Assignment 4 - MIS64060

Tim Pagliari tpagliar@kent.edu (mailto:tpagliar@kent.edu)

10/25/2020

Assignment details:

The dataset on American College and University Rankings contains information on 1302 American colleges and universities offering an undergraduate program. For each university, there are 17 measurements, including continuous measurements (such as tuition and graduation rate) and categorical measurements (such as location by state and whether it is a private or public school). Note that many records are missing some measurements.

- Remove all records with missing measurements from the dataset.
- For all the continuous measurements, run K-Means clustering. Make sure to normalize the measurements. How many clusters seem reasonable for describing these data? What was your optimal K?
- Compare the summary statistics for each cluster and describe each cluster in this context (e.g., “Universities with high tuition, low acceptance rate…”).
- Use the categorical measurements that were not used in the analysis (State and Private/Public) to characterize the different clusters. Is there any relationship between the clusters and the categorical information?
- What other external information can explain the contents of some or all of these clusters?
- Consider Tufts University, which is missing some information. Compute the Euclidean distance of this record from each of the clusters that you found above (using only the measurements that you have). Which cluster is it closest to? Impute the missing values for Tufts by taking the average of the cluster on those measurements.

Import and tidy the data

```
UniversitiesFull <- read.csv("/Users/tpagliar/GIT/MIS64060-tpagliar/Universities.csv"
, stringsAsFactors = TRUE)
UniversitiesFull <- dummy_columns(UniversitiesFull, select_columns = 'Public..1...Pri
vate..2.')
colnames(UniversitiesFull) <- c("Name","State","PublicPrivate","Applied","Accepted","
Enrolled","Top10","Top25","FullTimeUG","PartTimeUG","InStateTuition","OutStateTuition
","Room","Board","Fees","EstBookCost","EstPersCost","FacWPHD","StudFacRatio","GradRat
e","Public","Private")
Universities <- UniversitiesFull
summary(Universities)
```

```
##                 Name          State     PublicPrivate      Applied
##  Bethel College   :   4   NY     :101   Min.   :1.000   Min.   :   35.0
##  Concordia College:   4   PA     : 83   1st Qu.:1.000   1st Qu.:  695.8
##  Trinity College  :   4   CA     : 70   Median :2.000   Median : 1470.0
##  Columbia College :   3   TX     : 60   Mean   :1.639   Mean   : 2752.1
##  Union College    :   3   MA     : 56   3rd Qu.:2.000   3rd Qu.: 3314.2
##  Augustana College:   2   OH     : 52   Max.   :2.000   Max.   :48094.0
##  (Other)          :1282   (Other):880                   NA's   :10
##     Accepted          Enrolled          Top10           Top25
##  Min.   :   35.0   Min.   :  18.0   Min.   : 1.00   Min.   :  6.00
##  1st Qu.:  554.5   1st Qu.: 236.0   1st Qu.:13.00   1st Qu.: 36.75
##  Median : 1095.0   Median : 447.0   Median :21.00   Median : 50.00
##  Mean   : 1870.7   Mean   : 778.9   Mean   :25.67   Mean   : 52.35
##  3rd Qu.: 2303.0   3rd Qu.: 984.0   3rd Qu.:32.00   3rd Qu.: 66.00
##  Max.   :26330.0   Max.   :7425.0   Max.   :98.00   Max.   :100.00
##  NA's   :11        NA's   :5        NA's   :235     NA's   :202
##    FullTimeUG      PartTimeUG      InStateTuition  OutStateTuition
##  Min.   :   59   Min.   :    1.0   Min.   :  480   Min.   : 1044
##  1st Qu.:  966   1st Qu.:  131.2   1st Qu.: 2580   1st Qu.: 6111
##  Median : 1812   Median :  472.0   Median : 8050   Median : 8670
##  Mean   : 3693   Mean   : 1081.5   Mean   : 7897   Mean   : 9277
##  3rd Qu.: 4540   3rd Qu.: 1313.0   3rd Qu.:11600   3rd Qu.:11659
##  Max.   :31643   Max.   :21836.0   Max.   :25750   Max.   :25750
##  NA's   :3       NA's   :32        NA's   :30      NA's   :20
##       Room           Board           Fees         EstBookCost    EstPersCost
##  Min.   : 500   Min.   : 531   Min.   :   9.0   Min.   :  90   Min.   :  75
##  1st Qu.:1710   1st Qu.:1619   1st Qu.: 130.0   1st Qu.: 480   1st Qu.: 900
##  Median :2200   Median :1980   Median : 264.5   Median : 502   Median :1250
##  Mean   :2515   Mean   :2061   Mean   : 392.0   Mean   : 550   Mean   :1389
##  3rd Qu.:3040   3rd Qu.:2402   3rd Qu.: 480.0   3rd Qu.: 600   3rd Qu.:1794
##  Max.   :7400   Max.   :6250   Max.   :4374.0   Max.   :2340   Max.   :6900
##  NA's   :321    NA's   :498    NA's   :274      NA's   :48     NA's   :181
##     FacWPHD         StudFacRatio      GradRate          Public
##  Min.   :  8.00   Min.   : 2.30   Min.   :  8.00   Min.   :0.000
##  1st Qu.: 57.00   1st Qu.:11.80   1st Qu.: 47.00   1st Qu.:0.000
##  Median : 71.00   Median :14.30   Median : 60.00   Median :0.000
##  Mean   : 68.65   Mean   :14.86   Mean   : 60.41   Mean   :0.361
##  3rd Qu.: 82.00   3rd Qu.:17.60   3rd Qu.: 74.00   3rd Qu.:1.000
##  Max.   :105.00   Max.   :91.80   Max.   :118.00   Max.   :1.000
##  NA's   :32       NA's   :2       NA's   :98
##     Private
##  Min.   :0.000
##  1st Qu.:0.000
##  Median :1.000
##  Mean   :0.639
##  3rd Qu.:1.000
##  Max.   :1.000
```

```
##
```

```
Universities$PublicPrivate <- as.factor(Universities$PublicPrivate)
Universities$AcceptRate<-((Universities$Accepted/Universities$Applied)*100) #I'll use these to simplify the contribution to the selectivity and yield of each school.
Universities$EnrolledRate<-((Universities$Enrolled/Universities$Accepted)*100)#I'll use these to simplify the contribution to the selectivity and yield of each school.
Universities$AcceptRate<-((Universities$Accepted/Universities$Applied)*100)
Universities$EnrolledRate<-((Universities$Enrolled/Universities$Accepted)*100)
colMeans(is.na(Universities)*100) #we see there are many incomplete records, we'll remove them from our dataset
```

```
##            Name           State   PublicPrivate         Applied        Accepted
##       0.0000000       0.0000000       0.0000000       0.7680492       0.8448541
##        Enrolled           Top10           Top25      FullTimeUG      PartTimeUG
##       0.3840246      18.0491551      15.5145929       0.2304147       2.4577573
##  InStateTuition OutStateTuition            Room           Board            Fees
##       2.3041475       1.5360983      24.6543779      38.2488479      21.0445469
##     EstBookCost     EstPersCost         FacWPHD    StudFacRatio        GradRate
##       3.6866359      13.9016897       2.4577573       0.1536098       7.5268817
##          Public         Private      AcceptRate    EnrolledRate
##       0.0000000       0.0000000       0.9984639       0.9216590
```

