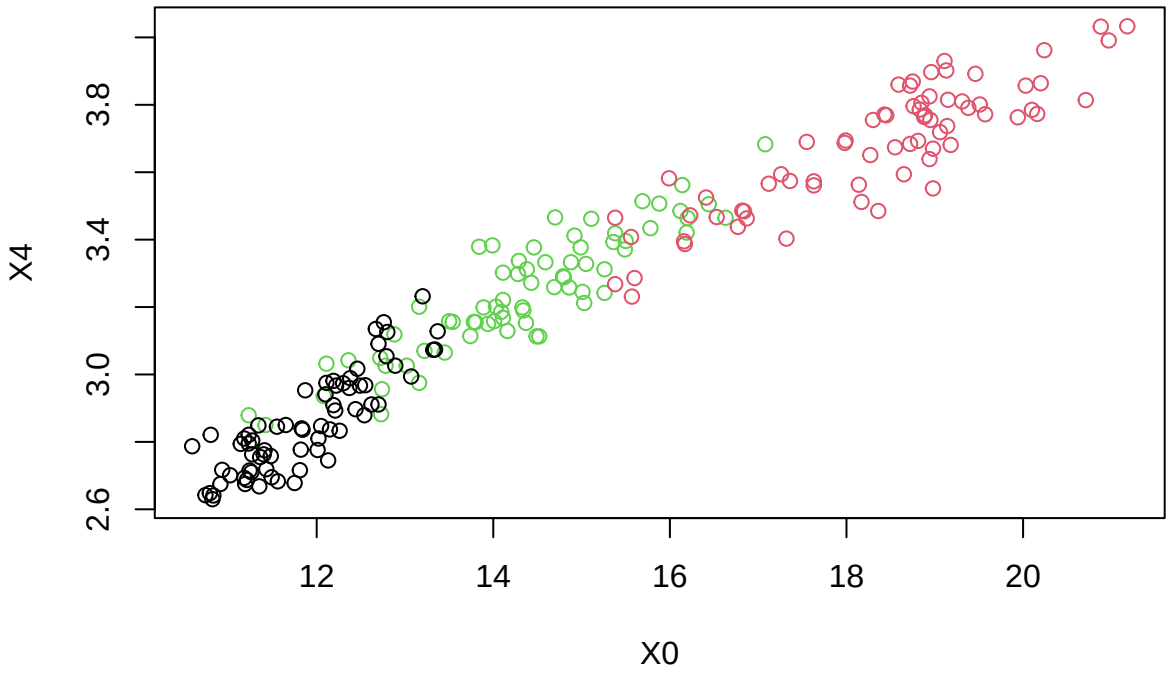
Problem 4

a)

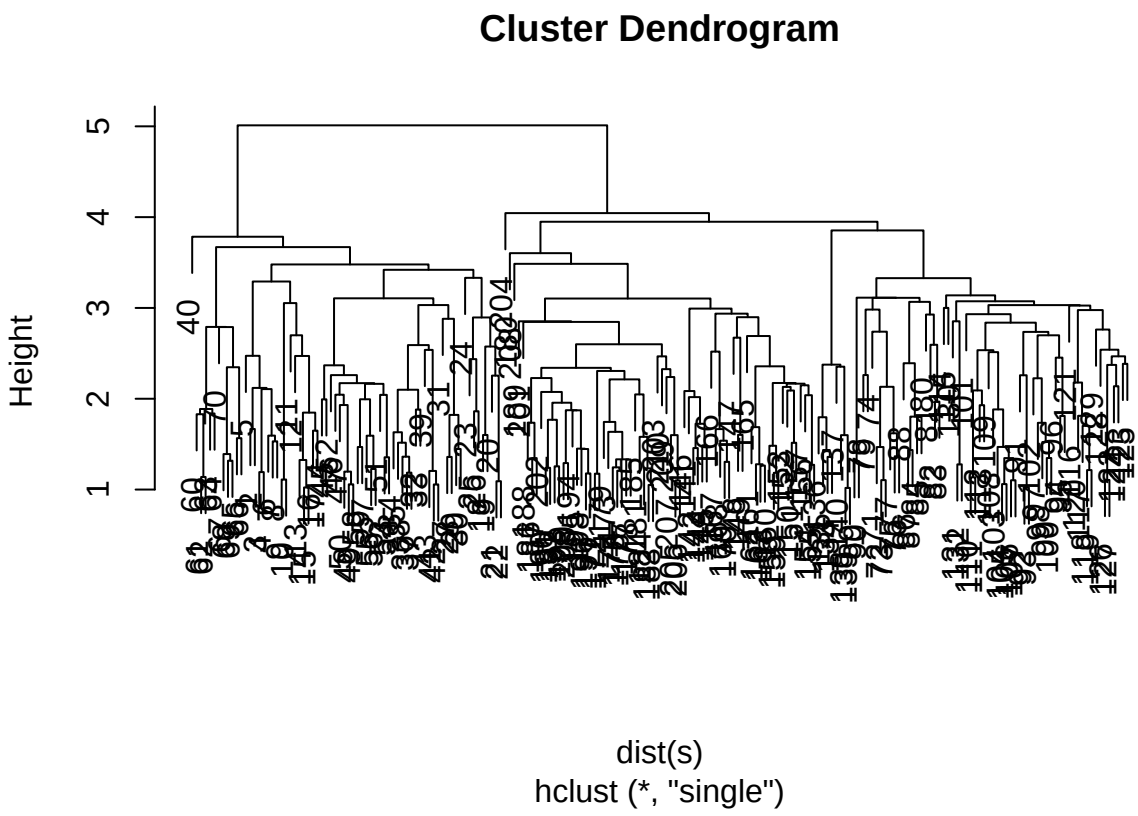
```
set.seed(2017)

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

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

