**מדעי הנתונים ובינה עסקית – מעבדה 2**

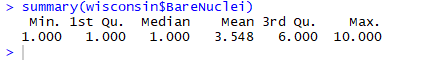
1. Commands:

wisconsinOG<-read.csv("C:\\Users\\Dan\\Desktop\\wisconsin.csv")

wisconsin<-na.omit(wisconsinOG)

summary(Wisconsin$BareNuclei)

Results:



2. Commands:

length(wisconsin$BareNuclei)

Results:

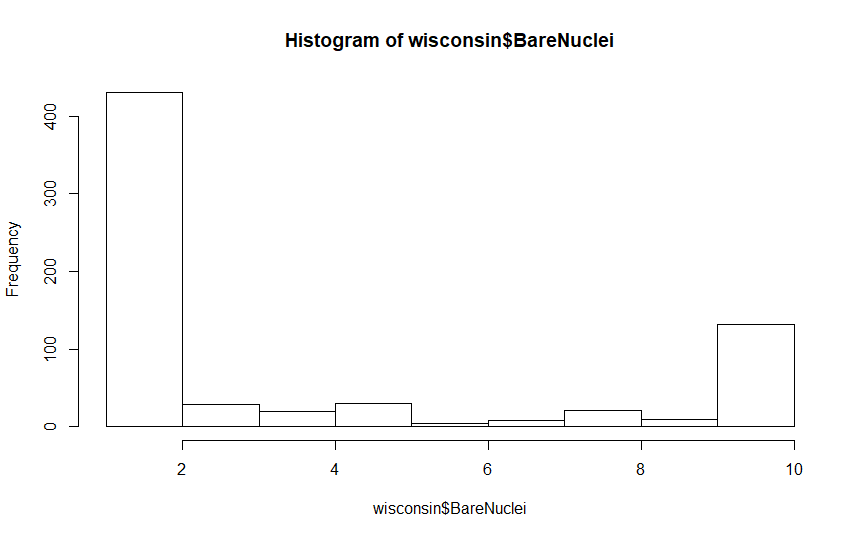
682

3.

3.1 Commands: (chosen column: BareNuclei)

hist(wisconsin$BareNuclei)

Results:



3.2 + 3.3 . נבצע Equal-Width discretization:

nb = wisconsin$BareNuclei

intervalWidth = (max(nb)-min(nb))/3

cutPoints = c(0,0)

i=1

while (i <= length(cutPoints)){

cutPoints[i] = min(nb) + i\*intervalWidth

i = i+1

}

i=1

for (x in nb){

if (x <= cutPoints[1]){

nb[i] = "LOW"

}else if (x > cutPoints[1] & x <= cutPoints[2]){

nb[i] = "MED"

}else{

nb[i] = "HIGH"

}

i = i+1

}

nb <- as.factor(nb)

wisconsin$BareNuclei = nb