```
Universities <- Universities[complete.cases(Universities),] #Include only complete data
summary(Universities) #we see that some schools have data that is a heavy outlier, ex: a 118% graduation rate or 103% faculty with PhDs. We'll filter these 'erroneous' rows below:
```

```
##                         Name          State     PublicPrivate    Applied
##  Trinity College          :  4   PA     : 42   1:128          Min.   :   77
##  Augustana College        :  2   NY     : 38   2:343          1st Qu.:  802
##  Monmouth College         :  2   OH     : 24                  Median : 1646
##  University of St. Thomas:  2   NC     : 23                  Mean   : 3147
##  Westminster College      :  2   MA     : 22                  3rd Qu.: 3862
##  Adams State College      :  1   TX     : 20                  Max.   :48094
##  (Other)                  :458   (Other):302
##     Accepted          Enrolled          Top10           Top25
##  Min.   :   61.0   Min.   :  27.0   Min.   : 1.00   Min.   :  9.00
##  1st Qu.:  635.5   1st Qu.: 264.0   1st Qu.:15.00   1st Qu.: 40.00
##  Median : 1227.0   Median : 443.0   Median :23.00   Median : 54.00
##  Mean   : 2063.0   Mean   : 780.7   Mean   :28.01   Mean   : 55.65
##  3rd Qu.: 2456.0   3rd Qu.: 896.5   3rd Qu.:36.00   3rd Qu.: 69.00
##  Max.   :26330.0   Max.   :6392.0   Max.   :96.00   Max.   :100.00
```

```
##
##    FullTimeUG      PartTimeUG      InStateTuition  OutStateTuition
##  Min.   :  249   Min.   :    1.0   Min.   :  608   Min.   : 1044
##  1st Qu.: 1018   1st Qu.:   81.5   1st Qu.: 3650   1st Qu.: 7290
##  Median : 1715   Median :  299.0   Median : 9858   Median :10100
##  Mean   : 3563   Mean   :  797.5   Mean   : 9407   Mean   :10575
##  3rd Qu.: 4056   3rd Qu.:  869.0   3rd Qu.:13246   3rd Qu.:13286
##  Max.   :31643   Max.   :21836.0   Max.   :20100   Max.   :20100
##
##       Room          Board           Fees         EstBookCost      EstPersCost
##  Min.   : 640   Min.   : 531   Min.   :  10.0   Min.   :  90.0   Min.   : 250
##  1st Qu.:1740   1st Qu.:1750   1st Qu.: 137.5   1st Qu.: 500.0   1st Qu.: 850
##  Median :2090   Median :2082   Median : 280.0   Median : 500.0   Median :1200
##  Mean   :2221   Mean   :2122   Mean   : 379.0   Mean   : 548.8   Mean   :1312
##  3rd Qu.:2663   3rd Qu.:2420   3rd Qu.: 486.0   3rd Qu.: 600.0   3rd Qu.:1600
##  Max.   :4816   Max.   :4541   Max.   :3247.0   Max.   :2340.0   Max.   :6800
##
##     FacWPHD        StudFacRatio      GradRate          Public
##  Min.   :  8.00   Min.   : 2.90   Min.   : 15.00   Min.   :0.0000
##  1st Qu.: 63.00   1st Qu.:11.30   1st Qu.: 53.00   1st Qu.:0.0000
##  Median : 76.00   Median :13.40   Median : 66.00   Median :0.0000
##  Mean   : 73.21   Mean   :13.96   Mean   : 65.56   Mean   :0.2718
##  3rd Qu.: 87.00   3rd Qu.:16.45   3rd Qu.: 79.00   3rd Qu.:1.0000
##  Max.   :103.00   Max.   :28.80   Max.   :118.00   Max.   :1.0000
##
##     Private         AcceptRate      EnrolledRate
##  Min.   :0.0000   Min.   : 15.45   Min.   :  9.975
##  1st Qu.:0.0000   1st Qu.: 65.14   1st Qu.: 30.931
##  Median :1.0000   Median : 77.69   Median : 38.440
##  Mean   :0.7282   Mean   : 73.65   Mean   : 41.449
##  3rd Qu.:1.0000   3rd Qu.: 84.86   3rd Qu.: 48.800
##  Max.   :1.0000   Max.   :100.00   Max.   :100.000
##
```

```
Universities <- filter(Universities, (GradRate <= 100) & (FacWPHD <= 100))#we see tha
t some schools have data that is a heavy outlier, ex: a 118% graduation rate or 103%
faculty with PhDs.
summary(Universities)
```

```
##                       Name         State     PublicPrivate    Applied
##  Trinity College         :  4   PA     : 42   1:127         Min.   :   77
##  Augustana College       :  2   NY     : 37   2:342         1st Qu.:  804
##  Monmouth College        :  2   OH     : 24                 Median : 1646
##  University of St. Thomas:  2   NC     : 23                 Mean   : 3151
##  Westminster College     :  2   MA     : 22                 3rd Qu.: 3877
##  Adams State College     :  1   TX     : 19                 Max.   :48094
```

```
##  (Other)                 :456   (Other):302
##     Accepted        Enrolled          Top10           Top25
##  Min.   :   61   Min.   :  27.0   Min.   : 1.00   Min.   :  9.00
##  1st Qu.:  638   1st Qu.: 265.0   1st Qu.:15.00   1st Qu.: 40.00
##  Median : 1227   Median : 443.0   Median :23.00   Median : 54.00
##  Mean   : 2063   Mean   : 782.4   Mean   :28.07   Mean   : 55.71
##  3rd Qu.: 2446   3rd Qu.: 902.0   3rd Qu.:36.00   3rd Qu.: 69.00
##  Max.   :26330   Max.   :6392.0   Max.   :96.00   Max.   :100.00
##
##    FullTimeUG      PartTimeUG      InStateTuition  OutStateTuition
##  Min.   :  249   Min.   :    1.0   Min.   :  608   Min.   : 1044
##  1st Qu.: 1022   1st Qu.:   82.0   1st Qu.: 3660   1st Qu.: 7320
##  Median : 1720   Median :  299.0   Median : 9888   Median :10100
##  Mean   : 3573   Mean   :  800.5   Mean   : 9425   Mean   :10590
##  3rd Qu.: 4106   3rd Qu.:  869.0   3rd Qu.:13252   3rd Qu.:13320
##  Max.   :31643   Max.   :21836.0   Max.   :20100   Max.   :20100
##
##       Room          Board           Fees         EstBookCost      EstPersCost
##  Min.   : 640   Min.   : 531   Min.   :  10.0   Min.   :  90.0   Min.   : 250
##  1st Qu.:1740   1st Qu.:1750   1st Qu.: 135.0   1st Qu.: 500.0   1st Qu.: 850
##  Median :2090   Median :2082   Median : 273.0   Median : 500.0   Median :1200
##  Mean   :2222   Mean   :2122   Mean   : 379.1   Mean   : 548.6   Mean   :1315
##  3rd Qu.:2666   3rd Qu.:2420   3rd Qu.: 486.0   3rd Qu.: 600.0   3rd Qu.:1600
##  Max.   :4816   Max.   :4541   Max.   :3247.0   Max.   :2340.0   Max.   :6800
##
##     FacWPHD        StudFacRatio      GradRate         Public
##  Min.   :  8.00   Min.   : 2.90   Min.   : 15.0   Min.   :0.0000
##  1st Qu.: 63.00   1st Qu.:11.30   1st Qu.: 53.0   1st Qu.:0.0000
##  Median : 76.00   Median :13.40   Median : 66.0   Median :0.0000
##  Mean   : 73.25   Mean   :13.95   Mean   : 65.5   Mean   :0.2708
##  3rd Qu.: 87.00   3rd Qu.:16.40   3rd Qu.: 79.0   3rd Qu.:1.0000
##  Max.   :100.00   Max.   :28.80   Max.   :100.0   Max.   :1.0000
##
##     Private         AcceptRate      EnrolledRate
##  Min.   :0.0000   Min.   : 15.45   Min.   :  9.975
##  1st Qu.:0.0000   1st Qu.: 65.11   1st Qu.: 30.931
##  Median :1.0000   Median : 77.60   Median : 38.440
##  Mean   :0.7292   Mean   : 73.58   Mean   : 41.485
##  3rd Qu.:1.0000   3rd Qu.: 84.83   3rd Qu.: 48.791
##  Max.   :1.0000   Max.   :100.00   Max.   :100.000
##
```

