MVA Project - PCA

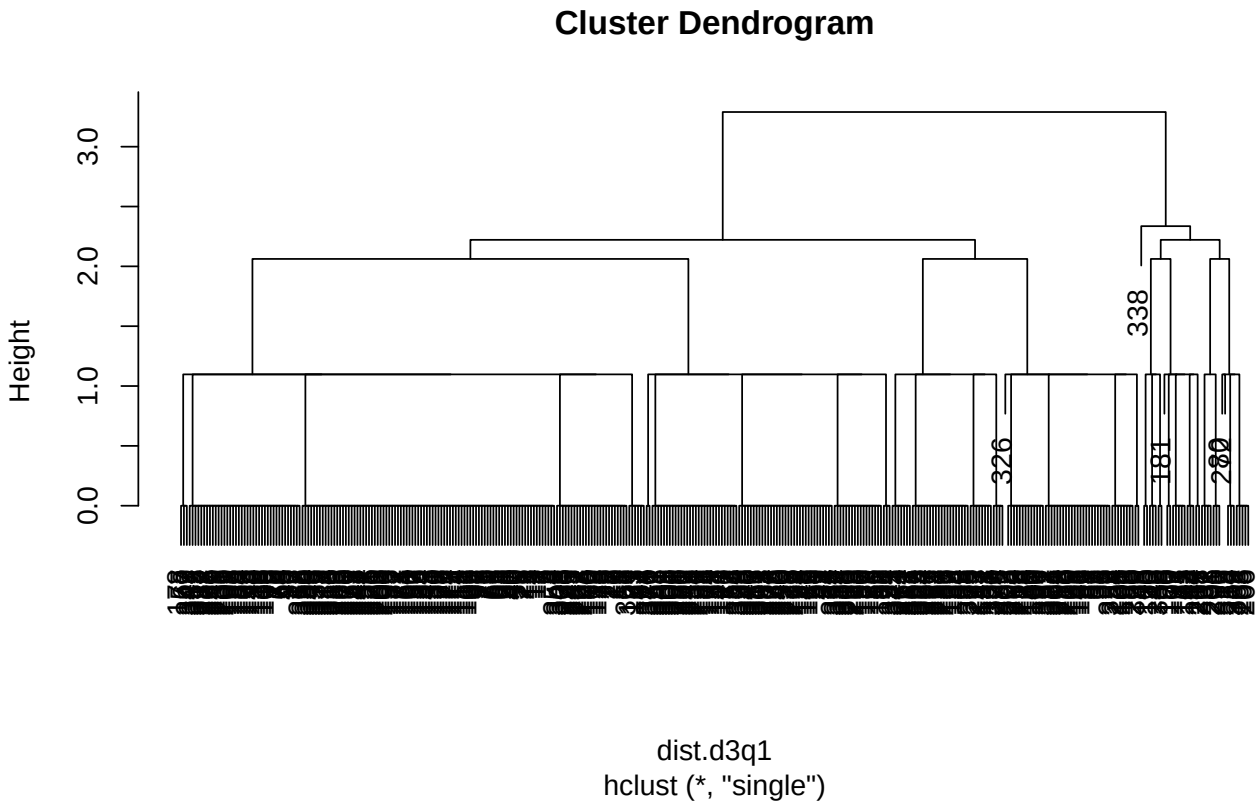
Chun-Jung Chen & Akshay Arora

Question 1: Deos Family conditions affect students' final grade in Math?

Cluster Result

- Clearly, we can't use dendrogram since there is too many rows

```
# Scaling

scale.d3q1 = scale(d3.q1[, -1])

dist.d3q1 = dist(scale.d3q1, method = "euclidean")
cluster.d3q1 = hclust(dist.d3q1, method = "single")

plot(cluster.d3q1)
```



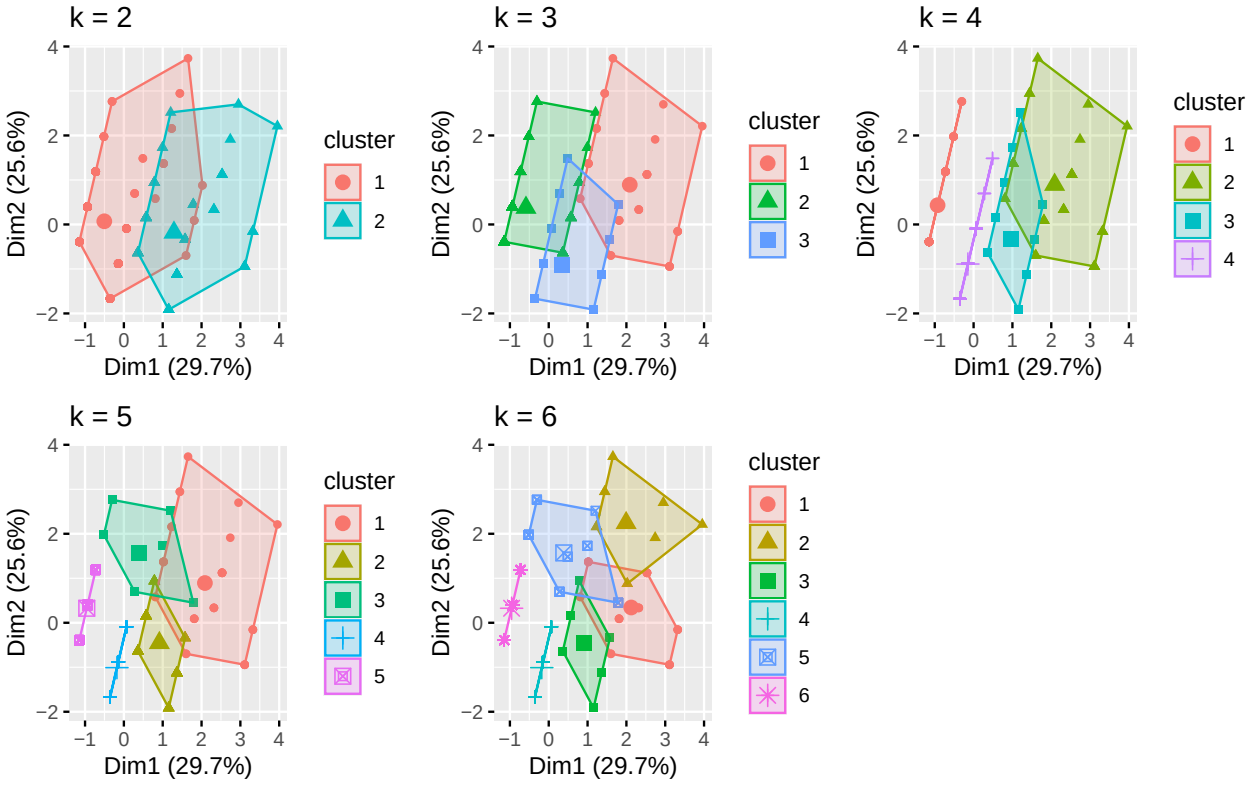
Kmeans

- We utilized Kmeans to separate dataset into 2 to 6 clusters and measure the variance of each cluster

