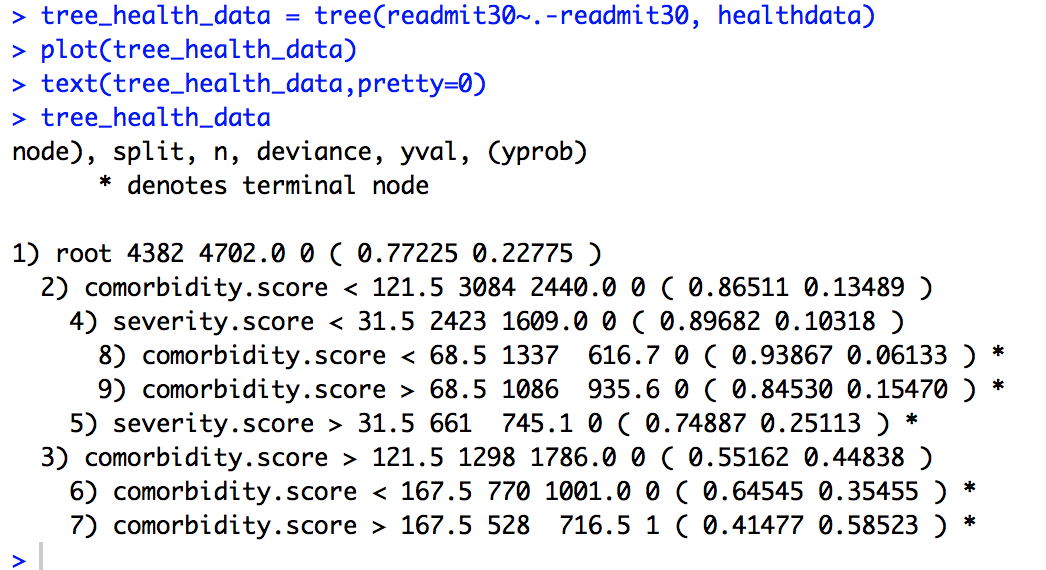
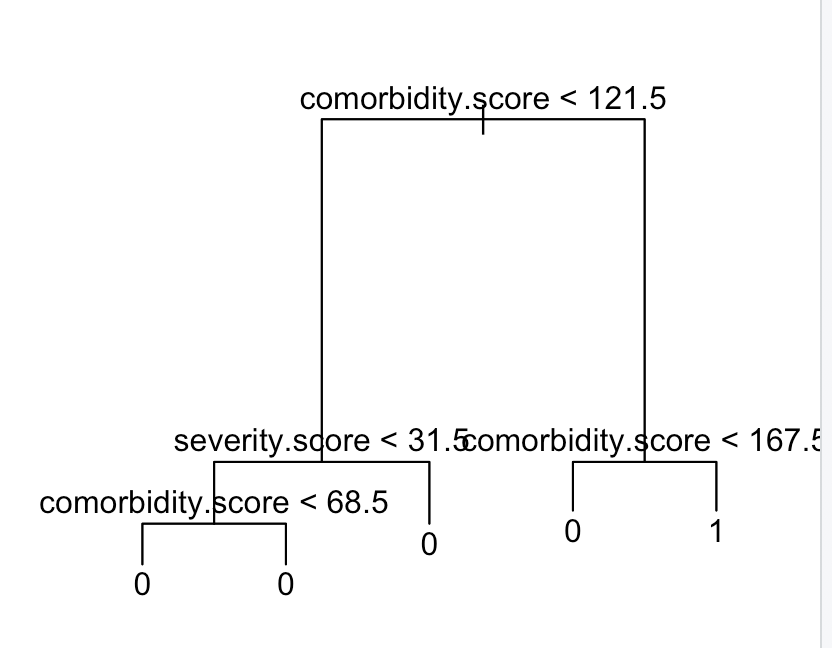
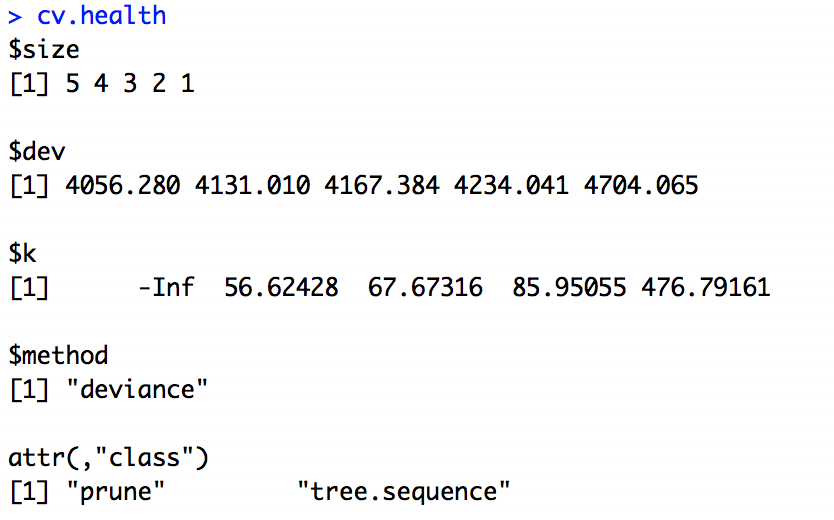
1.

