Business Analytics

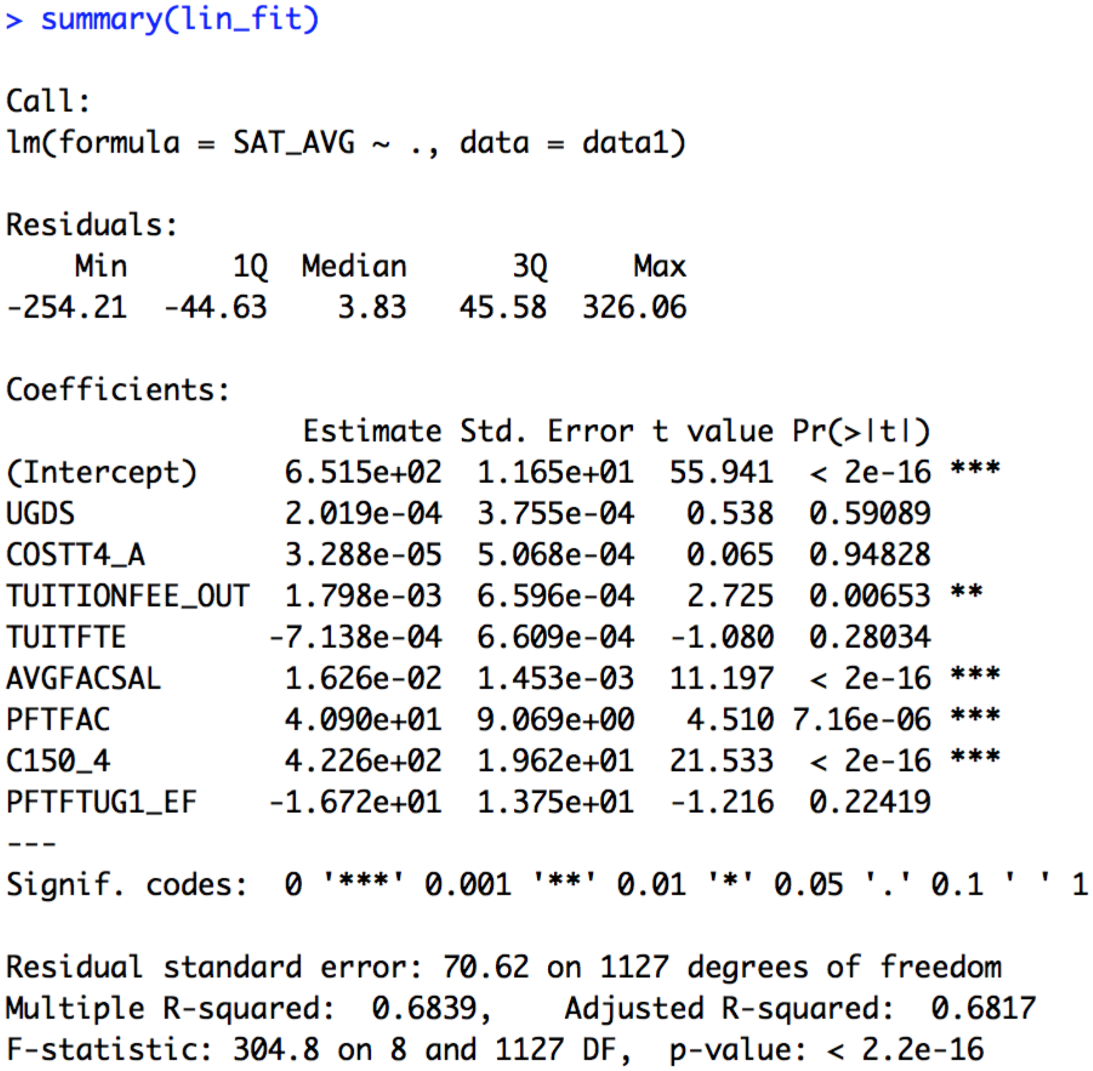
Assignment-4

Name-Subham Kedia

UNI-sk4355

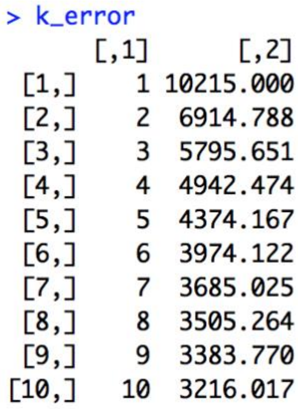
Answer 1

Part a.

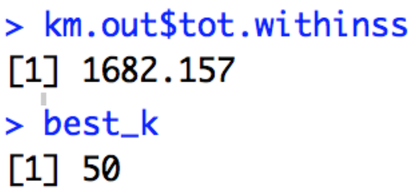


```
> summary(lin_fit)

Call:
lm(formula = SAT_AVG ~ ., data = data1)

Residuals:
    Min      1Q  Median      3Q     Max 
-254.21  -44.63    3.83   45.58  326.06 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     6.515e+02  1.165e+01  55.941  < 2e-16 ***
UGDS            2.019e-04  3.755e-04   0.538  0.59089    
COSTT4_A        3.288e-05  5.068e-04   0.065  0.94828    
TUITIONFEE_OUT  1.798e-03  6.596e-04   2.725  0.00653 ** 
TUITFTE        -7.138e-04  6.609e-04  -1.080  0.28034    
AVGFACSAL       1.626e-02  1.453e-03  11.197  < 2e-16 ***
PFTFAC          4.090e+01  9.069e+00   4.510 7.16e-06 ***
C150_4          4.226e+02  1.962e+01  21.533  < 2e-16 ***
PFTFTUG1_EF    -1.672e+01  1.375e+01  -1.216  0.22419    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 70.62 on 1127 degrees of freedom
Multiple R-squared:  0.6839,	Adjusted R-squared:  0.6817 
F-statistic: 304.8 on 8 and 1127 DF,  p-value: < 2.2e-16
```

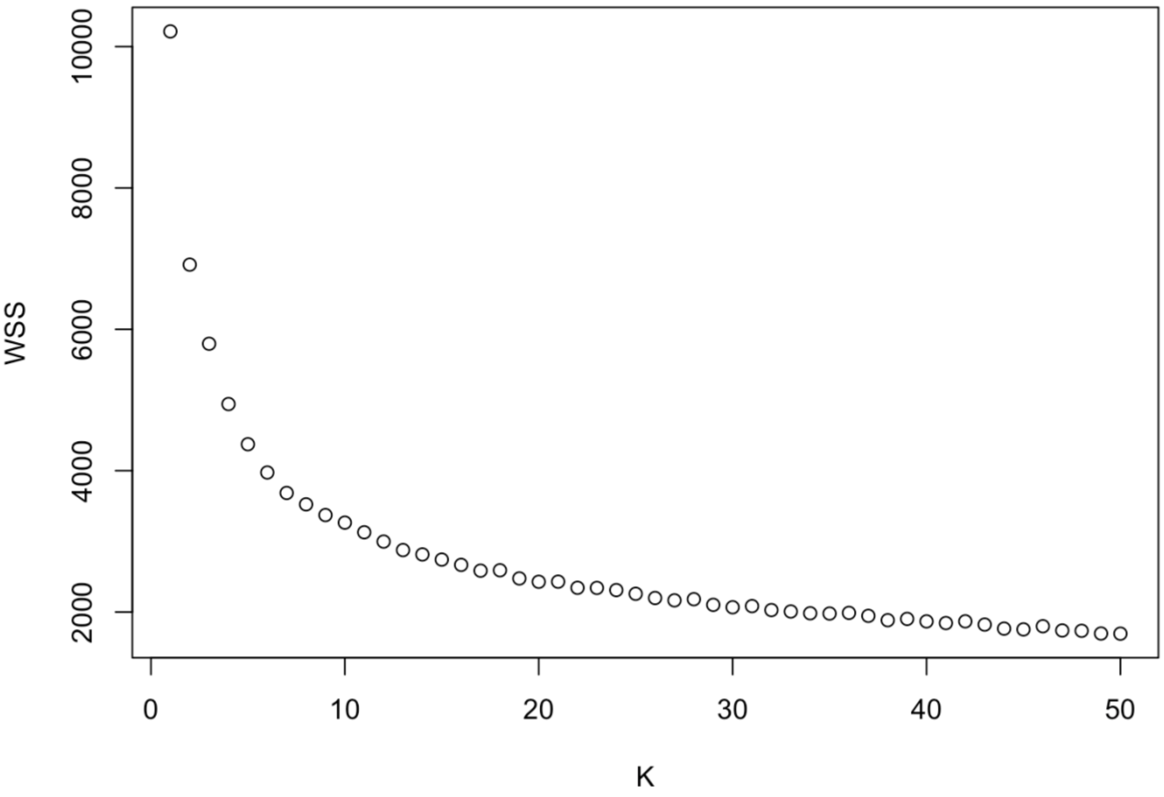
Part b.

