**Code is written in R Language**

#install some packages required

install.packages("tidyr")

install.packages("sqldf")

library(tidyr)

#Read data from CSV file

texas\_cities\_data<- read.csv("texas-cities.csv", header = T)

View(texas\_cities\_data)

cities\_to\_county<-read.csv("cities-to-county.csv", header = T)

Create dataframes

#View(texas\_cities\_data)

dataframe1<-data.frame(texas\_cities\_data)

dataframe\_county<-data.frame(cities\_to\_county)

Rename the column names of dataframe for user convinience

#View(dataframe1)

colnames(dataframe1) <- c("Name", "2010","2011","2012","2013","2014","2015","2016","2017")

View(dataframe1)

colnames(dataframe\_country)[2] <- "Name"

#Texas\_Cities\_Country<- merge(dataframe1, dataframe\_country, by = "Category", all. = TRUE)

Convert the upper case to lowercase in cities\_to\_county dataframe

levels(dataframe\_county$Name) <- tolower(levels(dataframe\_country$Name))

levels(dataframe\_cities$Name) <- tolower(levels(dataframe\_cities$Name))

View(dataframe\_cities)

Texas\_Cities\_County <- merge(dataframe\_cities,dataframe\_country,by.dataframe\_cities="Name",by.dataframe\_country="Name")

View(Texas\_Cities\_County)

dataframe\_cities\_county<-data.frame(Texas\_Cities\_Country)

View(dataframe\_cities\_county)

total <- merge(dataframe\_cities,dataframe\_country,by="Name")

View(total)

#Separate Cities, towns and villages

dataframe2<-data.frame(separate(data = dataframe1, col = Name, into = c("City\_town\_village", "State"), sep = ","))

View(dataframe2)

dataframe3<-data.frame(separate(data = dataframe2, col = City\_town\_village, into = c("City" ,"town","village","v1"), sep = " "))

View(dataframe3)

library(sqldf)

colnames(dataframe3)[1] <- "city\_name"

colnames(dataframe3)[2] <- "t\_1"

colnames(dataframe3)[3] <- "v\_1"

library(stringr)

library(dplyr)

dataframe4<-data.frame(dataframe3 %>%

filter(str\_detect(t\_1, "city")))

View(dataframe4)

dataframe4<-data.frame(dataframe3 %>%

filter(str\_detect(t\_1, "city")))

dataframe4[3:4] <- NULL

dataframe5[4] <- NULL

dataframe5<-data.frame(dataframe3 %>%

filter(str\_detect(v\_1, "city")))

View(dataframe5)

dataframe6<-data.frame(dataframe3 %>%

filter(str\_detect(v1, "city")))

View(dataframe6)

dataframe\_cities <- rbind(dataframe4, dataframe5, dataframe6)

View(dataframe\_cities)

View(dataframe3)

dataframe7<-data.frame(dataframe3 %>%

filter(str\_detect(t\_1, "town")))

dataframe10<-data.frame(dataframe3 %>%

filter(str\_detect(v\_1, "town")))

View(dataframe7)

dataframe10[4] <- NULL

dataframe10$city\_name <- paste(dataframe10$city\_name," ",dataframe10$t\_1)

dataframe10[2] <- NULL

colnames(dataframe10)[2] <- "t\_1"

dataframe\_town <- rbind(dataframe7, dataframe10)

View(dataframe\_town)

View(dataframe8)

dataframe7[3:4] <- NULL

dataframe8<-data.frame(dataframe3 %>%

filter(str\_detect(v\_1, "village")))

dataframe8[4] <- NULL

dataframe11<-data.frame(dataframe3 %>%

filter(str\_detect(t\_1, "village")))

View(dataframe8)

dataframe11[3] <- NULL

View(dataframe11)

dataframe8$city\_name <- paste(dataframe8$city\_name," ",dataframe8$t\_1)

dataframe8[2] <- NULL

colnames(dataframe8)[2] <- "t\_1"

dataframe\_village <- rbind(dataframe8, dataframe11)

View(dataframe\_village)

dataframe9$city\_name<-data.frame(within(dataframe5, city <- paste(city\_name, t\_1, sep=' ')))

View(dataframe9)

#dataframe5$city<-c("city\_name","t\_1")

dataframe5$city\_name <- paste(dataframe5$city\_name," ",dataframe5$t\_1)

