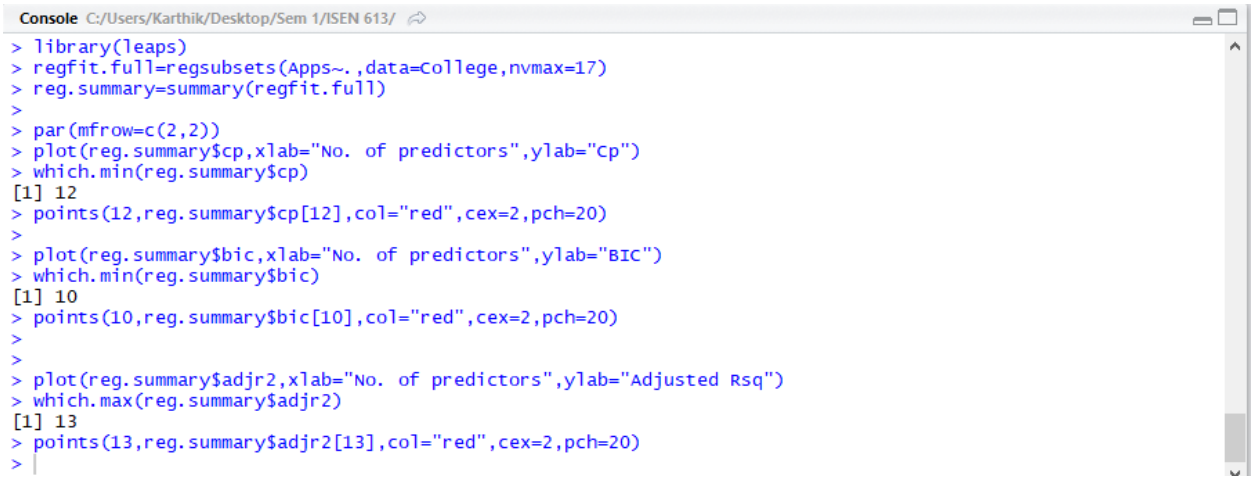
1)

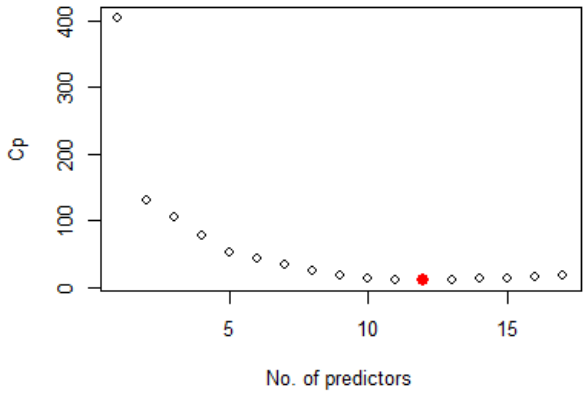
a)