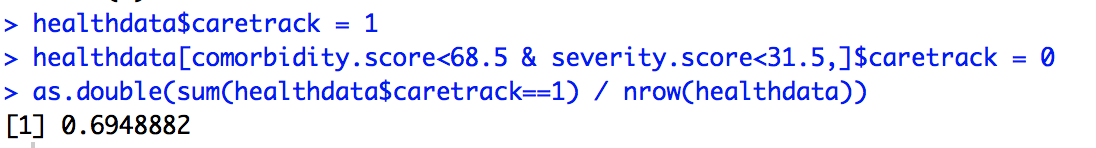
(a)



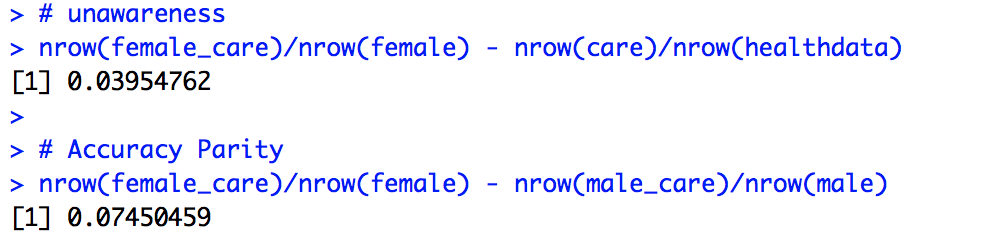
(b)



(c)

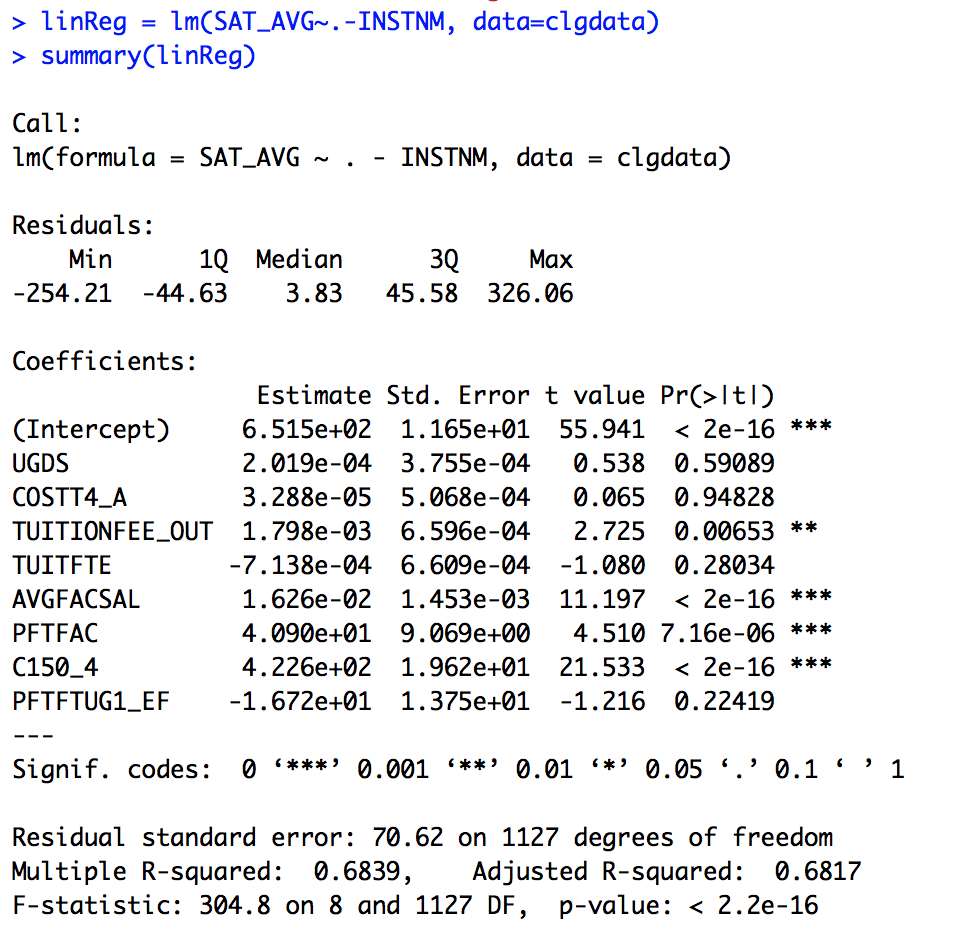


(d) The unawareness difference between male and female is about 0.0395, while the accuracy parity difference is about 0.0745

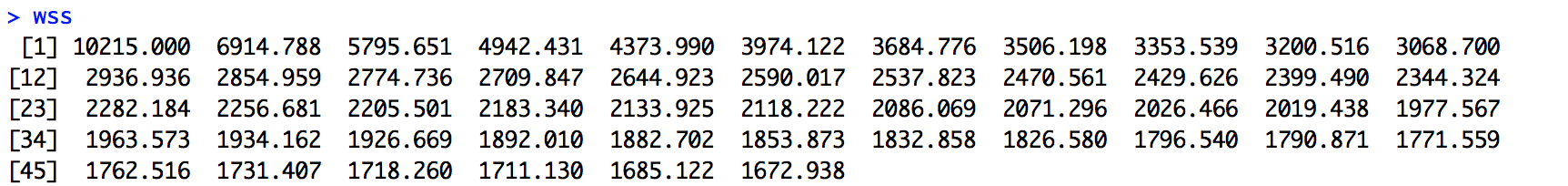


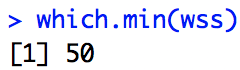
2.

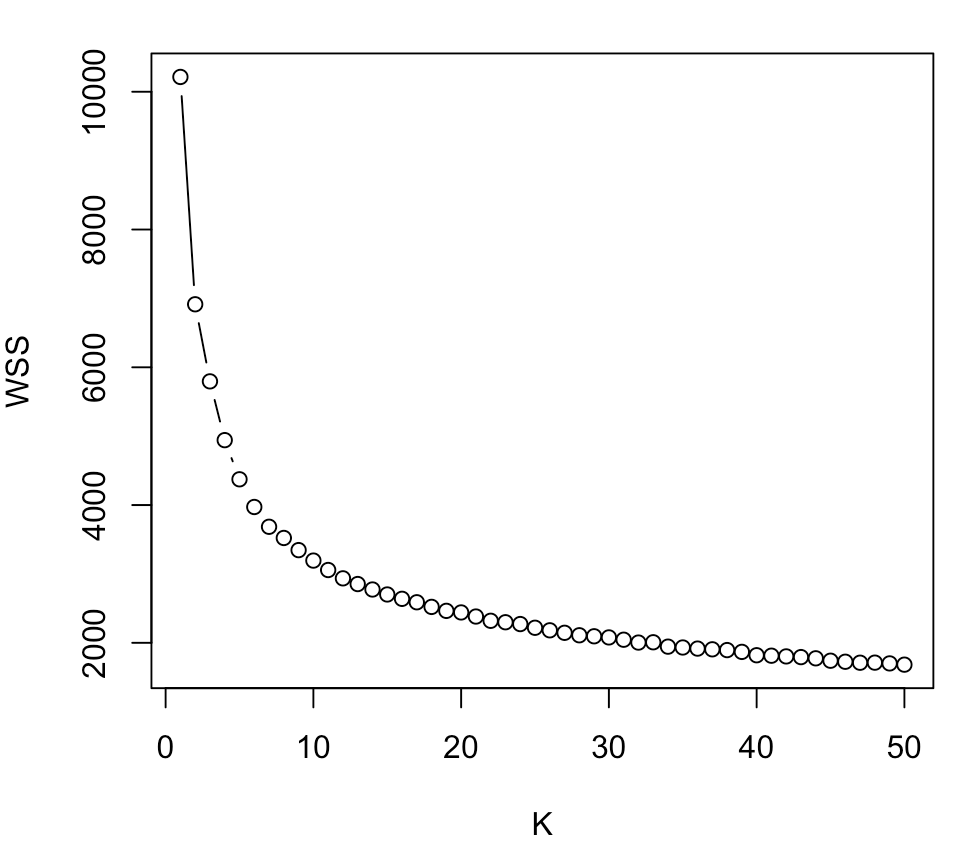
(a)