View(dataframe5)

dataframe5[2] <- NULL

dataframe5[13] <- NULL

dataframe6$city\_name <- paste(dataframe6$city\_name," ",dataframe6$t\_1," ", dataframe6$v\_1)

View(dataframe6)

dataframe6[2] <- NULL

dataframe6[2] <- NULL

colnames(dataframe4)[2] <- "v1"

colnames(dataframe5)[2] <- "v1"

#dataframe5[12] <- NULL

dataframe\_cities <- rbind(dataframe4, dataframe5, dataframe6)

View(dataframe\_cities)

colnames(dataframe\_cities) <- c("Name","Category","State","Y\_2010","Y\_2011","Y\_2012","Y\_2013","Y\_2014","Y\_2015","Y\_2016","Y\_2017")

colnames(dataframe\_town) <- c("Name","Category","State","Y\_2010","Y\_2011","Y\_2012","Y\_2013","Y\_2014","Y\_2015","Y\_2016","Y\_2017")

colnames(dataframe\_village) <- c("Name","Category","State","Y\_2010","Y\_2011","Y\_2012","Y\_2013","Y\_2014","Y\_2015","Y\_2016","Y\_2017")

**1)What was the largest city, town, and village in each year?**

**ANS: City: Houston**

**Town:Flower Mound**

**Village: The Hills**

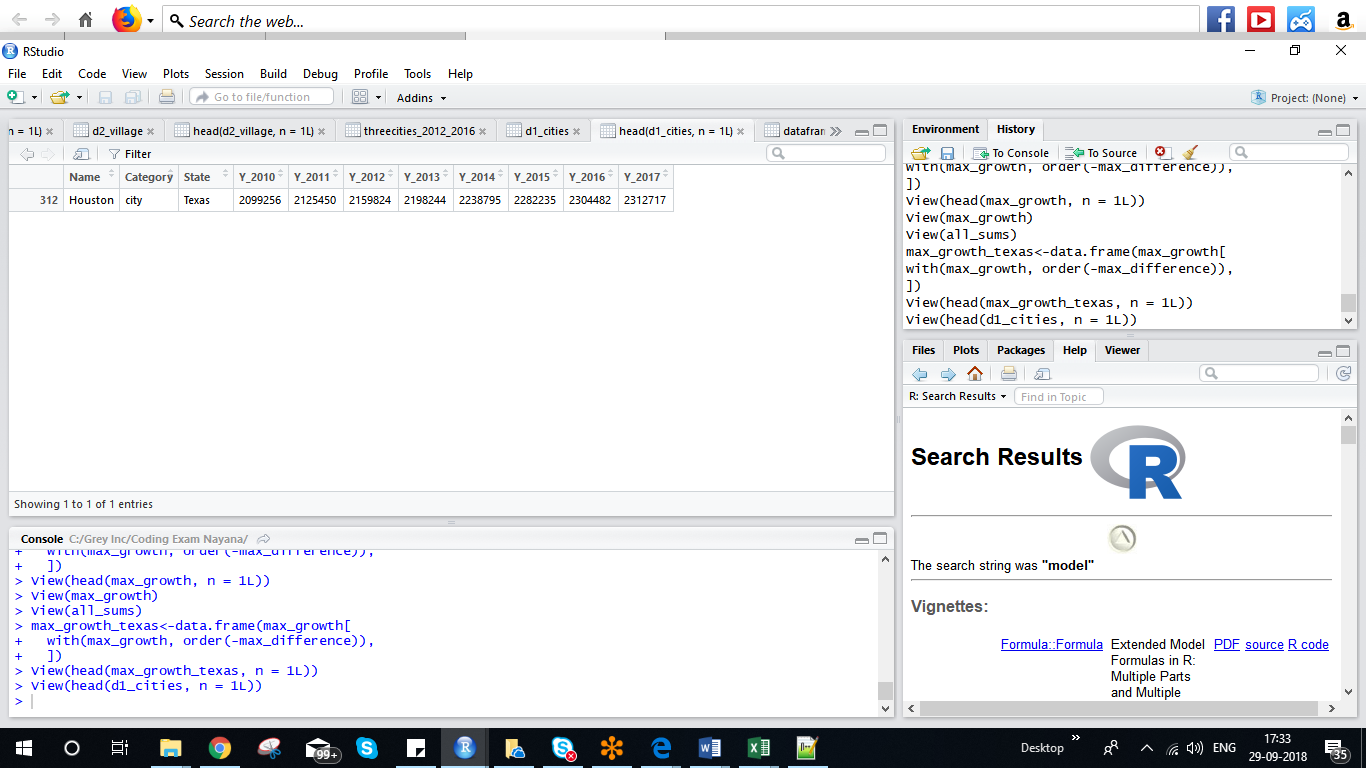
d1\_cities<-data.frame(dataframe\_cities[

with(dataframe\_cities, order(-Y\_2011)),

])

