

K means Clustering

ManojKumar Chalamala

10/29/2019

Install the required Packages

```
#install.packages("tidyverse")  
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.2.1 --
```

```
## v ggplot2 3.2.1      v purrr   0.3.2  
## v tibble  2.1.3      v dplyr   0.8.3  
## v tidyr   1.0.0      v stringr 1.4.0  
## v readr   1.3.1      v forcats 0.4.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()    masks stats::lag()
```

```
#install.packages("factoextra")  
library(factoextra)
```

```
## Welcome! Related Books: `Practical Guide To Cluster Analysis in R` at https://goo.gl/13EFCZ
```

```
library(readr)  
library(ISLR)  
library(caret)
```

```
## Loading required package: lattice
```

```
##  
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':  
##  
## lift
```

```
library(ggplot2)
```

Read the data

```
Univ <- read.csv("Universities.csv")
str(Univ)
```

```
## 'data.frame': 1302 obs. of 20 variables:
## $ College.Name : Factor w/ 1274 levels "Abilene Christian University",...: 8 1004
1005 1003 6 327 1084 7 43 88 ...
## $ State : Factor w/ 51 levels "AK","AL","AR",...: 1 1 1 1 2 2 2 2 2 ...
## $ Public..1...Private..2. : int 2 1 1 1 1 2 1 1 1 2 ...
## $ X..appli..rec.d : int 193 1852 146 2065 2817 345 1351 4639 7548 805 ...
## $ X..appl..accepted : int 146 1427 117 1598 1920 320 892 3272 6791 588 ...
## $ X..new.stud..enrolled : int 55 928 89 1162 984 179 570 1278 3070 287 ...
## $ X..new.stud..from.top.10.: int 16 NA 4 NA NA NA 18 NA 25 67 ...
## $ X..new.stud..from.top.25.: int 44 NA 24 NA NA 27 78 NA 57 88 ...
## $ X..FT.undergrad : int 249 3885 492 6209 3958 1367 2385 4051 16262 1376 ...
## $ X..PT.undergrad : int 869 4519 1849 10537 305 578 331 405 1716 207 ...
## $ in.state.tuition : int 7560 1742 1742 1742 1700 5600 2220 1500 2100 11660 ...
## $ out.of.state.tuition : int 7560 5226 5226 5226 3400 5600 4440 3000 6300 11660 ...
## $ room : int 1620 1800 2514 2600 1108 1550 NA 1960 NA 2050 ...
## $ board : int 2500 1790 2250 2520 1442 1700 NA NA NA 2430 ...
## $ add..fees : int 130 155 34 114 155 300 124 84 NA 120 ...
## $ estim..book.costs : int 800 650 500 580 500 350 300 500 600 400 ...
## $ estim..personal.. : int 1500 2304 1162 1260 850 NA 600 NA 1908 900 ...
## $ X..fac..w.PHD : int 76 67 39 48 53 52 72 48 85 74 ...
## $ stud..fac..ratio : num 11.9 10 9.5 13.7 14.3 32.8 18.9 18.7 16.7 14 ...
## $ Graduation.rate : int 15 NA 39 NA 40 55 51 15 69 72 ...
```

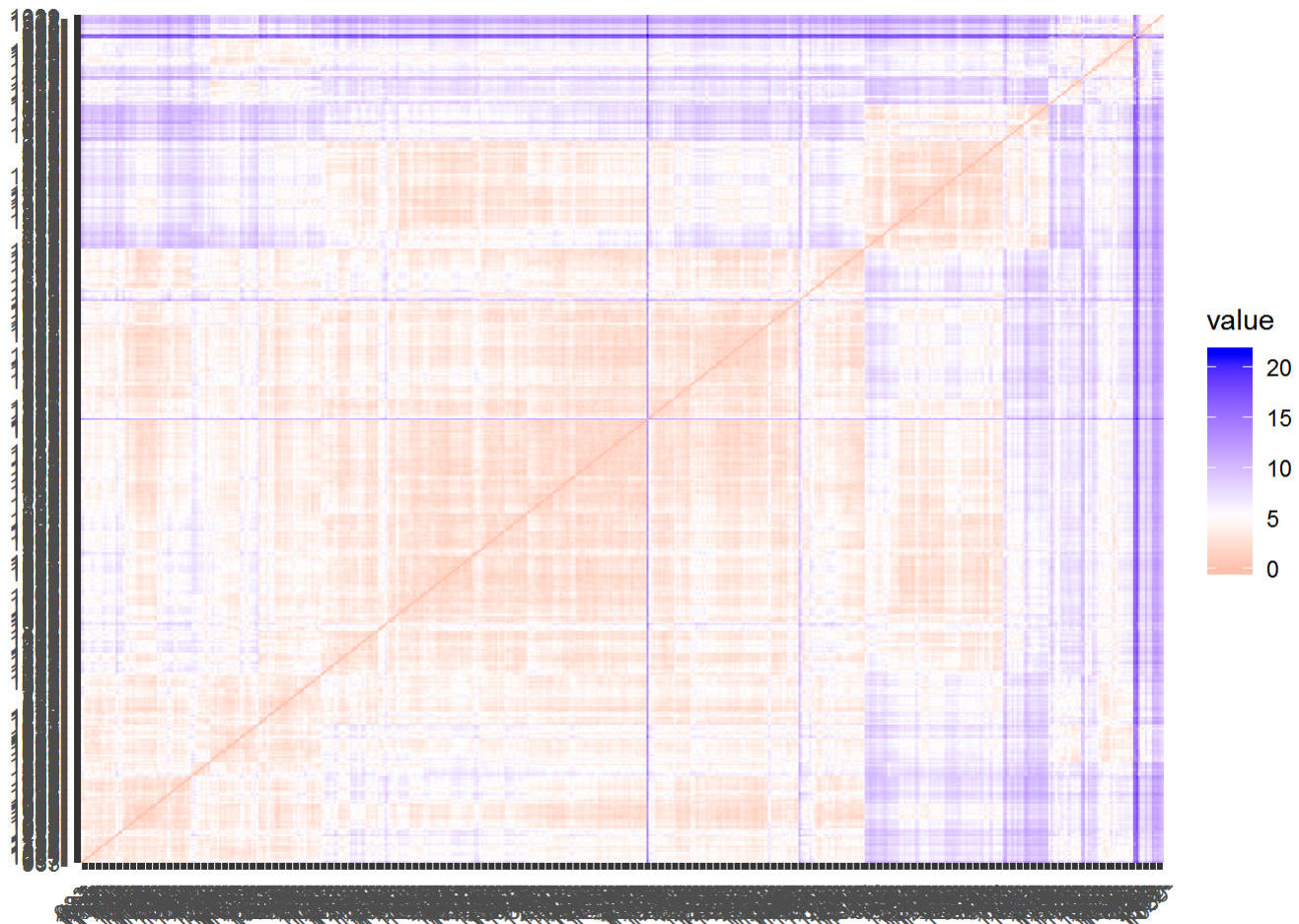
a. Remove all records with missing measurements from the dataset.

```
# remove all the missing values from the dataset by using na.omit
MyData <- na.omit(Univ)

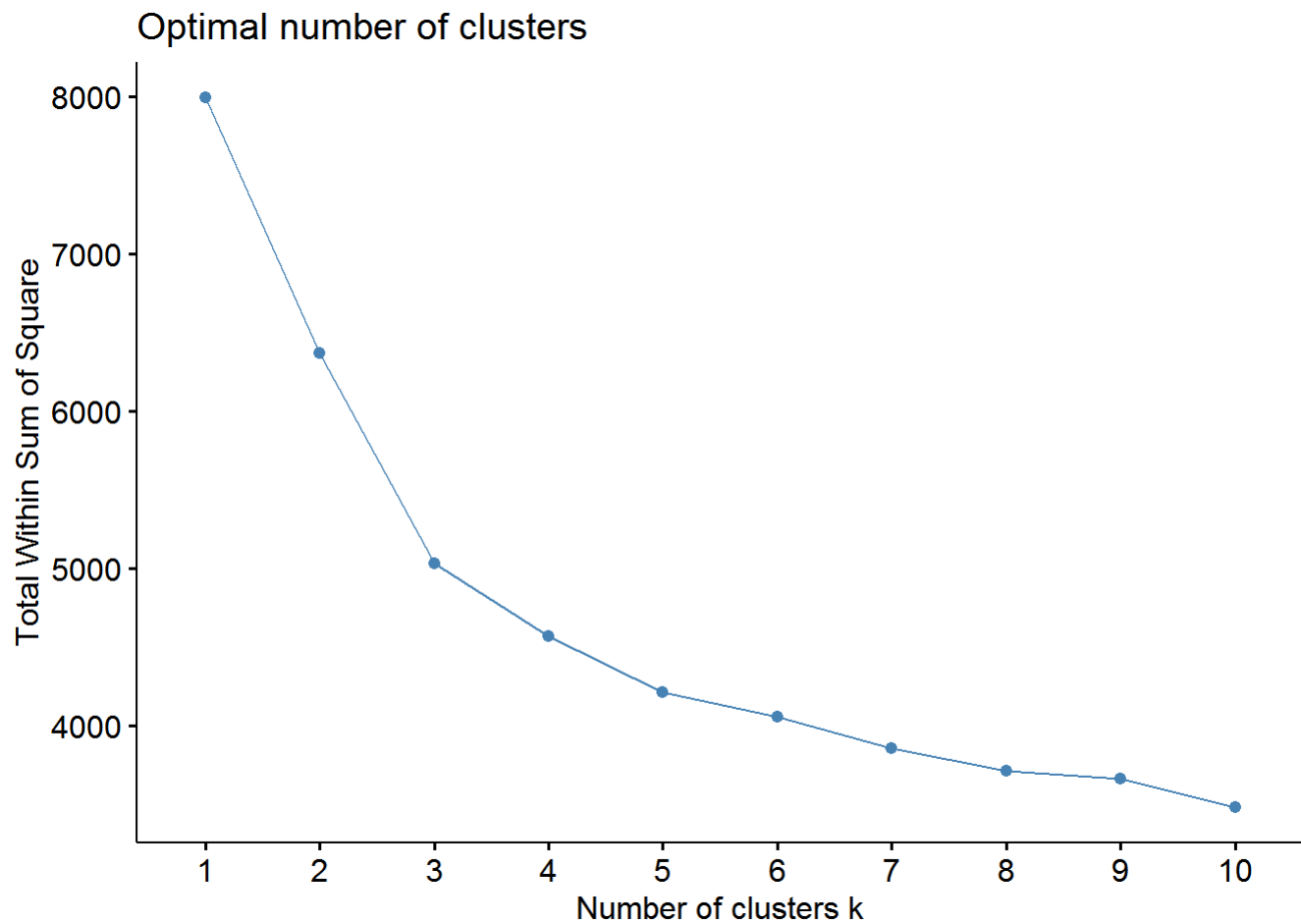
# remove all the categorical variables from the dataset
MyData <- MyData[, c(-1, -2, -3)]
```