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




```
# par(mfrow=c(1,2))
# plot(s, col=(km.out$cluster + 1), pch=20,cex=2)
```

b)

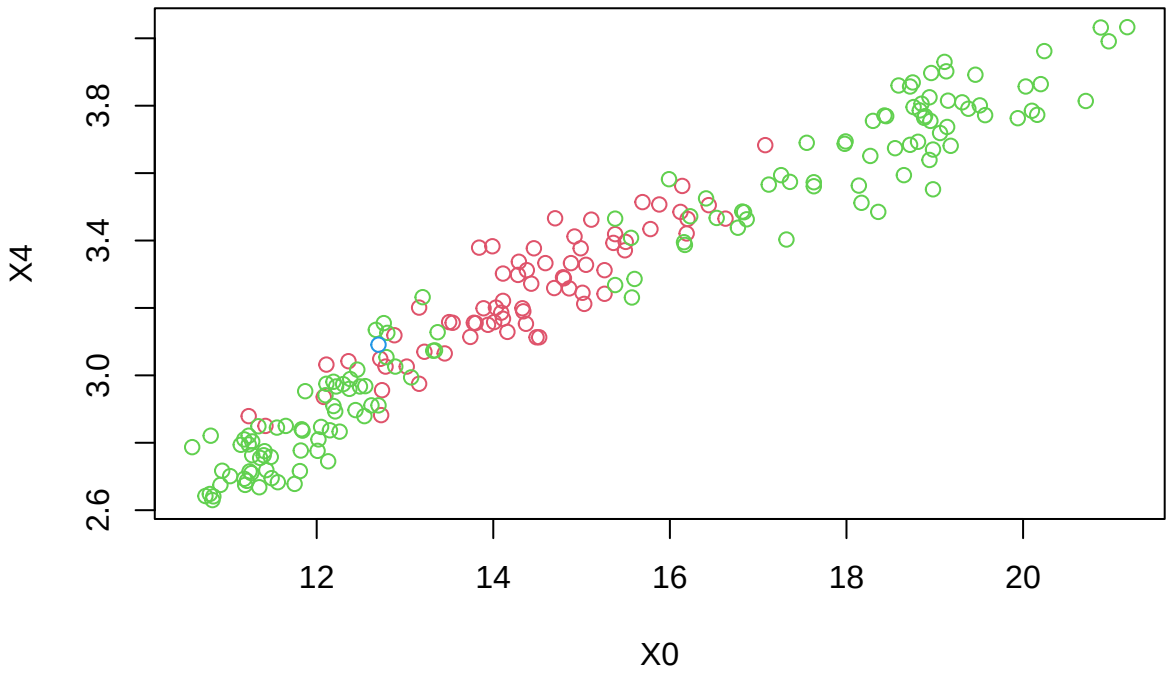
```
cl_sing <- hclust(dist(s),method="single")