View(d1\_cities)

View(head(d1\_cities, n = 1L))



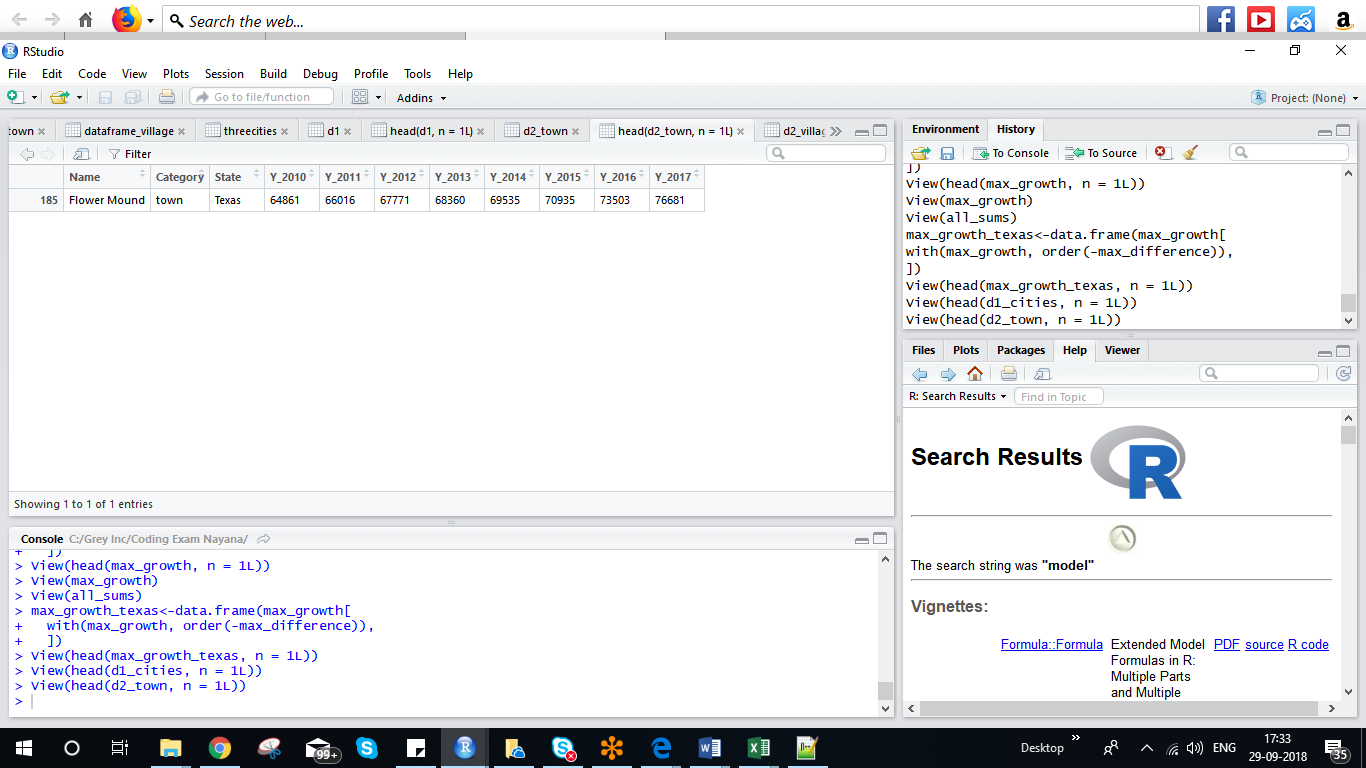
d2\_town<-data.frame(dataframe\_town[

with(dataframe\_town, order(-Y\_2011)),

])

View(d2\_town)

View(head(d2\_town, n = 1L))