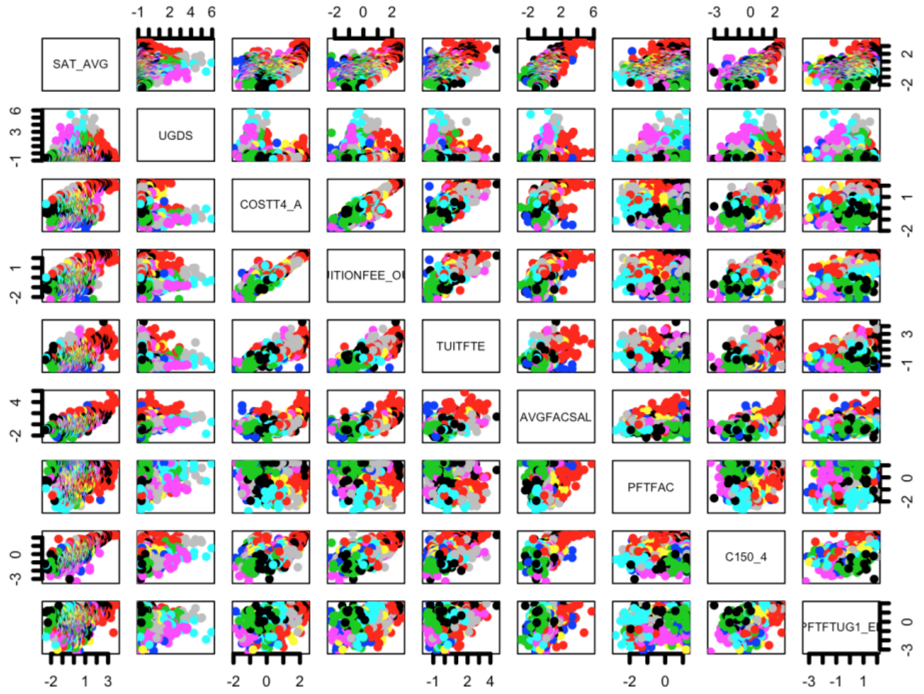


```
> k_error
      [,1]      [,2]
 [1,]    1 10215.000
 [2,]    2  6914.788
 [3,]    3  5795.651
 [4,]    4  4942.474
 [5,]    5  4374.167
 [6,]    6  3974.122
 [7,]    7  3685.025
 [8,]    8  3505.264
 [9,]    9  3383.770
[10,]   10  3216.017
```