plot(cl_sing)
```



c)

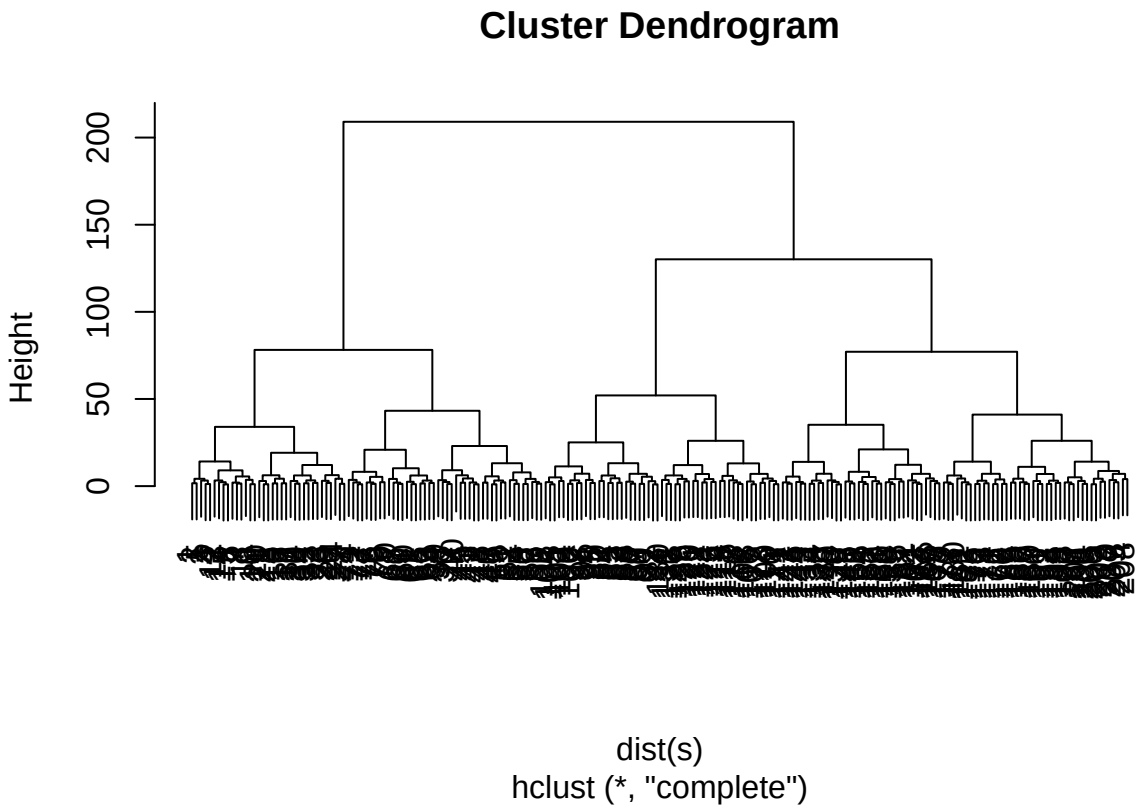
```
three <- cutree(cl_sing,3)
```

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



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.