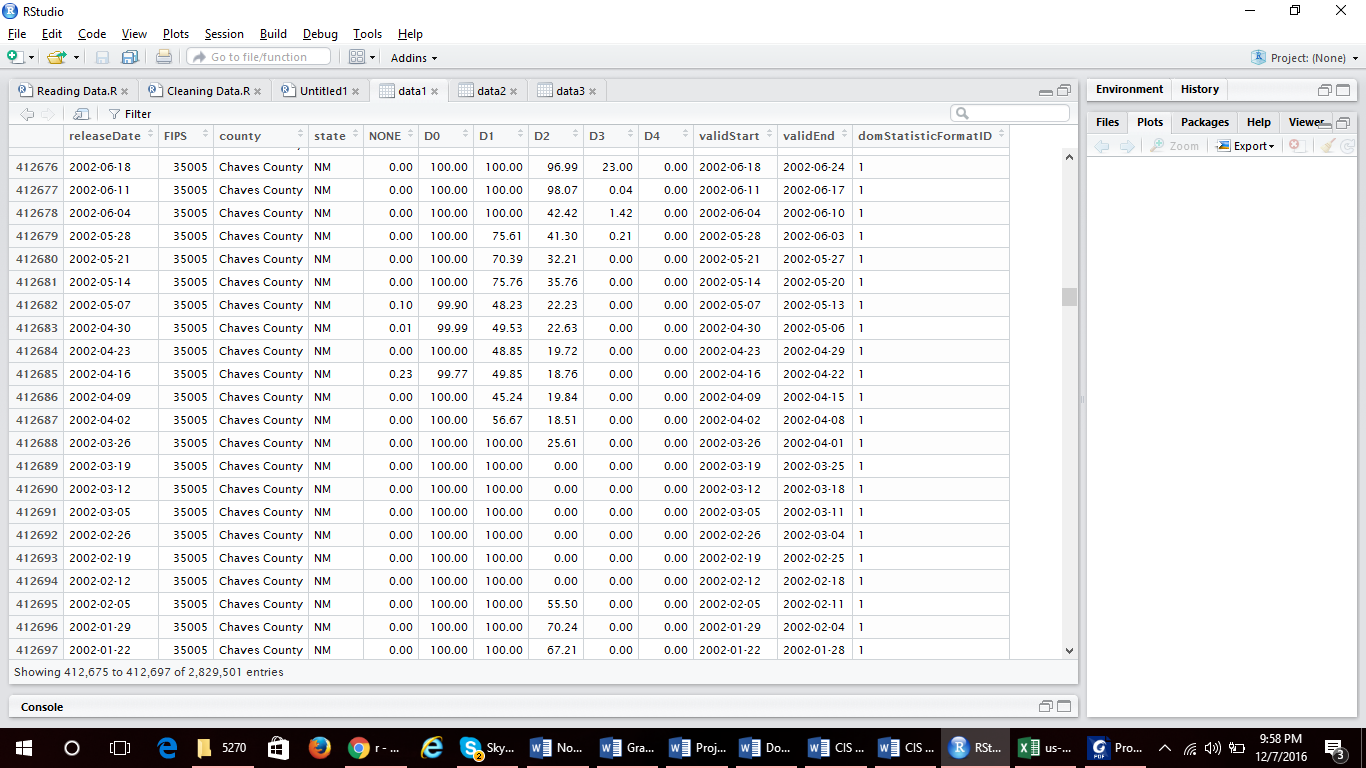


Screenshot

Before.

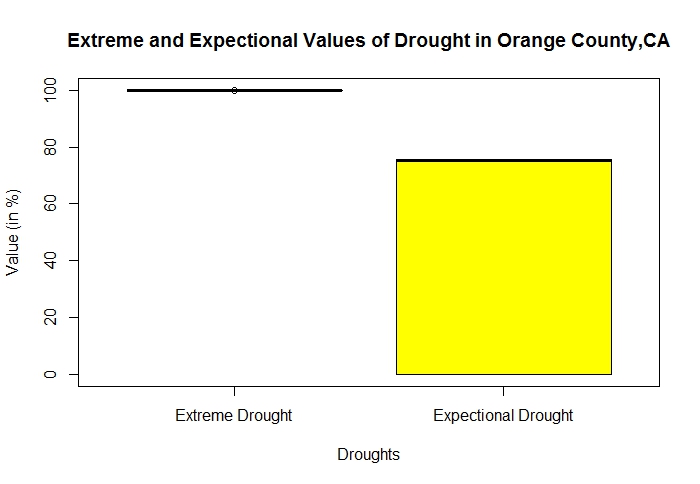
After.