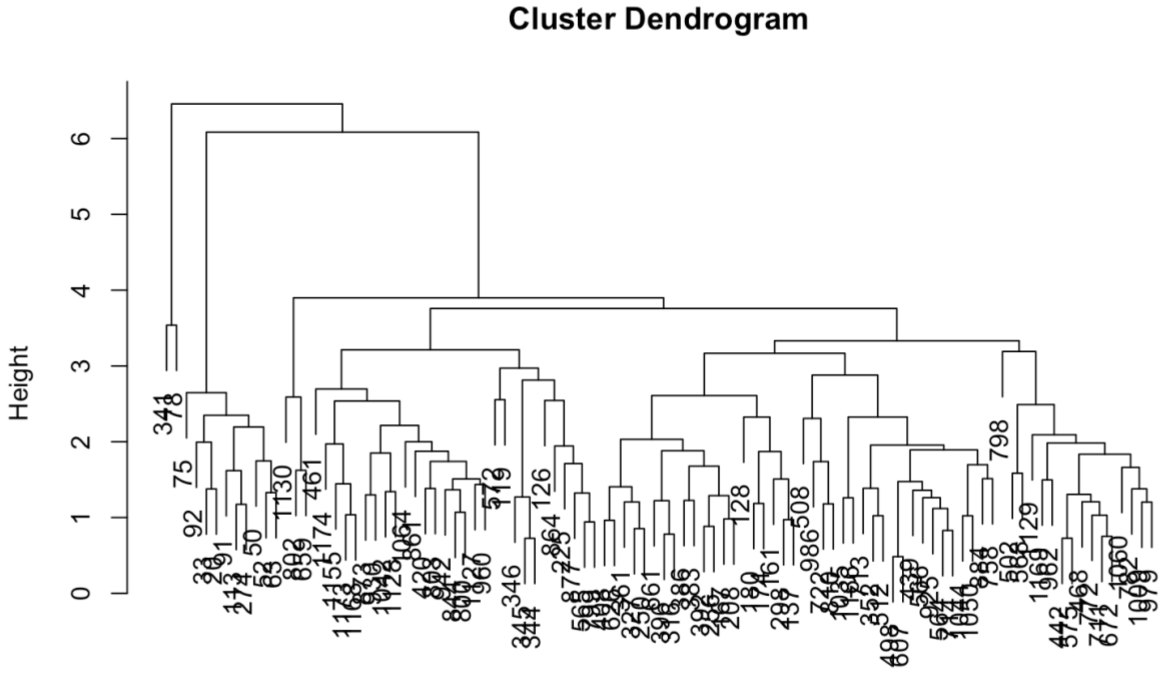


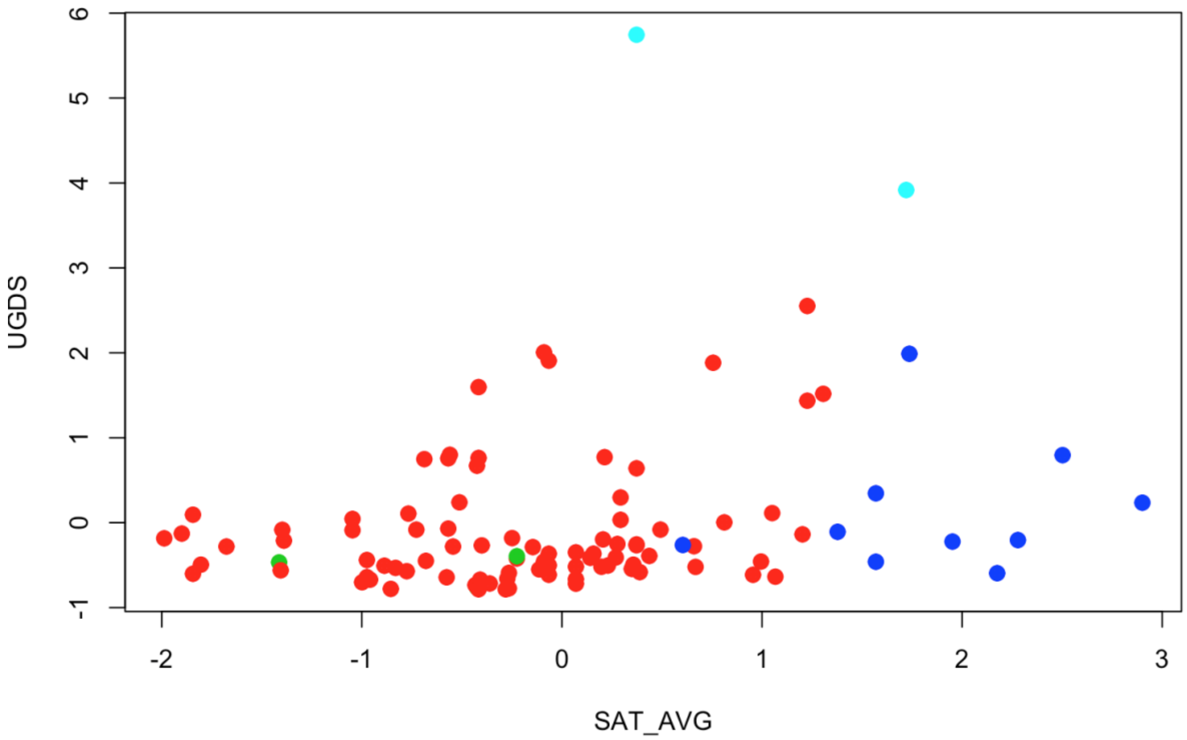
```
> km.out$tot.withinss
[1] 1682.157
> best_k
[1] 50
```



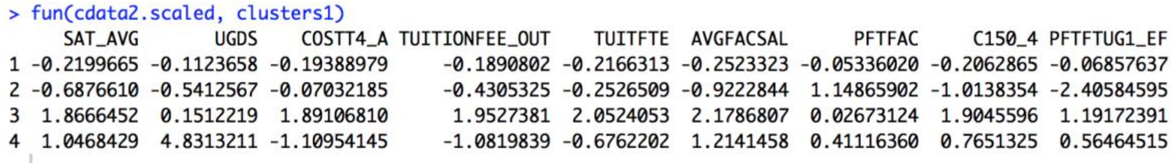


Part c.



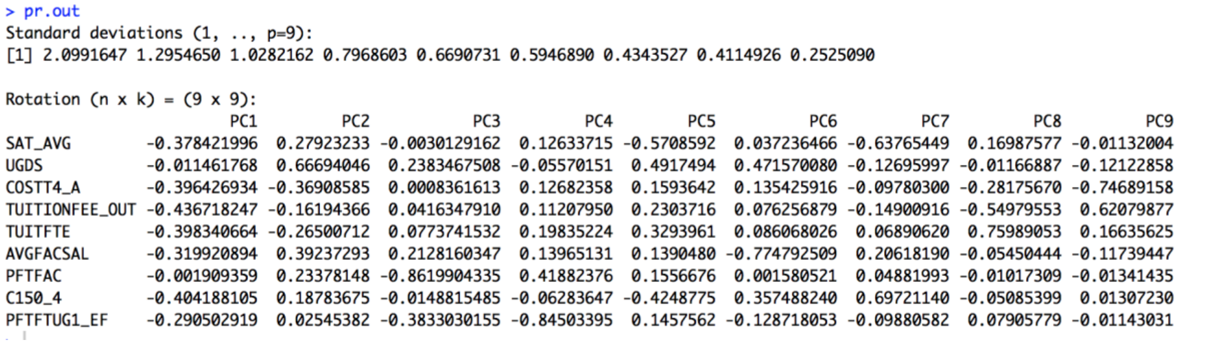


The center of the four clusters are as follows: -

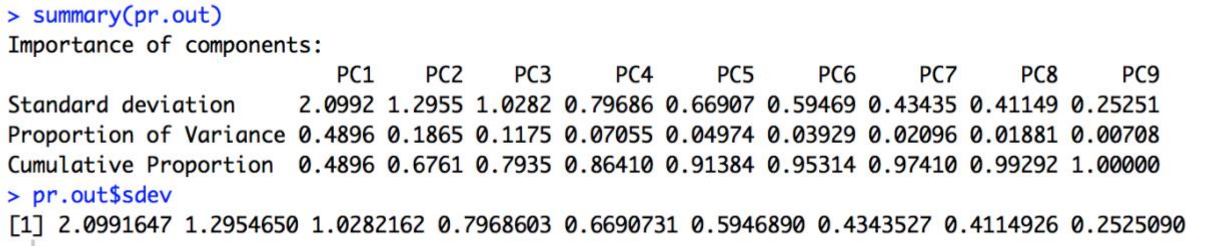


```
> fun(cdata2.scaled, clusters1)
     SAT_AVG       UGDS    COSTT4_A TUITIONFEE_OUT    TUITFTE  AVGFACSAL      PFTFAC     C150_4 PFTFTUG1_EF
1 -0.2199665 -0.1123658 -0.19388979     -0.1890802 -0.2166313 -0.2523323 -0.05336020 -0.2062865 -0.06857637
2 -0.6876610 -0.5412567 -0.07032185     -0.4305325 -0.2526509 -0.9222844  1.14865902 -1.0138354 -2.40584595
3  1.8666452  0.1512219  1.89106810      1.9527381  2.0524053  2.1786807  0.02673124  1.9045596  1.19172391
4  1.0468429  4.8313211 -1.10954145     -1.0819839 -0.6762202  1.2141458  0.41116360  0.7651325  0.56464515
```

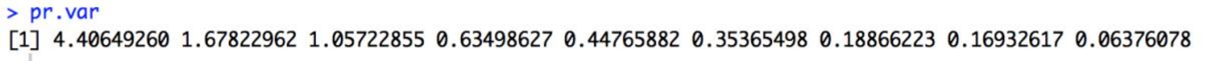
Part d.



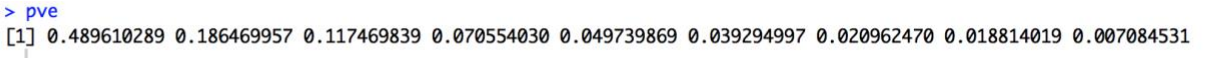
```
> pr.out
Standard deviations (1, .., p=9):
[1] 2.0991647 1.2954650 1.0282162 0.7968603 0.6690731 0.5946890 0.4343527 0.4114926 0.2525090

Rotation (n x k) = (9 x 9):
                     PC1         PC2           PC3         PC4        PC5          PC6         PC7         PC8         PC9
SAT_AVG       -0.378421996  0.27923233 -0.0030129162  0.12633715 -0.5708592  0.037236466 -0.63765449  0.16987577 -0.01132004
UGDS          -0.011461768  0.66694046  0.2383467508 -0.05570151  0.4917494  0.471570080 -0.12695997 -0.01166887 -0.12122858
COSTT4_A      -0.396426934 -0.36908585  0.0008361613  0.12682358  0.1593642  0.135425916 -0.09780300 -0.28175670 -0.74689158
TUITIONFEE_OUT -0.436718247 -0.16194366  0.0416347910  0.11207950  0.2303716  0.076256879 -0.14900916 -0.54979553  0.62079877
TUITFTE       -0.398340664 -0.26500712  0.0773741532  0.19835224  0.3293961  0.086068026  0.06890620  0.75989053  0.16635625
AVGFACSAL     -0.319920894  0.39237293  0.2128160347  0.13965131  0.1390480 -0.774792509  0.20618190 -0.05450444 -0.11739447
PFTFAC        -0.001909359  0.23378148 -0.8619904335  0.41882376  0.1556676  0.001580521  0.04881993 -0.01017309 -0.01341435
C150_4        -0.404188105  0.18783675 -0.0148815485 -0.06283647 -0.4248775  0.357488240  0.69721140 -0.05085399  0.01307230
PFTFTUG1_EF   -0.290502919  0.02545382 -0.3833030155 -0.84503395  0.1457562 -0.128718053 -0.09880582  0.07905779 -0.01143031
```