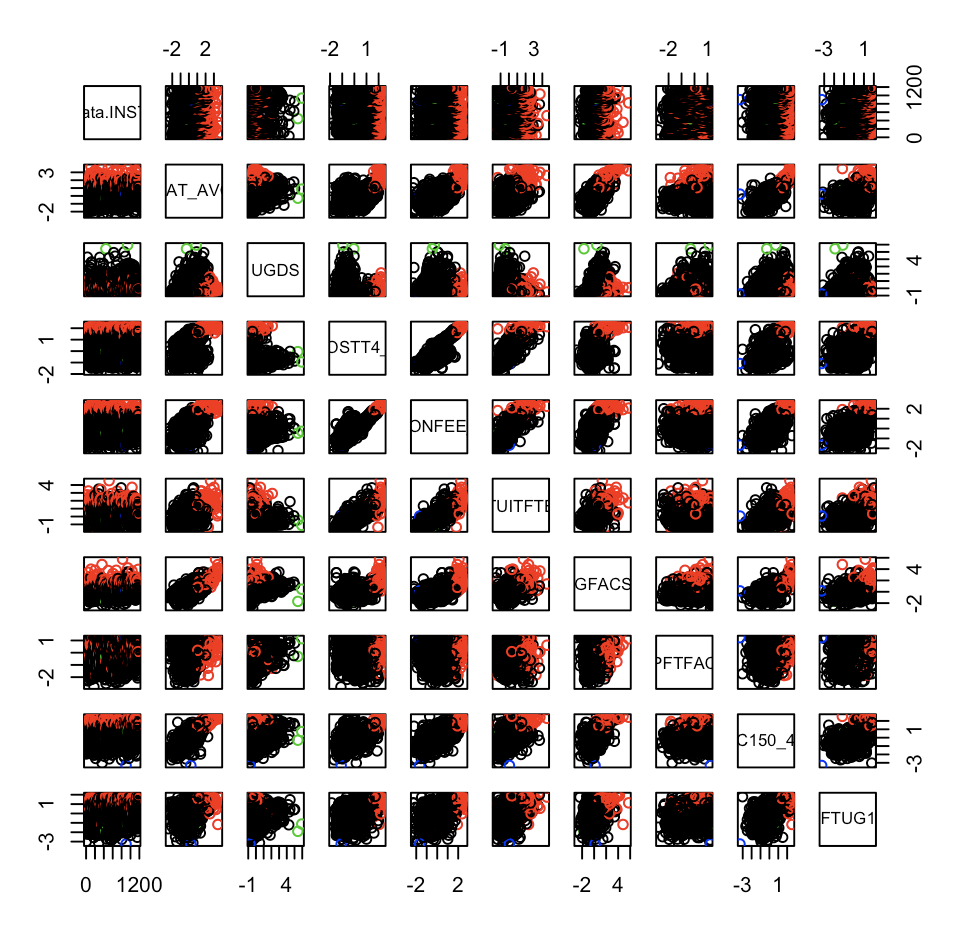
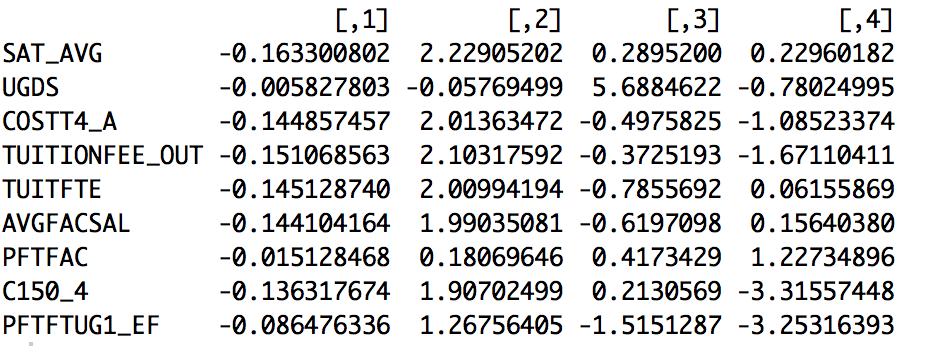
(b) I will choose k=10, because WSS[10] is quite small and k=10 is not large.



