**Set 4 R Script and Analysis**

#1.)Import the library to read an excel file

library(readxl)

library(readxl)

#2.)Create a data frame and view the data frame

Set\_4 <- read\_excel("~/Desktop/Nebraska education /Set 4.xlsx")

View(Set\_4)

Graduation <- read\_excel("Set 4.xlsx")

View(Graduation)

#3.)To view all the fields, present in the dataset

names(Graduation)

#4.)To view all the unique values, present for the field “Graduation Count”

unique(Graduation$`Graduation Count`)

#5.)Create a data frame and view the data frame

Grad <- Graduation

View(Grad)

#6.)To replace all the values “-1” to NA and view the unique values

Grad[Grad == '-1'] <- NA

View(Grad)

unique(Grad$`Graduation Count`)

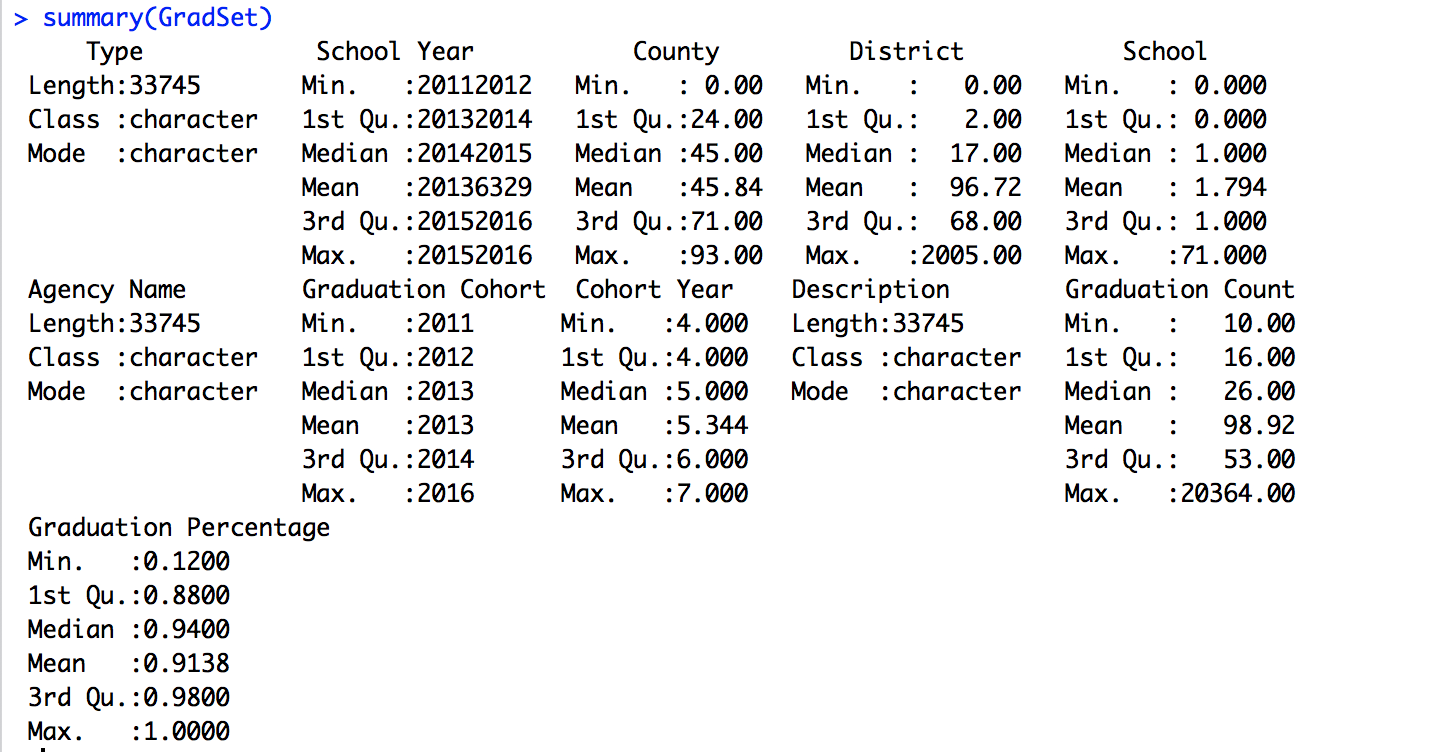
#7.) To omit all the “NA” values and create new data frame and view it.

GradSet <- na.omit(Grad)

View(GradSet)

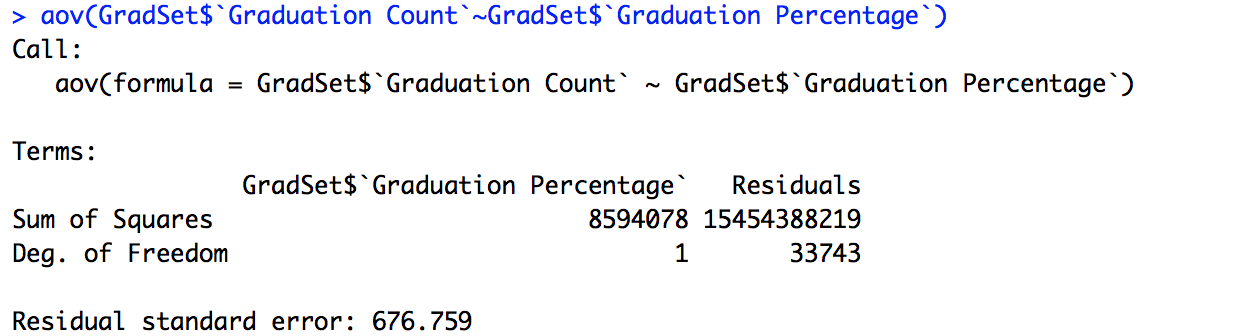
#8.)Ran the initial descriptive statistics

summary(GradSet)



#9.)Performed Anova on this set

aov(GradSet$`Graduation Count`~GradSet$`Graduation Percentage`)



#10.)To view the fields "Description", "Graduation Count", "Graduation Percentage”.

This helps to answer the research question.

GradSet[c("Description","Graduation Count","Graduation Percentage")]

#11.)Create a new data frame with only these fields.

GradCln <- GradSet[c("Description","Graduation Count","Graduation Percentage")]

View(GradCln)

#12.) To get the maximum and minimum

GradCln[GradCln$`Graduation Percentage`==max(GradCln$`Graduation Percentage`)| GradCln$`Graduation Percentage`==min(GradCln$`Graduation Percentage`),]

#13.) Create a new data frame and view.

mx <- GradCln[GradCln$`Graduation Percentage`==max(GradCln$`Graduation Percentage`)| GradCln$`Graduation Percentage`==min(GradCln$`Graduation Percentage`),]

View(mx)

#14.)Save the script

savehistory("~/Desktop/Set 4 R Script and analysis.Rhistory")