##Dealing with Missing Data

#Setting the working directory

getwd()

setwd("E://Advanced R")

getwd()

#Loading dataset into R

fin<-read.csv("Future 500.csv", na.strings = c(""))

fin

head(fin, 24)

str(fin)

fin$Expenses <- gsub(" Dollars", "", fin$Expenses)

head(fin)

fin$Expenses <-gsub(" Dollar","", fin$Expenses)

head(fin)

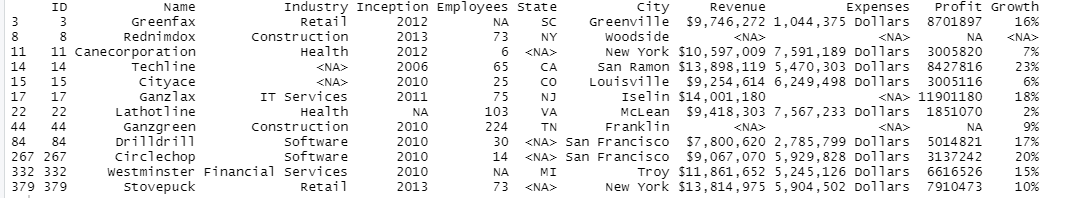
fin$Revenue <-gsub(",","", fin$Revenue)

fin

fin$Revenue<-gsub("\\$", "", fin$Revenue)

fin[!complete.cases(fin),]

**The code will generate a list of all the rows that has at least one NA in them**



**Now we will find out what value to put in row 332 for employees**

##Dealing with Missing Data

#Setting the working directory

getwd()

setwd("E://Advanced R")

getwd()

#Loading dataset into R

fin<-read.csv("Future 500.csv", na.strings = c(""))

fin

head(fin, 24)

str(fin)

fin$Expenses <- gsub(" Dollars", "", fin$Expenses)

head(fin)

fin$Expenses <-gsub(" Dollar","", fin$Expenses)

head(fin)

fin$Revenue <-gsub(",","", fin$Revenue)

fin

fin$Revenue<-gsub("\\$", "", fin$Revenue)

fin[!complete.cases(fin),]

median(fin[,"Employees"], na.rm =TRUE)

mean(fin[,"Employees"], na.rm =TRUE)

#See median in the retail industry, the below filter will subset iur data fir retail industry

median\_emp\_retail<-mean(fin[fin$Industry== "Retail","Employees"], na.rm =TRUE)

median\_emp\_retail

#Now we will place this value in row 332

fin[is.na(fin$Employees),]

#The code will give all rows with na in employee

fin[is.na(fin$Employees) & fin$Industry== "Financial Services", "Employees"] <- median\_emp\_retail

fin[332,]

#this will put the mean value in the row 332 for retail in employee and replaces with na

