Step 1: Import the CSV file into the workspace:

Commands:

data1 <- read.csv(file.choose(), header = TRUE)

data1

attach(data1)

Step 2: names(data1) //to know the attribute names of the dataset

Output:

> names(data1)

[1] "A1\_Score" "A2\_Score" "A3\_Score" "A4\_Score"

"A5\_Score" "A6\_Score"

[7] "A7\_Score" "A8\_Score" "A9\_Score" "A10\_Score"

"age" "gender"

[13] "ethnicity" "jundice" "austim" "contry\_of\_res"

"used\_app\_before" "result"

[19] "age\_desc" "relation" "Class.ASD"

Step 3: colSums(is.na(data1)) //to calculate the number of NAs/Missing values in every columns/ attributes.

After using the command, I’ve got that, there are 95 missing values(The highest number of missing values) both in relation column and ethnicity column.

Step 4 : table(data1$Class.ASD) //to calculate the number of yes & no in ASD(Autism Spectrum Disorder column)

Output:

NO YES

515 189

Step 5: prop.table(table(data1$Class.ASD))\*100 //To calculate the percentage

Output:

NO YES

73.15341 26.84659

Step 6: barplot(table(data1$Class.ASD),xlab="Class.ASD", ylab="No. of People", main="ASD") //visualization of ASD

Output:

