

Group Discussion 10 Final Answer

November 1, 2020

10.11

(a)

```
ch10num11 <- read.csv(here::here("group-discussion-10", "Ch10Ex11.csv"), header = FALSE)
dplyr::glimpse(ch10num11)
```

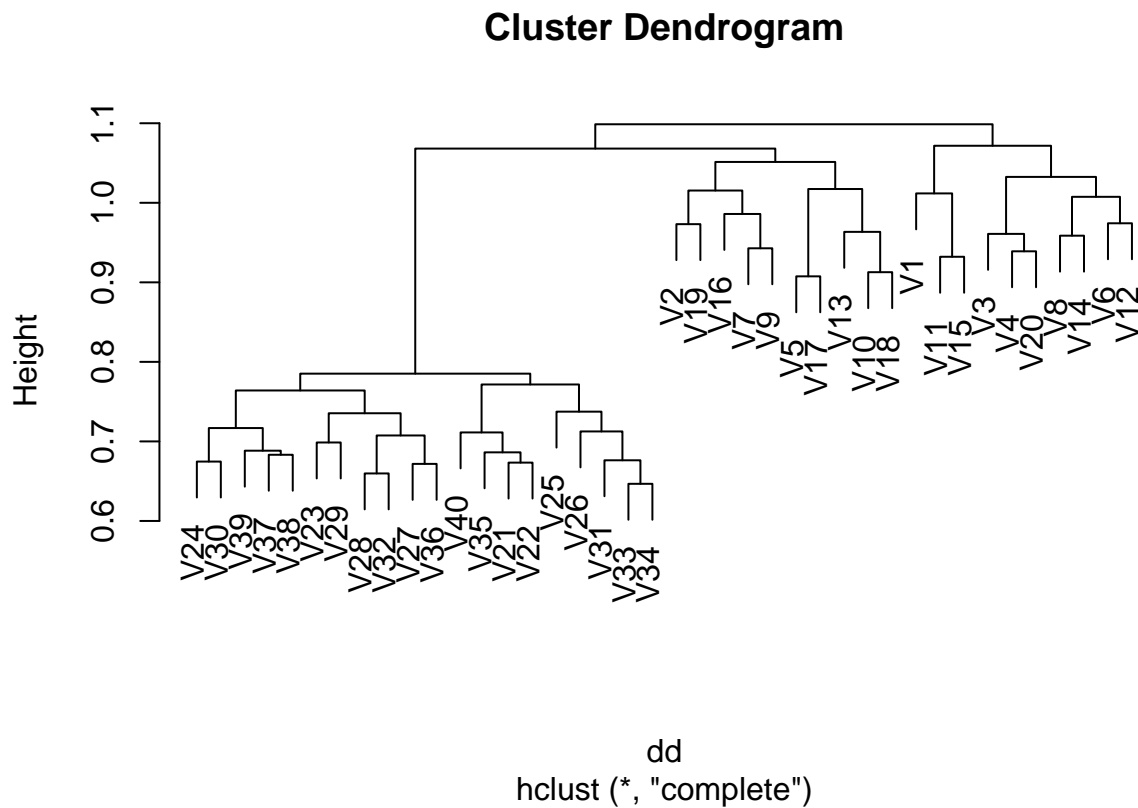
```
## Rows: 1,000
## Columns: 40
## $ V1 <dbl> -0.96193340, -0.29252570, 0.25878820, -1.15213200, 0.19578280, ...
## $ V2 <dbl> 0.441802800, -1.139267000, -0.972844800, -2.213168000, 0.593305...
## $ V3 <dbl> -0.97500510, 0.19583700, 0.58848580, -0.86152490, 0.28299210, ...
## $ V4 <dbl> 1.417504000, -1.281121000, -0.800258100, 0.630925300, 0.2471472...
## $ V5 <dbl> 0.8188148, -0.2514393, -1.8203980, 0.9517719, 1.9786680, -0.364...
## $ V6 <dbl> 0.31629370, 2.51199700, -2.05892400, -1.16572400, -0.87101800, ...
## $ V7 <dbl> -0.02496682, -0.92220620, -0.06476437, -0.39155860, -0.98971500...
## $ V8 <dbl> -0.06396600, 0.05954277, 1.59212400, 1.06361900, -1.03225300, ...
## $ V9 <dbl> 0.03149702, -1.40964500, -0.17311700, -0.35000900, -1.10965400,...
## $ V10 <dbl> -0.35031060, -0.65671220, -0.12108740, -1.48905800, -0.38514230...
## $ V11 <dbl> -0.7227299, -0.1157652, -0.1875790, -0.2432189, 1.6509570, -0.2...
## $ V12 <dbl> -0.2819547, 0.8259783, -1.5001630, -0.4330340, -1.7449090, 2.17...
## $ V13 <dbl> 1.33751500, 0.34644960, -1.22873700, -0.03879128, -0.37888530, ...
## $ V14 <dbl> 0.70197980, -0.56954860, 0.85598900, -0.05789677, -0.67982610, ...
## $ V15 <dbl> 1.0076160, -0.1315365, 1.2498550, -1.3977620, -2.1315840, 0.298...
## $ V16 <dbl> -0.46538280, 0.69022900, -0.89808150, -0.15618710, -0.23017180,...
## $ V17 <dbl> 0.63859510, -0.90903820, 0.87020580, -2.73598200, 0.46612430, 0...
## $ V18 <dbl> 0.28678070, 1.30264200, -0.22525290, 0.77561690, -1.80044900, ...
## $ V19 <dbl> -0.22707820, -1.67269500, 0.45028920, 0.61415620, 0.62629040, 0...