Normalizing Data

```
#columnsare <- c(3,)
ScaledValues <- scale(Universities[,4:20], center=TRUE,scale=TRUE)
UniversitiesScaled <- Universities
UniversitiesScaled[,4:20]<- ScaledValues
summary(UniversitiesScaled)
```

```
##                         Name          State     PublicPrivate    Applied
##  Trinity College          :  4   PA      : 42   1:127         Min.   :-0.7535
##  Augustana College        :  2   NY      : 37   2:342         1st Qu.:-0.5754
##  Monmouth College         :  2   OH      : 24                 Median :-0.3690
##  University of St. Thomas:  2   NC      : 23                 Mean   : 0.0000
##  Westminster College      :  2   MA      : 22                 3rd Qu.: 0.1778
##  Adams State College      :  1   TX      : 19                 Max.   :11.0156
##  (Other)                  :456   (Other):302
##     Accepted          Enrolled           Top10             Top25
##  Min.   :-0.7987   Min.   :-0.8236   Min.   :-1.4634   Min.   :-2.29652
##  1st Qu.:-0.5685   1st Qu.:-0.5641   1st Qu.:-0.7064   1st Qu.:-0.77253
##  Median :-0.3336   Median :-0.3700   Median :-0.2739   Median :-0.08428
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.00000
##  3rd Qu.: 0.1526   3rd Qu.: 0.1304   3rd Qu.: 0.4290   3rd Qu.: 0.65314
##  Max.   : 9.6786   Max.   : 6.1162   Max.   : 3.6730   Max.   : 2.17713
##
##    FullTimeUG        PartTimeUG       InStateTuition     OutStateTuition
##  Min.   :-0.7109   Min.   :-0.51638   Min.   :-1.59899   Min.   :-2.2136
##  1st Qu.:-0.5456   1st Qu.:-0.46407   1st Qu.:-1.04551   1st Qu.:-0.7582
##  Median :-0.3963   Median :-0.32392   Median : 0.08395   Median :-0.1136
##  Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000
##  3rd Qu.: 0.1139   3rd Qu.: 0.04421   3rd Qu.: 0.69402   3rd Qu.: 0.6331
##  Max.   : 6.0024   Max.   :13.58560   Max.   : 1.93592   Max.   : 2.2053
##
##       Room              Board              Fees          EstBookCost
##  Min.   :-2.2158   Min.   :-2.80533   Min.   :-1.0351   Min.   :-2.8047
##  1st Qu.:-0.6752   1st Qu.:-0.65660   1st Qu.:-0.6845   1st Qu.:-0.2970
##  Median :-0.1850   Median :-0.07139   Median :-0.2976   Median :-0.2970
##  Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.6217   3rd Qu.: 0.52440   3rd Qu.: 0.2997   3rd Qu.: 0.3146
##  Max.   : 3.6329   Max.   : 4.26308   Max.   : 8.0422   Max.   :10.9568
##
##   EstPersCost          FacWPHD          StudFacRatio        GradRate
##  Min.   :-1.5627   Min.   :-3.9607   Min.   :-2.8317   Min.   :-2.80658
##  1st Qu.:-0.6824   1st Qu.:-0.6224   1st Qu.:-0.6800   1st Qu.:-0.69465
##  Median :-0.1688   Median : 0.1667   Median :-0.1421   Median : 0.02785
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.00000
##  3rd Qu.: 0.4180   3rd Qu.: 0.8344   3rd Qu.: 0.6264   3rd Qu.: 0.75035
##  Max.   : 8.0474   Max.   : 1.6234   Max.   : 3.8027   Max.   : 1.91746
##
```

```
##      Public          Private         AcceptRate      EnrolledRate
##  Min.   :0.0000   Min.   :0.0000   Min.   : 15.45   Min.   :  9.975
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 65.11   1st Qu.: 30.931
##  Median :0.0000   Median :1.0000   Median : 77.60   Median : 38.440
##  Mean   :0.2708   Mean   :0.7292   Mean   : 73.58   Mean   : 41.485
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.: 84.83   3rd Qu.: 48.791
##  Max.   :1.0000   Max.   :1.0000   Max.   :100.00   Max.   :100.000
##
```