b. 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?

```
# Scaling the data frame (z-score)
MyData <- scale(MyData)
View(MyData)
distance <- get_dist(MyData)
fviz_dist(distance)
```

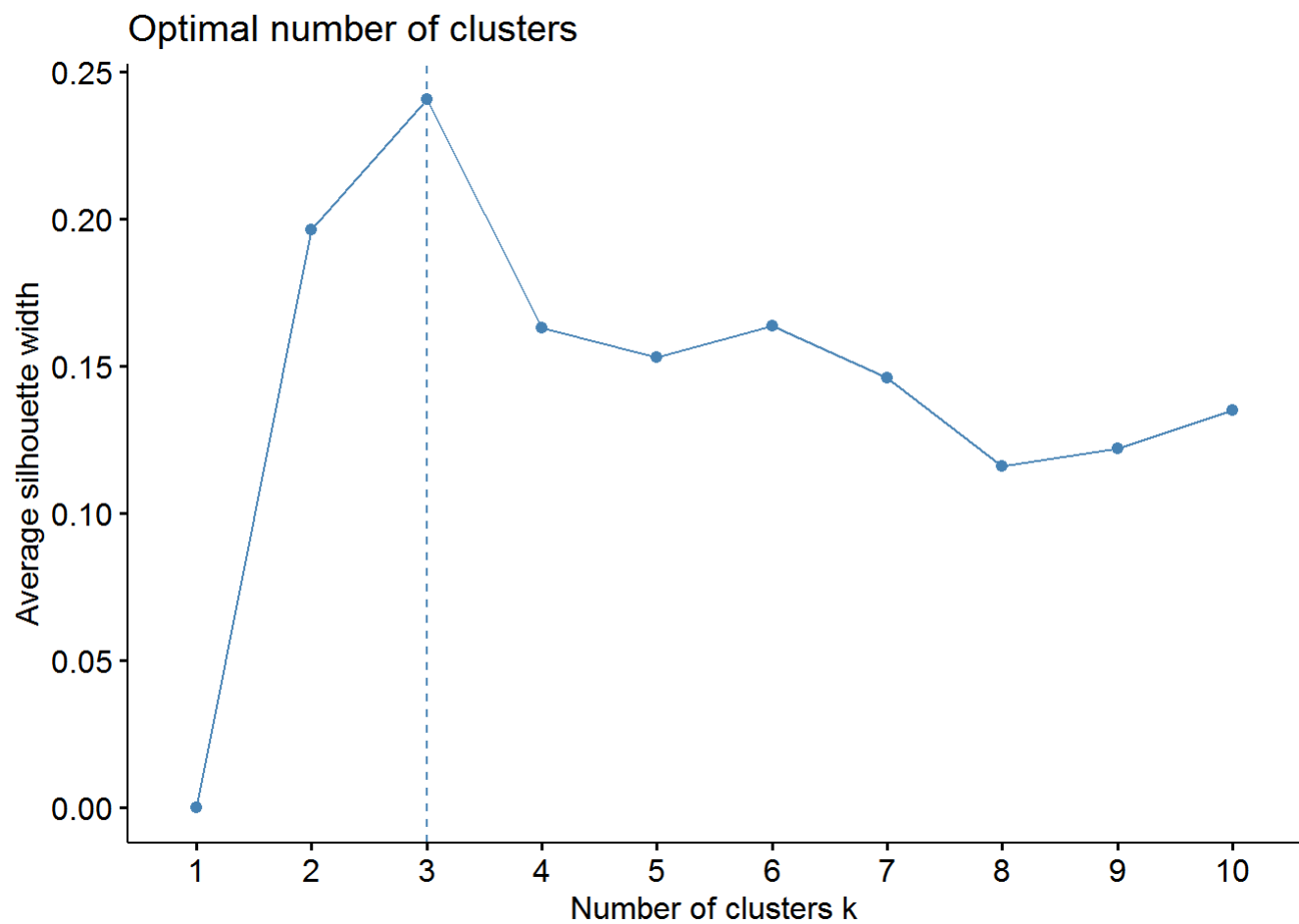


```
# Determining K
# WSS method - Let us use an "elbow chart" to determine k
fviz_nbclust(MyData, kmeans, method = "wss")
```



The chart shows that the elbow point 3 provides the best value for k.