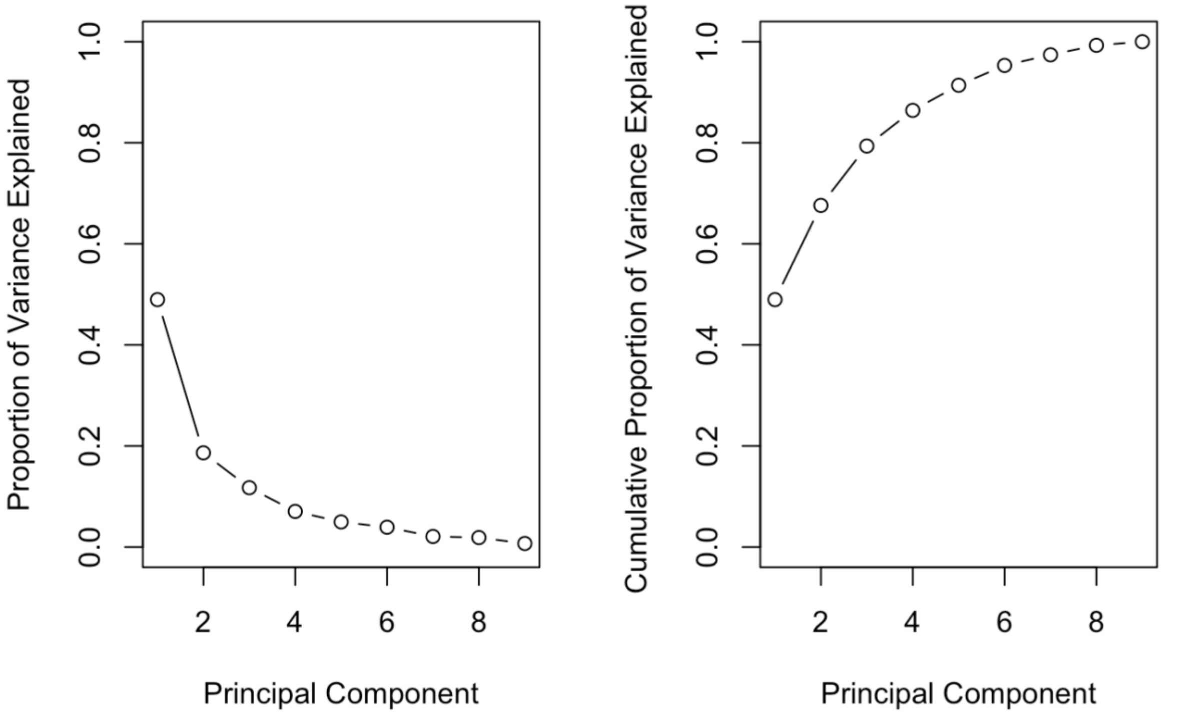
```
> summary(pr.out)
Importance of components:
                          PC1    PC2    PC3     PC4     PC5     PC6     PC7     PC8     PC9
Standard deviation     2.0992 1.2955 1.0282 0.79686 0.66907 0.59469 0.43435 0.41149 0.25251
Proportion of Variance 0.4896 0.1865 0.1175 0.07055 0.04974 0.03929 0.02096 0.01881 0.00708
Cumulative Proportion  0.4896 0.6761 0.7935 0.86410 0.91384 0.95314 0.97410 0.99292 1.00000
> pr.out$sdev
[1] 2.0991647 1.2954650 1.0282162 0.7968603 0.6690731 0.5946890 0.4343527 0.4114926 0.2525090
```



```
> pr.var
[1] 4.40649260 1.67822962 1.05722855 0.63498627 0.44765882 0.35365498 0.18866223 0.16932617 0.06376078
```



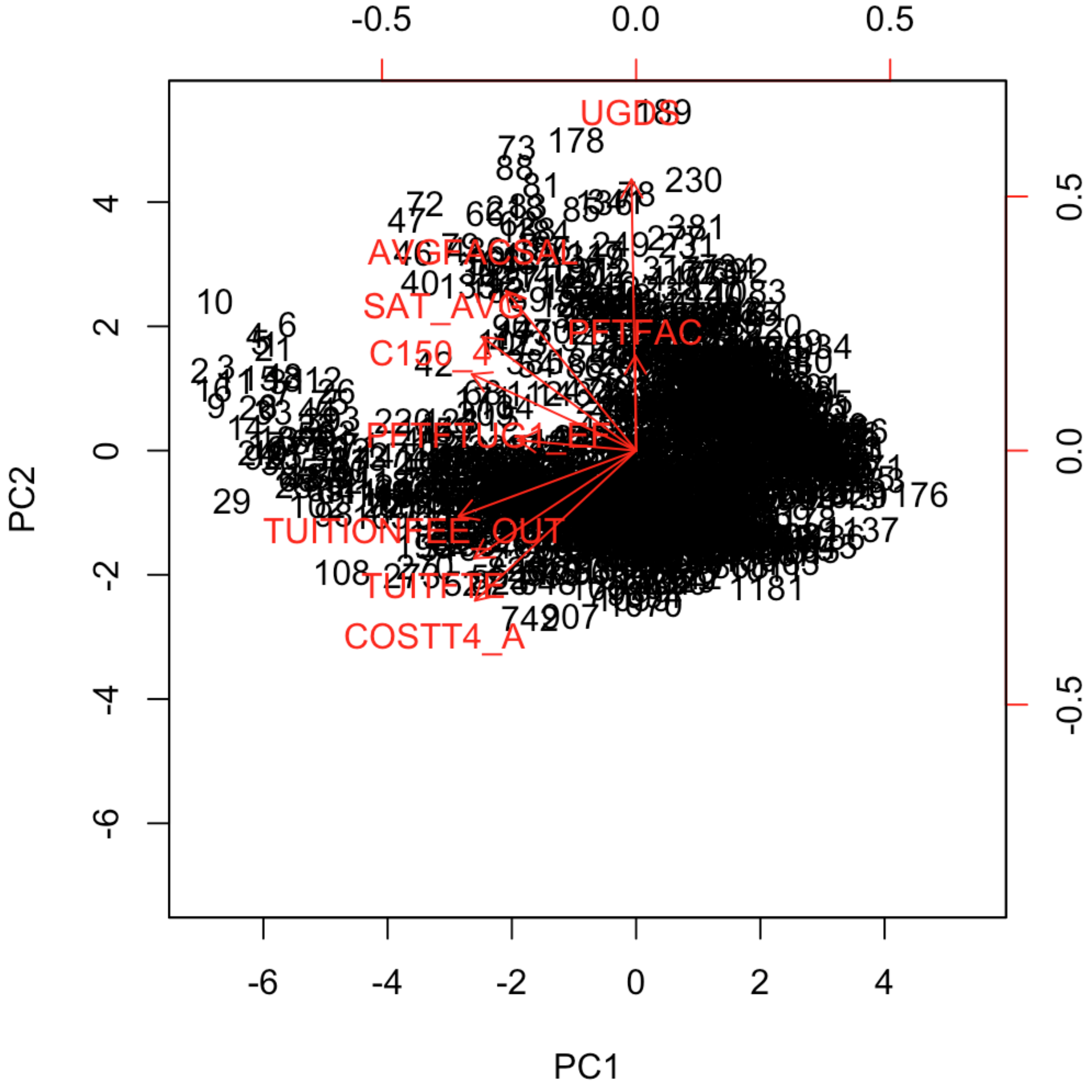
```
> pve
[1] 0.489610289 0.186469957 0.117469839 0.070554030 0.049739869 0.039294997 0.020962470 0.018814019 0.007084531
```