Clustering

We'll start with 5 clusters as a base, then will do some analysis to determine the optimal number of centers.

```
k5 <- kmeans(ScaledValues, centers = 5, nstart = 25) # k = 5, number of restarts = 25
k5$centers # output the centers
```

```
##      Applied   Accepted    Enrolled      Top10      Top25  FullTimeUG
## 1  2.2968753  2.5535643  2.69056311  0.2010037  0.3555161  2.71476761
## 2  0.5114509  0.1874549  0.04016794  1.7896448  1.5208325 -0.07927153
## 3  0.1848991  0.2219851  0.38819198 -0.5862947 -0.5000209  0.46792382
## 4 -0.5648207 -0.5646196 -0.54968870 -0.6159400 -0.7176700 -0.53242594
## 5 -0.3444641 -0.2992972 -0.37612694  0.1155148  0.2296478 -0.39292261
##   PartTimeUG InStateTuition OutStateTuition        Room       Board       Fees
## 1  1.8633078     -1.0074879      -0.4104606  0.05891951 -0.09311124  0.6351832
## 2 -0.3749079      1.5126248       1.6984791  1.20809050  1.03681846  0.1315288
## 3  0.3755535     -1.2855039      -1.0776176 -0.42887761 -0.70732803  0.5520826
## 4 -0.2533251     -0.2319552      -0.5032276 -0.50856743 -0.44742860 -0.4080688
## 5 -0.2739143      0.5402301       0.4575865  0.19775123  0.39395245 -0.1294033
##   EstBookCost EstPersCost     FacWPHD StudFacRatio   GradRate
## 1  0.23425051  0.89264019  0.76054464   0.49832557 -0.1373811
## 2  0.33303861 -0.47903561  1.10143572  -1.16197773  1.1782907
## 3 -0.02761874  0.39386243  0.05204093   1.01221429 -0.6411341
## 4 -0.03558822  0.09793485 -1.07848930   0.06991768 -0.5988026
## 5 -0.14053133 -0.32485893  0.32934584  -0.26377373  0.4506685
```