Silhouette Method - Let us now apply the Silhouette Method to determine the number of clusters
`fviz_nbclust(MyData, kmeans, method = "silhouette")`



```
# We see that 3 is the ideal number of clusters.  
  
# Run the Kmeans algorithm for clustering.  
# We will choose an initial value of k = 3, number of restarts =15  
k3 <- kmeans(MyData, centers = 3, nstart = 15)  
  
# Visualize the output  
k3$centers # output the centers
```

```

## X..appli..rec.d X..appl..accepted X..new.stud..enrolled
## 1 -0.35953828 -0.34918455 -0.3171053
## 2 0.05140256 -0.04367128 -0.1683551
## 3 1.98179657 2.22992267 2.4447222
## X..new.stud..from.top.10. X..new.stud..from.top.25. X..FT.undergrad
## 1 -0.5020886 -0.5128195 -0.2952142
## 2 0.8795798 0.8620961 -0.2324464
## 3 0.1334215 0.2545856 2.5228452
## X..PT.undergrad in.state.tuition out.of.state.tuition room
## 1 -0.1217682 -0.4036544 -0.5263964 -0.3588740
## 2 -0.3130216 1.0620416 1.1158839 0.6698444
## 3 1.7486849 -1.0500277 -0.4918168 -0.0388330
## board add..fees estim..book.costs estim..personal.. X..fac..w.PHD
## 1 -0.3938990 -0.05832646 -0.06621454 0.05935933 -0.5322257
## 2 0.7756859 -0.04496556 0.07122705 -0.39665857 0.7659627
## 3 -0.1745795 0.49531762 0.16358567 0.93858632 0.6840794
## stud..fac..ratio Graduation.rate
## 1 0.2810858 -0.4171456
## 2 -0.7036167 0.8426062
## 3 0.6139980 -0.2538234

```

```
k3$size #No. of Universities in each cluster
```

```
## [1] 275 150 46
```

```
k3 # mean of each cluster
```

```

## K-means clustering with 3 clusters of sizes 275, 150, 46
##
## Cluster means:
##   X..appli..rec.d X..appl..accepted X..new.stud..enrolled
## 1   -0.35953828   -0.34918455   -0.3171053
## 2    0.05140256   -0.04367128   -0.1683551
## 3    1.98179657    2.22992267    2.4447222
##   X..new.stud..from.top.10. X..new.stud..from.top.25. X..FT.undergrad
## 1   -0.5020886   -0.5128195   -0.2952142
## 2    0.8795798    0.8620961   -0.2324464
## 3    0.1334215    0.2545856    2.5228452
##   X..PT.undergrad in.state.tuition out.of.state.tuition   room
## 1   -0.1217682   -0.4036544   -0.5263964 -0.3588740
## 2   -0.3130216    1.0620416    1.1158839  0.6698444
## 3    1.7486849   -1.0500277   -0.4918168 -0.0388330
##   board   add..fees estim..book.costs estim..personal.. X..fac..w.PHD
## 1 -0.3938990 -0.05832646   -0.06621454    0.05935933   -0.5322257
## 2  0.7756859 -0.04496556    0.07122705   -0.39665857    0.7659627
## 3 -0.1745795  0.49531762    0.16358567    0.93858632    0.6840794
##   stud..fac..ratio Graduation.rate
## 1    0.2810858   -0.4171456
## 2   -0.7036167    0.8426062
## 3    0.6139980   -0.2538234
##
## Clustering vector:
##   1   3   10   12   22   26   32   38   39   46   49   50   63   77   78
##   1   1   2   1   1   1   1   1   1   1   3   3   3   2   2
##  79   81   90   92   95   96   97  108  110  112  120  121  122  123  126
##   2   2   2   1   2   1   2   2   3   2   2   1   1   2   1
## 127  130  134  139  140  146  148  149  151  152  153  154  156  158  160
##   1   1   1   1   2   1   2   2   2   2   2   1   3   2   2
## 161  164  168  169  172  174  176  181  186  188  190  194  208  210  220
##   2   2   3   1   1   2   1   1   2   2   2   3   1   2   1
## 228  230  235  236  239  244  245  246  247  248  250  251  252  258  259
##   2   3   1   1   1   1   1   1   2   1   1   1   2   1   1
## 260  262  263  264  265  268  269  270  272  275  277  287  294  297  298
##   1   1   1   1   1   1   1   1   1   1   2   1   2   2   2
## 302  304  312  317  319  320  321  322  323  324  326  329  330  331  332
##   1   2   1   1   1   1   2   3   3   1   1   2   2   2   1
## 333  336  340  349  352  359  360  362  365  366  368  369  370  376  377
##   1   1   1   1   1   1   2   2   2   2   1   1   1   1   1
## 382  387  390  395  398  400  403  405  412  415  418  421  428  433  436
##   1   1   1   2   1   1   1   2   1   1   3   2   2   2   2
## 438  441  442  443  444  451  452  454  456  457  460  462  463  464  467
##   2   3   1   2   2   1   2   1   2   1   1   1   1   2   3
## 477  479  481  483  484  494  500  505  511  513  514  515  520  522  525
##   3   2   2   2   2   2   1   3   1   2   2   1   1   1   1
## 526  527  528  537  538  541  543  544  551  552  556  558  563  564  565
##   2   2   1   2   2   1   3   1   1   1   3   1   2   1   1
## 567  571  572  577  578  579  582  584  587  589  591  595  598  599  605
##   2   2   2   1   1   1   1   3   1   1   1   1   1   1   1
## 606  607  608  614  615  616  617  618  626  627  629  630  633  638  642
##   1   3   2   2   1   1   1   1   1   1   1   1   1   1   1