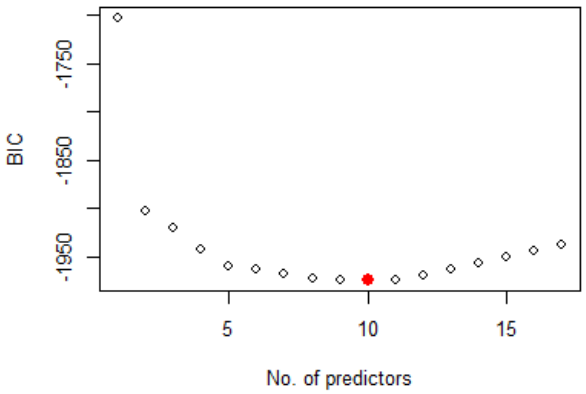


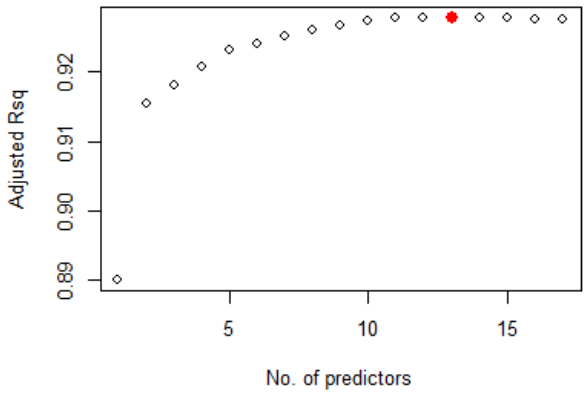
Cp -12 predictor model

BIC – 10 predictor model

AdjR2 – 13 predictor model







Coefficient estimates For BIC:

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> coef(regfit.full,10)
  (Intercept)    PrivateYes        Accept        Enroll     Top10perc     Top25perc       Outstate
-100.51668243 -575.07060789    1.58421887    -0.56220848   49.13908916  -13.86531103    -0.09466457
   Room.Board           PhD        Expend     Grad.Rate
     0.16373674  -10.01608705    0.07273776    7.33268904
>
```

Coefficient estimates For Cp :

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> coef(regfit.full,12)
  (Intercept)    PrivateYes        Accept        Enroll     Top10perc     Top25perc   F.Undergrad   P.Undergrad
-157.28685883 -511.78760196    1.58691470    -0.88265385   50.41131660  -14.74735373     0.05945481     0.04593068
     Outstate    Room.Board           PhD        Expend     Grad.Rate
    -0.09017643     0.14776586  -10.70502848    0.07246655    8.63961002
>
```

Coefficient estimates For AdjR2:

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> coef(regfit.full,13)
  (Intercept)    PrivateYes        Accept        Enroll     Top10perc     Top25perc   F.Undergrad   P.Undergrad
-440.74148270 -484.77261885    1.58542302   -0.87824288   50.41461998  -14.63667155    0.05762769    0.04642270
     Outstate    Room.Board           PhD     S.F.Ratio        Expend     Grad.Rate
   -0.08823311    0.14696204  -10.91804823   15.15475056    0.07786425    8.58578735
>
```

b)

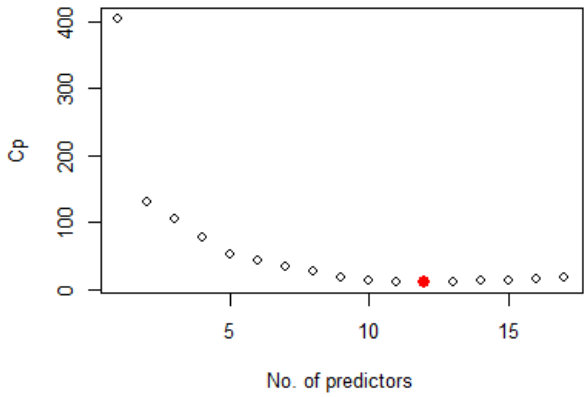
FORWARD:

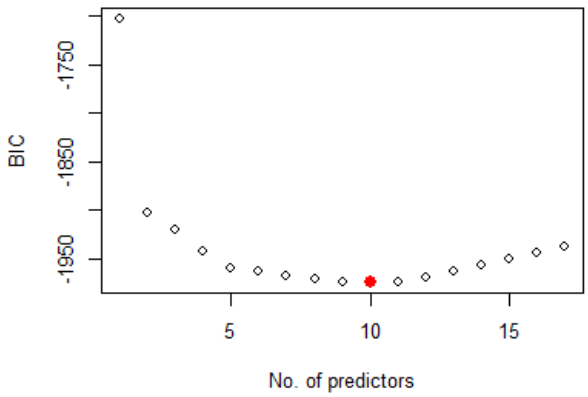
```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> regfit.fwd=regsubsets(Apps~.,data=College,nvmax=17,method="forward")
> reg.summary2=summary(regfit.fwd)
>
> par(mfrow=c(2,2))
> plot(reg.summary2$cp,xlab="No. of predictors",ylab="Cp")
> which.min(reg.summary2$cp)
[1] 12
> points(12,reg.summary2$cp[12],col="red",cex=2,pch=20)
>
>
> plot(reg.summary2$bic,xlab="No. of predictors",ylab="BIC")
> which.min(reg.summary2$bic)
[1] 10
> points(10,reg.summary2$bic[10],col="red",cex=2,pch=20)
>
>
> plot(reg.summary2$adjr2,xlab="No. of predictors",ylab="Adjusted Rsq")
> which.max(reg.summary2$adjr2)
[1] 13
> points(13,reg.summary2$adjr2[13],col="red",cex=2,pch=20)
>
```

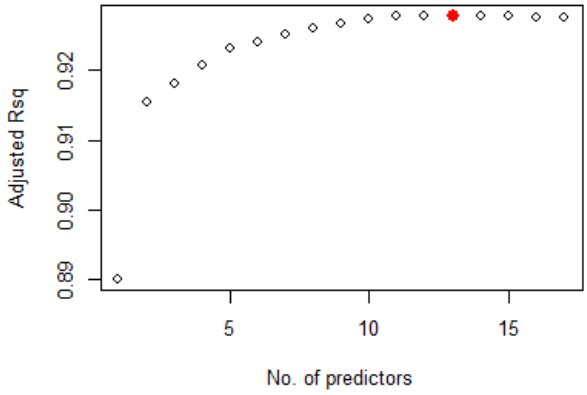
Cp -12 predictor model

BIC – 10 predictor model

AdjR2 – 13 predictor model







Coefficient estimates For BIC:

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> coef(regfit.full,10)
  (Intercept)    PrivateYes        Accept        Enroll     Top10perc     Top25perc      Outstate
-100.51668243 -575.07060789    1.58421887   -0.56220848   49.13908916  -13.86531103   -0.09466457
   Room.Board           PhD        Expend     Grad.Rate
    0.16373674  -10.01608705    0.07273776    7.33268904
>
```

