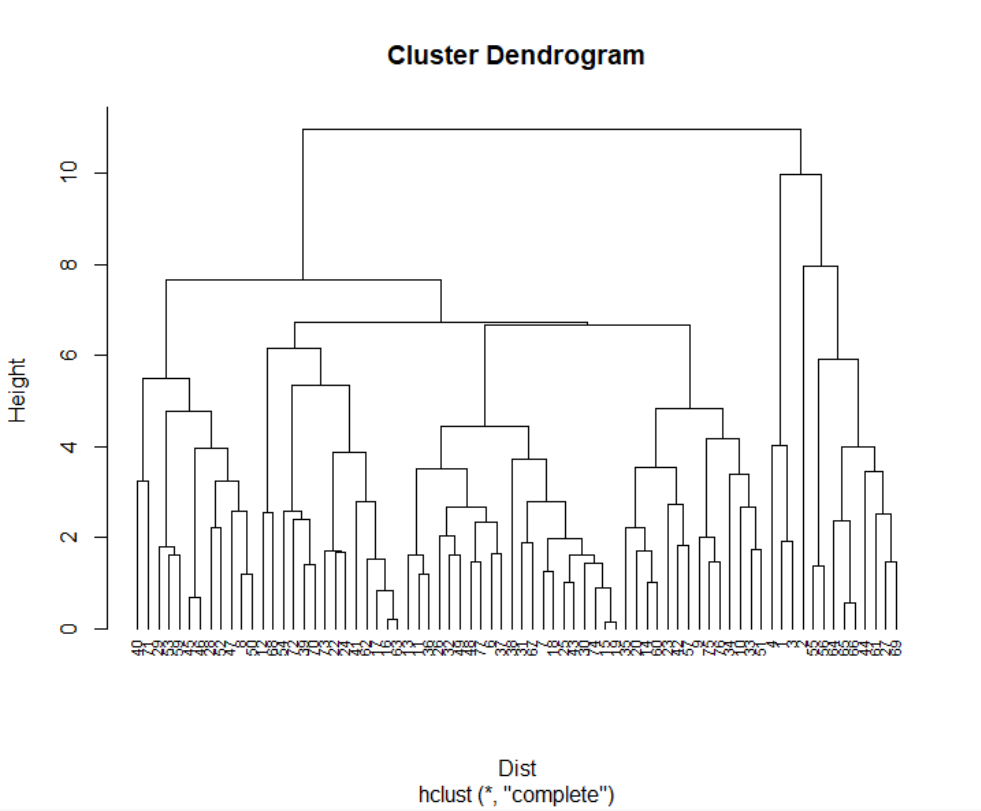
Chase Holland
11/30/2023

Assignment 5

```
> library(cluster)
> library(caret)
> library(dendextend)
> library(knitr)
> library(factoextra)
> library(readr)

> Cereals <- read.csv("C:/Users/choll/Downloads/Cereals.csv")
> NumData <- data.frame(Cereals[,4:16])
> NumData <- na.omit(NumData)
> CerealsNormalize <- scale(NumData)
```

Task 1

```
> Dist <- dist(CerealsNormalize, method = "euclidean")
> HClust <- hclust(Dist, method = "complete")
> plot(HClust, cex = 0.7, hang = -1)
```



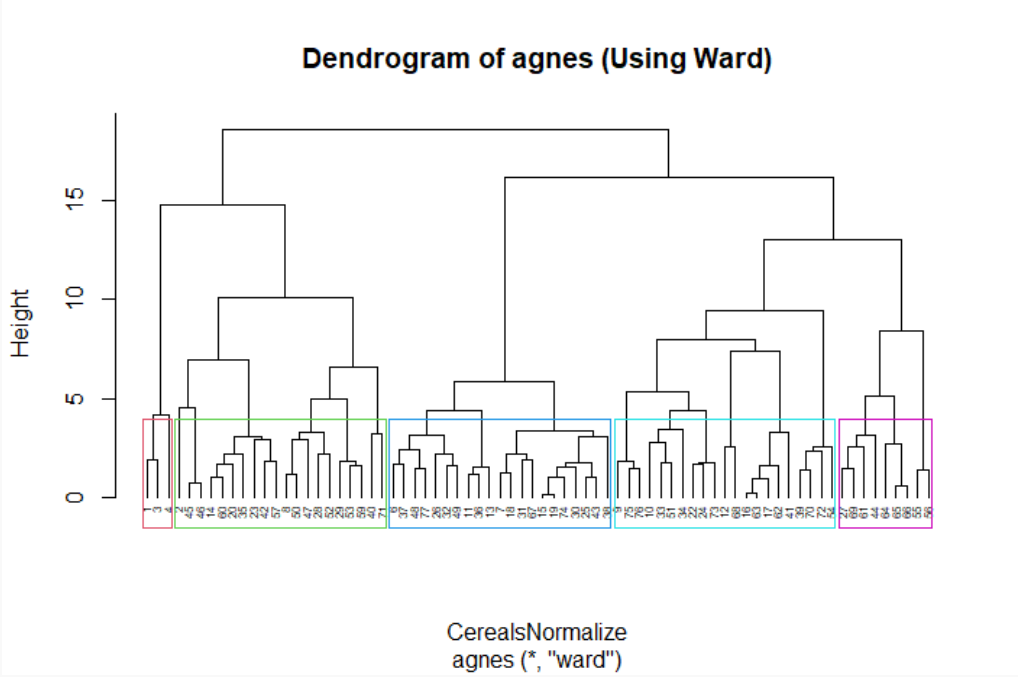

```
> SingleHClust <- agnes(CerealsNormalize, method = "single")
> CompleteHClust <- agnes(CerealsNormalize, method = "complete")
> AverageHClust <- agnes(CerealsNormalize, method = "average")
> wardHClust <- agnes(CerealsNormalize, method = "ward")

> print(SingleHClust$ac)
[1] 0.6067859

> print(CompleteHClust$ac)
[1] 0.8353712
```

```
> print(AverageHClust$ac)
[1] 0.7766075
> print(wardHClust$ac)
[1] 0.9046042
```

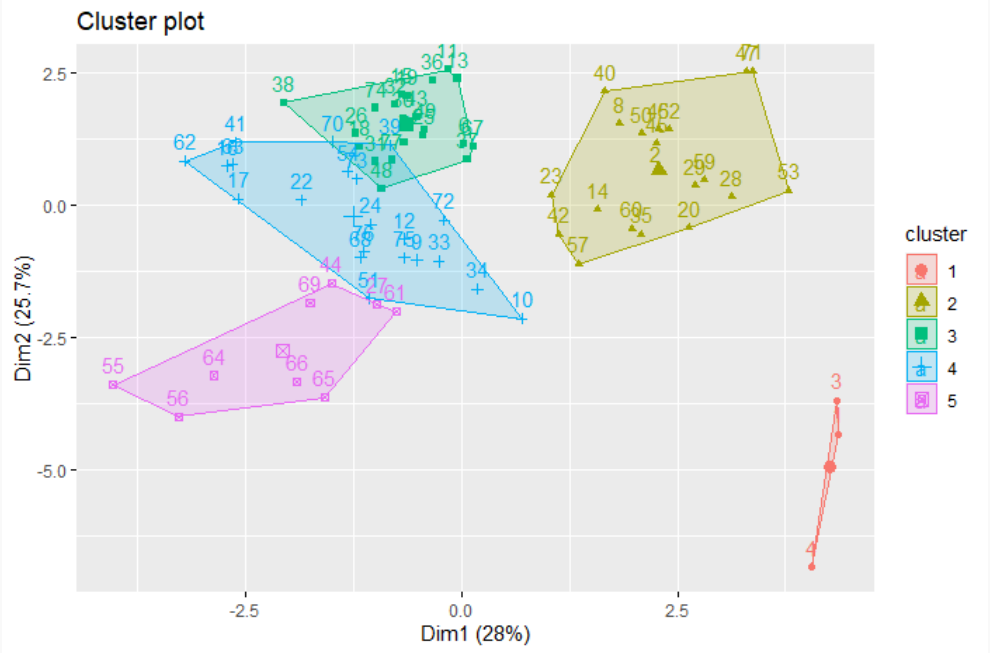
##Ward is the best option since its value is highest.