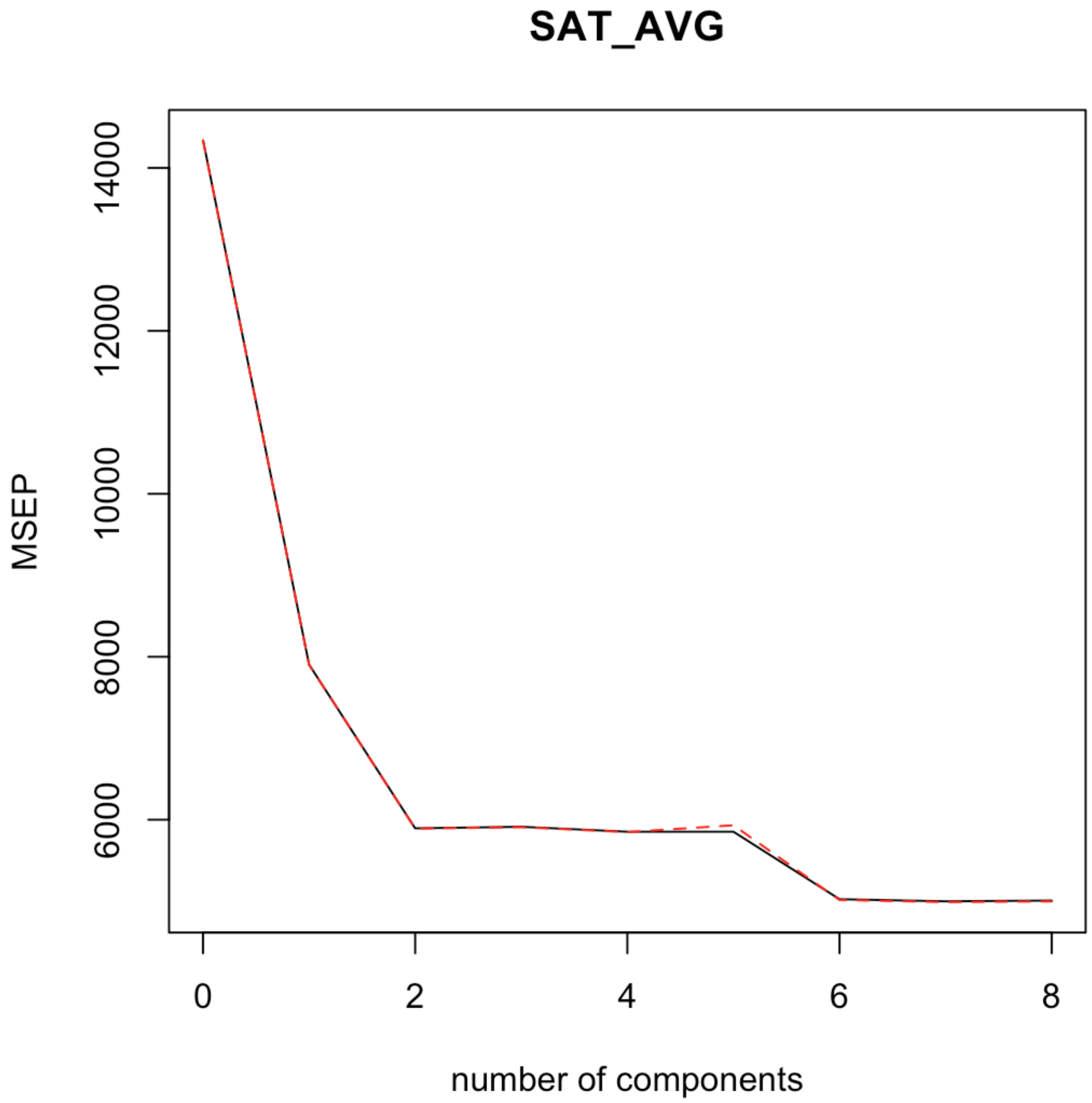
Part e.

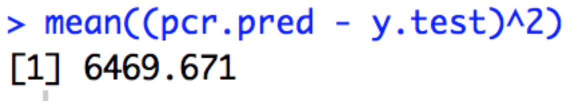
Interpreting the Biplot: -

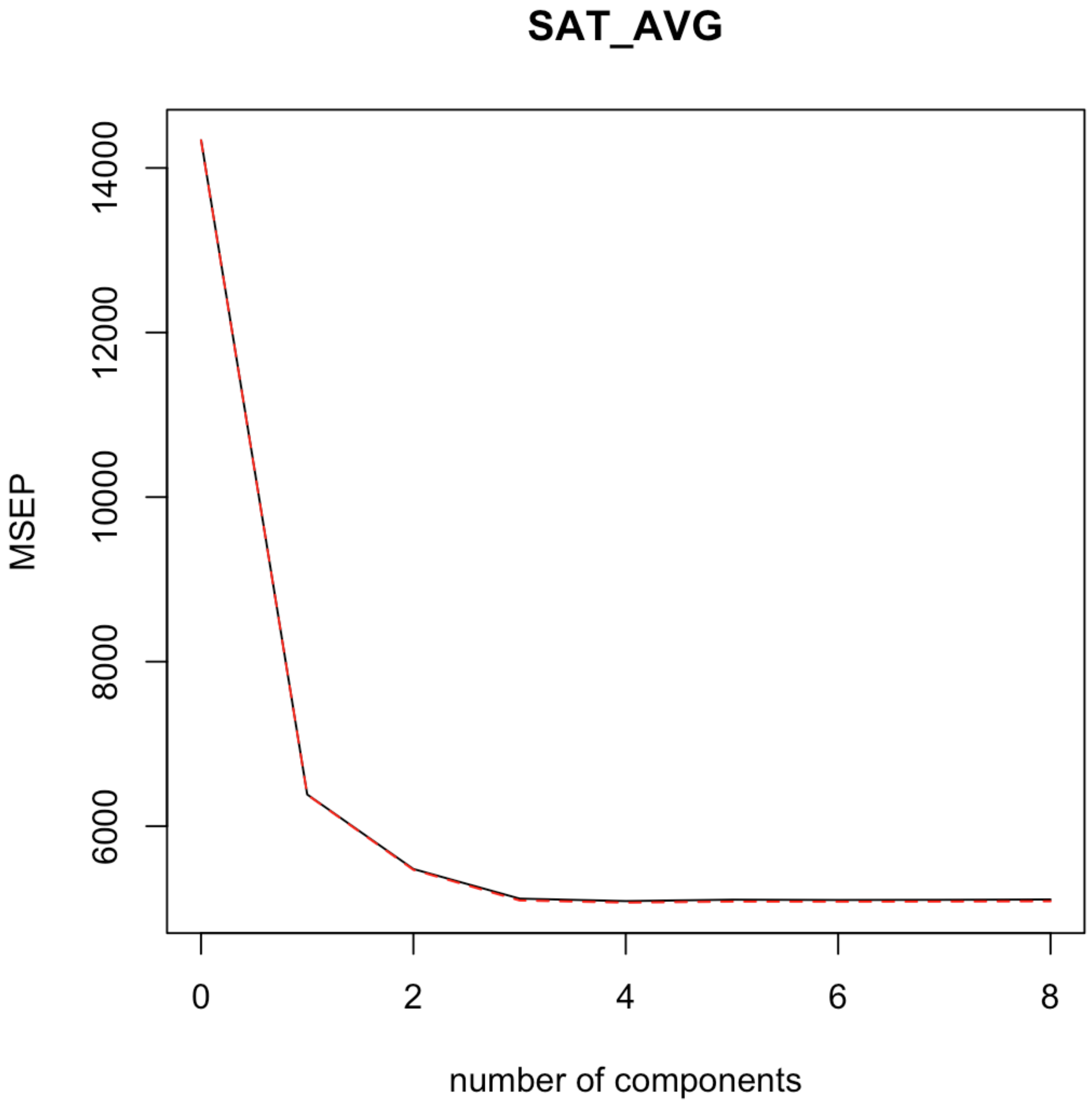
The bi plot represents the loading vectors for the principal component 1 and 2. For example if you look at the loading vector at COST4_A for PC1 it takes a value of -0.06 and for PC2 around -0.04.

