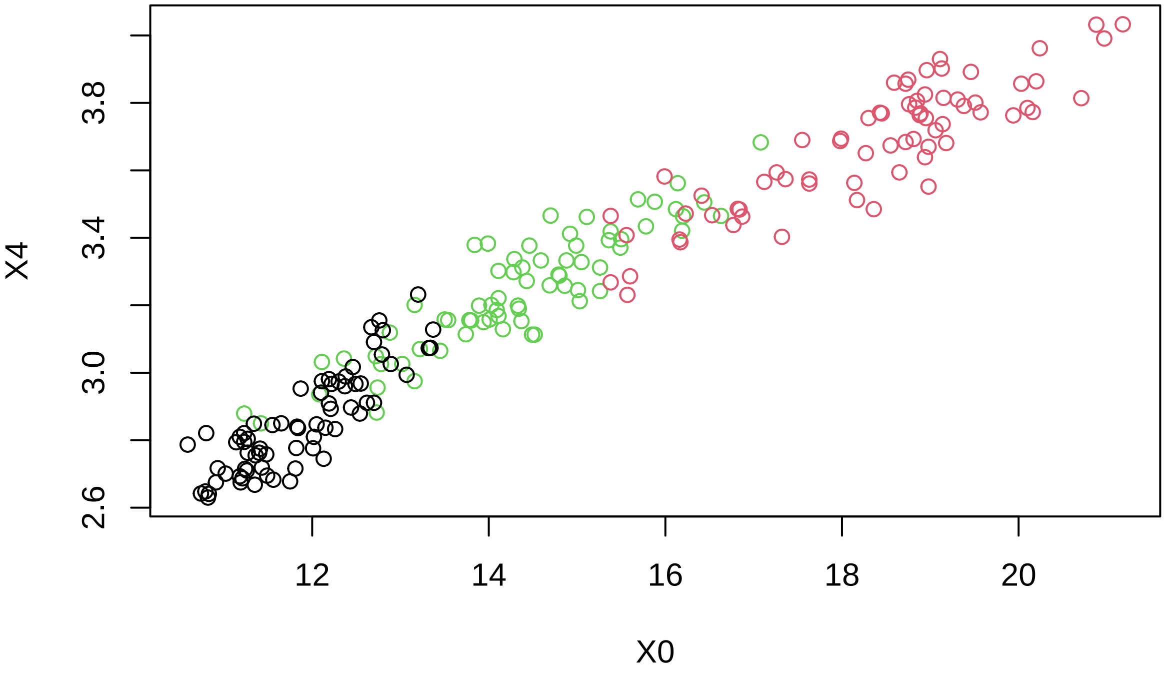
Problem 4

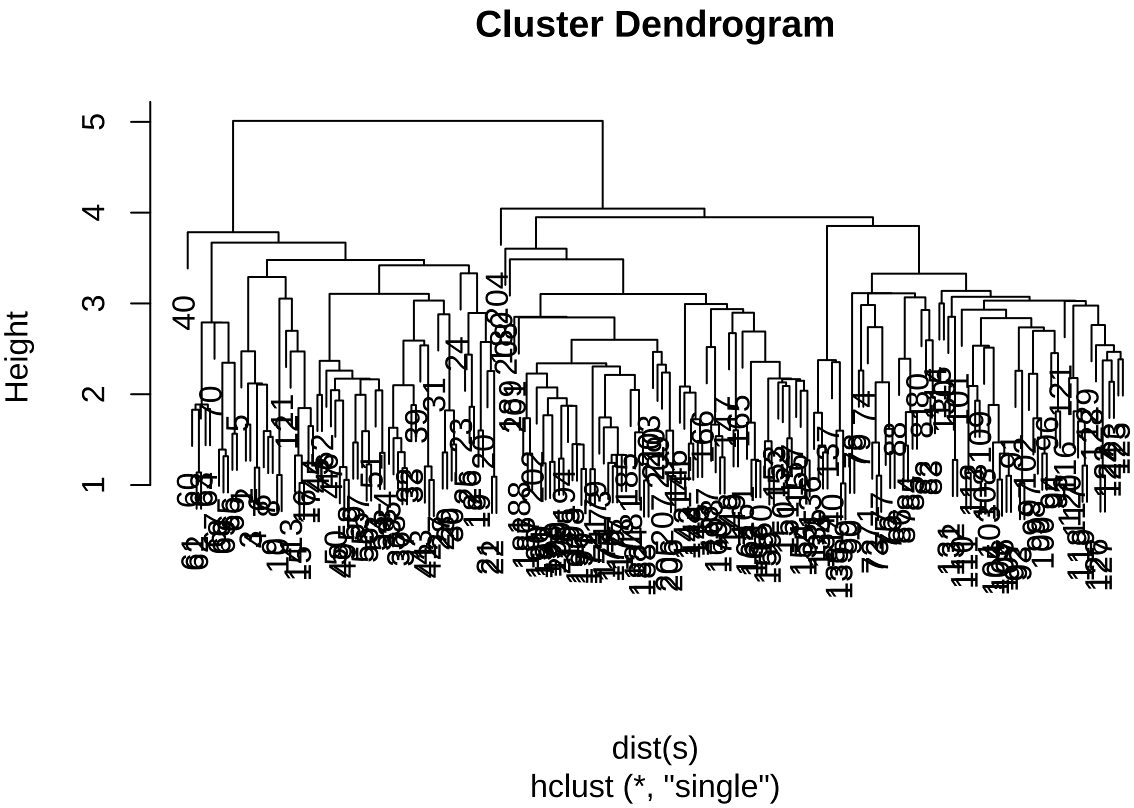
a)

```
set.seed(2017)

s <- read.csv("SeedData.csv")
attach(s)