Task 2

```
> pltree(wardHClust, cex = 0.5, hang = -1, main = "Dendrogram of agnes (Using
Ward)")
> rect.hclust(wardHClust, k = 5, border = 2:7)
```



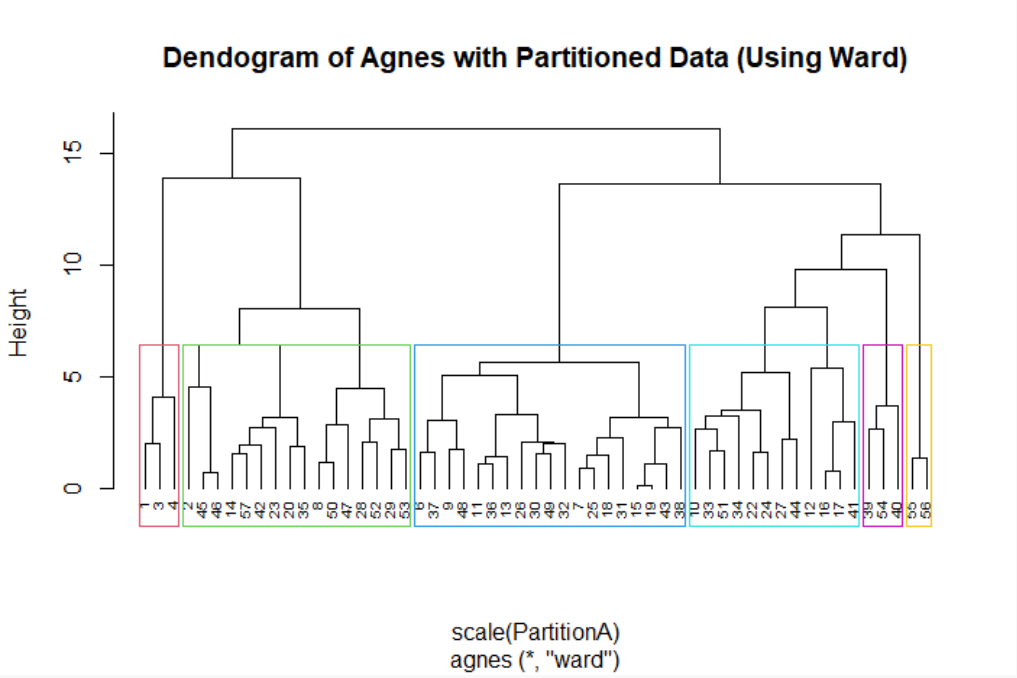

```
> SGroup <- cutree(wardHClust, k=5)
> DFrame2 <- as.data.frame(cbind(CerealsNormalize,SGroup))

> fviz_cluster(list(data = DFrame2, cluster = SGroup))
```



##Five clusters total

Task 3

```
> set.seed(123)
> PartitionA <- NumData[1:55,]
> PartitionB <- NumData[56:74,]

> SingleSL <- agnes(scale(PartitionA), method = "single")
> CompleteSL <- agnes(scale(PartitionA), method = "complete")
> AverageSL <- agnes(scale(PartitionA), method = "average")
> WardSL <- agnes(scale(PartitionA), method = "ward")
> cbind(single=SingleSL$ac , complete=CompleteSL$ac , average= AverageSL$ac , ward= WardSL$ac)

        single  complete   average      ward
[1,] 0.6564842 0.8120228 0.7449303 0.8808195

> pltree(WardSL, cex = 0.6, hang = -1, main = "Dendogram of Agnes with Partitioned Data (Using Ward)")
> rect.hclust(WardSL, k = 6, border = 2:7)
```




```
> Cut2 <- cutree(WardSL, k = 6)

> ResultSL <- as.data.frame(cbind(PartitionA, Cut2))
> ResultSL[ResultSL$Cut2==1,]

  calories protein fat sodium fiber carbo sugars potass vitamins shelf weight cups   rating Cut2
1       70       4   1    130    10     5      6    280        25     3      1 0.33 68.40297    1
3       70       4   1    260     9     7      5    320        25     3      1 0.33 59.42551    1
4       50       4   0    140    14     8      0    330        25     3      1 0.50 93.70491    1
```

```
> CentroidOne <- colMeans(ResultSL[ResultSL$Cut2==1,])
> ResultSL[ResultSL$Cut2==2,]
   calories protein fat sodium fiber carbo sugars potass vitamins shelf weight cups   rating Cut2
2       120       3   5     15   2.0   8.0      8    135         0     3   1.00 1.00 33.98368    2
8       130       3   2    210   2.0  18.0      8    100        25     3   1.33 0.75 37.03856    2
14      110       3   2    140   2.0  13.0      7    105        25     3   1.00 0.50 40.40021    2
20      110       3   3    140   4.0  10.0      7    160        25     3   1.00 0.50 40.44877    2
23      100       2   1    140   2.0  11.0     10    120        25     3   1.00 0.75 36.17620    2
28      120       3   2    160   5.0  12.0     10    200        25     3   1.25 0.67 40.91705    2
29      120       3   0    240   5.0  14.0     12    190        25     3   1.33 0.67 41.01549    2
35      120       3   3     75   3.0  13.0      4    100        25     3   1.00 0.33 45.81172    2
42      100       4   2    150   2.0  12.0      6     95        25     2   1.00 0.67 45.32807    2
45      150       4   3     95   3.0  16.0     11    170        25     3   1.00 1.00 37.13686    2
46      150       4   3    150   3.0  16.0     11    170        25     3   1.00 1.00 34.13976    2
47      160       3   2    150   3.0  17.0     13    160        25     3   1.50 0.67 30.31335    2
50      140       3   2    220   3.0  21.0      7    130        25     3   1.33 0.67 40.69232    2
52      130       3   2    170   1.5  13.5     10    120        25     3   1.25 0.50 30.45084    2
53      120       3   1    200   6.0  11.0     14    260        25     3   1.33 0.67 37.84059    2
57      100       4   1    135   2.0  14.0      6    110        25     3   1.00 0.50 49.51187    2
```

```
> CentroidTwo <- colMeans(ResultSL[ResultSL$Cut2==2,])
> ResultSL[ResultSL$Cut2==3,]
  calories protein fat sodium fiber carbo sugars potass vitamins shelf weight cups   rating Cut2
6      110       2   2    180   1.5  10.5     10     70        25     1      1 0.75 29.50954    3
```

```
