Final Project Part B

# Biostat213 Summer 2023

First: Peijun

Last: Liu

## Read data

Final<-read.csv("/Users/petra/Desktop/study/UCSF/BIOSTAT 213/ObesityDataSet.csv")

## Summary

[1pt] Print the structure of ‘hf’:

str(Final)

'data.frame': 2111 obs. of 17 variables:  
 $ Gender : chr "Female" "Female" "Male" "Male" ...  
 $ Age : num 21 21 23 27 22 29 23 22 24 22 ...  
 $ Height : num 1.62 1.52 1.8 1.8 1.78 1.62 1.5 1.64 1.78 1.72 ...  
 $ Weight : num 64 56 77 87 89.8 53 55 53 64 68 ...  
 $ family\_history\_with\_overweight: chr "yes" "yes" "yes" "no" ...  
 $ FAVC : chr "no" "no" "no" "no" ...  
 $ FCVC : num 2 3 2 3 2 2 3 2 3 2 ...  
 $ NCP : num 3 3 3 3 1 3 3 3 3 3 ...  
 $ CAEC : chr "Sometimes" "Sometimes" "Sometimes" "Sometimes" ...  
 $ SMOKE : chr "no" "yes" "no" "no" ...  
 $ CH2O : num 2 3 2 2 2 2 2 2 2 2 ...  
 $ SCC : chr "no" "yes" "no" "no" ...  
 $ FAF : num 0 3 2 2 0 0 1 3 1 1 ...  
 $ TUE : num 1 0 1 0 0 0 0 0 1 1 ...  
 $ CALC : chr "no" "Sometimes" "Frequently" "Frequently" ...  
 $ MTRANS : chr "Public\_Transportation" "Public\_Transportation" "Public\_Transportation" "Walking" ...  
 $ NObeyesdad : chr "Normal\_Weight" "Normal\_Weight" "Normal\_Weight" "Overweight\_Level\_I" ...

## Identify up to 4 categorical and up to 4 continuous variables of your choice

sapply(Final, class)

Gender Age   
 "character" "numeric"   
 Height Weight   
 "numeric" "numeric"   
family\_history\_with\_overweight FAVC   
 "character" "character"   
 FCVC NCP   
 "numeric" "numeric"   
 CAEC SMOKE   
 "character" "character"   
 CH2O SCC   
 "numeric" "character"   
 FAF TUE   
 "numeric" "numeric"   
 CALC MTRANS   
 "character" "character"   
 NObeyesdad   
 "character"

#4 categoricalvariables: 'Gender','family\_history\_with\_overweight','SMOKE','NObeyesdad'  
#4 continuous variables: 'Age','Height','Weight','FAF'

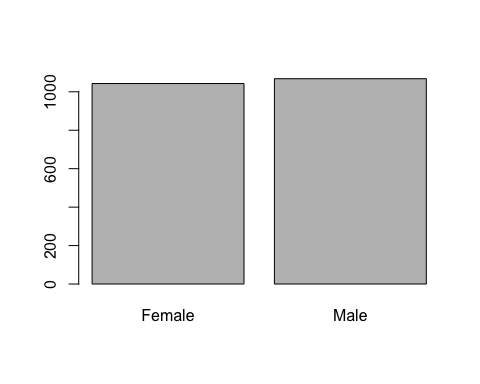
## For each categorical variable, print counts for each level

for (i4 in c('Gender','family\_history\_with\_overweight','SMOKE','NObeyesdad')){  
 print(table(Final[[i4]]))  
}

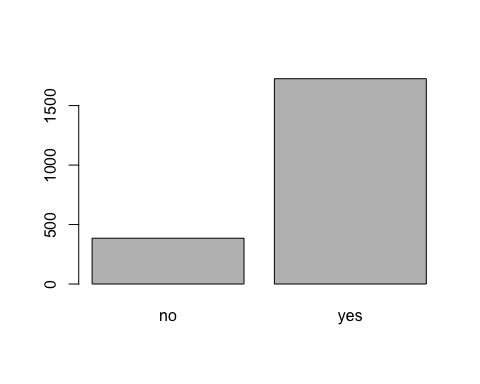
Female Male   
 1043 1068   
  
 no yes   
 385 1726   
  
 no yes   
2067 44   
  
Insufficient\_Weight Normal\_Weight Obesity\_Type\_I Obesity\_Type\_II   
 272 287 351 297   
 Obesity\_Type\_III Overweight\_Level\_I Overweight\_Level\_II   
 324 290 290

### For each categorical variable, make a barplot of counts

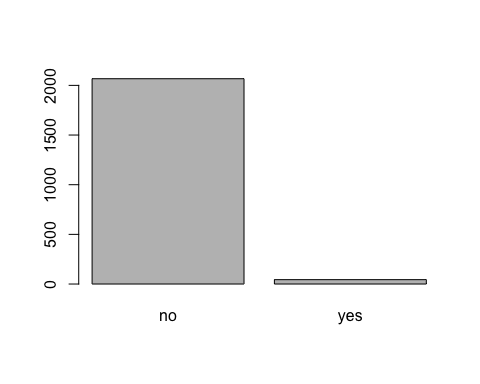
barplot(table(Final$Gender))



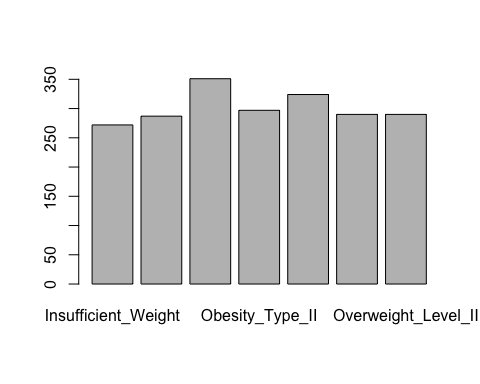
barplot(table(Final$family\_history\_with\_overweight))



barplot(table(Final$SMOKE))



barplot(table(Final$NObeyesdad))



### For each continuous variable, print mean, median, and standard deviation

for (i5 in c('Age','Height','Weight','FAF')){  
 cat(i5," : ","mean",mean(Final[[i5]])," ","median",median(Final[[i5]])," ","standard deviation",sd(Final[[i5]]),"\n")  
}

Age : mean 24.3126 median 22.77789 standard deviation 6.345968   
Height : mean 1.701677 median 1.700499 standard deviation 0.09330482   
Weight : mean 86.58606 median 83 standard deviation 26.19117   
FAF : mean 1.010298 median 1 standard deviation 0.8505924

## For each continuous variable, plot its distribution using a boxplot or a histogram (pick one type of plot and use it for all continuous variables)

par(mfrow = c(1, 4))  
hist(Final$Age, col = "#052049bb", border = "white", main = "")  
hist(Final$Height, col = "#052049bb", border = "white", main = "")  
hist(Final$Weight, col = "#052049bb", border = "white", main = "")  
hist(Final$FAF, col = "#052049bb", border = "white", main = "")