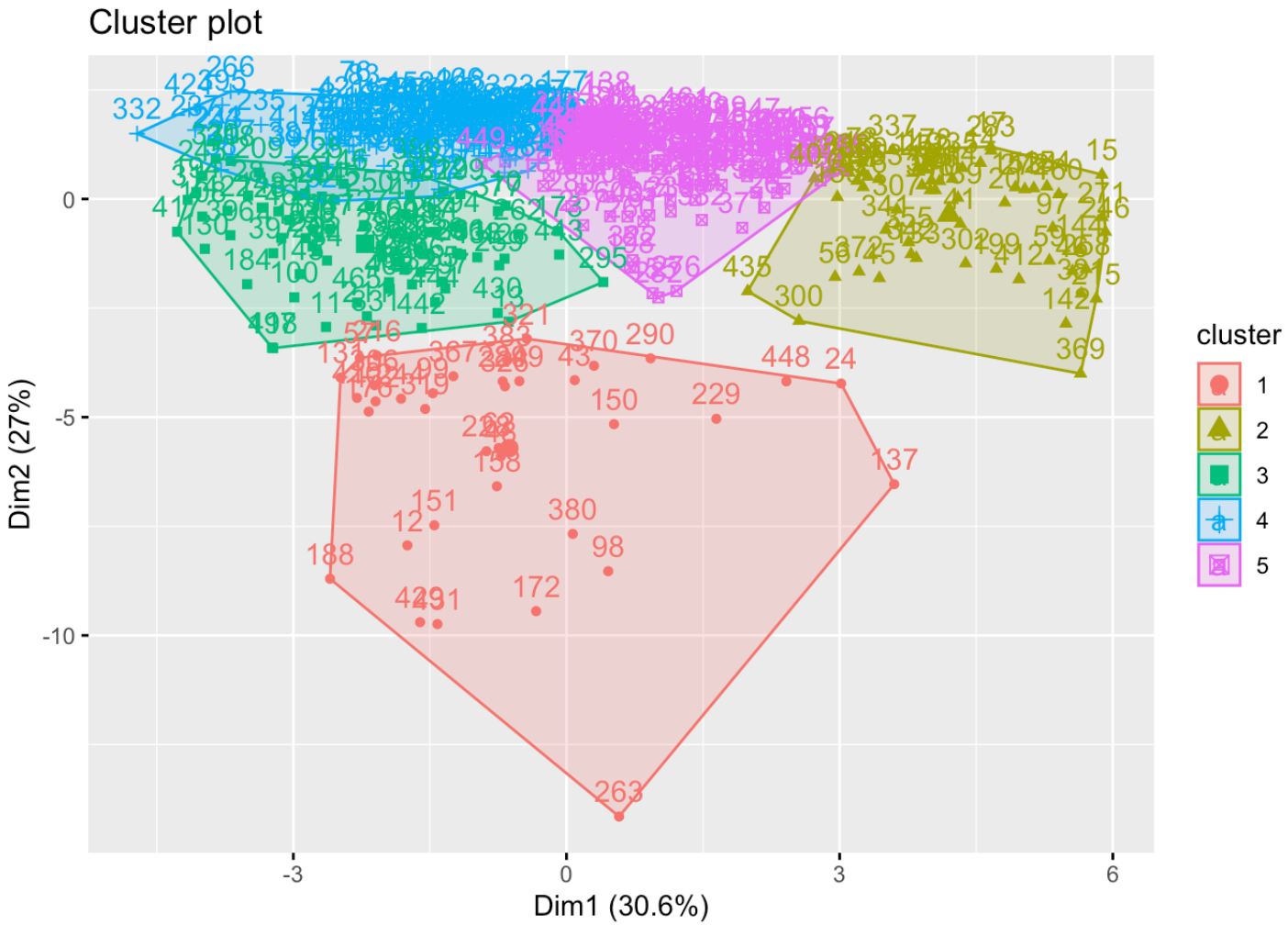
```
k5$size # Number of schools in each cluster
```

```
## [1]  37  60  81 138 153
```

```
k5$cluster[120] # Identify the cluster of the 120th observation as an example
```