7       110       2   0    125   1.0  11.0      14      30       25      2       1 1.00 33.17409    3
9        90       2   1    200   4.0  15.0       6     125       25      1       1 0.67 49.12025    3
11      120       1   2    220   0.0  12.0      12      35       25      2       1 0.75 18.04285    3
13      120       1   3    210   0.0  13.0       9      45       25      2       1 0.75 19.82357    3
15      110       1   1    180   0.0  12.0      13      55       25      2       1 1.00 22.73645    3
18      110       1   0     90   1.0  13.0      12      20       25      2       1 1.00 35.78279    3
19      110       1   1    180   0.0  12.0      13      65       25      2       1 1.00 22.39651    3
25      110       2   1    125   1.0  11.0      13      30       25      2       1 1.00 32.20758    3
26      110       1   0    200   1.0  14.0      11      25       25      1       1 0.75 31.43597    3
30      110       1   1    135   0.0  13.0      12      25       25      2       1 0.75 28.02576    3
31      100       2   0     45   0.0  11.0      15      40       25      1       1 0.88 35.25244    3
32      110       1   1    280   0.0  15.0       9      45       25      2       1 0.75 23.80404    3
36      120       1   2    220   1.0  12.0      11      45       25      2       1 1.00 21.87129    3
37      110       3   1    250   1.5  11.5      10      90       25      1       1 0.75 31.07222    3
38      110       1   0    180   0.0  14.0      11      35       25      1       1 1.33 28.74241    3
43      110       2   1    180   0.0  12.0      12      55       25      2       1 1.00 26.73451    3
48      100       2   1    220   2.0  15.0       6      90       25      1       1 1.00 40.10596    3
49      120       2   1    190   0.0  15.0       9      40       25      2       1 0.67 29.92429    3

> CentroidThree <- colMeans(ResultSL[ResultSL$Cut2==3,])
> ResultSL[ResultSL$Cut2==4,]
   calories protein fat sodium fiber carbo sugars potass vitamins shelf weight cups    rating Cut2
10       90       3   0    210     5    13       5     190       25     3       1 0.67 53.31381    4
12      110       6   2    290     2    17       1     105       25     1       1 1.25 50.76500    4
16      110       2   0    280     0    22       3      25       25     1       1 1.00 41.44502    4
17      100       2   0    290     1    21       2      35       25     1       1 1.00 45.86332    4
22      110       2   0    220     1    21       3      30       25     3       1 1.00 46.89564    4
24      100       2   0    190     1    18       5      80       25     3       1 0.75 44.33086    4
27      100       3   0      0     3    14       7     100       25     2       1 0.80 58.34514    4
33      100       3   1    140     3    15       5      85       25     3       1 0.88 52.07690    4
34      110       3   0    170     3    17       3      90       25     3       1 0.25 53.37101    4
41      110       2   1    260     0    21       3      40       25     2       1 1.50 39.24111    4
44      100       4   1      0     0    16       3      95       25     2       1 1.00 54.85092    4
51       90       3   0    170     3    18       2      90       25     3       1 1.00 59.64284    4

> CentroidFour <- colMeans(ResultSL[ResultSL$Cut2==4,])
> centroids <- rbind(CentroidOne, CentroidTwo, CentroidThree, CentroidFour)
> x2 <- as.data.frame(rbind(centroids[,-14], PartitionB))

> Dist1 <- get_dist(x2)
> Matrix1 <- as.matrix(Dist1)
> dataframe1 <- data.frame(data=seq(1,nrow(PartitionB),1), Clusters =rep(0,nr
ow(PartitionB)))
> for(i in 1:nrow(PartitionB))
+ {dataframe1[i,2] <- which.min(Matrix1[i+4, 1:4])}
> dataframe1
