### Part 2 ###

#data preparation

mat=read.table("student-mat.csv",sep=";",header=TRUE)

por=read.table("student-por.csv",sep=";",header=TRUE)

mat=mat[,-(31:32)]

por=por[,-(31:32)]

Xmat=model.matrix(G3~.,mat)[,-1]

Xpor=model.matrix(G3~.,por)[,-1]

Ymat=mat$G3

Ypor=por$G3

set.seed(1)

cvMat=cv.glmnet(Xmat,Ymat,alpha=1)

plot(cvMat)

#cvMat$lambda.min

coefMat = coef(cvMat, s = "lambda.min")

sqrt(min(cvMat$cvm))/mean(Ymat)

#cvMat$lambda.1se

coef(cvMat, s = "lambda.1se")

set.seed(1)

cvPor=cv.glmnet(Xpor,Ypor,alpha=1)

plot(cvPor)

#cvPor$lambda.min

coefPor = coef(cvPor, s = "lambda.min")

sqrt(min(cvPor$cvm))/mean(Ypor)

#cvPor$lambda.1se

coef(cvPor, s = "lambda.1se")

### Part 3 ###

#data preparation

bank=read.table("bank.csv",sep=";",header=TRUE)

Xbank=model.matrix(y~.,bank)[,-1]

Ybank=bank$y

#cross validation

set.seed(1)

cvBank = cv.glmnet(Xbank, Ybank, family = "binomial", type.measure = "class")

min(cvBank$cvm) #misclassification error = 0.09666003

bankPred = predict(cvBank, newx = Xbank, s = "lambda.min", type = "class")

#predict(cvBank, newx = Xbank, s = "lambda.min", type = "response")

table(bank$y,bankPred) #confusion matrix

### Part 1 ###

#data preparation

ewcs=read.table("EWCS\_2016.csv",sep=",",header=TRUE)

ewcs[,][ewcs[, ,] == -999] <- NA

kk=complete.cases(ewcs)

ewcs=ewcs[kk,]

#PCA

pcaEW = prcomp(ewcs,scale=T)

biplot(pcaEW,scale=0)