km.out <- kmeans(s,3,nstart=20)

plot(X4~X0,col=km.out$cluster)
```



b)

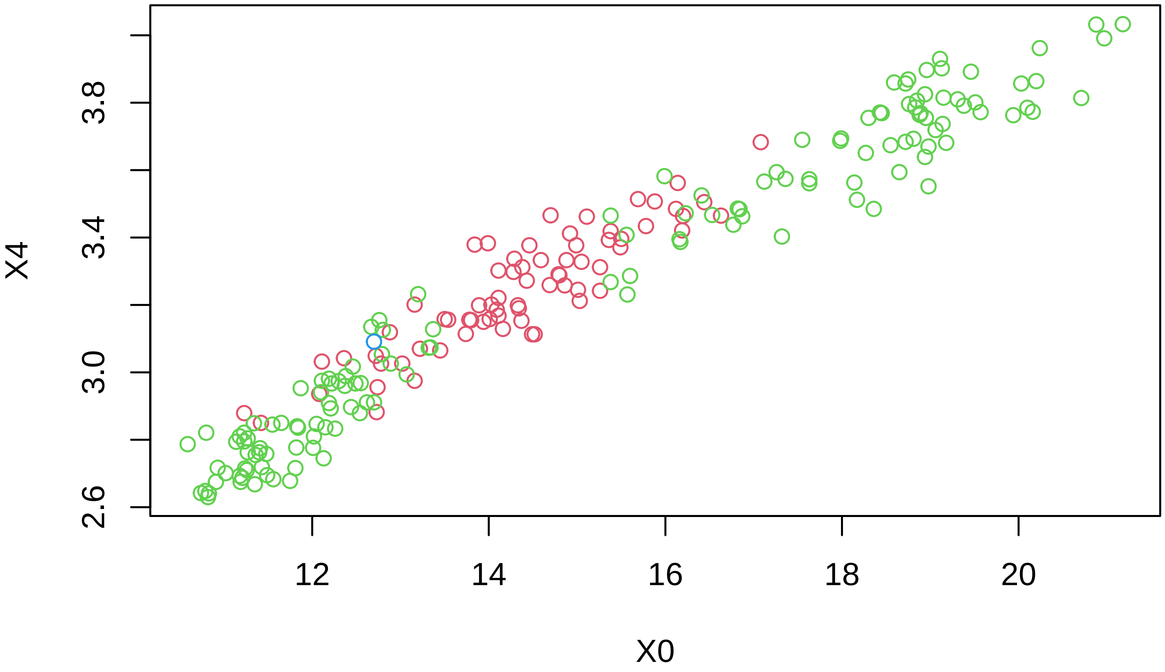
```
cl_sing <- hclust(dist(s),method="single")
plot(cl_sing)
```