## $ V20 <dbl> -0.22004520, -0.52550400, 0.55144040, 2.01919400, -0.09772305, ...
## $ V21 <dbl> -1.24257300, 0.79797000, 0.14629430, 1.08113900, -0.29971080, 0...
## $ V22 <dbl> -0.1085056, -0.6897930, 0.1297400, -1.0766180, -0.5295591, 1.06...
## $ V23 <dbl> -1.8642620, 0.8995305, 1.3042290, -0.2434181, -2.0235670, 1.230...
## $ V24 <dbl> -0.50051220, 0.42858120, -1.66190800, 0.51348220, -0.51084020, ...
## $ V25 <dbl> -1.32500800, -0.67611410, -1.63037600, -0.51285780, 0.04600274,...
## $ V26 <dbl> 1.06341100, -0.53409490, -0.07742528, 2.55167600, 1.26803000, ...
## $ V27 <dbl> -0.29637120, -1.73250700, 1.30618200, -2.31430100, -0.74398680,...
## $ V28 <dbl> -0.12164570, -1.60344700, 0.79260020, -1.27647000, 0.22313190, ...
## $ V29 <dbl> 0.08516605, -1.08362000, 1.55946500, -1.22927100, 0.85846280, ...
## $ V30 <dbl> 0.62417640, 0.03342185, -0.68851160, 1.43439600, 0.27472610, 0....
## $ V31 <dbl> -0.5095915, 1.7007080, -0.6154720, -0.2842774, -0.6929984, 1.15...
## $ V32 <dbl> -0.216725500, 0.007289556, 0.009999363, 0.198945600, -0.8457072...
```

```
## $ V33 <dbl> -0.05550597, 0.09906234, 0.94581000, -0.09183320, -0.17749680, ...
## $ V34 <dbl> -0.4844491, 0.5638533, -0.3185212, 0.3496279, -0.1664908, 0.182...
## $ V35 <dbl> -0.52158110, -0.25727520, -0.11788950, -0.29890970, 1.48315500,...
## $ V36 <dbl> 1.94913500, -0.58178050, 0.62136620, 1.51369600, -1.68794600, -...
## $ V37 <dbl> 1.32433500, -0.16988710, -0.07076396, 0.67118470, -0.14142960, ...
## $ V38 <dbl> 0.46814710, -0.54230360, 0.40168180, 0.01085530, 0.20077850, -1...
## $ V39 <dbl> 1.06110000, 0.31293890, -0.01622713, -1.04368900, -0.67594210, ...
## $ V40 <dbl> 1.65597000, -1.28437700, -0.52655320, 1.62527500, 2.22061100, 0...
```