```

```

## 646 647 651 654 655 656 658 659 660 662 666 667 670 673 674
## 3 1 1 2 2 3 1 1 1 1 1 1 1 1 1
## 676 677 681 682 684 687 688 689 691 693 696 697 698 702 703
## 1 2 3 3 1 1 1 1 1 1 1 1 1 2 1
## 704 705 710 712 713 716 717 720 723 724 726 732 733 736 737
## 1 1 1 3 1 2 1 1 3 1 1 2 1 1 1
## 738 739 742 744 745 746 750 751 754 756 757 769 771 772 777
## 1 1 2 1 2 1 1 3 1 1 1 2 1 1 2
## 778 782 783 789 792 793 794 795 801 803 804 814 815 823 824
## 2 2 2 1 2 2 2 1 2 2 1 1 2 2 2
## 825 826 828 831 833 834 836 837 838 839 840 841 843 844 845
## 2 1 1 2 3 3 1 1 1 1 1 1 1 1 1
## 846 848 851 860 869 870 872 874 875 878 879 882 885 889 891
## 2 2 2 1 1 1 3 2 1 1 1 2 1 2 1
## 892 893 894 896 898 901 903 904 907 911 912 916 917 928 929
## 2 1 1 2 3 1 3 2 1 2 1 3 1 1 1
## 931 932 933 939 943 945 950 952 955 958 959 963 965 967 969
## 1 3 1 1 1 2 2 2 2 2 2 2 2 2 2
## 971 974 975 977 978 979 986 987 988 989 991 992 994 996 997
## 2 1 2 1 1 2 2 1 2 2 2 1 1 2 2
## 1001 1009 1010 1014 1017 1020 1021 1023 1024 1025 1026 1027 1029 1030 1031
## 1 1 1 1 1 1 2 3 1 2 3 2 2 2 1
## 1032 1033 1035 1036 1037 1039 1041 1043 1047 1048 1051 1052 1053 1055 1059
## 1 1 1 1 1 3 2 2 3 1 1 1 1 1 1
## 1060 1061 1064 1065 1075 1079 1081 1084 1087 1089 1090 1095 1096 1098 1101
## 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1
## 1102 1105 1107 1110 1111 1115 1117 1118 1121 1125 1127 1131 1132 1138 1139
## 1 1 2 1 1 3 2 2 1 1 1 1 1 3 1
## 1143 1146 1152 1154 1156 1158 1163 1164 1166 1168 1172 1176 1177 1181 1185
## 1 1 1 1 1 1 2 2 1 3 3 1 3 3 1
## 1188 1189 1192 1194 1195 1196 1198 1204 1206 1212 1214 1218 1221 1222 1223
## 1 2 1 2 1 2 1 1 3 1 3 1 1 2 3
## 1227 1231 1232 1236 1237 1238 1239 1246 1253 1257 1258 1262 1268 1269 1273
## 1 1 1 1 2 1 2 2 2 2 1 2 2 2 1
## 1274 1275 1284 1285 1292 1302
## 1 1 1 1 1 1
##
## Within cluster sum of squares by cluster:
## [1] 2562.342 1424.892 1044.680
## (between_SS / total_SS = 37.0 %)
##
## Available components:
##
## [1] "cluster" "centers" "totss" "withinss"
## [5] "tot.withinss" "betweenss" "size" "iter"
## [9] "ifault"

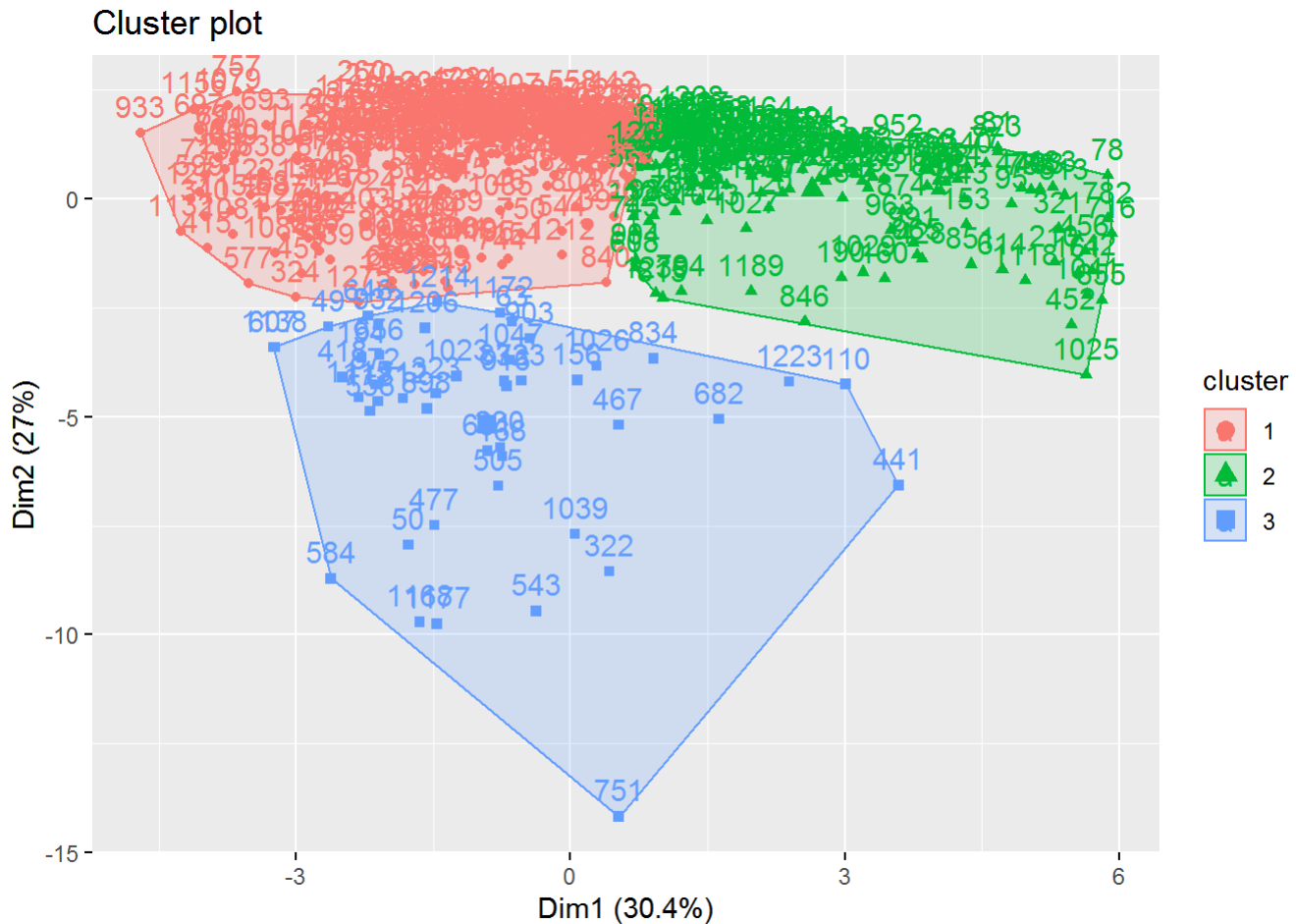
```

```
k3$cluster[325] # Identify the cluster of the 325th observation as an example
```

```
## 911
## 2
```



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



c. Compare the summary statistics for each cluster and describe each cluster in this context (e.g., “Universities with high tuition, low acceptance rate...”)

```
# mean of each cluster
k3$centers
```

```
## X..appli..rec.d X..appl..accepted X..new.stud..enrolled
## 1 -0.35953828 -0.34918455 -0.3171053
## 2 0.05140256 -0.04367128 -0.1683551
## 3 1.98179657 2.22992267 2.4447222
## X..new.stud..from.top.10. X..new.stud..from.top.25. X..FT.undergrad
## 1 -0.5020886 -0.5128195 -0.2952142
## 2 0.8795798 0.8620961 -0.2324464
## 3 0.1334215 0.2545856 2.5228452
## X..PT.undergrad in.state.tuition out.of.state.tuition room
## 1 -0.1217682 -0.4036544 -0.5263964 -0.3588740
## 2 -0.3130216 1.0620416 1.1158839 0.6698444
## 3 1.7486849 -1.0500277 -0.4918168 -0.0388330
## board add..fees estim..book.costs estim..personal.. X..fac..w.PHD
## 1 -0.3938990 -0.05832646 -0.06621454 0.05935933 -0.5322257
## 2 0.7756859 -0.04496556 0.07122705 -0.39665857 0.7659627
## 3 -0.1745795 0.49531762 0.16358567 0.93858632 0.6840794
## stud..fac..ratio Graduation.rate
## 1 0.2810858 -0.4171456
## 2 -0.7036167 0.8426062
## 3 0.6139980 -0.2538234
```

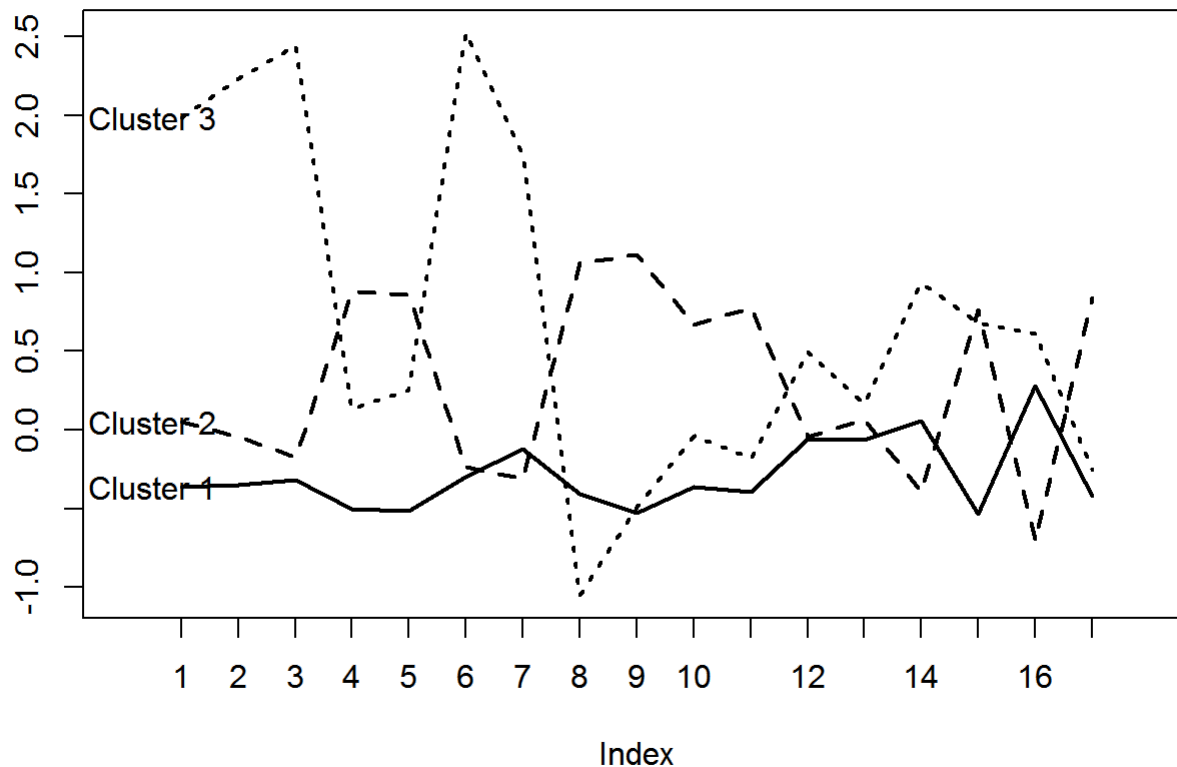
Based on mean of each cluster compare the statistics of each cluster and described each cluster with High, Low and medium context

