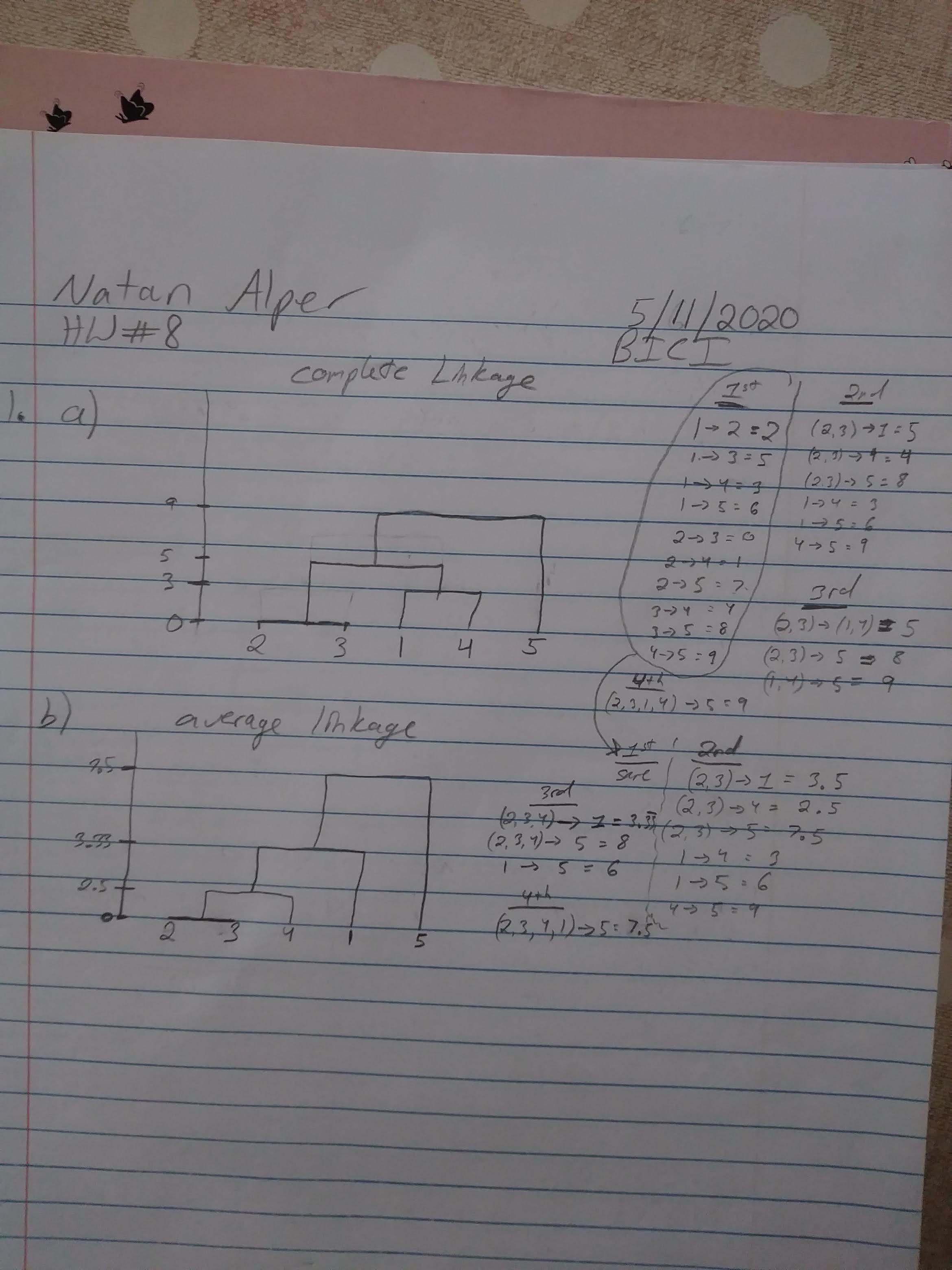
Aaron Zazulia & Natan Alper 5/11/2020

Business Intelligence & Consumer Insights- Professor Kovtun

HW #8

**1** (**a** & **b**)



**2**

### Q2 ###

## a ##

> data = data[,c(1,3,5,11,12,13)]

