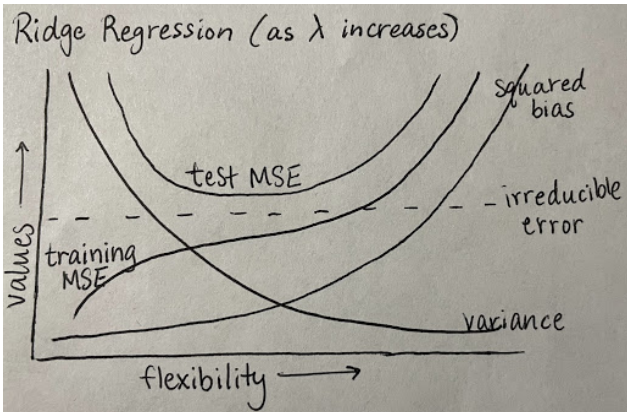
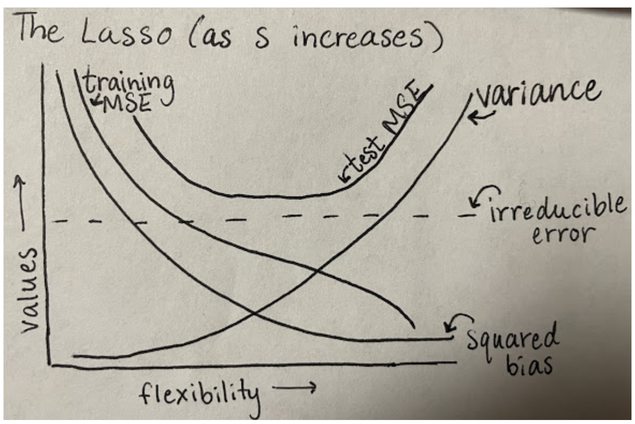
Neha Maddali

Problem 1:

a. When λ = 0, the bias of βˆ lasso is small but there will be a higher variance
b. When λ = ∞ the variance of βˆ lasso is small but the bias is high
c. 

d. 

Problem 2:

a. See RScript
b. B0 = 3, B1 = 2, B2 = -3, B3 = 0.3, see RScript for response vector Y
c. There are 10 models of size 9

	p	adjr2	AIC	BIC
1	2	0.5848648	198.3050078	203.515348
2	3	0.9329541	16.9567012	24.772212
3	4	0.9479154	-7.3303663	3.090314
4	5	0.9476882	-5.9422318	7.083619
5	6	0.9472503	-4.1668054	11.464216
6	7	0.9468046	-2.3950978	15.841094
7	8	0.9466350	-1.1578797	19.683482
8	9	0.9460658	0.8102997	24.256831
9	10	0.9454749	2.7948688	28.846571
10	11	0.9449490	4.6374815	33.294354

Model 3 is the best model based on the RSS, AIC and BIC criteria. When the models are of the same size, AIC/BIC will lead you to pick the same model because AIC/BIC makes each predictor pay a price for being in the model. In this problem, all the models have the same size so they have the same penalties and will choose the same best model.

Moreover, because the penalties are the same, only the RSS matters in the AIC/BIC formulas.

d. BIC and Adjusted R^2: model 3

BIC: model 3

```
(Intercept) poly(x, 10, raw = TRUE)1 poly(x, 10, raw = TRUE)2 poly(x, 10, raw = TRUE)7
 3.07627412               2.35623596              -3.16514887               0.01046843
```

e. Forward BIC and Adjusted R^2: model3

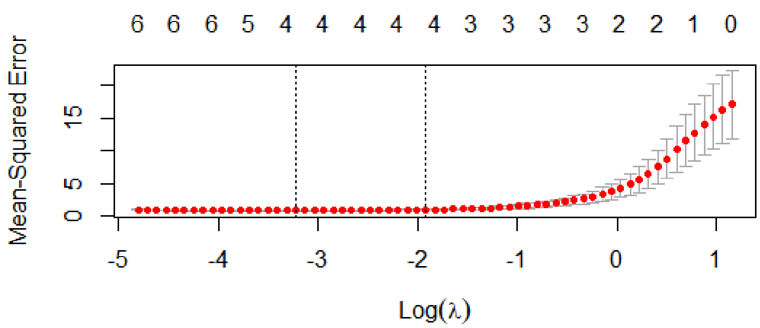
```
(Intercept) poly(x, 10, raw = TRUE)1 poly(x, 10, raw = TRUE)2 poly(x, 10, raw = TRUE)7
 3.07627412               2.35623596              -3.16514887               0.01046843
```

Backward BIC and Adjusted R^2: model 3

```
(Intercept) poly(x, 10, raw = TRUE)1 poly(x, 10, raw = TRUE)2 poly(x, 10, raw = TRUE)7
 3.07627412               2.35623596              -3.16514887               0.01046843
```