   data Clusters
1     1        1
2     2        2
3     3        2
4     4        3
5     5        3
6     6        2
7     7        2
8     8        2
9     9        3
10   10        4
11   11        2
12   12        3
13   13        2
14   14        4
15   15        4
16   16        3
17   17        4
18   18        4
19   19        3

> cbind(DFrame2$SGroup[56:74], dataframe1$Clusters)
      [,1] [,2]
 [1,]    2    1
 [2,]    2    2
```

```
 [3,]    5    2
 [4,]    4    3
 [5,]    4    3
 [6,]    5    2
 [7,]    5    2
 [8,]    5    2
 [9,]    3    3
[10,]    4    4
[11,]    5    2
[12,]    4    3
[13,]    2    2
[14,]    4    4
[15,]    4    4
[16,]    3    3
[17,]    4    4
[18,]    4    4
[19,]    3    3

> table(DFrame2$SGroup[56:74] == dataframe1$Clusters)

FALSE  TRUE
    9    10
```

##The model would be considered partially stable given that there are 9 False values and 10 true values.

Task 4

```
> HealthyCereals <- Cereals
> HealthyCerealsRD <- na.omit(HealthyCereals)
> clust <- cbind(HealthyCerealsRD, SGroup)
> clust[clust$SGroup==1,]
                      name mfr type calories protein fat sodium fiber carbo
sugars potass vitamins shelf weight cups   rating SGroup
1                  100%_Bran   N    C       70       4   1    130    10     5
6    280       25     3      1 0.33 68.40297      1
3                   All-Bran   K    C       70       4   1    260     9     7
5    320       25     3      1 0.33 59.42551      1
4 All-Bran_with_Extra_Fiber   K    C       50       4   0    140    14     8
0    330       25     3      1 0.50 93.70491      1

> clust[clust$SGroup==2,]
                                    name mfr type calories protein fat sodiu
m fiber carbo sugars potass vitamins shelf weight cups   rating SGroup
2                        100%_Natural_Bran   Q    C      120       3   5     1
5   2.0   8.0      8    135        0     3   1.00 1.00 33.98368      2
8                                  Basic_4   G    C      130       3   2    21
0   2.0  18.0      8    100       25     3   1.33 0.75 37.03856      2
14                                Clusters   G    C      110       3   2    14
0   2.0  13.0      7    105       25     3   1.00 0.50 40.40021      2
20                       Cracklin'_Oat_Bran   K    C      110       3   3    14
0   4.0  10.0      7    160       25     3   1.00 0.50 40.44877      2
23                    Crispy_Wheat_&_Raisins   G    C      100       2   1    14
0   2.0  11.0     10    120       25     3   1.00 0.75 36.17620      2
28 Fruit_&_Fibre_Dates,_Walnuts,_and_Oats   P    C      120       3   2    16
0   5.0  12.0     10    200       25     3   1.25 0.67 40.91705      2
29                             Fruitful_Bran   K    C      120       3   0    24
0   5.0  14.0     12    190       25     3   1.33 0.67 41.01549      2
35                       Great_Grains_Pecan   P    C      120       3   3     7
5   3.0  13.0      4    100       25     3   1.00 0.33 45.81172      2
40                    Just_Right_Fruit_&_Nut   K    C      140       3   1    17
0   2.0  20.0      9     95      100     3   1.30 0.75 36.47151      2
42                                     Life   Q    C      100       4   2    15
0   2.0  12.0      6     95       25     2   1.00 0.67 45.32807      2
```

```
45       Muesli_Raisins,_Dates,_&_Almonds   R    C      150       4   3     9
5   3.0  16.0     11    170       25     3   1.00 1.00 37.13686      2
46     Muesli_Raisins,_Peaches,_&_Pecans   R    C      150       4   3    15
0   3.0  16.0     11    170       25     3   1.00 1.00 34.13976      2
47                  Mueslix_Crispy_Blend   K    C      160       3   2    15