(b)

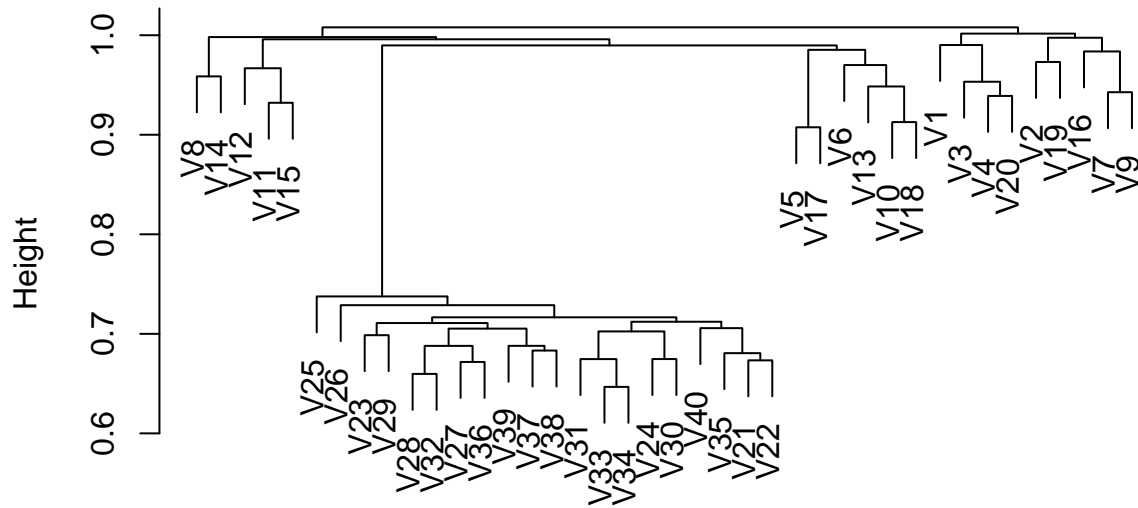
```
dd <- as.dist(1 - cor(ch10num11))

plot(hclust(dd, method = "complete"))
```



```
plot(hclust(dd, method = "average"))
```

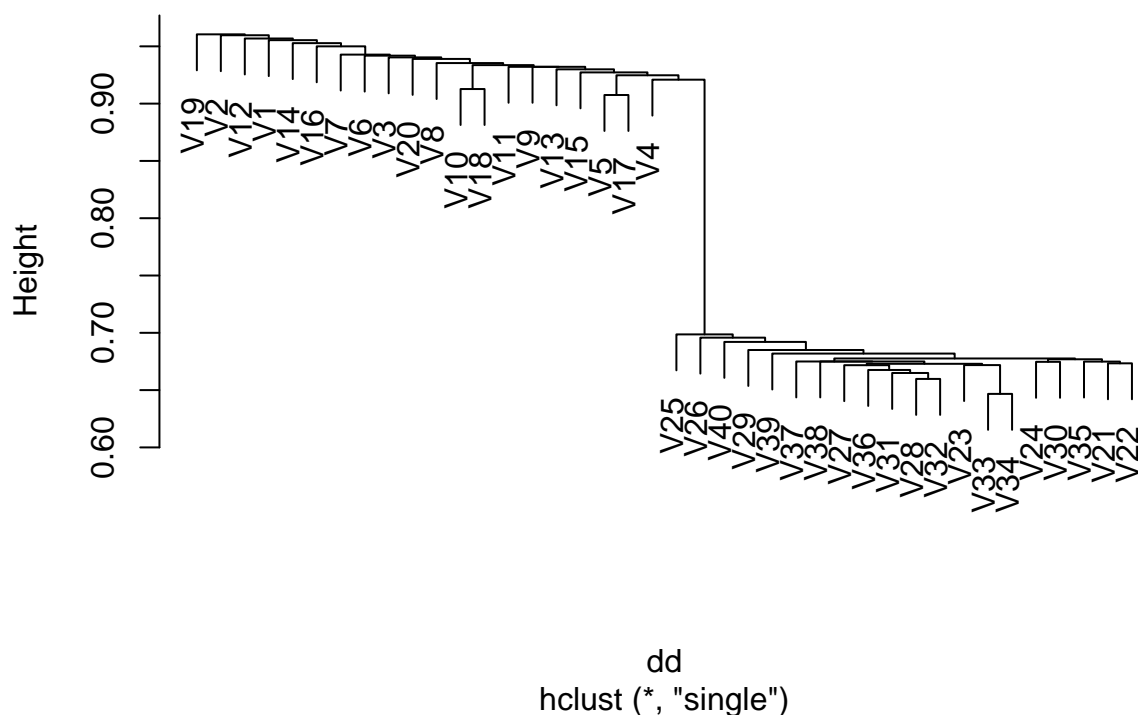
Cluster Dendrogram



dd
hclust (*, "average")

```
plot(hclust(dd, method = "single"))
```

Cluster Dendrogram



When using linkage types “complete” and “single”, we see two groups clustered. When using type “average”, there are three groups clustered.

(c)

We can use PCA. Then we can identify which genes have the largest effect by the principal component loadings (the rotation matrix of `prcomp` output).

```
pr_out <- prcomp(t(ch10num11), scale = TRUE)
genes_ranked <- order(abs(rowSums(pr_out$rotation))), decreasing = TRUE)
genes_ranked[1:10]
```

```
## [1] 889 676 755 960 907 19 475 673 374 174
```