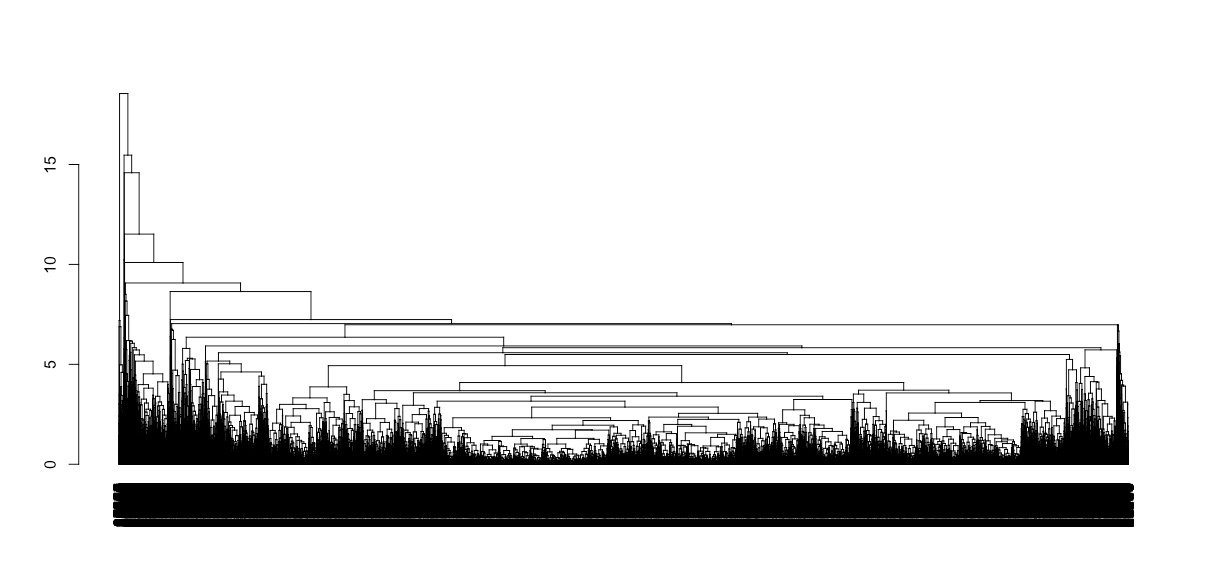
> for(k in 1:ncol(data)){

+ data[,k]=(data[,k]-mean(data[,k]))/sd(data[,k])

+ }

> Manh.Distances= dist(data,method="manhattan")

> Clusters=hclust(Manh.Distances,method="average")

> plot(as.dendrogram(Clusters))

## b ##

> Groups=cutree(Clusters,k=2)

## c ##

Split=cut(as.dendrogram(Clusters),h=1) ## Divides the tree by branches below height=1 ("lower") and branches above height=1 ("upper")

length(Split$lower)

plot(Split$lower[[1]]) ## We see which observations get grouped together first left-to-right

**3**

### Q3a ###

> cerQuant <- Cereal[,(seq(4,15,1))]

> Cereal[,17] <- NA

>

> Cereal$V17[Cereal$rating<=50] <- 0

> Cereal$V17[Cereal$rating>50] <- 1

>

> # Standardize

> for(k in 1:ncol(cerQuant)){

+ cerQuant[,k]=(cerQuant[,k]-mean(cerQuant[,k]))/sd(cerQuant[,k])

+ }

>

> eucl.Distances = dist(CerQuant, method = "euclidean")

> Clusters=hclust(eucl.Distances,method="single")

>

> Groups=cutree(Clusters,k=2)

>

> # Group 1 that are >50

> sum(Cereal$rating[Groups==1]>50)

[1] 20

> # Group 1 that are <=50

> sum(Cereal$rating[Groups==1]<=50)

[1] 56

>

> ## Therefore Group 1 should be predicted as 0 (<=50)

> Pred=rep(NA,77)

> Pred[Groups==1]=0

>

> # Group 2 that are >50

> sum(Cereal$rating[Groups==2]>50)

[1] 1

> # Group 2 that are <=50

> sum(Cereal$rating[Groups==2]<=50)

[1] 0

>

> ## Therefore Group 2 should be predicted as " >50"

> Pred[Groups==2]=1

>

> missclass=sum(Cereal$V17!=Pred)/77

> missclass

[1] 0.2597403

### Q3b ###

> FPR=c()

> TPR=c()

>

> # If we want to compute the FPR for ">50" class then we need to see the percentage of "<=50" individuals that are predicted to be ">50":

> under50obs=(1:77)[Cereal$V17==0]

> FPR=sum(Cereal$V17[under50obs]!=Pred[under50obs])/length(Pred[under50obs])

> FPR # All values that were predicted to be >50 were correct

[1] 0

>

> # If we want to compute the TPR for ">50" class then we need to see the percentage of ">50" individuals that are predicted to be ">50":

> over50obs=(1:77)[Cereal$V17==1]

> TPR=sum(Cereal$V17[over50obs]==Pred[over50obs])/length(Pred[over50obs])

> TPR

[1] 0.04761905