Coefficient estimates For Cp :

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> coef(regfit.full,12)
  (Intercept)    PrivateYes        Accept        Enroll     Top10perc     Top25perc   F.Undergrad   P.Undergrad
-157.28685883 -511.78760196    1.58691470   -0.88265385   50.41131660  -14.74735373    0.05945481    0.04593068
     Outstate    Room.Board           PhD        Expend     Grad.Rate
   -0.09017643    0.14776586  -10.70502848    0.07246655    8.63961002
>
```

Coefficient estimates For AdjR2:

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> coef(regfit.full,13)
  (Intercept)    PrivateYes        Accept        Enroll     Top10perc     Top25perc   F.Undergrad   P.Undergrad
-440.74148270 -484.77261885    1.58542302   -0.87824288   50.41461998  -14.63667155    0.05762769    0.04642270
     Outstate    Room.Board           PhD     S.F.Ratio        Expend     Grad.Rate
   -0.08823311    0.14696204  -10.91804823   15.15475056    0.07786425    8.58578735
>
```

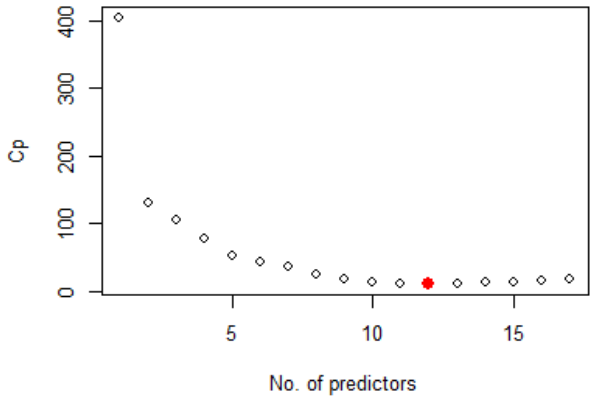
BACKWARD:

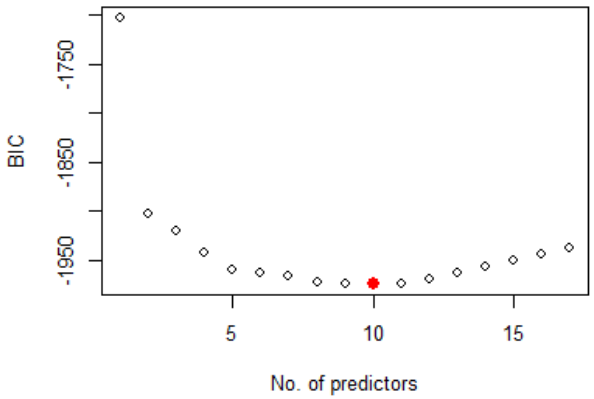
```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> regfit.bwd=regsubsets(Apps~.,data=College,nvmax=17,method="backward")
> reg.summary2=summary(regfit.bwd)
>
> par(mfrow=c(2,2))
> plot(reg.summary2$cp,xlab="No. of predictors",ylab="Cp")
> which.min(reg.summary2$cp)
[1] 12
> points(12,reg.summary2$cp[12],col="red",cex=2,pch=20)
>
>
> plot(reg.summary2$bic,xlab="No. of predictors",ylab="BIC")
> which.min(reg.summary2$bic)
[1] 10
> points(10,reg.summary2$bic[10],col="red",cex=2,pch=20)
>
>
> plot(reg.summary2$adjr2,xlab="No. of predictors",ylab="Adjusted Rsq")
> which.max(reg.summary2$adjr2)
[1] 13
> points(13,reg.summary2$adjr2[13],col="red",cex=2,pch=20)
>
```

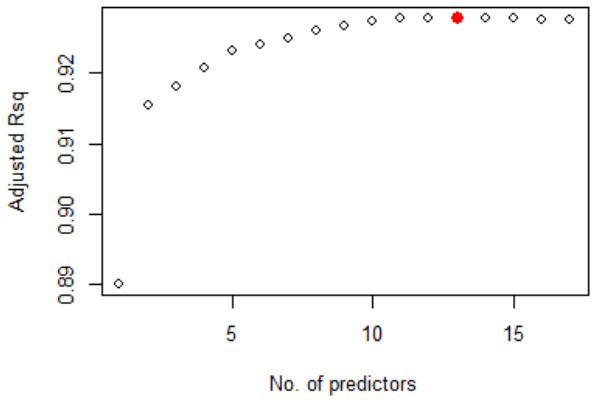
Cp -12 predictor model

BIC – 10 predictor model

AdjR2 – 13 predictor model







Coefficient estimates For BIC:

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> coef(regfit.full,10)
  (Intercept)    PrivateYes        Accept        Enroll     Top10perc     Top25perc      Outstate
-100.51668243 -575.07060789    1.58421887   -0.56220848   49.13908916  -13.86531103   -0.09466457
   Room.Board           PhD        Expend     Grad.Rate
    0.16373674  -10.01608705    0.07273776    7.33268904
>
```

Coefficient estimates For Cp :

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> coef(regfit.full,12)
  (Intercept)    PrivateYes        Accept        Enroll     Top10perc     Top25perc   F.Undergrad   P.Undergrad
-157.28685883 -511.78760196    1.58691470   -0.88265385   50.41131660  -14.74735373    0.05945481    0.04593068
     Outstate    Room.Board           PhD        Expend     Grad.Rate
   -0.09017643    0.14776586  -10.70502848    0.07246655    8.63961002
>
```

Coefficient estimates For AdjR2:

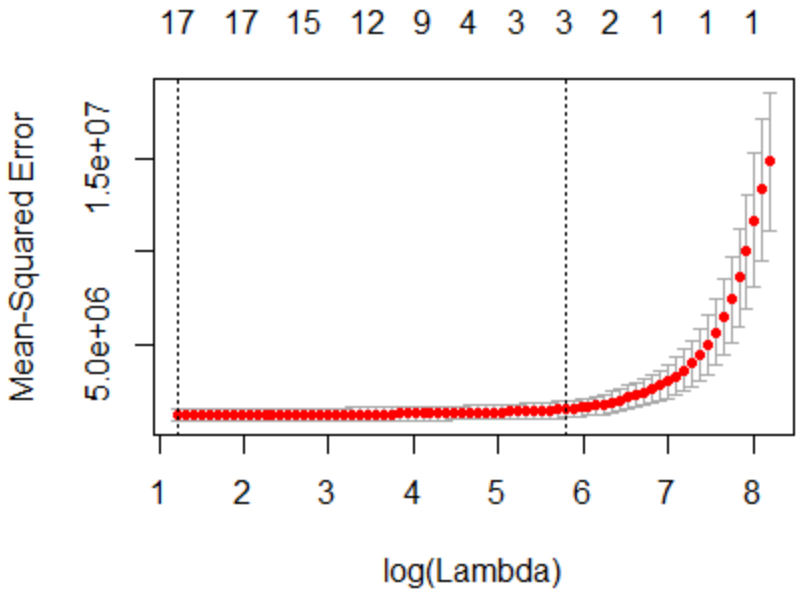
```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> coef(regfit.full,13)
  (Intercept)    PrivateYes        Accept        Enroll     Top10perc     Top25perc   F.Undergrad   P.Undergrad
-440.74148270 -484.77261885    1.58542302   -0.87824288   50.41461998  -14.63667155    0.05762769    0.04642270
     Outstate    Room.Board           PhD     S.F.Ratio        Expend     Grad.Rate
  -0.08823311    0.14696204  -10.91804823   15.15475056    0.07786425    8.58578735
>
```

The results are the same for both best sub set selection and the forward and backward stepwise selection methods.

c)