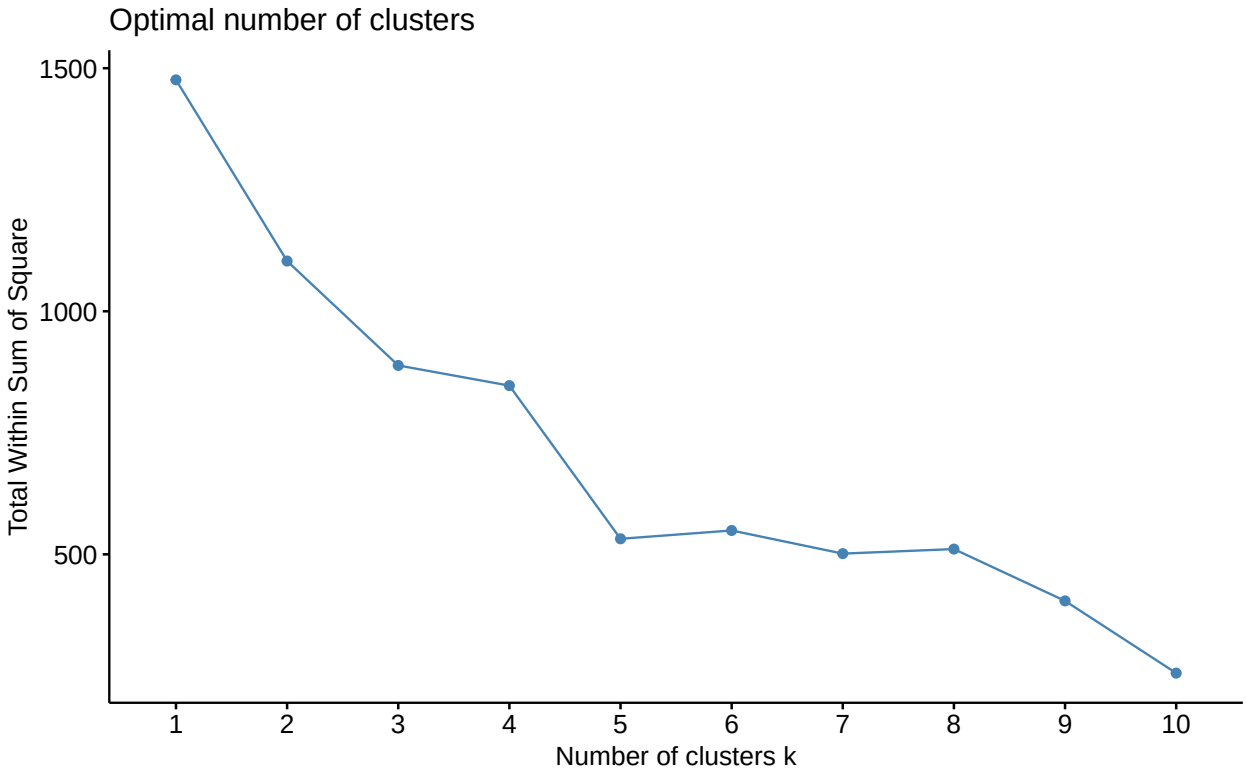
- Variance
  - Accroding to the variance of each cluster, we choose to separate our dataset into 5 groups.

```
##
## clu.Var2 74.2
## clu.Var3 51.8
## clu.Var4 36.8
## clu.Var5 30.2
## clu.Var6 27.3
```

- Cluster plot



- According to the plot below, we choose cluster 5 as the number to separate datasets.



Analysis of cluster 5

- we compare means in 5 groups
  - According to the table, group 1 has the largest mean comparing to other groups.
  - Majority of students tend to have family size with less than 3, living together, highly family relationship but without family support.

```
##             Mean
## Group1 11.157895
## Group2 11.075949
## Group3  9.857143
## Group4 10.575000
## Group5 10.000000
```

Question 2: Does parents' jobs and education level influence students' first period of grade in Math?

Cluster Result

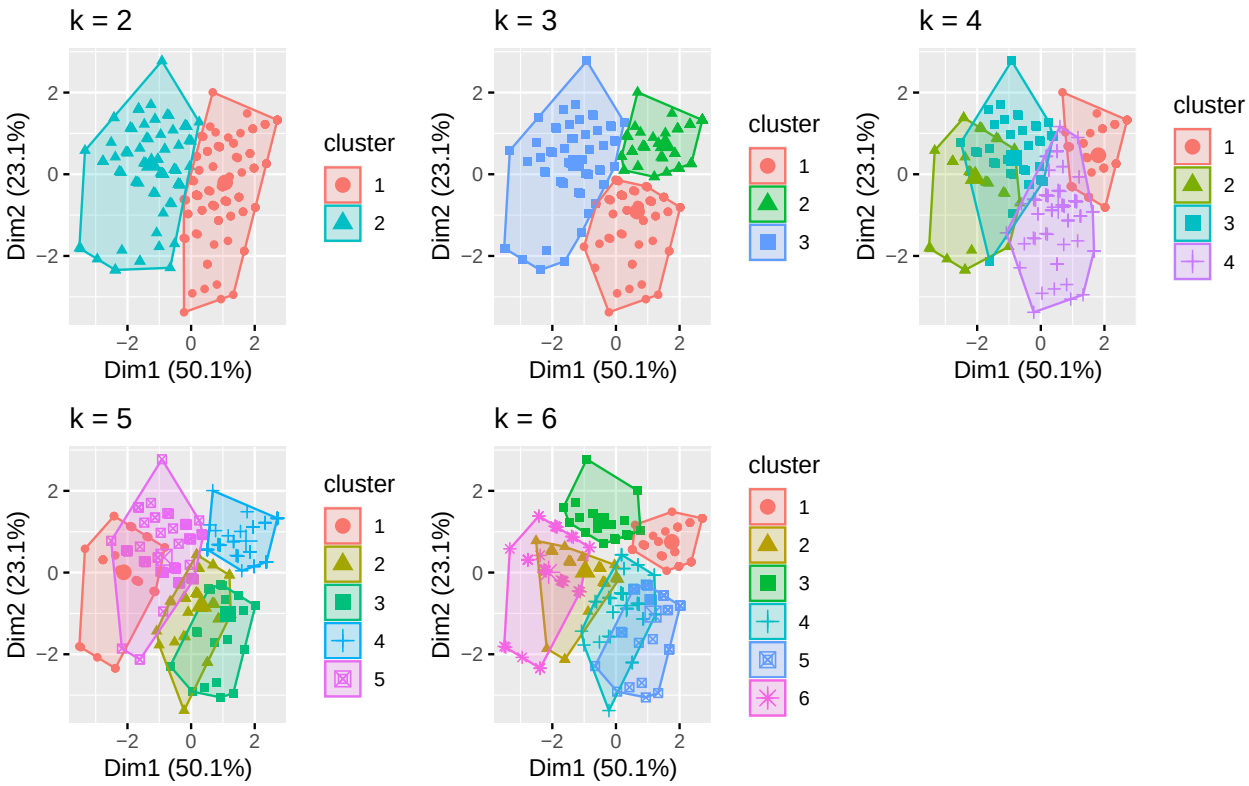
```
# Scaling

scale.d3q2 = scale(d3.q2[, -1])

dist.d3q2 = dist(scale.d3q2, method = "euclidean")
cluster.d3q2 = hclust(dist.d3q2, method = "single")
```