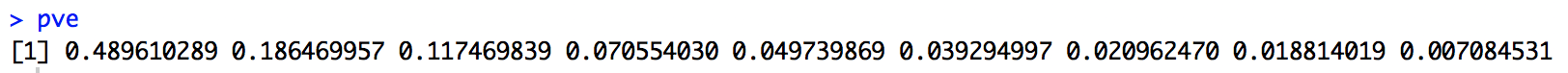


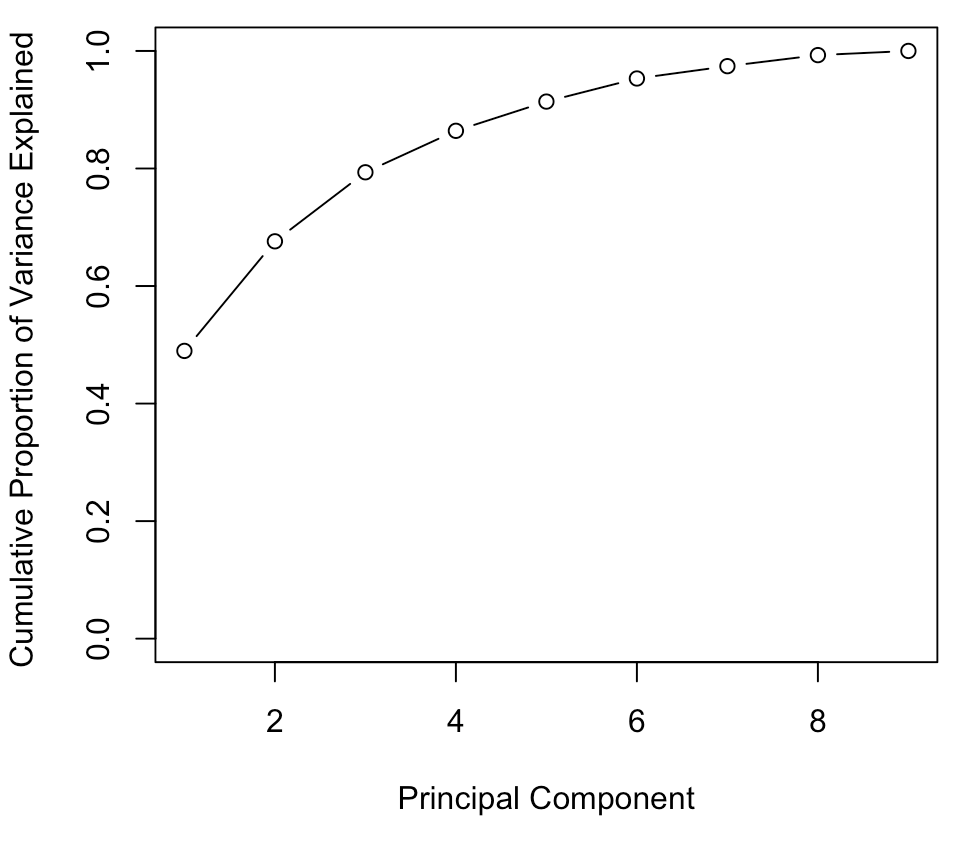


(c) Four centroids are as below



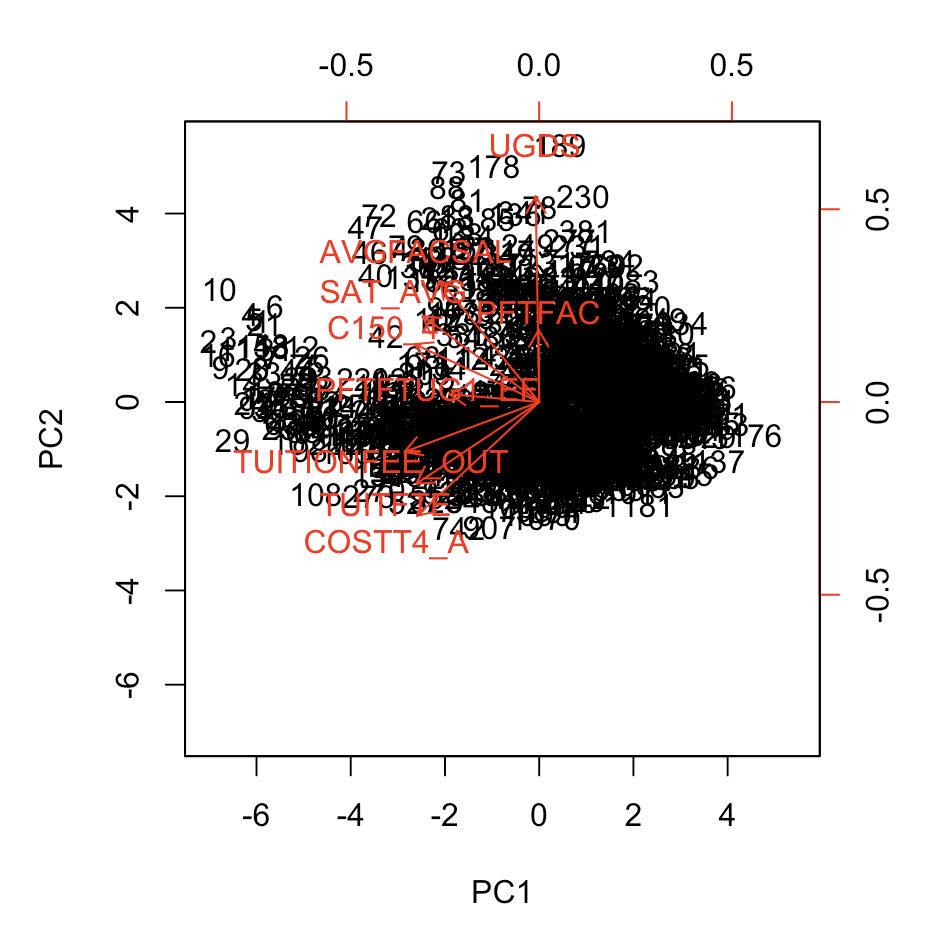
(d) PVE values and cumulative PVE plot is as below.