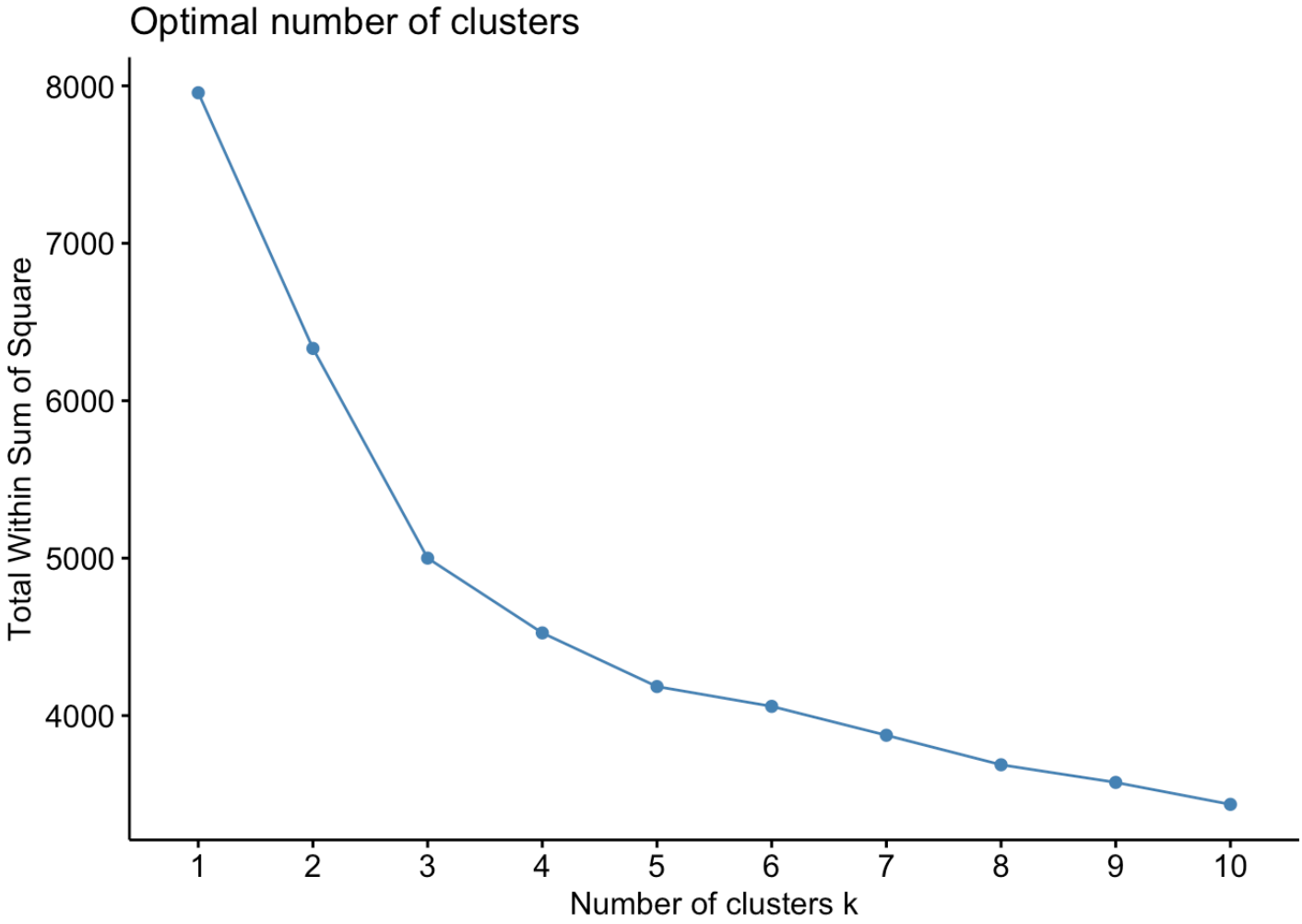
```
## [1] 4
```

```
fviz_cluster(k5, data = ScaledValues) # Visualize the output
```

