#display dataframeView(df)  
   
#size dim(df)  
   
#check datatypesstr(df)  
   
#check for missing valuescolSums([is.na](http://is.na/)(df))  
   
#drop null valuesdf <- na.omit(df) #it will delete all rows  
   
df <- read.csv("C:/Users/yusufs/Downloads/Data Set- Inc5000 Company List\_2014.csv")  
   
#drop null values method-2drops <- c("X\_input")  
   
names(df)  
   
df <- df[ ,!(names(df) %in% drops)]  
   
View(df)  
   
#drop null values method-3df <- read.csv("C:/Users/yusufs/Downloads/Data Set- Inc5000 Company List\_2014.csv")df <- df[ ,!(names(df) %in% c("X\_input"))]  
   
View(df)  
   
#check duplicatesdf <- df[!duplicated(df$id),]  
   
   
#round of to 2 decimal places#df$growth <- round(df$growth,digit=2)  
   
View(df)  
   
   
#checking outliersinstall.packages("tidyverse") #one time processlibrary(ggplot2)  
   
#drawing boxplot and labeling outliersggplot(df, aes(x=revenue, y=growth)) +   geom\_boxplot(outlier.colour = "red", outlier.shape = 1)+   scale\_x\_continuous(labels = scales::comma)+coord\_cartesian(ylim = c(0, 1000))  
   
#Calcualting IQR growthQ1\_growth <- quantile(df$growth,0.25)Q3\_growth <- quantile(df$growth,0.75)  
   
IQR\_growth <- Q3\_growth - Q1\_growth  
   
#using IQR functionIQR\_growth\_new <- IQR(df$growth) #calculate using IQR  
   
#Calculate IQR for revenue columnIQR\_revenue <- IQR(df$revenue)  
   
Q1\_revenue <- quantile(df$revenue,0.25)Q3\_revenue <- quantile(df$revenue,0.75)  
   
   
#removing outliersno\_outliers <- subset(df, df$growth> (Q1\_growth - 1.5\*IQR\_growth) & df$growth< (Q3\_growth + 1.5\*IQR\_growth))  
   
   
no\_outliers <- subset(no\_outliers, no\_outliers$revenue> (Q1\_revenue - 1.5\*IQR\_revenue) & no\_outliers$revenue< (Q3\_revenue + 1.5\*IQR\_revenue)