Category 2) Contradicting Values – Already Clean

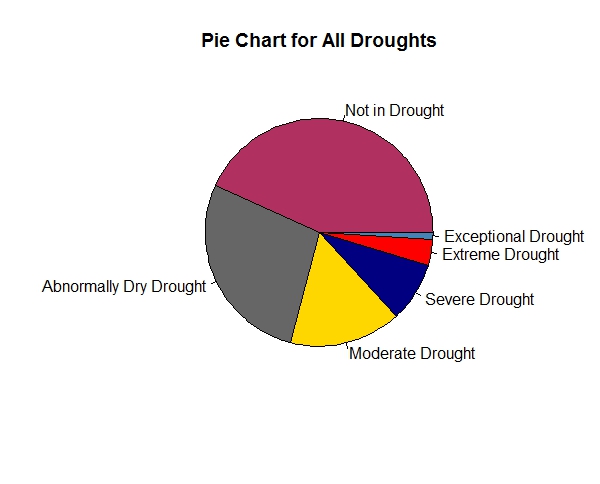
Category 3) Uniqueness Violation

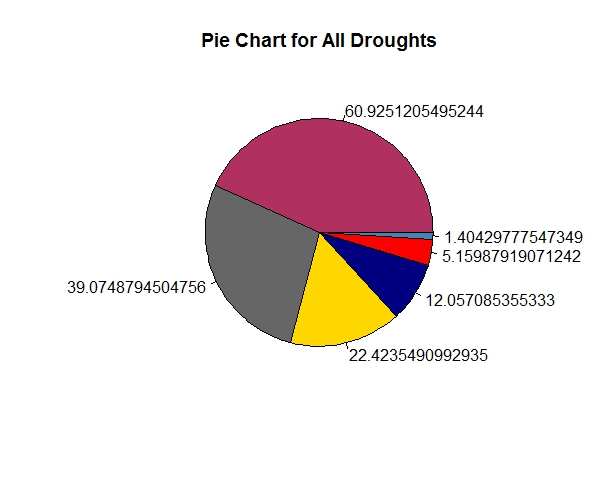
**Analysis and Visualization**

**Q1) What is the Extreme and Exceptional Value of Drought in Orange County, California?**

****

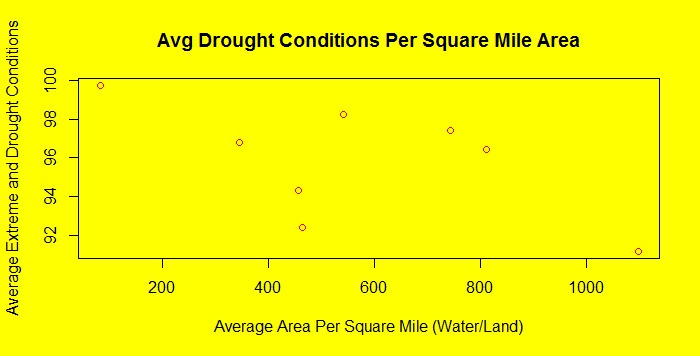
In the first visualization of the project, we are analyzing the two major factors of droughts that are the extreme and exceptional conditions of droughts in Orange County, California. Here we can understand that in Orange County there is a 100% Extreme Condition of drought and a variable % of Exceptional Condition of drought. California is known for its drought situation and even though this is just a case of one county in California it puts a major light on wat the situation is like in California.

**Q2) What are the mean values for all different conditions of droughts?**

****

Here in the second visualization of the project, we analyze the mean drought percentages of all the data set we have. We understand from the data that almost on an average 61% counties in the United States don’t have any sort of drought conditions. The other readings from the findings are 40% on an average have abnormally dry drought condition whereas almost 22%, 12%, 5%, 1% have Moderate, Severe, Extreme and Exceptional conditions of drought respectively. Pie chart shows us a well-organized chart with different color codes for different conditions.

3) What is comparison of Average Area Per Square Mile (Water/Land) vs Average Extreme and Exceptional Drought Conditions



Here in the third and final visualization, we have analyzed the Average drought conditions per square mile area in the whole of data. We here can understand that there are various different aspects of Drought conditions per the square miles in the county of United States. Values go as high as 99% per close to 110 sq/mile. We can understand that the drought condition is mutually exclusive as even if the sq/ml for instance is around 750 sq/ml there is a possibility of 97% of extreme and exceptional drought conditions in all the counties.

CODE FOR ALL ANALYSIS/VISUALIZATION

Script 1 – Reading Data.R

# Reading 1st File

data1<-read.csv("us-droughts.csv")

View(data1)

# Reading 2nd File

data2<-read.csv("county\_info\_2016.csv")

View(data2)

Script 2 – Cleaning Data.R

data1<-read.csv("us-droughts.csv")

View(data1)

#Removing 13th Coloumn

data3<-data1[-13]

View(data3)

Script 3 – Analysis.R

#mydata<-subset(data3, data3$state=='CA' & data3$county=='Orange County' & data3$D3>90 )

#View(mydata)

boxplot(mydata$D3, mydata$D4, names = c('Extreme Drought', 'Exceptional Drought') ,main= 'Extreme and Expectional Values of Drought in Orange County,CA' , xlab= 'Droughts', ylab= 'Value (in %)', col = 'yellow')

Script 4 – Analysis2.R

#dnonemean<-mean(data3$NONE)

#View(dnonemean)

#d0mean<-mean(data3$D0)

#d1mean<-mean(data3$D1)

#d2mean<-mean(data3$D2)

#d3mean<-mean(data3$D3)

#d4mean<-mean(data3$D4)

#allmean<-c(dnonemean,d0mean, d1mean, d2mean, d3mean, d4mean)

#View(allmean)

#pc<-c('maroon', 'grey40', 'gold', 'navy', 'red','steelblue')

#datalabels<-c('Not in Drought', 'Abnormally Dry Drought', 'Moderate Drought', 'Severe Drought', 'Extreme Drought', 'Exceptional Drought')

pie(allmean, labels =datalabels, main = 'Pie Chart for All Droughts', col=pc, clockwise = FALSE)

Script 5 – Analysis3.R

#newdata<-mean(data2$AWATER\_SQMI)

#newdata2<-mean(data2$ALAND\_SQMI)

#df\_1<-data.frame(newdata,newdata2)

#View(df\_1)

#newmean<-mean(mydatanew$D3)

#newmean2<-mean(mydatanew$D4)

#df\_2<-data.frame(newmean,newmean2)

#View(df\_2)

#par(bg= 'Yellow')

#plot(df\_1$newdatasqml, df\_2$newmean, xlab = "Average Area Per Square Mile (Water/Land)" , ylab = "Average Extreme and Drought Conditions", main = "Avg Drought Conditions Per Square Mile Area", col='Red'