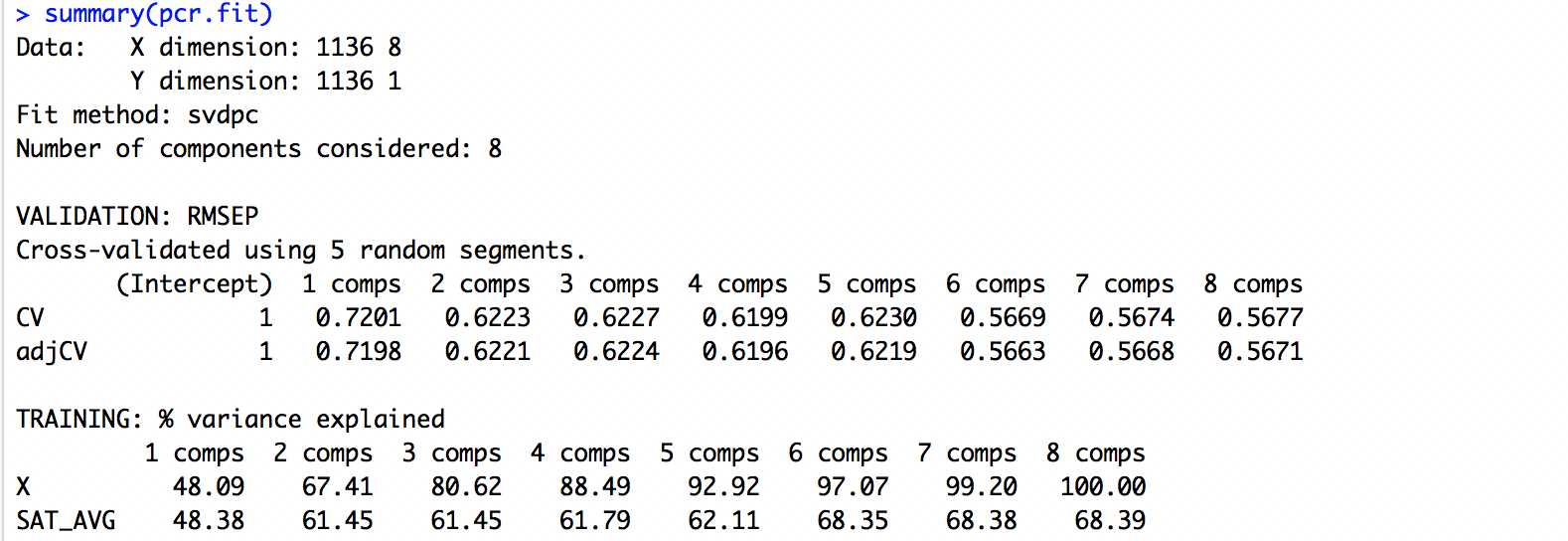
(e)

