Question 9

a) I set the seed to 1 and then split the college dataset 4/5 train and 1/5 test

```r
# a)
rand = sample(c(TRUE,TRUE,TRUE,TRUE,FALSE), nrow(College), replace=TRUE)
train = College[rand, ]
test = College[!rand, ]
```

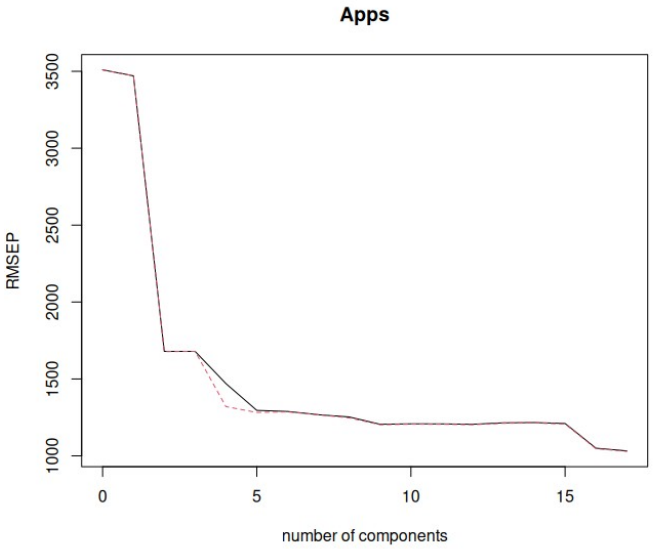
b) I created a linear model on the training data, and then ran a prediction on the test data. I got a test error of 1804734.

```r
# b)
lm = lm(Apps ~ ., data = train)

pred_lm = predict(lm, test, type = "response")
error_lm = mean((test$Apps - pred_lm)^2)
```

e) I then created a PCR model with cross validation to choose the M value. Here is the code and analytic plot of RMSEP for the number of components. The best number of components to use is 5 according to a rough visual estimate of the one standard error rule. The test error for 5 components is 5362426

```r
# c)
pcr_model <- pcr(Apps ~ ., data = train, scale = TRUE, validation = "CV")
print(summary(pcr_model))
validationplot(pcr_model)
pred_pcr = predict(pcr_model, test, ncomp = 10)
error_pcr = mean((test$Apps - pred_pcr)^2)
```



f) I repeated this process for PLS. Here is the diagnostic plot for best M by RMSEP. The best choice for M is 6, with a test error of 1991644. Plot shown below