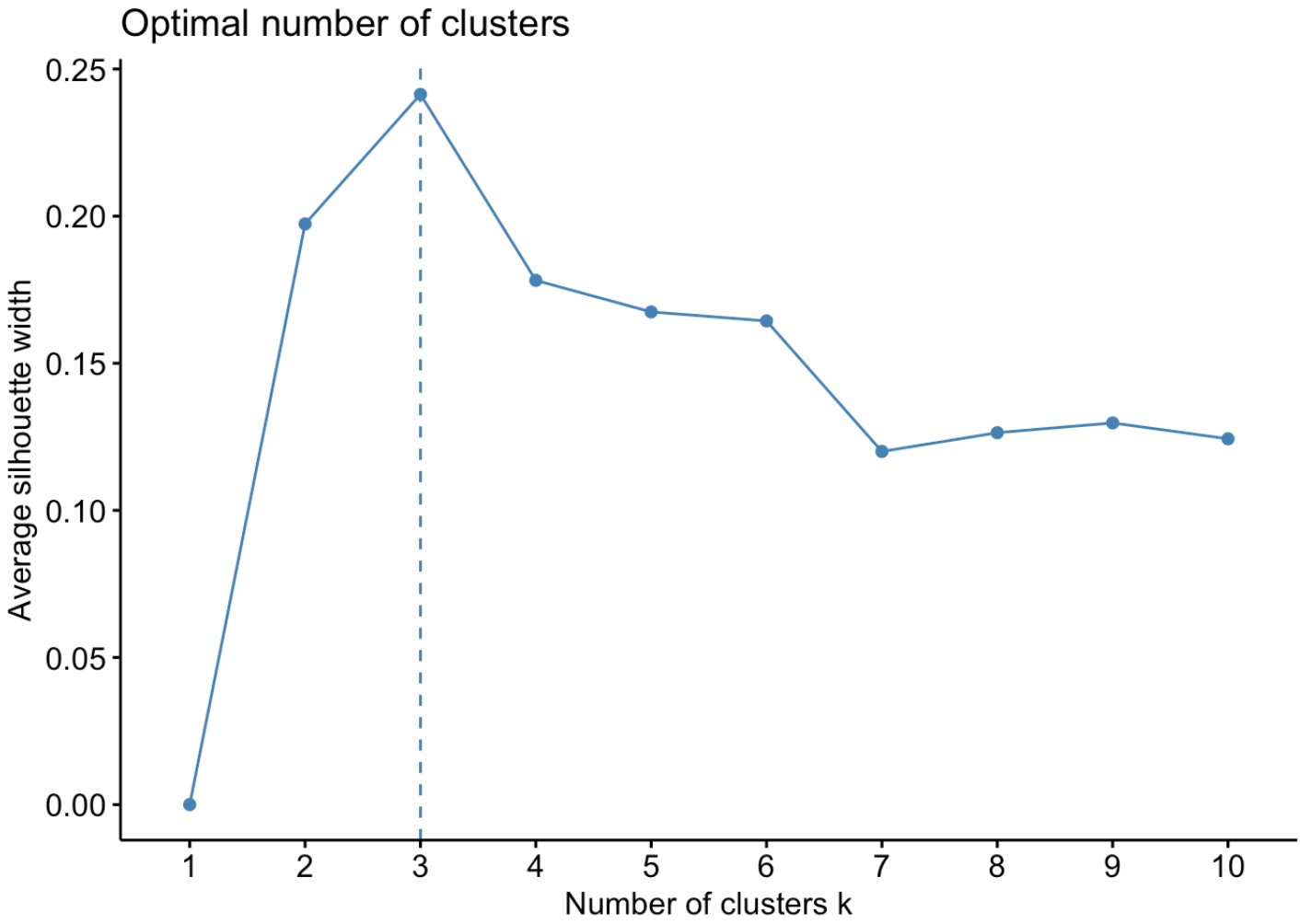


Optimal Clustering

```
set.seed(321)
fviz_nbclust(ScaledValues, kmeans, method = "wss")
```




```
#It looks like 3 or 4 is the optimal n-cluster here....
fviz_nbclust(ScaledValues, kmeans, method = "silhouette")
```




```
#We'll run kmeans with 3 centers as recommended by the silhouette
k3 <- kmeans(ScaledValues, centers = 3, nstart = 25) # k = 4, number of restarts = 25
k3$centers # output the centers
```

```
##      Applied   Accepted   Enrolled      Top10      Top25 FullTimeUG PartTimeUG
## 1 -0.3608515 -0.3511253 -0.3175707 -0.5032469 -0.5139401 -0.2953121 -0.1210245
## 2  0.0503146 -0.0437909 -0.1699120  0.8759144  0.8583053 -0.2343273 -0.3144979
## 3  1.9775059  2.2266486  2.4387744  0.1304181  0.2513016  2.5167238  1.7437904
##   InStateTuition OutStateTuition        Room      Board        Fees EstBookCost
## 1     -0.4044062      -0.5277099 -0.36010260 -0.3957617 -0.05844569 -0.06753063
## 2      1.0592086       1.1122770  0.66770532  0.7740830 -0.04512856  0.07242614
## 3     -1.0538778      -0.4951468 -0.04016933 -0.1754245  0.49402081  0.16460744
##   EstPersCost    FacWPHD StudFacRatio   GradRate
## 1  0.06312818 -0.5403354    0.2812986 -0.4263361
## 2 -0.40142642  0.7720498   -0.7006448  0.8533508
## 3  0.93434720  0.6892196    0.6152653 -0.2524534
```

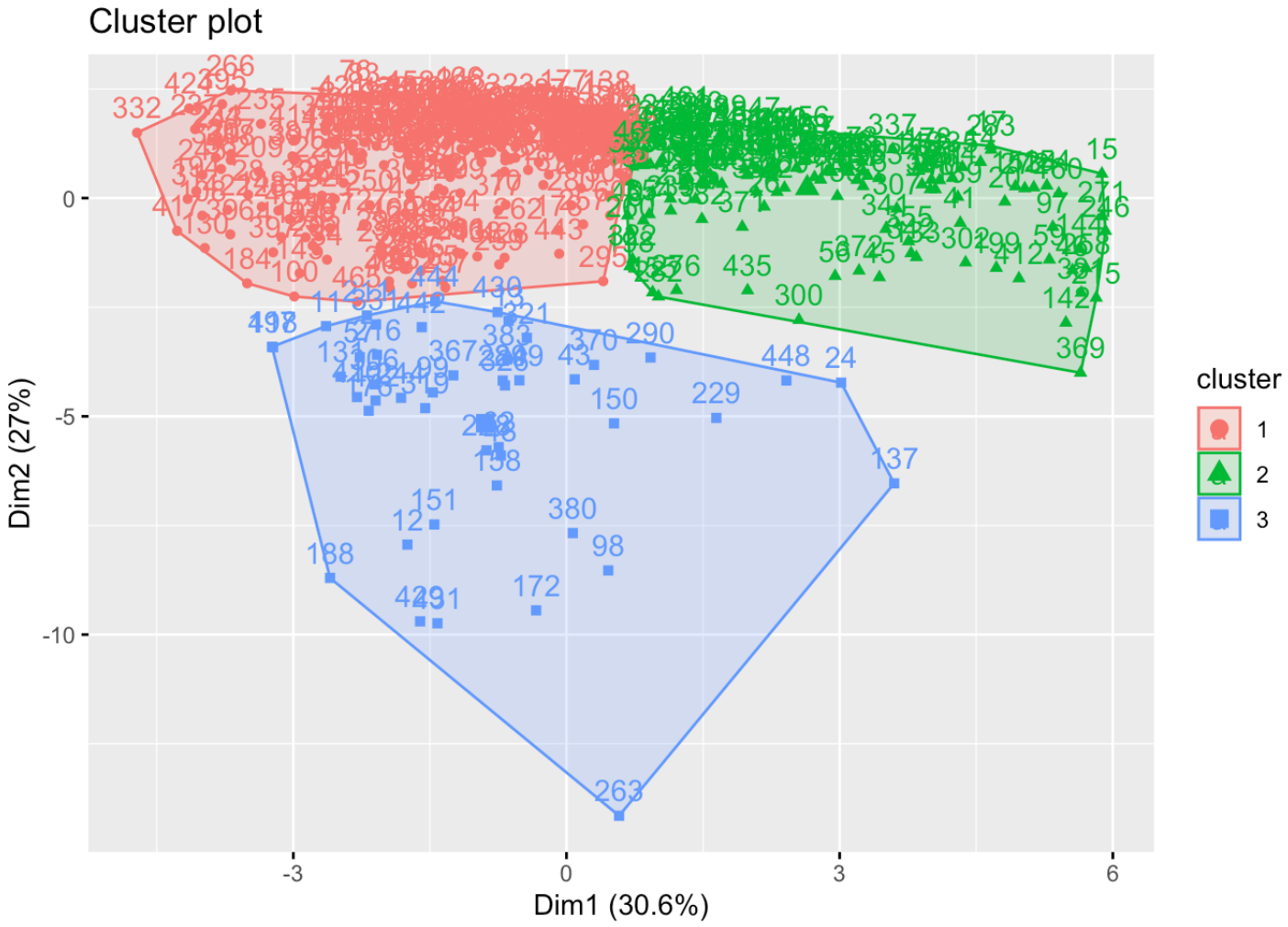
```
k3$size # Number of schools in each cluster
```

```
## [1] 273 150  46
```

```
k3$cluster[120] # Identify the cluster of the 120th observation as an example
```

```
## [1] 1
```

```
kmeansplot = fviz_cluster(k3, data = ScaledValues) # Visualize the output
kmeansplot
```