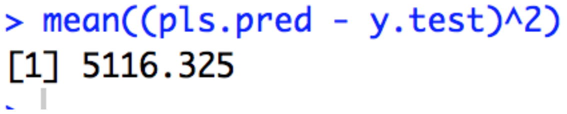


Part f.




```
> mean((pcr.pred - y.test)^2)
[1] 6469.671
```





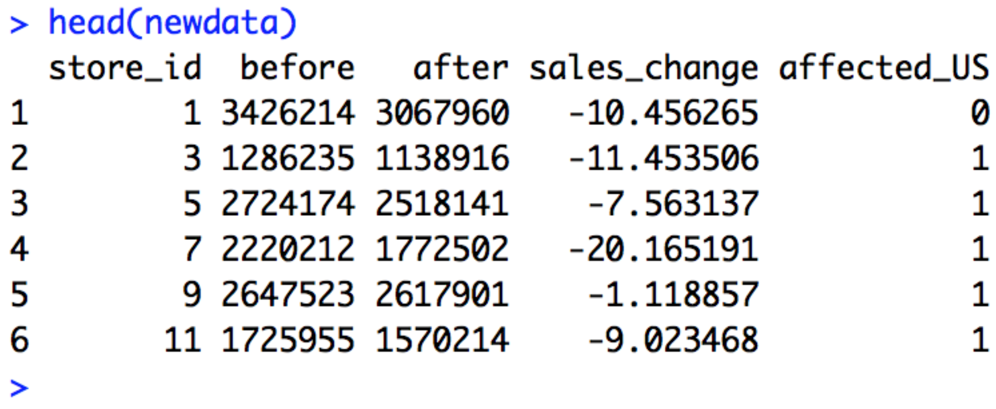
```
> mean((pls.pred - y.test)^2)
[1] 5116.325
```

Comparing the two mean squared errors, we can see that PLS is performing better.

Answer 2

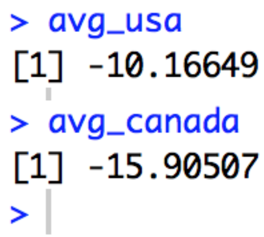
Part a.

For each store, below is the head of a dataset containing the percent change in sales before and after the BOPS initiative began.



```
> head(newdata)
  store_id  before   after sales_change affected_US
1        1 3426214 3067960   -10.456265           0
2        3 1286235 1138916   -11.453506           1
3        5 2724174 2518141    -7.563137           1
4        7 2220212 1772502   -20.165191           1
5        9 2647523 2617901    -1.118857           1
6       11 1725955 1570214    -9.023468           1
>
```

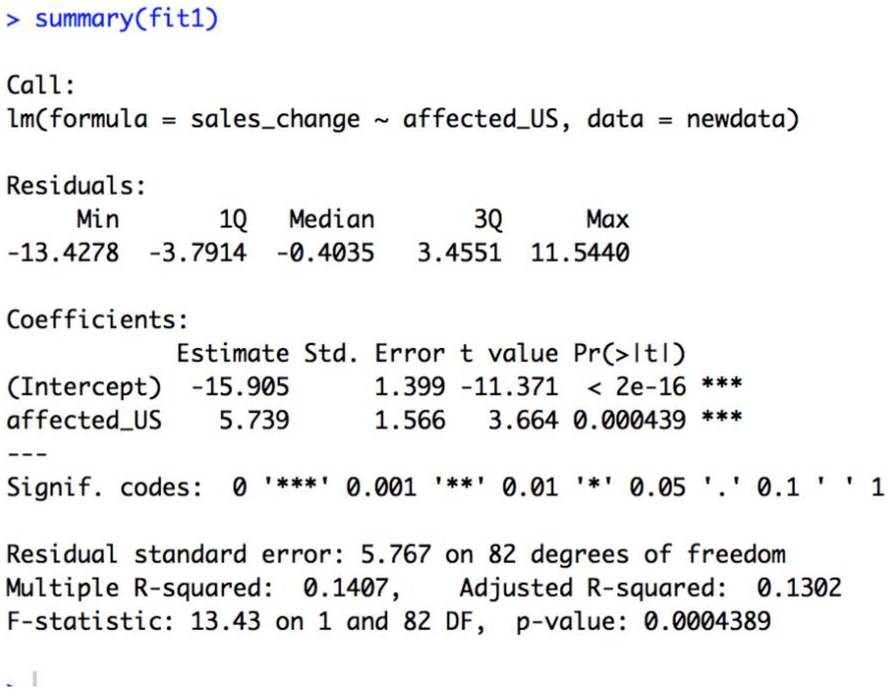
The average percent change for stores in USA and Canada are: -



```
> avg_usa
[1] -10.16649
> avg_canada
[1] -15.90507
>
```

Part b.

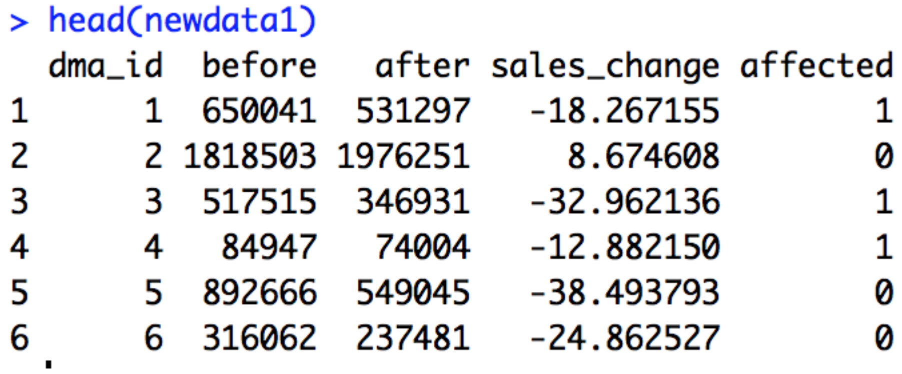
The effect of BOPS, and the standard error are: -



```
> summary(fit1)

Call:
lm(formula = sales_change ~ affected_US, data = newdata)

Residuals:
     Min       1Q   Median       3Q      Max 
-13.4278  -3.7914  -0.4035   3.4551  11.5440 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -15.905      1.399 -11.371  < 2e-16 ***
affected_US    5.739      1.566   3.664 0.000439 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.767 on 82 degrees of freedom
Multiple R-squared:  0.1407,	Adjusted R-squared:  0.1302 
F-statistic: 13.43 on 1 and 82 DF,  p-value: 0.0004389
```

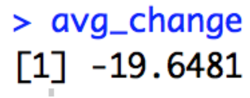
Part c.