ComparedData <- read.csv("ClusterComparison.csv") # Comparison of each cluster is stored in this excel

ComparedData #Displays the comparison of each cluster for all columns

```
## Columns Cluster.1 Cluster.2 Cluster.3
## 1 appli. rec'd # LOW MEDIUM HIGH
## 2 appl. accepted # LOW MEDIUM HIGH
## 3 new stud. enrolled % LOW MEDIUM HIGH
## 4 new stud. from top 10% # LOW HIGH MEDIUM
## 5 new stud. from top 25% # LOW HIGH MEDIUM
## 6 FT undergrad # LOW MEDIUM HIGH
## 7 PT undergrad MEDIUM LOW HIGH
## 8 n-state tuition o MEDIUM HIGH LOW
## 9 ut-of-state tuition LOW HIGH MEDIUM
## 10 room LOW HIGH MEDIUM
## 11 board LOW HIGH MEDIUM
## 12 add. fees estim. LOW MEDIUM HIGH
## 13 book costs estim. LOW MEDIUM HIGH
## 14 personal $ % MEDIUM LOW HIGH
## 15 % fac. w/PHD LOW HIGH MEDIUM
## 16 stud./fac. ratio MEDIUM LOW HIGH
## 17 Graduation rate LOW HIGH MEDIUM
```

```
# plot an empty scatter plot
windows()
plot(c(0), xaxt = 'n', ylab = "", type = "l",
ylim = c(min(k3$centers), max(k3$centers)), xlim = c(0, 18))
# label x-axes
axis(1, at = c(1:17), labels = names(k3$centers))
# plot centroids
for (i in c(1:3))
lines(k3$centers[i,], lty = i, lwd = 2, col = ifelse(i %in% c(1, 2, 3),
"black", "dark grey"))
# name clusters
text(x = 0.5, y = k3$centers[, 1], labels = paste("Cluster", c(1:3)))
```



The above graph displays the comparison of summary stats of each cluster

d. 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?

```

MyData1 <- Univ
MyData1 <- na.omit(MyData1)
View(MyData1)

MyData2 <- cbind(MyData1[,c(1:3)],k3$cluster)

MyData2$Public..1...Private..2. <- factor(MyData2$Public..1...Private..2.,levels = c(1,2),labels
= c("public","private"))

MyData2$k3$cluster` <- factor(MyData2$k3$cluster`,levels = c(1,2,3),labels = c("Low ","Medium"
, "High"))

head(MyData2)

```

```

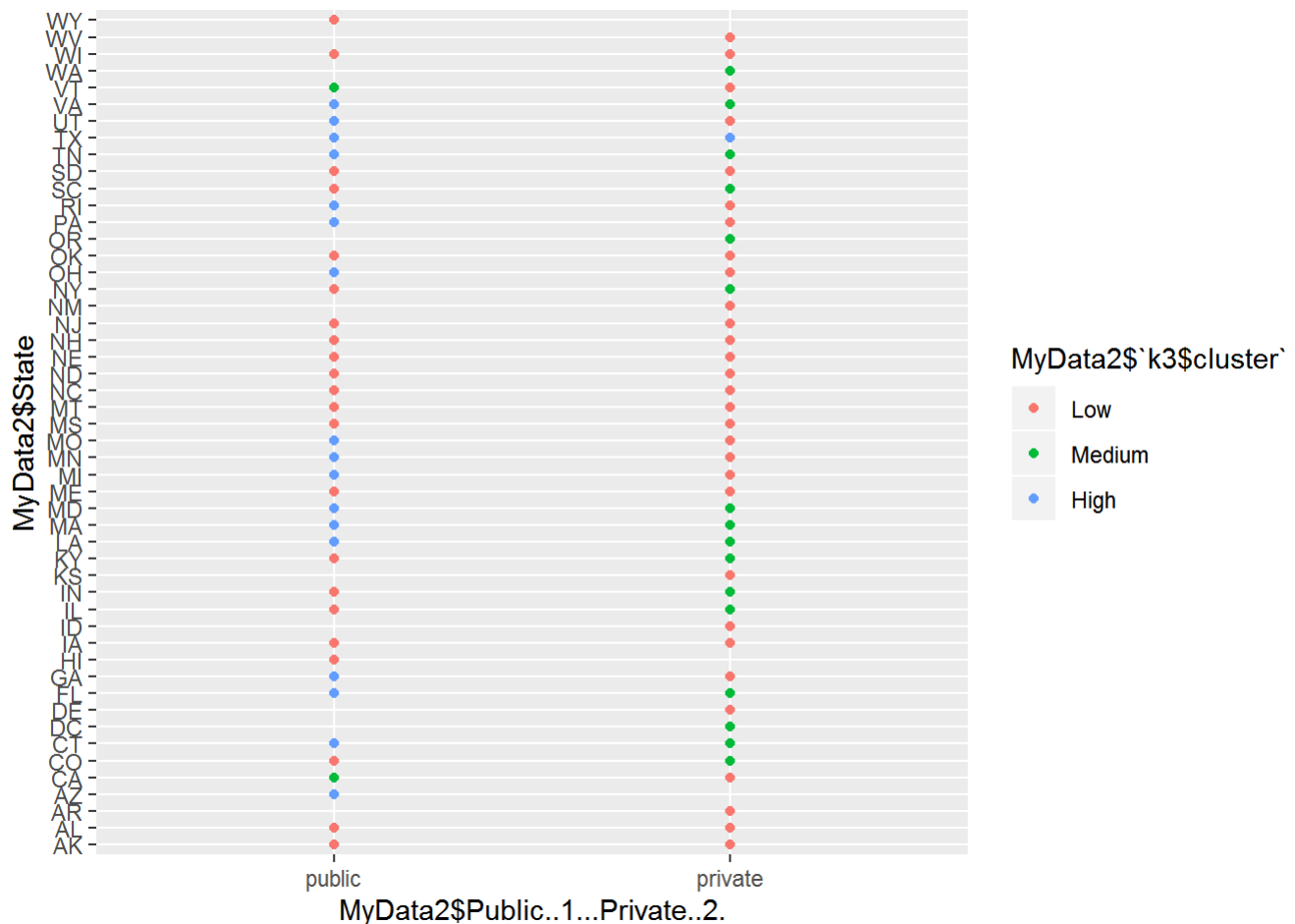
##              College.Name State Public..1...Private..2.
## 1      Alaska Pacific University      AK      private
## 3      University of Alaska Southeast      AK      public
## 10     Birmingham-Southern College      AL      private
## 12           Huntingdon College      AL      private
## 22           Talladega College      AL      private
## 26 University of Alabama at Birmingham      AL      public
##      k3$cluster
## 1      Low
## 3      Low
## 10     Medium
## 12     Low
## 22     Low
## 26     Low

```

```

ggplot(MyData2, aes(x=MyData2$Public..1...Private..2., y=MyData2$State, color=MyData2$k3$cluster`)) + geom_point()

```



Based on the graph,

Cluster 1 shows equivalent proportion of both Public and Private Universities

Cluster 2 shows more Public Universities when compared with private

Cluster 3 shows more Private Universities when compared with public

e. What other external information can explain the contents of some or all of these clusters?

```
MyData2[MyData2$k3$cluster=="Low ", ]
```