0   3.0  17.0     13    160       25     3   1.50 0.67 30.31335      2
50               Nutri-Grain_Almond-Raisin   K    C      140       3   2    22
0   3.0  21.0      7    130       25     3   1.33 0.67 40.69232      2
52                  Oatmeal_Raisin_Crisp   G    C      130       3   2    17
0   1.5  13.5     10    120       25     3   1.25 0.50 30.45084      2
53                 Post_Nat._Raisin_Bran   P    C      120       3   1    20
0   6.0  11.0     14    260       25     3   1.33 0.67 37.84059      2
57                     Quaker_Oat_Squares   Q    C      100       4   1    13
5   2.0  14.0      6    110       25     3   1.00 0.50 49.51187      2
59                           Raisin_Bran   K    C      120       3   1    21
0   5.0  14.0     12    240       25     2   1.33 0.75 39.25920      2
60                       Raisin_Nut_Bran   G    C      100       3   2    14
0   2.5  10.5      8    140       25     3   1.00 0.50 39.70340      2
71                     Total_Raisin_Bran   G    C      140       3   1    19
0   4.0  15.0     14    230      100     3   1.50 1.00 28.59278      2

> clust[clust$SGroup==3,]
                    name mfr type calories protein fat sodium fiber carbo s
ugars potass vitamins shelf weight cups   rating SGroup
6  Apple_Cinnamon_Cheerios   G    C      110       2   2    180   1.5  10.5
10     70       25     1      1 0.75 29.50954      3
7              Apple_Jacks   K    C      110       2   0    125   1.0  11.0
14     30       25     2      1 1.00 33.17409      3
11            Cap'n'Crunch   Q    C      120       1   2    220   0.0  12.0
12     35       25     2      1 0.75 18.04285      3
13   Cinnamon_Toast_Crunch   G    C      120       1   3    210   0.0  13.0
9     45       25     2      1 0.75 19.82357      3
15             Cocoa_Puffs   G    C      110       1   1    180   0.0  12.0
13     55       25     2      1 1.00 22.73645      3
18               Corn_Pops   K    C      110       1   0     90   1.0  13.0
12     20       25     2      1 1.00 35.78279      3
19           Count_Chocula   G    C      110       1   1    180   0.0  12.0
13     65       25     2      1 1.00 22.39651      3
25             Froot_Loops   K    C      110       2   1    125   1.0  11.0
13     30       25     2      1 1.00 32.20758      3
26          Frosted_Flakes   K    C      110       1   0    200   1.0  14.0
11     25       25     1      1 0.75 31.43597      3
30          Fruity_Pebbles   P    C      110       1   1    135   0.0  13.0
12     25       25     2      1 0.75 28.02576      3
31            Golden_Crisp   P    C      100       2   0     45   0.0  11.0
15     40       25     1      1 0.88 35.25244      3
32          Golden_Grahams   G    C      110       1   1    280   0.0  15.0
9     45       25     2      1 0.75 23.80404      3
36         Honey_Graham_Ohs   Q    C      120       1   2    220   1.0  12.0
11     45       25     2      1 1.00 21.87129      3
37      Honey_Nut_Cheerios   G    C      110       3   1    250   1.5  11.5
10     90       25     1      1 0.75 31.07222      3
38              Honey-comb   P    C      110       1   0    180   0.0  14.0
11     35       25     1      1 1.33 28.74241      3
43            Lucky_Charms   G    C      110       2   1    180   0.0  12.0
12     55       25     2      1 1.00 26.73451      3
48    Multi-Grain_Cheerios   G    C      100       2   1    220   2.0  15.0
6      90       25     1      1 1.00 40.10596      3
49         Nut&Honey_Crunch   K    C      120       2   1    190   0.0  15.0
9     40       25     2      1 0.67 29.92429      3
67                    Smacks   K    C      110       2   1     70   1.0   9.0