For each DMA, below is the head of a dataset containing the percent change in sales before and after the BOPS initiative began.



```
> head(newdata1)
  dma_id  before   after sales_change affected
1      1  650041  531297   -18.267155        1
2      2 1818503 1976251     8.674608        0
3      3  517515  346931   -32.962136        1
4      4   84947   74004   -12.882150        1
5      5  892666  549045   -38.493793        0
6      6  316062  237481   -24.862527        0
```

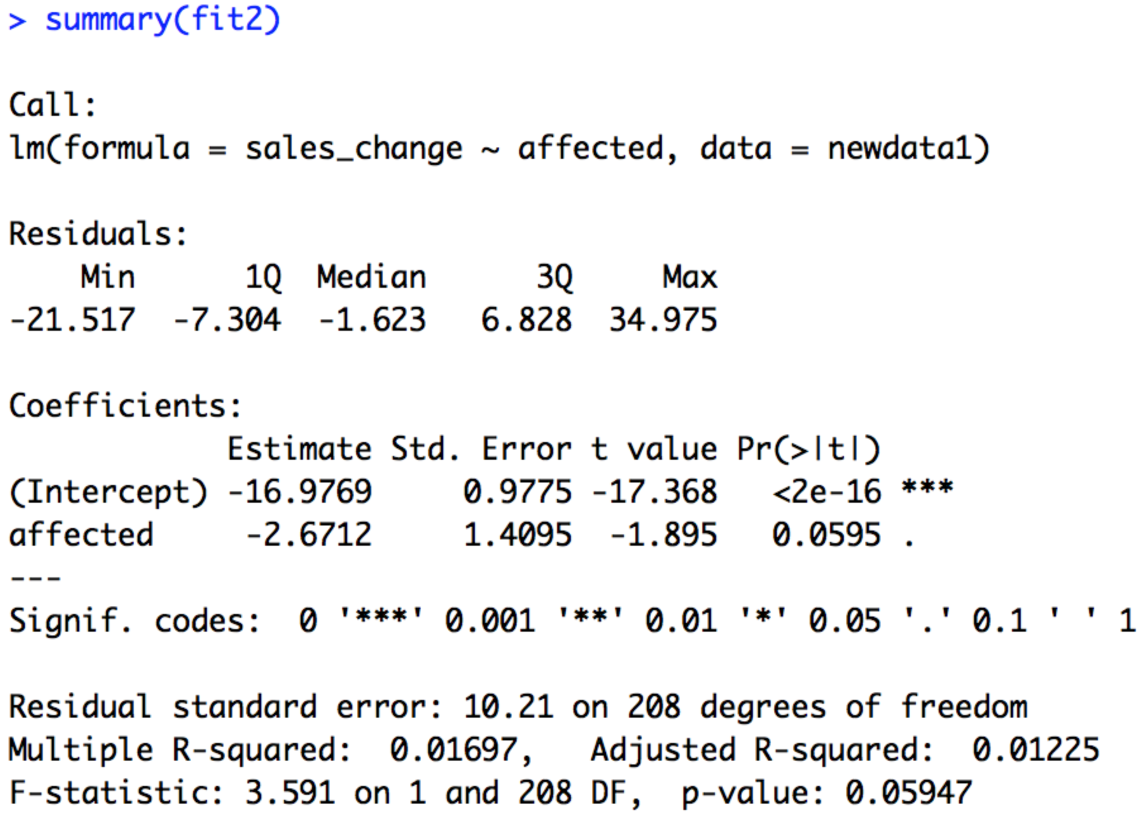
The average percent change for DMAs close to stores with BOPS is as follows: -



```
> avg_change
[1] -19.6481
```

Part d.

The effect of BOPS, and the standard error: -

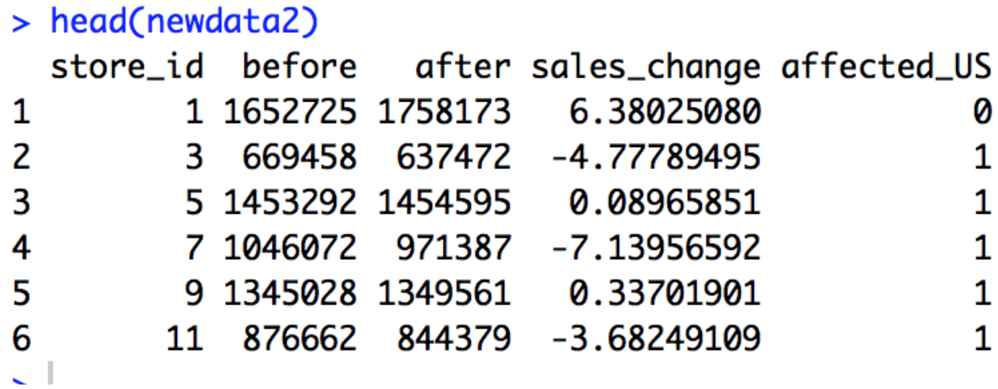


```
> summary(fit2)

Call:
lm(formula = sales_change ~ affected, data = newdata1)

Residuals:
    Min      1Q  Median      3Q     Max 
-21.517  -7.304  -1.623   6.828  34.975 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -16.9769     0.9775 -17.368   <2e-16 ***
affected     -2.6712     1.4095  -1.895   0.0595 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.21 on 208 degrees of freedom
Multiple R-squared:  0.01697,	Adjusted R-squared:  0.01225 
F-statistic: 3.591 on 1 and 208 DF,  p-value: 0.05947
```

Part e.

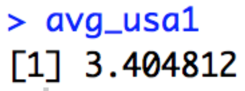
Below, is the complete analysis for the 13 weeks of data before and after the BOPS initiative began:

For each store, below is the head of a dataset containing the percent change in sales before and after the BOPS initiative began.

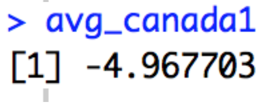


```
> head(newdata2)
  store_id  before   after sales_change affected_US
1        1 1652725 1758173   6.38025080           0
2        3  669458  637472  -4.77789495           1
3        5 1453292 1454595   0.08965851           1
4        7 1046072  971387  -7.13956592           1
5        9 1345028 1349561   0.33701901           1
6       11  876662  844379  -3.68249109           1
```