- Cluster 1 has low enrollment and lower student profile, medium student:faculty ratio and few faculty with PhDs
- Cluster 2 has high performing students, high selectivity and a relatively high cost. It also has a low student:faculty ratio and many PhD faculty. Most importantly, it has a very high Graduation Rate.
- Cluster 3 has high enrollment, low selectivity, low tuition but high costs, large class sizes and medium

performance by its students.

```
#now lets check the categorical variables
Universities$ClusterNum <- as.factor(k3$cluster)
with(Universities, table(ClusterNum,PublicPrivate))
```

```
##           PublicPrivate
## ClusterNum   1   2
##          1  83 190
##          2   3 147
##          3  41   5
```

```
with(Universities, table(ClusterNum,State))
```

```
##           State
## ClusterNum AK AL AR AZ CA CO CT DC DE FL GA HI IA ID IL IN KS KY LA MA MD ME MI
##          1  2  3  4  0  3  5  3  0  1  3  4  1 16  2  7  8  7  4  2  7  1  4  7
##          2  0  1  0  0 10  1  6  4  0  4  2  0  2  0  6  7  0  2  2 12  1  2  4
##          3  0  0  0  2  2  0  1  0  1  1  1  0  0  0  2  0  0  0  1  3  1  0  2
##           State
## ClusterNum MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA
##          1  6 12  5  2 16  5  5  4  9  2  0 17 13  5  1 19  1  7  4 11 13  1  8
##          2  4  2  0  0  3  0  1  1  3  0  0 18  7  0  4 20  2  2  0  3  2  0  4
##          3  1  1  0  0  4  0  1  1  1  0  0  2  4  1  0  3  1  0  0  1  4  1  3
##           State
## ClusterNum VT WA WI WV WY
##          1  5  0  5  2  1
##          2  2  2  4  0  0
##          3  0  0  0  0  0
```

We can see that group 2 - the high performers - are almost entirely private (98%); and that group 3 - the large schools - are 89% public. A correlation by state is not obvious.

Tufts University

```
UniversitiesTufts<-Universities
UniversitiesTufts[nrow(UniversitiesTufts)+1,]<-UniversitiesFull[which(UniversitiesFul
l$Name == "Tufts University"),]
#we'll normalize it as before

ScaledValuesT <- scale(UniversitiesTufts[,4:20], center=TRUE,scale=TRUE)
UniversitiesScaledT <- UniversitiesTufts
UniversitiesScaledT[,4:20]<- ScaledValuesT
ScaledValuesT[which(UniversitiesTufts$Name == "Tufts University"),]
```

```
##         Applied        Accepted        Enrolled           Top10           Top25
##       1.0912480       0.6139581       0.4601778       1.7192776       1.6786552
##      FullTimeUG      PartTimeUG  InStateTuition OutStateTuition            Room
##       0.2188536              NA       1.8547253       2.1005471       1.1399312
##           Board            Fees     EstBookCost     EstPersCost         FacWPHD
##       1.4188042       0.3469945       0.3142207      -0.5671244       1.5570294
##    StudFacRatio        GradRate
##      -0.9342687       1.4678934
```

```
clustercenters<-k3$centers
distance1 <- rbind(ScaledValuesT[which(UniversitiesTufts$Name == "Tufts University"),
],clustercenters[1,])
distance2 <- rbind(ScaledValuesT[which(UniversitiesTufts$Name == "Tufts University"),
],k3$centers[2,])
distance3 <- rbind(ScaledValuesT[which(UniversitiesTufts$Name == "Tufts University"),
],k3$centers[3,])
dist(distance1,method="euclidean") #distance from Tufts to Cluster 1 center
```

```
##          1
## 2 6.630903
```

```
dist(distance2,method="euclidean") #distance from Tufts to Cluster 2 center
```

```
##          1
## 2 2.734784
```

```
dist(distance3,method="euclidean") #distance from Tufts to Cluster 3 center
```

```
##          1
## 2 6.890573
```

Tufts is closest (at at 2.73 units) to **Cluster 2 , the high performer cluster**. Tufts is missing their Part-Time UG, so we can impute this from the cluster 2 average (-0.3144979).