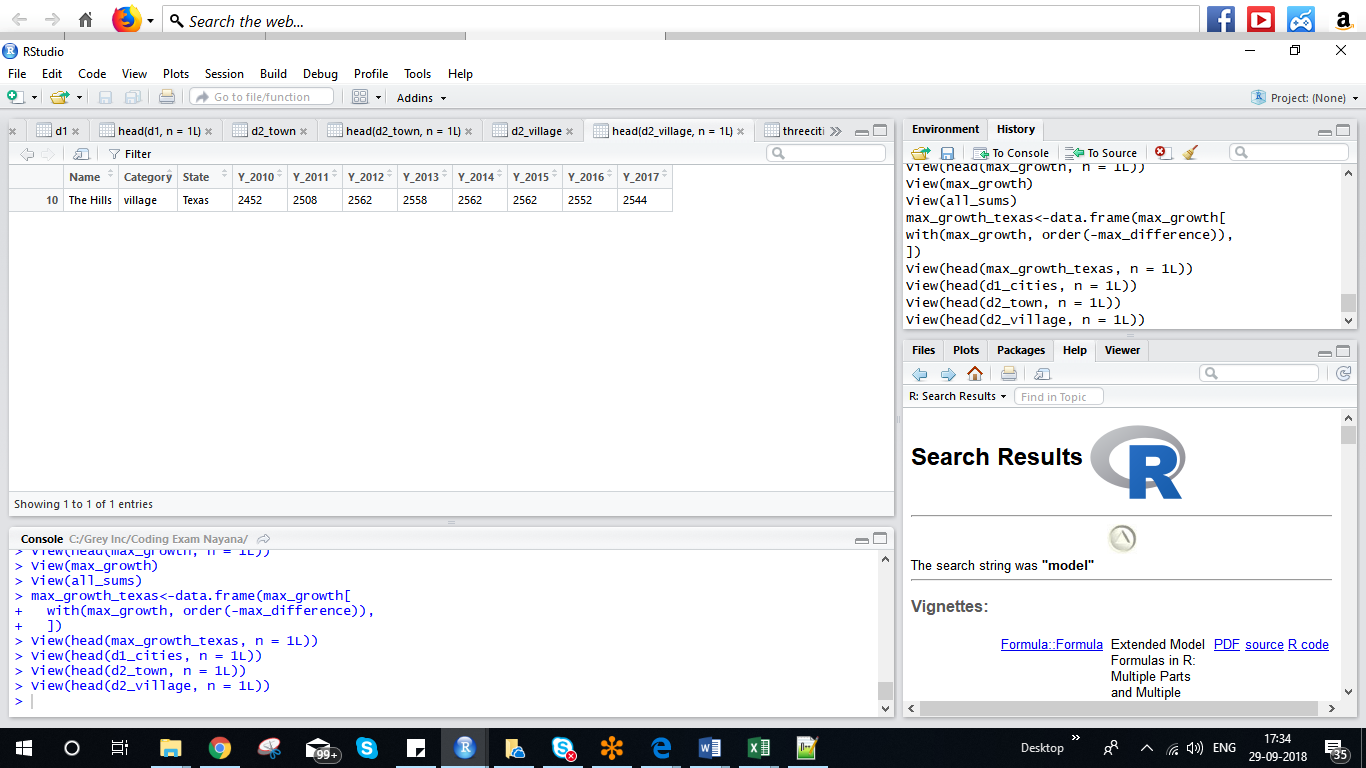
d2\_village<-data.frame(dataframe\_village[

with(dataframe\_village, order(-Y\_2011)),

])

View(d2\_village)

View(head(d2\_village, n = 1L))