##	College.Name	State
## 1	Alaska Pacific University	AK
## 3	University of Alaska Southeast	AK
## 12	Huntingdon College	AL
## 22	Talladega College	AL
## 26	University of Alabama at Birmingham	AL
## 32	Arkansas College (Lyon College)	AR
## 38	Hendrix College	AR
## 39	John Brown University	AR
## 46	Harding University	AR
## 92	Fresno Pacific College	CA
## 96	Southern California College	CA
## 121	Concordia University	CA
## 122	Adams State College	CO
## 126	Fort Lewis College	CO
## 127	Mesa State College	CO
## 130	University of Southern Colorado	CO
## 134	Western State College of Colorado	CO
## 139	Central Connecticut State University	CT
## 146	Sacred Heart University	CT
## 154	Eastern Connecticut State University	CT
## 169	Wesley College	DE
## 172	Bethune Cookman College	FL
## 176	Embry Riddle Aeronautical University	FL
## 181	Florida Southern College	FL
## 208	Brewton-Parker College	GA
## 220	Morehouse College	GA
## 235	Covenant College	GA
## 236	Savannah Coll. of Art and Design	GA
## 239	University of Hawaii at Manoa	HI
## 244	Central College	IA
## 245	Clarke College	IA
## 246	Coe College	IA
## 248	Dordt College	IA
## 250	Graceland College	IA
## 251	Grand View College	IA
## 258	Morningside College	IA
## 259	Mount Mercy College	IA
## 260	Mount Saint Clare College	IA
## 262	Simpson College	IA
## 263	Saint Ambrose University	IA
## 264	University of Northern Iowa	IA
## 265	University of Dubuque	IA
## 268	Wartburg College	IA
## 269	Teikyo Westmar University	IA
## 270	William Penn College	IA
## 272	Albertson College	ID
## 275	Northwest Nazarene College	ID
## 287	Eureka College	IL
## 302	McKendree College	IL
## 312	Quincy University	IL
## 317	Southern Illinois University at Edwardsville	IL
## 319	Saint Xavier University	IL

## 320	Trinity College	IL
## 324	Western Illinois University	IL
## 326	Anderson University	IN
## 332	University of Evansville	IN
## 333	Franklin College	IN
## 336	Hanover College	IN
## 340	University of Southern Indiana	IN
## 349	Manchester College	IN
## 352	Oakland City College	IN
## 359	Saint Mary-of-the-Woods College	IN
## 368	Baker University	KS
## 369	Bethany College	KS
## 370	Bethel College	KS
## 376	Kansas Wesleyan University	KS
## 377	McPherson College	KS
## 382	Tabor College	KS
## 387	Benedictine College	KS
## 390	Bellarmino College	KY
## 398	Georgetown College	KY
## 400	Kentucky Wesleyan College	KY
## 403	Murray State University	KY
## 412	Centenary College of Louisiana	LA
## 415	Grambling State University	LA
## 442	Bradford College	MA
## 451	Gordon College	MA
## 454	Lesley College	MA
## 457	Bridgewater State College	MA
## 460	North Adams State College	MA
## 462	Westfield State College	MA
## 463	Worcester State College	MA
## 500	Salisbury State University	MD
## 511	University of Maine at Presque Isle	ME
## 515	University of Maine at Farmington	ME
## 520	Thomas College	ME
## 522	University of Maine at Machias	ME
## 525	Adrian College	MI
## 528	Andrews University	MI
## 541	Madonna University	MI
## 544	Michigan Technological University	MI
## 551	Spring Arbor College	MI
## 552	University of Detroit Mercy	MI
## 558	Center for Creative Studies	MI
## 564	College of Saint Benedict	MN
## 565	College of Saint Catherine	MN
## 577	Saint Cloud State University	MN
## 578	Saint John's University	MN
## 579	Saint Mary's College of Minnesota	MN
## 582	University of Minnesota at Morris	MN
## 587	Central Methodist College	MO
## 589	Columbia College	MO
## 591	Drury College	MO
## 595	Lindenwood College	MO
## 598	Missouri Valley College	MO
## 599	Missouri Western State College	MO

## 605	Southeast Missouri State University	MO
## 606	Southwest Baptist University	MO
## 615	Webster University	MO
## 616	Westminster College	MO
## 617	William Jewell College	MO
## 618	William Woods University	MO
## 626	Millsaps College	MS
## 627	Mississippi College	MS
## 629	Mississippi State University	MS
## 630	Mississippi Valley State University	MS
## 633	University of Mississippi	MS
## 638	Montana State University-Billings	MT
## 642	Rocky Mountain College	MT
## 647	University of North Carolina at Asheville	NC
## 651	Bennett College	NC
## 658	Elon College	NC
## 659	Fayetteville State University	NC
## 660	Gardner Webb University	NC
## 662	Guilford College	NC
## 666	Lenoir-Rhyne College	NC
## 667	Livingstone College	NC
## 670	Methodist College	NC
## 673	North Carolina Wesleyan College	NC
## 674	Pembroke State University	NC
## 676	Queens College	NC
## 684	University of North Carolina at Greensboro	NC
## 687	Western Carolina University	NC
## 688	University of North Carolina at Wilmington	NC
## 689	Wingate College	NC
## 691	Dickinson State University	ND
## 693	University of Mary	ND
## 696	University of North Dakota	ND
## 697	Valley City State University	ND
## 698	North Dakota State University	ND
## 703	Dana College	NE
## 704	Doane College	NE
## 705	Hastings College	NE
## 710	Peru State College	NE
## 713	Wayne State College	NE
## 717	Franklin Pierce College	NH
## 720	Notre Dame College	NH
## 724	Keene State College	NH
## 726	Daniel Webster College	NH
## 733	Georgian Court College	NJ
## 736	Jersey City State College	NJ
## 737	Monmouth College	NJ
## 738	Montclair State University	NJ
## 739	New Jersey Institute of Technology	NJ
## 744	Rutgers State University at Newark	NJ
## 746	Saint Peter's College	NJ
## 750	Rutgers State University at Camden	NJ
## 754	Stockton College of New Jersey	NJ
## 756	College of Santa Fe	NM
## 757	College of the Southwest	NM

## 771	Canisius College	NY
## 772	Cazenovia College	NY
## 789	Elmira College	NY
## 795	Houghton College	NY
## 804	Marist College	NY
## 814	Nyack College	NY
## 826	St. Bonaventure University	NY
## 828	St. John Fisher College	NY
## 836	SUNY College at Brockport	NY
## 837	SUNY College at Buffalo	NY
## 838	SUNY College at Cortland	NY
## 839	SUNY College at Fredonia	NY
## 840	SUNY College at Geneseo	NY
## 841	SUNY College at New Paltz	NY
## 843	SUNY College at Oswego	NY
## 844	SUNY College at Plattsburgh	NY
## 845	SUNY College at Potsdam	NY
## 860	SUNY College at Purchase	NY
## 869	Ashland University	OH
## 870	Baldwin-Wallace College	OH
## 875	Cedarville College	OH
## 878	College of Mount St. Joseph	OH
## 879	Franciscan University of Steubenville	OH
## 885	Heidelberg College	OH
## 891	Malone College	OH
## 893	Mount Union College	OH
## 894	Muskingum College	OH
## 901	Tiffin University	OH
## 907	Ursuline College	OH
## 912	Xavier University	OH
## 917	Antioch University	OH
## 928	Oklahoma Baptist University	OK
## 929	Oklahoma Christian University	OK
## 931	University of Sci. and Arts of Oklahoma	OK
## 933	Oklahoma Panhandle State University	OK
## 939	Oral Roberts University	OK
## 943	George Fox College	OR
## 974	Eastern College	PA
## 977	Gannon University	PA
## 978	Geneva College	PA
## 987	King's College	PA
## 992	Lincoln University	PA
## 994	Marywood College	PA
## 1001	Bloomsburg Univ. of Pennsylvania	PA
## 1009	Millersville University of Penn.	PA
## 1010	Shippensburg University of Penn.	PA
## 1014	Philadelphia Coll. of Textiles and Sci.	PA
## 1017	Seton Hill College	PA
## 1020	Saint Vincent College	PA
## 1024	Thiel College	PA
## 1031	Waynesburg College	PA
## 1032	Westminster College	PA
## 1033	Wilkes University	PA
## 1035	York College of Pennsylvania	PA