Kmeans

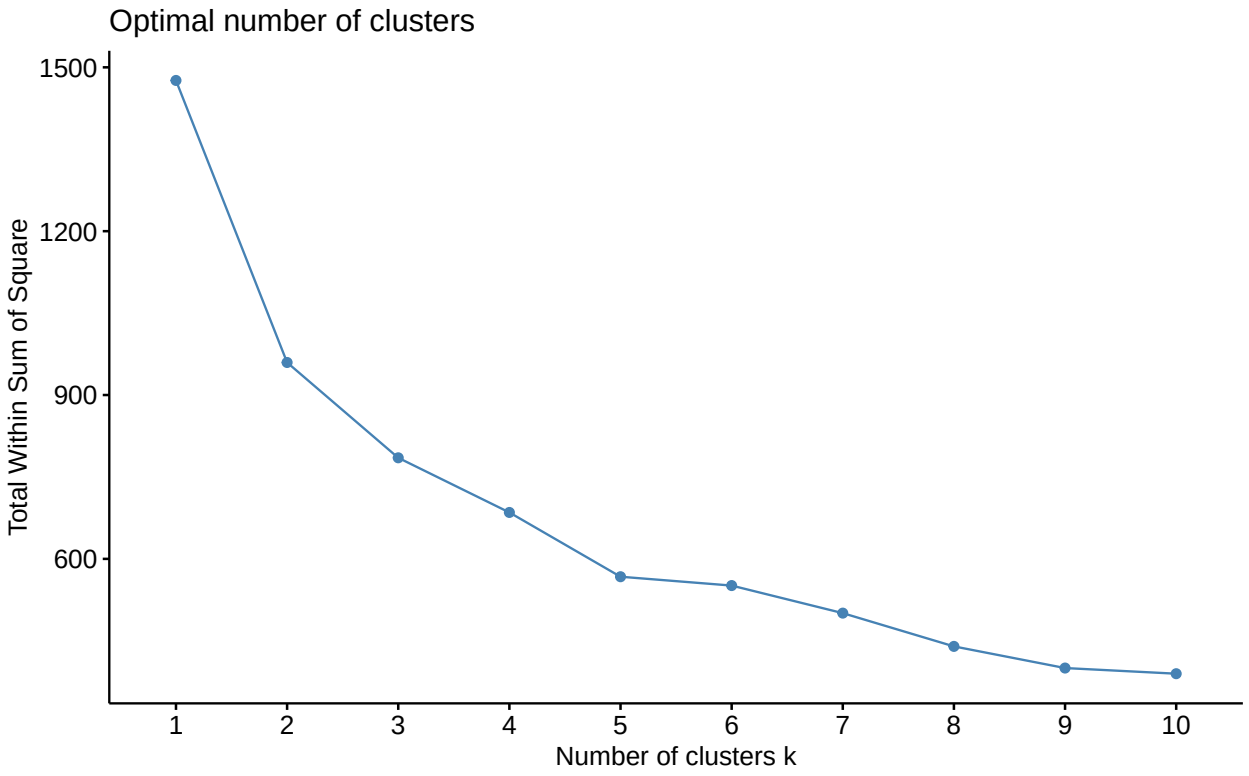
- We utilized Kmeans to separate dataset into 2 to 6 clusters and measure the variance of each cluster
- Variance
  - Accroding to the variance of each cluster, we choose to separate our dataset into 5 groups.

```
##
## clu.Var2 65.0
## clu.Var3 53.2
## clu.Var4 45.2
## clu.Var5 38.6
## clu.Var6 35.3
```