15     40       25     2      1 0.75 31.23005      3
74                      Trix   G    C      110       1   1    140   0.0  13.0
12     25       25     2      1 1.00 27.75330      3
```

```
77     Wheaties_Honey_Gold   G    C      110       2   1    200   1.0  16.0
8     60        25     1      1 0.75 36.18756      3

> clust[clust$SGroup==4,]
                         name mfr type calories protein fat sodium fiber car
bo sugars potass vitamins shelf weight cups   rating SGroup
9                    Bran_Chex   R    C       90       2   1    200     4
15      6    125        25     1      1 0.67 49.12025      4
10                 Bran_Flakes   P    C       90       3   0    210     5
13      5    190        25     3      1 0.67 53.31381      4
12                     Cheerios   G    C      110       6   2    290     2
17      1    105        25     1      1 1.25 50.76500      4
16                    Corn_Chex   R    C      110       2   0    280     0
22      3     25        25     1      1 1.00 41.44502      4
17                  Corn_Flakes   K    C      100       2   0    290     1
21      2     35        25     1      1 1.00 45.86332      4
22                      Crispix   K    C      110       2   0    220     1
21      3     30        25     3      1 1.00 46.89564      4
24                  Double_Chex   R    C      100       2   0    190     1
18      5     80        25     3      1 0.75 44.33086      4
33            Grape_Nuts_Flakes   P    C      100       3   1    140     3
15      5     85        25     3      1 0.88 52.07690      4
34                   Grape-Nuts   P    C      110       3   0    170     3
17      3     90        25     3      1 0.25 53.37101      4
39 Just_Right_Crunchy__Nuggets   K    C      110       2   1    170     1
17      6     60       100     3      1 1.00 36.52368      4
41                         Kix   G    C      110       2   1    260     0
21      3     40        25     2      1 1.50 39.24111      4
51            Nutri-grain_Wheat   K    C       90       3   0    170     3
18      2     90        25     3      1 1.00 59.64284      4
54                   Product_19   K    C      100       3   0    320     1
20      3     45       100     3      1 1.00 41.50354      4
62                    Rice_Chex   R    C      110       1   0    240     0
23      2     30        25     1      1 1.13 41.99893      4
63                Rice_Krispies   K    C      110       2   0    290     0
22      3     35        25     1      1 1.00 40.56016      4
68                    Special_K   K    C      110       6   0    230     1
16      3     55        25     1      1 1.00 53.13132      4
70            Total_Corn_Flakes   G    C      110       2   1    200     0
21      3     35       100     3      1 1.00 38.83975      4
72            Total_Whole_Grain   G    C      100       3   1    200     3
16      3    110       100     3      1 1.00 46.65884      4
73                      Triples   G    C      110       2   1    250     0
21      3     60        25     3      1 0.75 39.10617      4
75                   Wheat_Chex   R    C      100       3   1    230     3
17      3    115        25     1      1 0.67 49.78744      4
76                     Wheaties   G    C      100       3   1    200     3
17      3    110        25     1      1 1.00 51.59219      4

> mean(clust[clust$SGroup==1,"rating"])
[1] 73.84446
> mean(clust[clust$SGroup==2,"rating"])
[1] 38.26161
> mean(clust[clust$SGroup==3,"rating"])
[1] 28.84825
> mean(clust[clust$SGroup==4,"rating"])
[1] 46.46513

##The cluster with the highest rating is Cluster 1, so it should be chosen
```