The models picked by forward and backward for BIC and adjusted R^2 are the same model as in part d: model 3

f. Optimal lambda = 0.03991416




```
                                   s0
(Intercept)                 3.040337084
poly(x, 10, raw = TRUE)1    2.250503962
poly(x, 10, raw = TRUE)2   -3.105904642
poly(x, 10, raw = TRUE)3    .
poly(x, 10, raw = TRUE)4    .
poly(x, 10, raw = TRUE)5    0.040046945
poly(x, 10, raw = TRUE)6    .
poly(x, 10, raw = TRUE)7    0.002560369
poly(x, 10, raw = TRUE)8    .
poly(x, 10, raw = TRUE)9    .
poly(x, 10, raw = TRUE)10   .
```

The lasso method brough X4-X10 to either 0 or relatively close to 0. This makes sense as the true model only includes B0, B1, B2 and B3

g. B7 = 8, B0 = 3

Best subset selection BIC: model 1

```
(Intercept) poly(x, 10, raw = TRUE)7
    2.95894                  8.00077
```

Best subset selection Adjusted R^2: model 4

```
             (Intercept) poly(x, 10, raw = TRUE)1 poly(x, 10, raw = TRUE)2
               3.0762524                0.2914016               -0.1617671
poly(x, 10, raw = TRUE)3 poly(x, 10, raw = TRUE)7
              -0.2526527                8.0091338
```

From best subset selection, Model 7 was the best model for both criteria of BIC and Adjusted R^2. From the lasso, the best lambda is 6.061588. The plot also didn't have a similar curve as from part f. The coefficients predicted were the same however.

```
                                   s0
(Intercept)              3.947394330
poly(x, 10, raw = TRUE)1  .
poly(x, 10, raw = TRUE)2  .
poly(x, 10, raw = TRUE)3  .
poly(x, 10, raw = TRUE)4  .
poly(x, 10, raw = TRUE)5  .
poly(x, 10, raw = TRUE)6  .
poly(x, 10, raw = TRUE)7  7.750481484
poly(x, 10, raw = TRUE)8  .
poly(x, 10, raw = TRUE)9  0.002855533
poly(x, 10, raw = TRUE)10 .
```

In this problem, best subset selection (BIC) gave the best model since we know the true model uses B0 and B7. The other coefficients in adjusted R^2's model that aren't B7 are very close to 0. The lasso was interesting because not all the predictors were zero which could be because of the true model.

Problem 3:

a. See RScript

b.
```
 (Intercept)        AtBat         Hits        HmRun
 275.7770532   -0.4167917   -1.3213883    6.1168274
        Runs          RBI        Walks        Years
   1.4110950    0.9770503    3.8167417  -17.2703971
      CAtBat        CHits       CHmRun        CRuns
  -0.6058889    2.9999826    3.1742113   -0.9256397
        CRBI       CWalks      LeagueN    DivisionW
  -0.5020963    0.3179721  116.3375939 -145.0774039
     PutOuts      Assists       Errors   NewLeagueN
   0.1958674    0.6714270   -4.6552691  -67.8057776
```

Lambda = 0.013

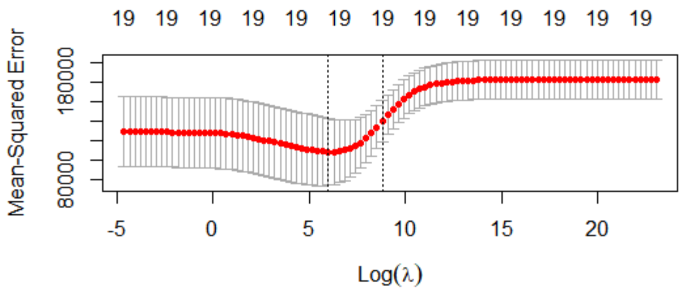
c.
```
  (Intercept)         AtBat          Hits         HmRun
 5.318652e+02  4.277932e-08  1.440236e-07  8.287785e-07
         Runs           RBI         Walks         Years
 2.883493e-07  3.011837e-07  4.089362e-07  1.612558e-06
       CAtBat         CHits        CHmRun         CRuns
 5.838270e-09  2.193834e-08  1.735773e-07  4.359863e-08
         CRBI        CWalks       LeagueN     DivisionW
 4.566257e-08  5.533716e-08  8.287113e-07 -1.022218e-05
      PutOuts       Assists        Errors    NewLeagueN
 1.577008e-08  3.279094e-09 -8.430214e-09 -2.550574e-07
```

Lambda = 10^10

d. As lambda increases, the regression coefficients move closer to 0 because the penalty term will dominate and push the coefficients to 0. This can be seen in part c flexibility of the ridge regression fit model decreases so there will be a decrease in the variance but and increase in bias

e. l_2 norm = 39577.44 when lambda = 0.013. I would expect this to be the larger l_2 norm because a larger lambda like 10^10 means the estimates are closer to 0 so the sum of the estimates squares would be smaller.