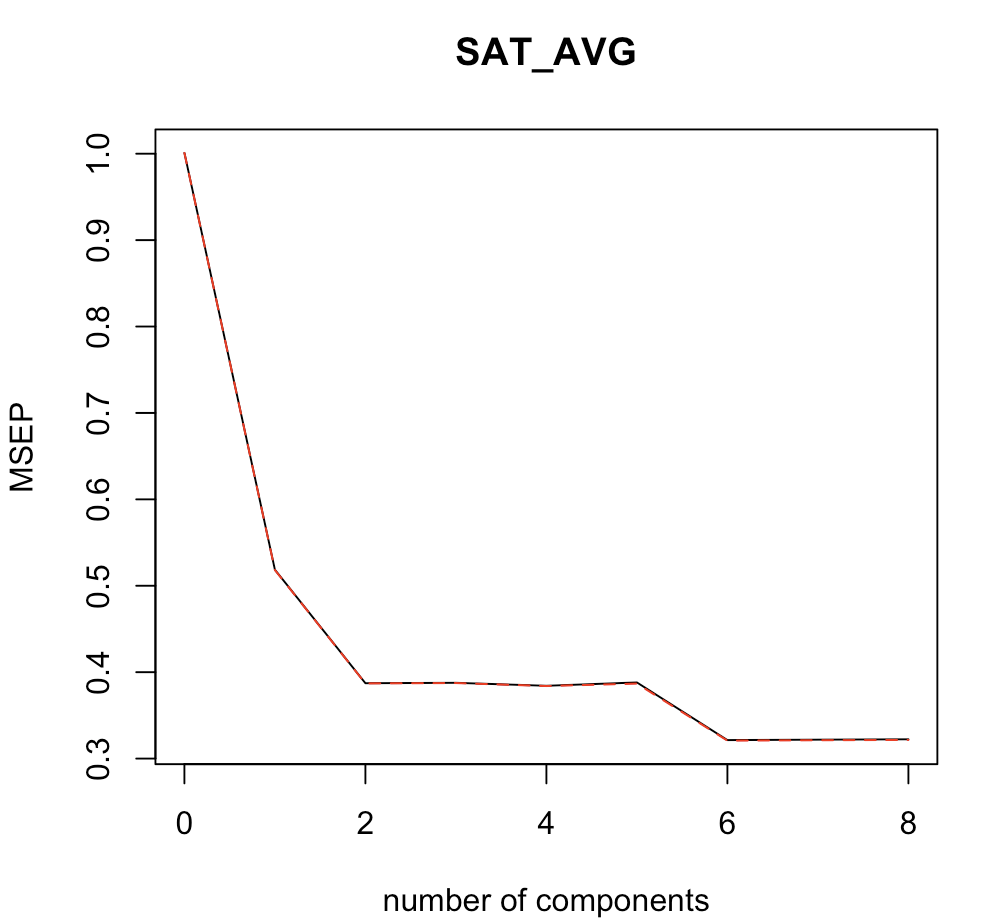
PC1 indicates a linear combination of COSTT4\_A, TUITFTE, TUITIONFEE\_OUT, PFTFTUG1\_EF, C150\_4, SAT\_AVG and AVGFACSAL, and obviously that PFTFTUG1\_EF plays an important part on it. While PC2 indicates another linear combination of UGDS, PFTFTUG1\_EF, C150\_4, SAT\_AVG, AVGFACSAL, COSTT4\_A, TUITFTE and TUITIONFEE\_OUT, and obviously that UGDS, PFTFTUG1\_EF play an important part on it. In addition, PC1 and PC2 are orthogonal.