**2)What are the three fastest growing cities in Texas between 2012-2016?**

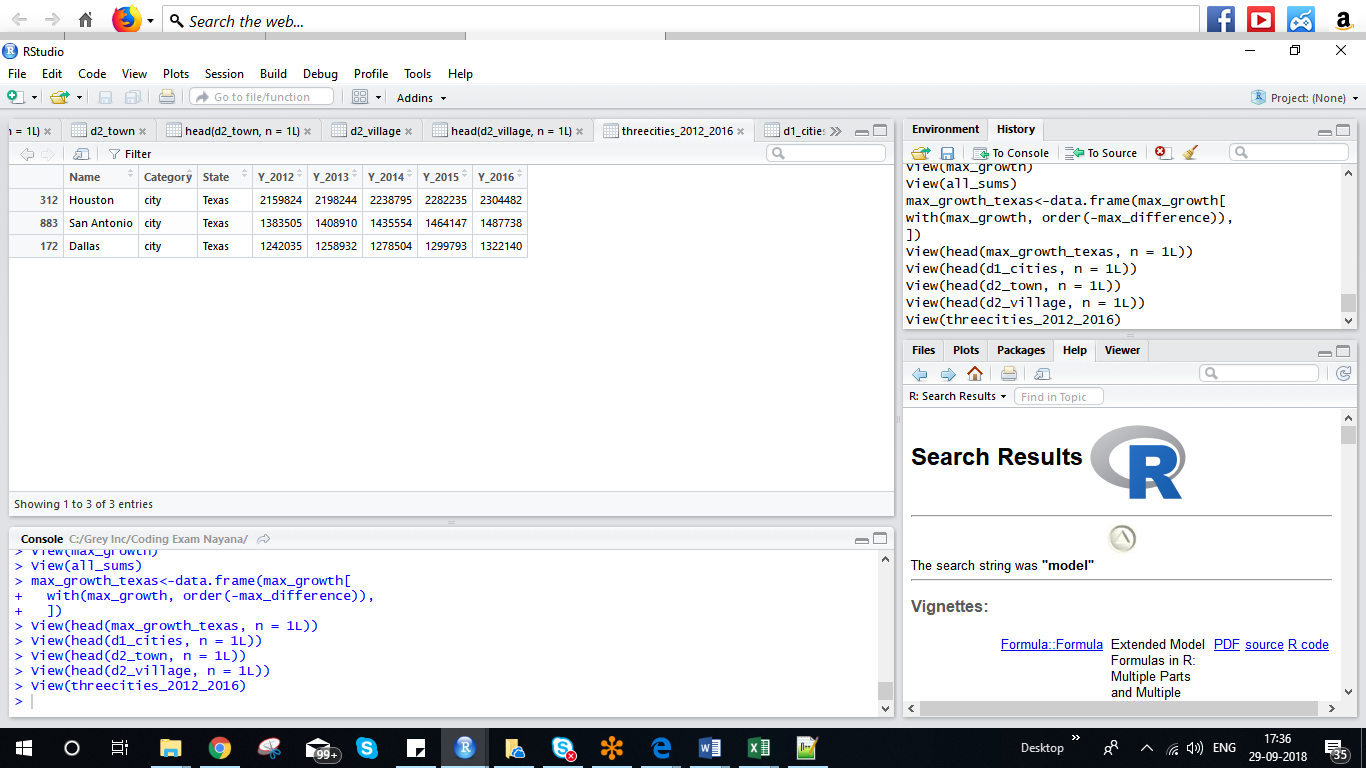
**ANS: Houston, San Antonio, Dallas**

threecities<-data.frame((head(d1\_cities, n = 3L)))

View(threecities)

threecities\_2012\_2016 <- data.frame(threecities[c(1:3,6:10)])

View(threecities\_2012\_2016)



**3) Which county had the most population in each year?**

Merge Data frames based on city as a unique field

Texas\_Cities\_County<- merge(dataframe\_cities, dataframe\_county, by = "Name")

Texas\_Cities\_County<-data.frame(Texas\_Cities\_County[

with(Texas\_Cities\_County, order(-Y\_2011)),

])

View(Texas\_Cities\_County)

View(head(Texas\_Cities\_County, n = 1L))

**5) Which county experienced the highest rate of growth from 2014-2016?**

Texas\_Cities\_County<-data.frame(Texas\_Cities\_County[

with(Texas\_Cities\_County, order(-Y\_2011)),

])

View(Texas\_Cities\_County)

df<-data.frame(head(Texas\_Cities\_County, n = 1L))

county\_2014\_2016 <- data.frame(df[c(1:3,8:10)])

**6) What percentage of Texans lived in towns in 2017?**

**Ans:1.8%**

sum\_2017\_cities<-sum(dataframe\_cities$Y\_2017)