c)

```r
three <- cutree(cl_sing,3)
```

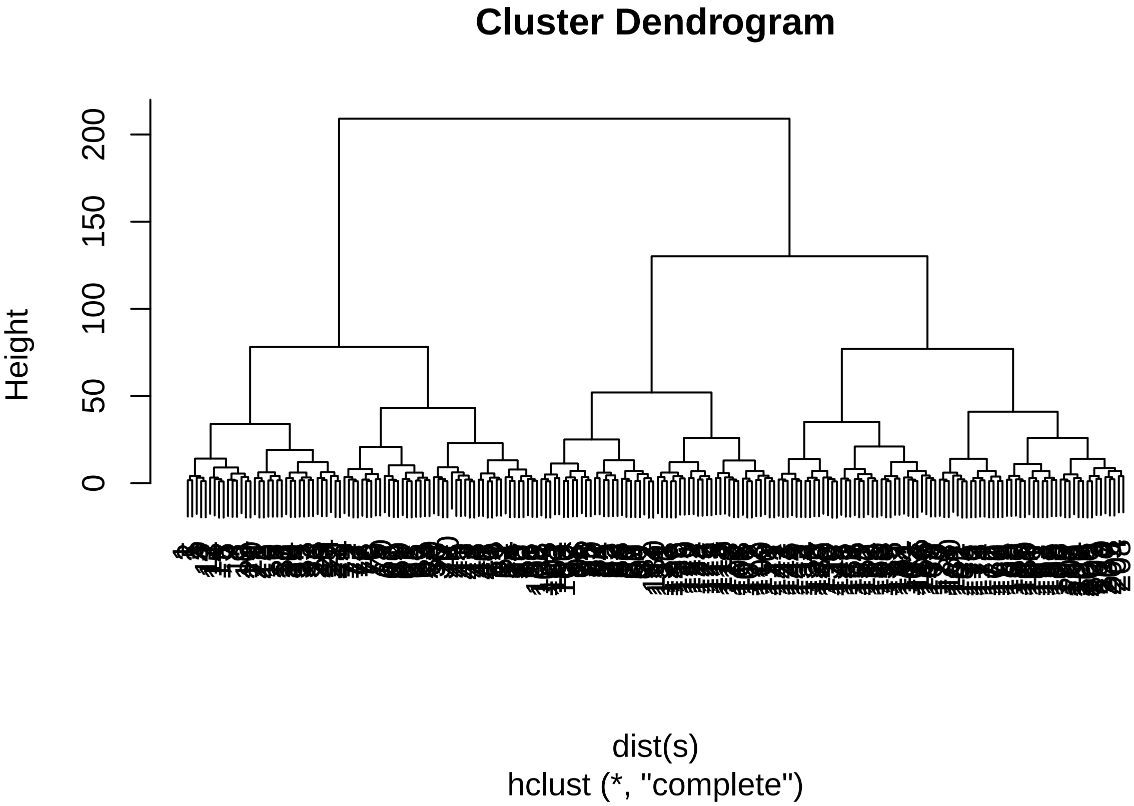
```r
plot(X4~X0,col=three+1)
```



The hierarchical clustering has not done as well at separating the classes in terms of the ability to separate along these two variables. K means clearly broke values into three separate clusters while this graph show that the red group is sorouned on the high and low ends by green observations that have been put in the same category. This may be acceptable if this differentiate is unimportant.

d)

```
cl_sing <- hclust(dist(s),method="complete")
plot(cl_sing)
```



I prefer complete clustering because it leads to a cleaner tree. There is a lot of low level clustering and then the tree breaks clearly into 3 parts.

e)

If I had the class labels I would use multinomial logistic regression to classify the seeds based on the predictors becase p is small and the observations don't appear to be totally seperable.