## 1036	Allentown Coll. of St. Francis de Sales	PA
## 1037	La Roche College	PA
## 1048	Roger Williams University	RI
## 1051	Central Wesleyan College	SC
## 1052	The Citadel	SC
## 1053	Claflin College	SC
## 1055	Coker College	SC
## 1059	Erskine College	SC
## 1061	Lander University	SC
## 1064	Newberry College	SC
## 1075	Augustana College	SD
## 1079	Huron University	SD
## 1081	Northern State University	SD
## 1084	South Dakota State University	SD
## 1087	Belmont University	TN
## 1089	Carson-Newman College	TN
## 1090	Christian Brothers University	TN
## 1095	Freed-Hardeman University	TN
## 1096	King College	TN
## 1098	Lambuth University	TN
## 1101	Lincoln Memorial University	TN
## 1102	Maryville College	TN
## 1105	Milligan College	TN
## 1110	Tennessee Wesleyan College	TN
## 1111	Trevecca Nazarene College	TN
## 1121	Angelo State University	TX
## 1125	East Texas Baptist University	TX
## 1127	Hardin-Simmons University	TX
## 1131	Incarnate Word College	TX
## 1132	Lamar University	TX
## 1139	Our Lady of the Lake University	TX
## 1143	Schreiner College	TX
## 1146	Southwestern Adventist College	TX
## 1152	Prairie View A. and M. University	TX
## 1154	Texas Christian University	TX
## 1156	Texas College	TX
## 1158	Texas Lutheran College	TX
## 1166	University of St. Thomas	TX
## 1176	Texas A&M University at Galveston	TX
## 1185	Westminster College of Salt Lake City	UT
## 1188	Bridgewater College	VA
## 1192	Emory & Henry College	VA
## 1195	Hampton University	VA
## 1198	Lynchburg College	VA
## 1204	Randolph-Macon College	VA
## 1212	Mary Washington College	VA
## 1218	Virginia State University	VA
## 1221	Virginia Wesleyan College	VA
## 1227	Castleton State College	VT
## 1231	Johnson State College	VT
## 1232	Lyndon State College	VT
## 1236	Southern Vermont College	VT
## 1238	Trinity College	VT
## 1258	Carroll College	WI

```

## 1273                Viterbo College      WI
## 1274      University of Wisconsin-Stout  WI
## 1275      Univ. of Wisconsin at Eau Claire  WI
## 1284                Wisconsin Lutheran College  WI
## 1285                Alderson-Broaddus College  WV
## 1292                Marshall University      WV
## 1302                University of Wyoming    WY
##      Public..1...Private..2. k3$cluster
## 1                private      Low
## 3                public      Low
## 12               private      Low
## 22               private      Low
## 26               public      Low
## 32               private      Low
## 38               private      Low
## 39               private      Low
## 46               private      Low
## 92               private      Low
## 96               private      Low
## 121              private      Low
## 122              public      Low
## 126              public      Low
## 127              public      Low
## 130              public      Low
## 134              public      Low
## 139              public      Low
## 146              private      Low
## 154              public      Low
## 169              private      Low
## 172              private      Low
## 176              private      Low
## 181              private      Low
## 208              private      Low
## 220              private      Low
## 235              private      Low
## 236              private      Low
## 239              public      Low
## 244              private      Low
## 245              private      Low
## 246              private      Low
## 248              private      Low
## 250              private      Low
## 251              private      Low
## 258              private      Low
## 259              private      Low
## 260              private      Low
## 262              private      Low
## 263              private      Low
## 264              public      Low
## 265              private      Low
## 268              private      Low
## 269              private      Low
## 270              private      Low
## 272              private      Low

```

## 275	private	Low
## 287	private	Low
## 302	private	Low
## 312	private	Low
## 317	public	Low
## 319	private	Low
## 320	private	Low
## 324	public	Low
## 326	private	Low
## 332	private	Low
## 333	private	Low
## 336	private	Low
## 340	public	Low
## 349	private	Low
## 352	private	Low
## 359	private	Low
## 368	private	Low
## 369	private	Low
## 370	private	Low
## 376	private	Low
## 377	private	Low
## 382	private	Low
## 387	private	Low
## 390	private	Low
## 398	private	Low
## 400	private	Low
## 403	public	Low
## 412	private	Low
## 415	public	Low
## 442	private	Low
## 451	private	Low
## 454	private	Low
## 457	public	Low
## 460	public	Low
## 462	public	Low
## 463	public	Low
## 500	public	Low
## 511	public	Low
## 515	public	Low
## 520	private	Low
## 522	public	Low
## 525	private	Low
## 528	private	Low
## 541	private	Low
## 544	public	Low
## 551	private	Low
## 552	private	Low
## 558	private	Low
## 564	private	Low
## 565	private	Low
## 577	public	Low
## 578	private	Low
## 579	private	Low
## 582	public	Low

## 587	private	Low
## 589	private	Low
## 591	private	Low
## 595	private	Low
## 598	private	Low
## 599	public	Low
## 605	public	Low
## 606	private	Low
## 615	private	Low
## 616	private	Low
## 617	private	Low
## 618	private	Low
## 626	private	Low
## 627	private	Low
## 629	public	Low
## 630	public	Low
## 633	public	Low
## 638	public	Low
## 642	private	Low
## 647	public	Low
## 651	private	Low
## 658	private	Low
## 659	public	Low
## 660	private	Low
## 662	private	Low
## 666	private	Low
## 667	private	Low
## 670	private	Low
## 673	private	Low
## 674	public	Low
## 676	private	Low
## 684	public	Low
## 687	public	Low
## 688	public	Low
## 689	private	Low
## 691	public	Low
## 693	private	Low
## 696	public	Low
## 697	public	Low
## 698	public	Low
## 703	private	Low
## 704	private	Low
## 705	private	Low
## 710	public	Low
## 713	public	Low
## 717	private	Low
## 720	private	Low
## 724	public	Low
## 726	private	Low
## 733	private	Low
## 736	public	Low
## 737	private	Low
## 738	public	Low
## 739	public	Low

## 744	public	Low
## 746	private	Low
## 750	public	Low
## 754	public	Low
## 756	private	Low
## 757	private	Low
## 771	private	Low
## 772	private	Low
## 789	private	Low
## 795	private	Low
## 804	private	Low
## 814	private	Low
## 826	private	Low
## 828	private	Low
## 836	public	Low
## 837	public	Low
## 838	public	Low
## 839	public	Low
## 840	public	Low
## 841	public	Low
## 843	public	Low
## 844	public	Low
## 845	public	Low
## 860	public	Low
## 869	private	Low
## 870	private	Low
## 875	private	Low
## 878	private	Low
## 879	private	Low
## 885	private	Low
## 891	private	Low
## 893	private	Low
## 894	private	Low
## 901	private	Low
## 907	private	Low
## 912	private	Low
## 917	private	Low
## 928	private	Low
## 929	private	Low
## 931	public	Low
## 933	public	Low
## 939	private	Low
## 943	private	Low
## 974	private	Low
## 977	private	Low
## 978	private	Low
## 987	private	Low
## 992	public	Low
## 994	private	Low
## 1001	public	Low
## 1009	public	Low
## 1010	public	Low
## 1014	private	Low
## 1017	private	Low

## 1020	private	Low
## 1024	private	Low
## 1031	private	Low
## 1032	private	Low
## 1033	private	Low
## 1035	private	Low
## 1036	private	Low
## 1037	private	Low
## 1048	private	Low
## 1051	private	Low
## 1052	public	Low
## 1053	private	Low
## 1055	private	Low
## 1059	private	Low
## 1061	public	Low
## 1064	private	Low
## 1075	private	Low
## 1079	private	Low
## 1081	public	Low
## 1084	public	Low
## 1087	private	Low
## 1089	private	Low
## 1090	private	Low
## 1095	private	Low
## 1096	private	Low
## 1098	private	Low
## 1101	private	Low
## 1102	private	Low
## 1105	private	Low
## 1110	private	Low
## 1111	private	Low
## 1121	public	Low
## 1125	private	Low
## 1127	private	Low
## 1131	private	Low
## 1132	public	Low
## 1139	private	Low
## 1143	private	Low
## 1146	private	Low
## 1152	public	Low
## 1154	private	Low
## 1156	private	Low
## 1158	private	Low
## 1166	private	Low
## 1176	public	Low
## 1185	private	Low
## 1188	private	Low
## 1192	private	Low
## 1195	private	Low
## 1198	private	Low
## 1204	private	Low
## 1212	public	Low
## 1218	public	Low
## 1221	private	Low