The average percent change for stores in USA and Canada are: -

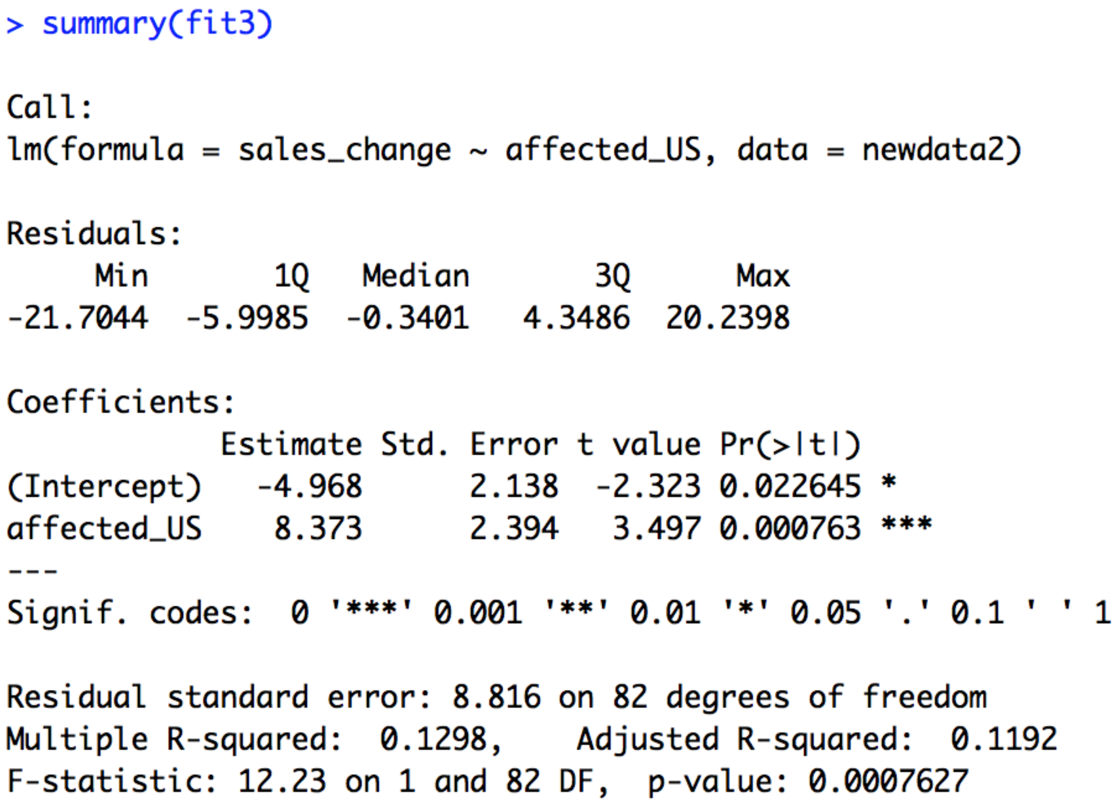


```
> avg_usa1
[1] 3.404812
```



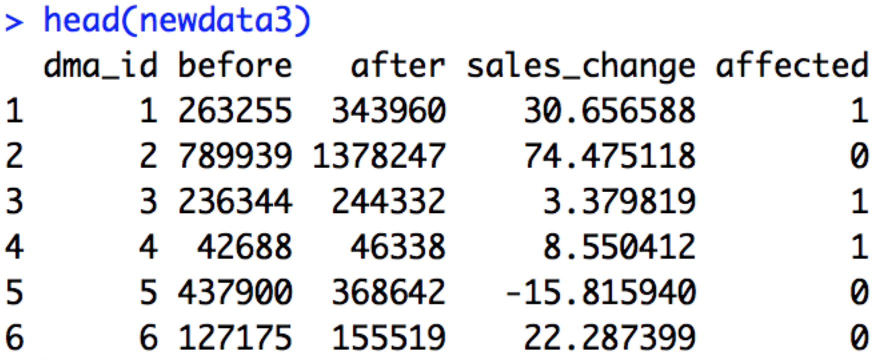
```
> avg_canada1
[1] -4.967703
```

The effect of BOPS, and the standard error are: -



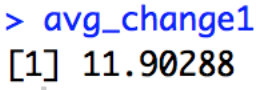
```
> summary(fit3)

Call:
lm(formula = sales_change ~ affected_US, data = newdata2)

Residuals:
     Min       1Q   Median       3Q      Max 
-21.7044  -5.9985  -0.3401   4.3486  20.2398 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   -4.968      2.138  -2.323 0.022645 *  
affected_US    8.373      2.394   3.497 0.000763 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.816 on 82 degrees of freedom
Multiple R-squared:  0.1298,	Adjusted R-squared:  0.1192 
F-statistic: 12.23 on 1 and 82 DF,  p-value: 0.0007627
```

For each DMA, below is the head of a dataset containing the percent change in sales before and after the BOPS initiative began.



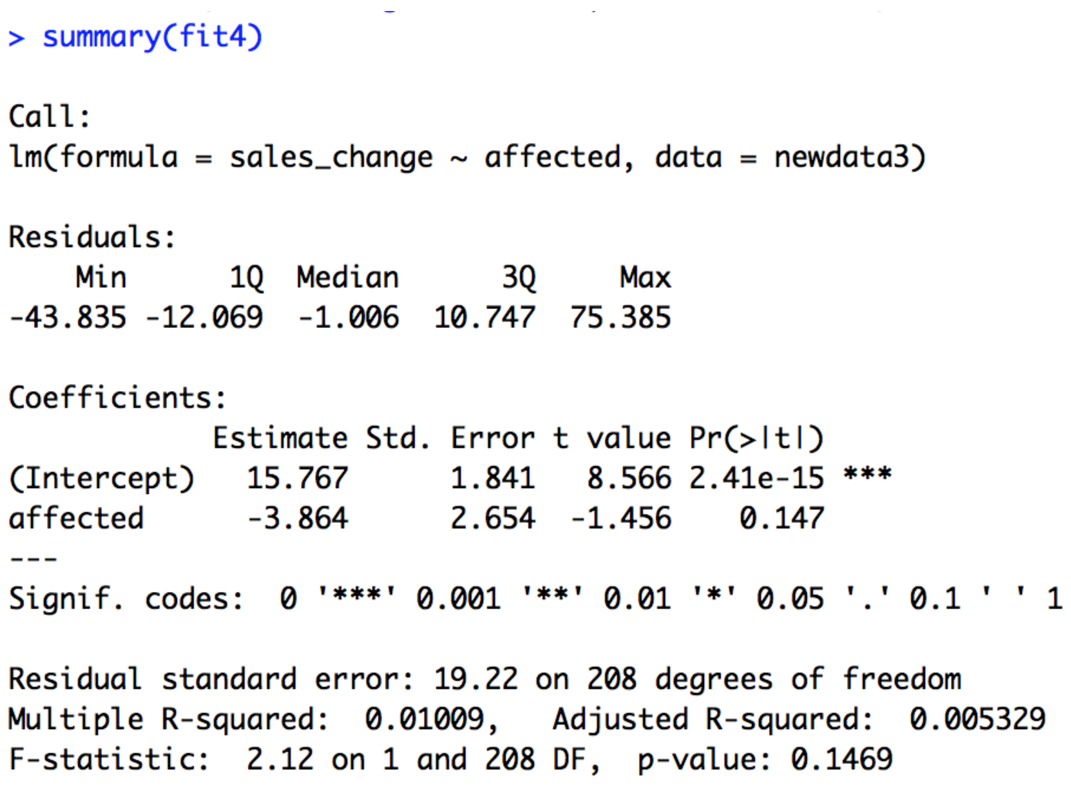
```
> head(newdata3)
  dma_id before   after sales_change affected
1      1 263255  343960    30.656588        1
2      2 789939 1378247    74.475118        0
3      3 236344  244332     3.379819        1
4      4  42688   46338     8.550412        1
5      5 437900  368642   -15.815940        0
6      6 127175  155519    22.287399        0
```

The average percent change for DMAs close to stores with BOPS is as follows: -



```
> avg_change1
[1] 11.90288
```

The effect of BOPS, and the standard error: -



```
> summary(fit4)

Call:
lm(formula = sales_change ~ affected, data = newdata3)

Residuals:
    Min      1Q  Median      3Q     Max 
-43.835 -12.069  -1.006  10.747  75.385 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   15.767      1.841   8.566 2.41e-15 ***
affected      -3.864      2.654  -1.456    0.147    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.22 on 208 degrees of freedom
Multiple R-squared:  0.01009,	Adjusted R-squared:  0.005329 
F-statistic:  2.12 on 1 and 208 DF,  p-value: 0.1469
```