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> x=model.matrix(Apps~.,data=College)[,-1]
> y=College$Apps
> grid=10^seq(10,-2,length=100)
> lasso.mod=glmnet(x,y,alpha=1,lambda=grid)
> dim(coef(lasso.mod))
[1]  18 100
> set.seed(1)
> cv.out=cv.glmnet(x,y,alpha=1)
>
> plot(cv.out)
> bestlam=cv.out$lambda.min
> bestlam
[1] 3.403063
> out=glmnet(x,y,alpha=1)
> lasso.coeff=predict(out,type="coefficients",s=bestlam)[1:18,]
> lasso.coeff
  (Intercept)    PrivateYes        Accept        Enroll     Top10perc     Top25perc   F.Undergrad
-481.69122766 -489.47698922    1.56285991   -0.69952897   47.20524294  -12.12210806    0.03356097
  P.Undergrad      Outstate    Room.Board         Books      Personal           PhD      Terminal
    0.04415215   -0.08184648    0.14813763    0.01201765    0.02785918   -8.24433269   -3.21033519
     S.F.Ratio   perc.alumni        Expend     Grad.Rate
  14.04536901   -0.13535398    0.07662786    8.06878113
> lasso.coeff[lasso.coeff!=0]
  (Intercept)    PrivateYes        Accept        Enroll     Top10perc     Top25perc   F.Undergrad
-481.69122766 -489.47698922    1.56285991   -0.69952897   47.20524294  -12.12210806    0.03356097
  P.Undergrad      Outstate    Room.Board         Books      Personal           PhD      Terminal
    0.04415215   -0.08184648    0.14813763    0.01201765    0.02785918   -8.24433269   -3.21033519
     S.F.Ratio   perc.alumni        Expend     Grad.Rate
  14.04536901   -0.13535398    0.07662786    8.06878113
>
```

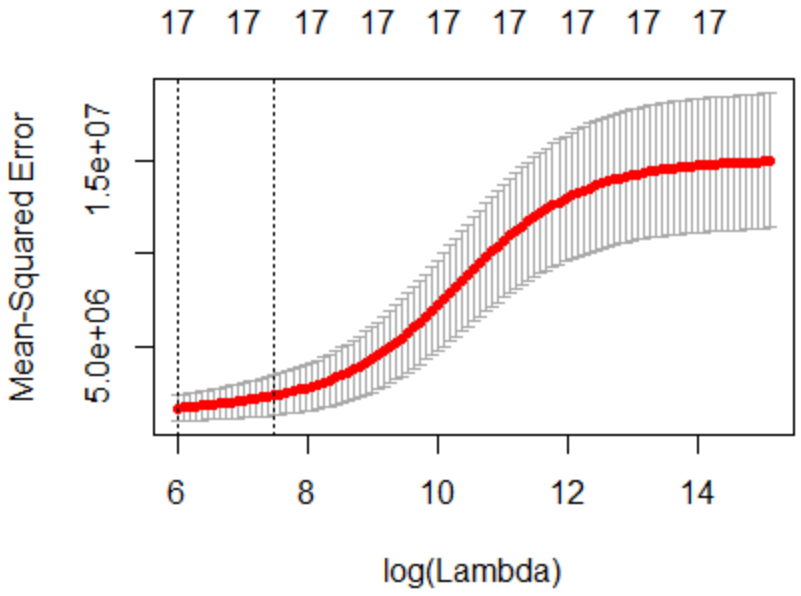
Best lambda: 3.403



4)

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> x=model.matrix(Apps~.,College)[,-1]
> y=College$Apps
> grid=10^seq(10,-2,length=100)
> lasso.mod=glmnet(x,y,alpha=0,lambda=grid)
> dim(coef(lasso.mod))
[1]  18 100
> set.seed(1)
> cv.out=cv.glmnet(x,y,alpha=0)
> plot(cv.out)
> bestlam=cv.out$lambda.min
> bestlam
[1] 400.4766
> out=glmnet(x,y,alpha=0)
> predict(out,type="coefficients",s=bestlam)[1:18,]
  (Intercept)    PrivateYes        Accept        Enroll     Top10perc     Top25perc   F.Undergrad
-1.514927e+03 -5.293325e+02  9.780751e-01  4.666917e-01  2.497314e+01  1.056473e+00  7.662859e-02
  P.Undergrad      Outstate    Room.Board         Books      Personal           PhD      Terminal
 2.445939e-02 -2.136542e-02  1.997980e-01  1.352799e-01 -8.966624e-03 -3.771159e+00 -4.713593e+00
    S.F.Ratio   perc.alumni        Expend     Grad.Rate
 1.282837e+01 -8.831661e+00  7.527598e-02  1.136663e+01
>
```

Best lambda: 400.4766



e-i)

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> set.seed(1)
> train=sample(777,388)
> regfit.full=regsubsets(Apps~.,data=College[train,],nvmax=17)
> reg.summary=summary(regfit.full)
>
```

If Cp is used:

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> which.min(reg.summary$cp)
[1] 10
> coef(regfit.full,id=10)
  (Intercept)    PrivateYes        Accept        Enroll     Top10perc     Top25perc      Outstate    Room.Board
  84.95670099 -691.04103152    1.67873705   -0.86164941   66.92631417  -22.35416377   -0.09482472    0.24520032
          PhD        Expend     Grad.Rate
 -10.14399113    0.03783190    6.45828153
>
> lm_bestfit=lm(Apps~Private+Accept+Enroll+Top10perc+Top25perc+Outstate+Room.Board+PhD+Expend+Grad.Rate,data=College,s
ubset=train)
> lm_pred=predict(lm_bestfit,College[-train,])
> test_error_cp=mean((Apps[-train]-lm_pred)^2)
> test_error_cp
[1] 1078371
>
```

If BIC is used (Chosen in the first question as the best model):