f. λ_{min}^{ridge} = 403.7017



g. $\lambda_{1se}^{ridge} = 6579.332$

h. $\lambda_{min}^{lasso} = 6.135907$, $\lambda_{1se}^{lasso} = 132.1941$

i. λ_{min}^{ridge} test error = 139020.3, λ_{1se}^{ridge} test error = 165523.3
   λ_{min}^{lasso} test error = 144217.6, λ_{1se}^{lasso} = 157773.7
   The ridge regression using the minimum lambda was the best model which could have been because the all the predictors influenced the response.

j. λ_{min}^{ridge}

	s0
(Intercept)	22.14199384
AtBat	0.09202637
Hits	0.79612092
HmRun	0.80036765
Runs	1.03423625
RBI	0.87793980
Walks	1.54008233
Years	1.82022134
CAtBat	0.01134633
CHits	0.05434354
CHmRun	0.38690022
CRuns	0.10821779
CRBI	0.11395518
CWalks	0.06088065
LeagueN	19.66599802
DivisionW	-72.32536945
PutOuts	0.15296151
Assists	0.02456701
Errors	-1.15478574
NewLeagueN	9.44765833

λ_{1se}^{ridge}

	s0
(Intercept)	342.454584686
AtBat	0.055333136
Hits	0.211556250
HmRun	0.769972808
Runs	0.349986914
RBI	0.359997480
Walks	0.442625783
Years	1.618528505
CAtBat	0.004665257
CHits	0.017500453
CHmRun	0.131097056
CRuns	0.035107946
CRBI	0.036259312
CWalks	0.036996438
LeagueN	0.609025156
DivisionW	-10.288271440
PutOuts	0.026487928
Assists	0.004086746
Errors	-0.042674529
NewLeagueN	0.820596873

λ_{min}^{lasso}

```
                       s0
(Intercept)   41.8141746
AtBat         -0.6006949
Hits           3.5685719
HmRun          .
Runs           .
RBI            .
Walks          3.0277367
Years         -4.0742639
CAtBat         .
CHits          .
CHmRun         0.1608128
CRuns          0.3585243
CRBI           0.4081870
CWalks        -0.1445588
LeagueN       26.5127706
DivisionW   -118.3247590
PutOuts        0.2476295
Assists        .
Errors        -0.5820076
NewLeagueN     .
```

λ_{1se}^{lasso}

```
                       s0
(Intercept) 316.02183375
AtBat         .
Hits          0.72112214
HmRun         .
Runs          .
RBI           .
Walks         0.56907852
Years         .
CAtBat        .
CHits         .
CHmRun        .
CRuns         0.09666601
CRBI          0.25371287
CWalks        .
LeagueN       .
DivisionW     .
PutOuts       .
Assists       .
Errors        .
NewLeagueN    .
```

Ridge regression models have all the predictors while lasso can zero them out when needed. Using lse for ridge regression made the coefficients smaller but the intercept larger. Using lse for lasso did the same and also got rid of some of the predictors compared to min for lasso.

k. Focus on the Runs, Walks, CRuns, and CRBI

Problem 4:

a. See RScript
b. Test MSE = 1116181

   Model 11:

```
 (Intercept)    PrivateYes        Accept        Enroll
-29.00118556 -360.38223219    1.78053679   -1.47299845
   Top10perc     Top25perc   F.Undergrad      Outstate
 65.66874627  -21.23295783    0.09723797   -0.10980932
  Room.Board           PhD        Expend     Grad.Rate
  0.22121228  -11.27016098    0.04329090    7.24452155
```

c. Ridge regression applies a penalty to the coefficients. So scale is important for regularized models. Without scaling, the penalty could have different affects on each coefficient and the coefficient would vary in size
d. Optimal lambda = 0.01
e. Test MSE = 1134677. There isn't improvement over the model compared to part b
f. Optimal lambda = 0.01
   number of non-zero coefficients = 17 non-zero coefficients and intercept
   test MSE = 1133422
g. Using the least squares based from test MSE, the best model was found.

```
 (Intercept)    PrivateYes        Accept        Enroll
-29.00118556 -360.38223219    1.78053679   -1.47299845
   Top10perc     Top25perc   F.Undergrad      Outstate
 65.66874627  -21.23295783    0.09723797   -0.10980932
  Room.Board           PhD        Expend     Grad.Rate
  0.22121228  -11.27016098    0.04329090    7.24452155
```

   The test errors for lasso and ridge regression were not that different. This could mean that each predictor was important to the response so lasso didn't zero out any coefficients leading to a similar model of ridge regression. Best subset performed the best though as it had the lowest test MSE and 11 coeffificents and an intercept. The final model shows that important predictors are PrivateYes and Top10perc.
   Best subset test MSE = 1116181
   Ridge regression test MSE = 1134677