theURL="https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data"

data=read.table(theURL, header=FALSE,sep=",")

dim(data)

View(data)

names(data)=c("age","workclass","fnlwgt","education","edunum","maritalstatus","occupation","relationship","race","sex","capitalgain","capitalloss","hoursperweek","nativecountry","salary")

names(data)

#handling the "?”

check<-(data==" ?")

answer<-which(check,arr.ind=TRUE)

delete<-unique(answer[,1])

data=data[-delete,]

age=data[,1];workclass=data[,2];fnlwgt=data[,3]

education=data[,4];edunum=data[,5];maritalstatus=data[,6]

occupation=data[,7];relationship=data[,8];race=data[,9]

sex=data[,10];capitalgain=data[,11];capitalloss=data[,12]

hoursperweek=data[,13];nativecountry=data[,14];salary=data[,15]

install.packages("caret")

require("caret")

data$age=as.character(data$age)

age=data$age

agdummy=dummyVars(~age,data)

agmatrix=predict(agdummy, newdata=data)

#check

#head(agmatrix)

#dim(agmatrix)

wcdummy=dummyVars(~workclass,data)

wcmatrix=predict(wcdummy, newdata=data)

#check

#head(wcmatrix)

#dim(wcmatrix)

#delete "?" although we delete all the ? but there's still a level of ?

wcmatrix1=wcmatrix[,-1]

edudummy=dummyVars(~education,data)

edumatrix=predict(edudummy, newdata=data)

#check

#head(edumatrix)

#dim(edumatrix)

data$edunum=as.character(data$edunum)

edunum=data$edunum

endummy=dummyVars(~edunum,data)

enmatrix=predict(endummy, newdata=data)

#check

#head(enmatrix)

#dim(enmatrix)

msdummy=dummyVars(~maritalstatus,data)

msmatrix=predict(msdummy, newdata=data)

#check

#head(msmatrix)

#dim(msmatrix)

ocdummy=dummyVars(~occupation,data)

ocmatrix=predict(ocdummy, newdata=data)

#check

#head(ocmatrix)

#dim(ocmatrix)

#delete "?"

ocmatrix1=ocmatrix[,-1]

rsdummy=dummyVars(~relationship,data)

rsmatrix=predict(rsdummy, newdata=data)

#check

#head(rsmatrix)

#dim(rsmatrix)

rcdummy=dummyVars(~race,data)

rcmatrix=predict(rcdummy, newdata=data)

#check

#head(rcmatrix)

#dim(rcmatrix)

sexdummy=dummyVars(~sex,data)

sexmatrix=predict(sexdummy, newdata=data)

#check

#head(sexmatrix)

#dim(sexmatrix)

data$capitalgain=as.character(data$capitalgain)

capitalgain=data$capitalgain

cgdummy=dummyVars(~capitalgain,data)

cgmatrix=predict(cgdummy, newdata=data)

#check

#head(cgmatrix)

#dim(cgmatrix)

data$capitalloss=as.character(data$capitalloss)

capitalloss=data$capitalloss

cldummy=dummyVars(~capitalloss,data)

clmatrix=predict(cldummy, newdata=data)

#check

#head(clmatrix)

#dim(clmatrix)

data$hoursperweek=as.character(data$hoursperweek)

hoursperweek=data$hoursperweek

hwdummy=dummyVars(~hoursperweek,data)

hwmatrix=predict(hwdummy, newdata=data)

#check

#head(hwmatrix)

#dim(hwmatrix)

ncdummy=dummyVars(~nativecountry,data)

ncmatrix=predict(ncdummy, newdata=data)

#check

#head(ncmatrix)

#dim(ncmatrix)

#delete "?"

ncmatrix1=ncmatrix[,-1]

sadummy=dummyVars(~salary,data)

samatrix=predict(sadummy, newdata=data)

#check

#head(samatrix)

#dim(samatrix)

sparse=cbind(agmatrix,wcmatrix1,edumatrix,enmatrix,msmatrix,ocmatrix1,rsmatrix,rcmatrix,sexmatrix,cgmatrix,clmatrix,hwmatrix,ncmatrix1,samatrix)

dim(sparse)

#outlierl

#mean var etc.

#plot and graph