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> which.min(reg.summary$bic)
[1] 7
> coef(regfit.full,id=7)
  (Intercept)    PrivateYes        Accept        Enroll     Top10perc     Top25perc      Outstate    Room.Board
-115.30390947 -555.03473306    1.68351720   -0.87949983   73.84825907  -25.95810534   -0.07981525    0.24821390
>
> lm_bestfit=lm(Apps~Private+Accept+Enroll+Top10perc+Top25perc+Outstate+Room.Board,data=College,subset=train)
> lm_pred=predict(lm_bestfit,College[-train,])
>
> test_error_bic=mean((Apps[-train]-lm_pred)^2)
> test_error_bic
[1] 1165374
>
```

If AdjR2 is used:

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> which.max(reg.summary$adjr2)
[1] 10
> coef(regfit.full,id=10)
  (Intercept)    PrivateYes        Accept        Enroll     Top10perc     Top25perc      Outstate    Room.Board
  84.95670099 -691.04103152    1.67873705   -0.86164941   66.92631417  -22.35416377   -0.09482472    0.24520032
          PhD        Expend     Grad.Rate
 -10.14399113    0.03783190    6.45828153
>
> lm_bestfit=lm(Apps~Private+Accept+Enroll+Top10perc+Top25perc+Outstate+Room.Board+PhD+Expend+Grad.Rate,data=College,s
ubset=train)
> lm_pred=predict(lm_bestfit,College[-train,])
>
> test_error_adjr2=mean((Apps[-train]-lm_pred)^2)
> test_error_adjr2
[1] 1078371
>
```

e - ii)

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> x=model.matrix(Apps~.,data=College)[,-1]
> y=College$Apps
>
>
> set.seed(1)
> train=sample(1:nrow(x),nrow(x)/2)
> test=(-train)
> y.test=y[test]
> set.seed(1)
> cv.out=cv.glmnet(x[train,],y[train],alpha=1)
> bestlam=cv.out$lambda.min
> bestlam
[1] 24.62086
> lasso.mod=glmnet(x[train,],y[train],alpha=1,lambda=bestlam)
> lasso.pred=predict(lasso.mod,s=bestlam,newx=x[test,])
> mean((lasso.pred-y.test)^2)
[1] 1034786
>
```

e-iii)

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> x=model.matrix(Apps~.,data=College)[,-1]
> y=College$Apps
>
>
> set.seed(1)
> train=sample(1:nrow(x),nrow(x)/2)
> test=(-train)
> y.test=y[test]
> set.seed(1)
> cv.out=cv.glmnet(x[train,],y[train],alpha=0)
> bestlam=cv.out$lambda.min
> bestlam
[1] 450.7435
> ridge.mod=glmnet(x[train,],y[train],alpha=0,lambda=bestlam)
> ridge.pred=predict(ridge.mod,s=bestlam,newx=x[test,])
> mean((ridge.pred-y.test)^2)
[1] 1038427
>
```

e-iv)

Model	Test Error
Best Subset Selection (Cp)	1078371
Best Subset Selection (BIC) (CHOSEN as best among best subset selection in Q1)	1165374
Best Subset Selection (AdjR2)	1078371
Lasso Regression	**1034786**
Ridge Regression	1038427

Lasso Coefficient estimates:

```
Console C:/Users/Karthik/Desktop/Sem 1/ISEN 613/
> lasso.coeff=predict(lasso.mod,type="coefficients",s=bestlam)[1:18,]
> lasso.coeff
  (Intercept)    PrivateYes        Accept        Enroll     Top10perc     Top25perc   F.Undergrad   P.Undergrad
-3.674800e+02 -5.305394e+02  1.556370e+00 -4.041862e-01  5.032803e+01 -9.825451e+00 -1.798594e-02  0.000000e+00
     Outstate    Room.Board         Books      Personal           PhD      Terminal     S.F.Ratio   perc.alumni
-5.752475e-02  1.963418e-01  1.927388e-02  3.556499e-03 -4.740559e+00 -2.924054e+00  0.000000e+00 -2.264210e+00
       Expend     Grad.Rate
 3.257759e-02  3.412244e+00
>
```

Lasso regression seems to give the lowest test error and hence is the optimal model.