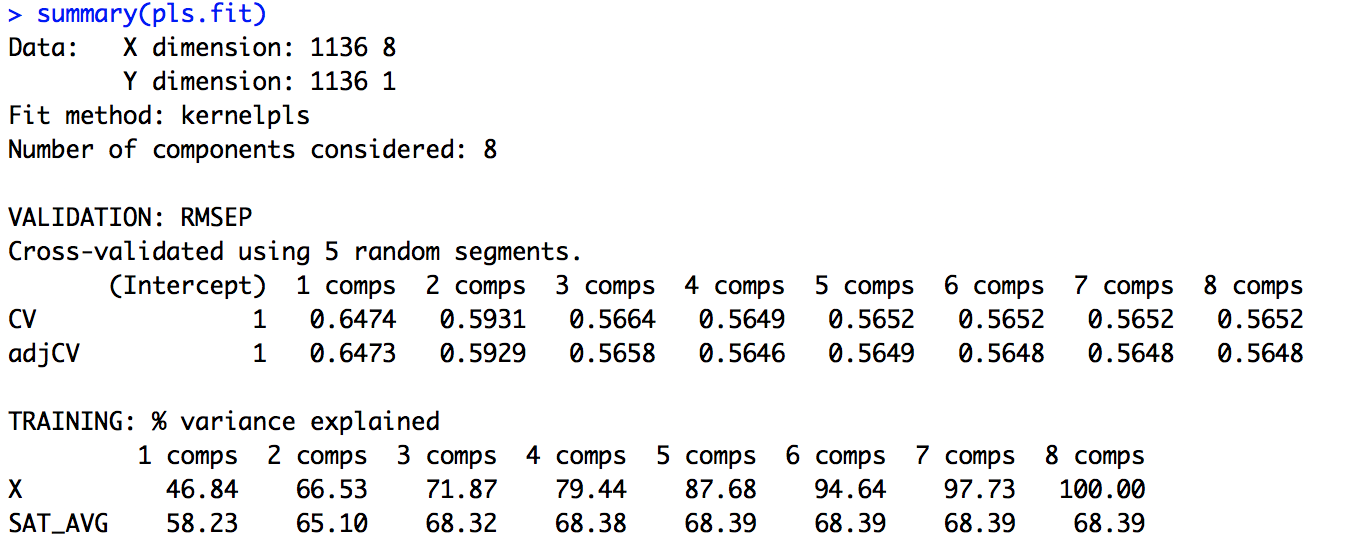
(f)

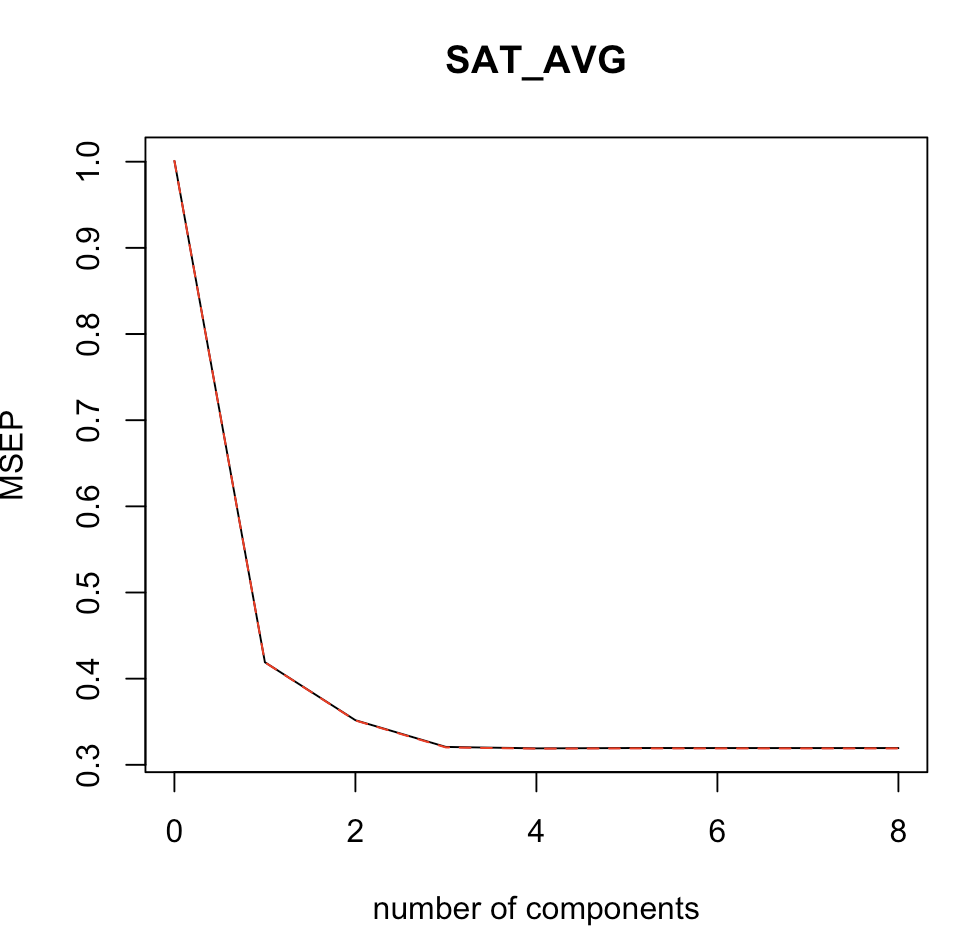
Below is the summary for PCR, I will choose M=5 because it explains 92% of the variance.





Below is the summary for PLS, I will choose M=6 because it explains 94% of the variance.





PCR works better because it only requires M=5 while PCR asks for M=6