- Cluster plot



- According to the plot below, we choose cluster 5 as the number to separate datasets.



Analysis of cluster 5

- We compare means in 5 group
  - According to the table, group 4 has the largest mean in First Math score comparing to other groups.
  - Majority of students tend to have parents with higher education but they are unemployeed.

```
##          Mean
## grp1 10.27273
## grp2 11.20290
## grp3 11.40000
## grp4 11.95890
## grp5 10.16418
```

Question 3: Does student's learning conditions really impact students' final grade math score and Portuguese scores in average?

Cluster Result

```
# Scaling

scale.d3q3 = scale(d3.q3[, c(2, 4, 5, 6)])

dist.d3q3 = dist(scale.d3q3, method = "euclidean")
cluster.d3q3 = hclust(dist.d3q3, method = "single")
```

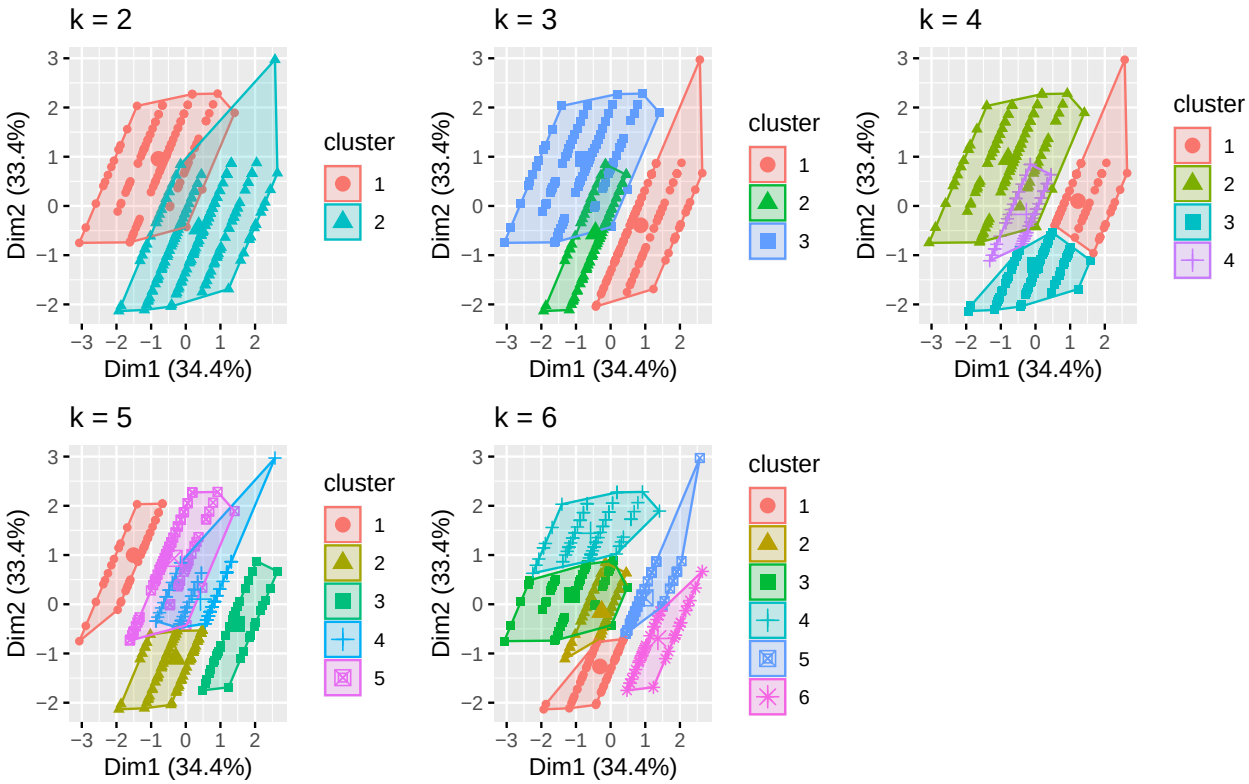
Kmeans

- We utilized Kmeans to separate dataset into 2 to 6 clusters and measure the variance of each cluster

