1. Load the Alzheimer's disease data using the commands:

library(AppliedPredictiveModeling)

data(AlzheimerDisease)

Which of the following commands will create non-overlapping training and test sets with about 50% of the observations assigned to each?

adData = data.frame(diagnosis,predictors)

trainIndex = createDataPartition(diagnosis, p = 0.50,list=FALSE)

training = adData[trainIndex,]

testing = adData[-trainIndex,]

1. Load the cement data using the commands:

library(AppliedPredictiveModeling)

data(concrete)

library(caret)

set.seed(1000)

inTrain = createDataPartition(mixtures$CompressiveStrength, p = 3/4)[[1]]

training = mixtures[ inTrain,]

testing = mixtures[-inTrain,]

Make a plot of the outcome (CompressiveStrength) versus the index of the samples. Color by each of the variables in the data set (you may find the cut2() function in the Hmisc package useful for turning continuous covariates into factors). What do you notice in these plots?