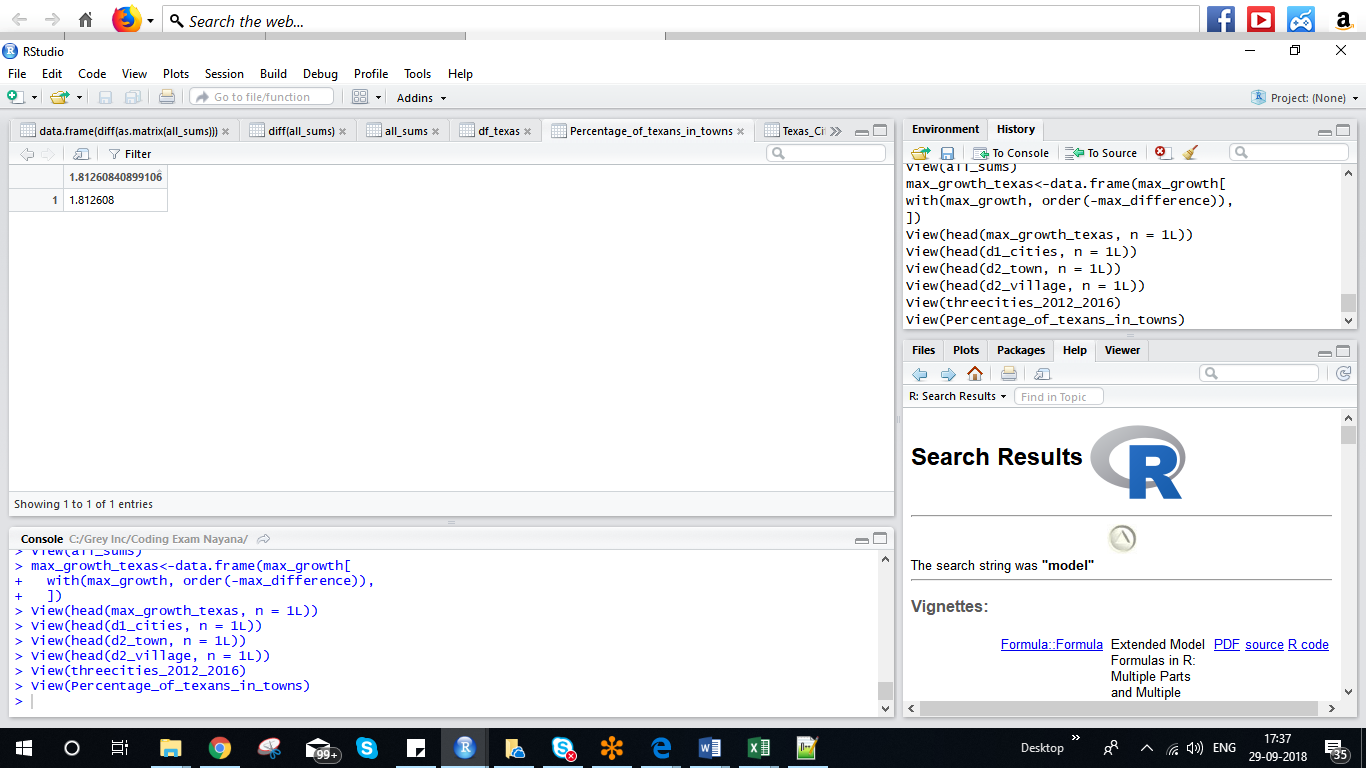
sum\_2017\_towns<-sum(dataframe\_town$Y\_2017)

sum\_2017\_villages<-sum(dataframe\_village$Y\_2017)

Total\_2017\_population<- sum\_2017\_cities+sum\_2017\_towns+sum\_2017\_villages

Percentage\_of\_texans\_in\_towns<-(sum\_2017\_towns/Total\_2017\_population)\*100

View(Percentage\_of\_texans\_in\_towns)



**7) Which year did Texas grow the most?**

**Ans:2015**

dataframe\_village <- rbind(dataframe8, dataframe11)

Texas\_total\_population<-rbind(dataframe\_cities,dataframe\_town, dataframe\_village )

df\_texas<-data.frame(Texas\_total\_population)

View(df\_texas)

all\_sums<-data.frame(colSums(Filter(is.numeric, df\_texas)))

View(all\_sums)

colnames(all\_sums)<-c("Total\_population")

View(apply( all\_sums , 2 , diff ))

all\_sums[ , diff := value - shift(value), by = Total\_population]

max\_growth<-data.frame(diff(all\_sums$Total\_population))

colnames(max\_growth)<-c("max\_difference")

View(max\_growth)

max\_growth\_texas<-data.frame(max\_growth[

with(max\_growth, order(-max\_difference)),

])

View(head(max\_growth\_texas, n = 1L))