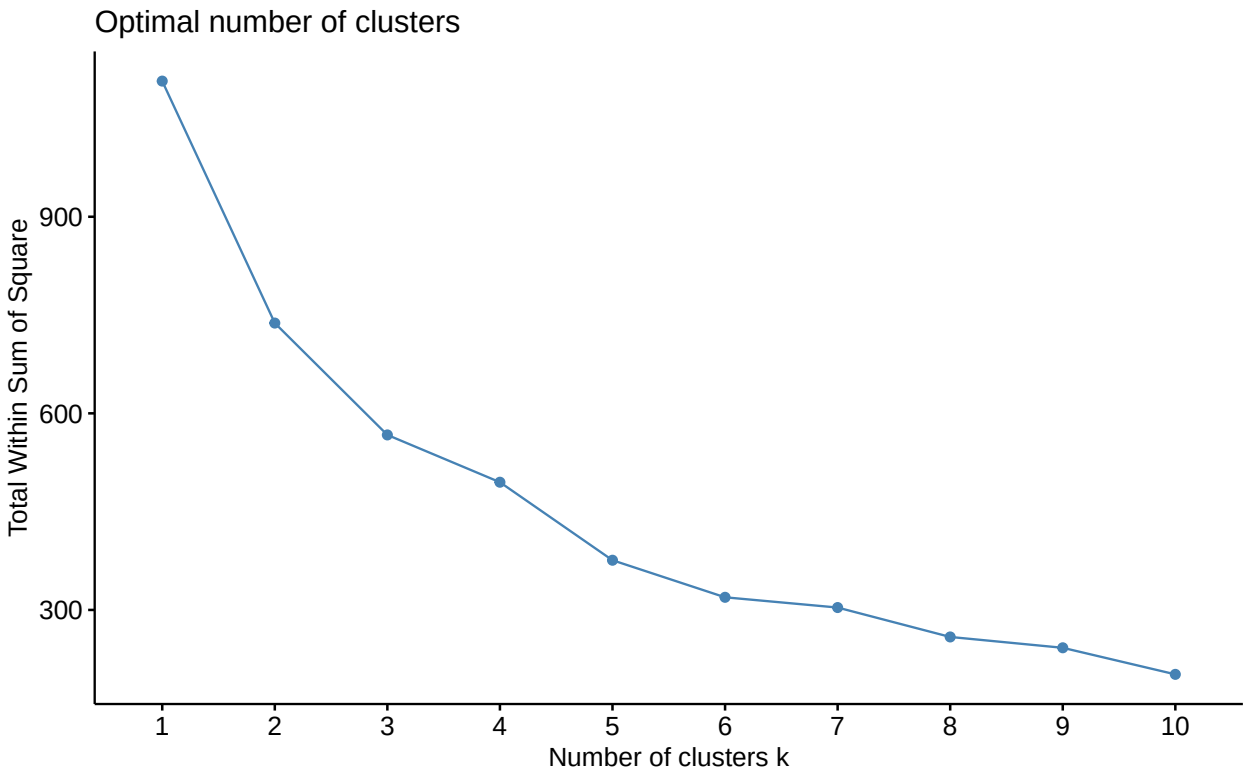
- Variance
  - Accroding to the variance of each cluster, we choose to separate out dataset into 5 groups.

```
##
## clu.Var2 66.6
## clu.Var3 51.2
## clu.Var4 41.0
## clu.Var5 34.5
## clu.Var6 28.6
```

- Cluster plot



- According to the plot below, we choose cluster 6 as the number to separate datasets.



Analysis of cluster 6

- we compare means in 6 groups
  - According to the table, group 4 has the largest mean in average of Final Math and Portuguese score comparing to other groups.
  - Majority of students tend to have Internet assess, parents who don't have romantic relationship, and less freetime.

```
##          Mean
## grp1 11.28261
## grp2 11.74219
## grp3 10.23913
## grp4 11.36301
## grp5 12.37342
## grp6 11.55128
```