setwd(<folder path>) # setting the desired folder as our working directory

# reading the data in R

train\_data=read.csv("Additional File 1.csv")

train\_data=train\_data[,-4]

# taking a look at the first few rows of the data to get a feel of the data

head(train\_data)

# adjusting the p-values using BY method , this code can be repeated with method="bonferroni"

pv\_train=train\_data$pvalue

# storing the new adjusted p values in pv\_adj\_train

pv\_adj\_train=round(p.adjust(pv\_train,method="BY"),9)

# here my objective is to create a new column in the data called sig\_adj which will declare significance by

# looking at the adjusted p-values, then we can compare sig\_adj with the column significance as a check

n1=nrow(train\_data)

sig\_adj=rep(".",n1)

# replacing the NAs in the adjusted p-values

pv\_adj\_train[is.na(pv\_adj\_train)] = 999

# determining significance by looking at the adjusted p-values

for(i in 1:n1)

{

if(pv\_adj\_train[i]<0.05){sig\_adj[i]="\*"}

}

# inserting this new column of significance into the dataframe

train\_data\_adj=cbind(train\_data,pv\_adj\_train,sig\_adj)

head(train\_data\_adj)

tail(train\_data\_adj)

# count the no. of significances

sum(sig\_adj=="\*")

# counting the no of relevant codes (significant and OR > 1)

train\_data\_relevant=subset(train\_data\_adj, ORest\_pre>1)

sum(train\_data\_relevant$sig\_adj=="\*")