## 1227	public	Low
## 1231	public	Low
## 1232	public	Low
## 1236	private	Low
## 1238	private	Low
## 1258	private	Low
## 1273	private	Low
## 1274	public	Low
## 1275	public	Low
## 1284	private	Low
## 1285	private	Low
## 1292	private	Low
## 1302	public	Low

Cluster 1 shows equivalent proportion of both Public and Private Universities

```
MyData2[MyData2$k3$cluster=="Medium",]
```


##	College.Name	State	Public..1...Private..2.
## 10	Birmingham-Southern College	AL	private
## 77	Claremont McKenna College	CA	private
## 78	Harvey Mudd College	CA	private
## 79	Pitzer College	CA	private
## 81	Scripps College	CA	private
## 90	Occidental College	CA	private
## 95	Pepperdine University	CA	private
## 97	St. Mary's College of California	CA	private
## 108	University of San Francisco	CA	private
## 112	Westmont College	CA	public
## 120	Loyola Marymount University	CA	private
## 123	Colorado College	CO	private
## 140	Connecticut College	CT	private
## 148	Saint Joseph College	CT	private
## 149	Trinity College	CT	private
## 151	University of Bridgeport	CT	private
## 152	University of Hartford	CT	private
## 153	Wesleyan University	CT	private
## 158	Catholic University of America	DC	private
## 160	George Washington University	DC	private
## 161	Georgetown University	DC	private
## 164	Trinity College	DC	private
## 174	Florida Institute of Technology	FL	private
## 186	Rollins College	FL	private
## 188	Stetson University	FL	private
## 190	University of Miami	FL	private
## 210	Emory University	GA	private
## 228	Spelman College	GA	private
## 247	Cornell College	IA	private
## 252	Grinnell College	IA	private
## 277	Augustana College	IL	private
## 294	Illinois Wesleyan University	IL	private
## 297	Knox College	IL	private
## 298	Lake Forest College	IL	private
## 304	Monmouth College	IL	private
## 321	University of Chicago	IL	private
## 329	Butler University	IN	private
## 330	DePauw University	IN	private
## 331	Earlham College	IN	private
## 360	Saint Mary's College	IN	private
## 362	Taylor University	IN	private
## 365	Valparaiso University	IN	private
## 366	Wabash College	IN	private
## 395	Centre College	KY	private
## 405	Transylvania University	KY	private
## 421	Loyola University	LA	private
## 428	Tulane University	LA	private
## 433	Amherst College	MA	private
## 436	Merrimack College	MA	private
## 438	Bentley College	MA	private
## 443	Brandeis University	MA	private
## 444	Clark University	MA	private

## 452	Harvard University	MA	private
## 456	Massachusetts Institute of Technology	MA	private
## 464	Mount Holyoke College	MA	private
## 479	Wellesley College	MA	private
## 481	Wheaton College	MA	private
## 483	Williams College	MA	private
## 484	Worcester Polytechnic Institute	MA	private
## 494	Goucher College	MD	private
## 513	Bowdoin College	ME	private
## 514	Colby College	ME	private
## 526	Albion College	MI	private
## 527	Alma College	MI	private
## 537	Hillsdale College	MI	private
## 538	Hope College	MI	private
## 563	Carleton College	MN	private
## 567	University of St. Thomas	MN	private
## 571	Hamline University	MN	private
## 572	Macalester College	MN	private
## 608	Saint Louis University	MO	private
## 614	Washington University	MO	private
## 654	Davidson College	NC	private
## 655	Duke University	NC	private
## 677	Salem College	NC	private
## 702	Creighton University	NE	private
## 716	Dartmouth College	NH	private
## 732	Drew University	NJ	private
## 742	Princeton University	NJ	private
## 745	Seton Hall University	NJ	private
## 769	Alfred University	NY	private
## 777	Clarkson University	NY	private
## 778	Colgate University	NY	private
## 782	Columbia University	NY	private
## 783	Barnard College	NY	private
## 792	Hartwick College	NY	private
## 793	Hobart and William Smith Colleges	NY	private
## 794	Hofstra University	NY	private
## 801	Le Moyne College	NY	private
## 803	Manhattanville College	NY	private
## 815	Pace University	NY	private
## 823	Sarah Lawrence College	NY	private
## 824	Skidmore College	NY	private
## 825	Siena College	NY	private
## 831	St. Lawrence University	NY	private
## 846	Syracuse University	NY	private
## 848	Union College	NY	private
## 851	University of Rochester	NY	private
## 874	Case Western Reserve University	OH	private
## 882	Denison University	OH	private
## 889	Kenyon College	OH	private
## 892	Marietta College	OH	private
## 896	Oberlin College	OH	private
## 904	University of Dayton	OH	private
## 911	Wittenberg University	OH	private
## 945	Linfield College	OR	private

## 950	Pacific University	OR	private
## 952	Reed College	OR	private
## 955	University of Portland	OR	private
## 958	Albright College	PA	private
## 959	Allegheny College	PA	private
## 963	Bucknell University	PA	private
## 965	Carnegie Mellon University	PA	private
## 967	Chatham College	PA	private
## 969	College Misericordia	PA	private
## 971	Dickinson College	PA	private
## 975	Elizabethtown College	PA	private
## 979	Gettysburg College	PA	private
## 986	Juniata College	PA	private
## 988	Lafayette College	PA	private
## 989	La Salle University	PA	private
## 991	Lehigh University	PA	private
## 996	Messiah College	PA	private
## 997	Moravian College	PA	private
## 1021	Susquehanna University	PA	private
## 1025	University of Pennsylvania	PA	private
## 1027	University of Scranton	PA	private
## 1029	Villanova University	PA	private
## 1030	Washington and Jefferson College	PA	private
## 1041	Brown University	RI	private
## 1043	Providence College	RI	private
## 1060	Furman University	SC	private
## 1065	Presbyterian College	SC	private
## 1107	Rhodes College	TN	private
## 1117	University of the South	TN	private
## 1118	Vanderbilt University	TN	private
## 1163	Trinity University	TX	private
## 1164	University of Dallas	TX	private
## 1189	College of William and Mary	VA	public
## 1194	Hampden - Sydney College	VA	private
## 1196	Hollins College	VA	private
## 1222	Washington and Lee University	VA	private
## 1237	Saint Michael's College	VT	private
## 1239	University of Vermont	VT	public
## 1246	Seattle University	WA	private
## 1253	Whitman College	WA	private
## 1257	Beloit College	WI	private
## 1262	Lawrence University	WI	private
## 1268	Ripon College	WI	private
## 1269	St. Norbert College	WI	private
##	k3\$cluster		
## 10	Medium		
## 77	Medium		
## 78	Medium		
## 79	Medium		
## 81	Medium		
## 90	Medium		
## 95	Medium		
## 97	Medium		
## 108	Medium		

## 112	Medium
## 120	Medium
## 123	Medium
## 140	Medium
## 148	Medium
## 149	Medium
## 151	Medium
## 152	Medium
## 153	Medium
## 158	Medium
## 160	Medium
## 161	Medium
## 164	Medium
## 174	Medium
## 186	Medium
## 188	Medium
## 190	Medium
## 210	Medium
## 228	Medium
## 247	Medium
## 252	Medium
## 277	Medium
## 294	Medium
## 297	Medium
## 298	Medium
## 304	Medium
## 321	Medium
## 329	Medium
## 330	Medium
## 331	Medium
## 360	Medium
## 362	Medium
## 365	Medium
## 366	Medium
## 395	Medium
## 405	Medium
## 421	Medium
## 428	Medium
## 433	Medium
## 436	Medium
## 438	Medium
## 443	Medium
## 444	Medium
## 452	Medium
## 456	Medium
## 464	Medium
## 479	Medium
## 481	Medium
## 483	Medium
## 484	Medium
## 494	Medium
## 513	Medium
## 514	Medium
## 526	Medium

## 527	Medium
## 537	Medium
## 538	Medium
## 563	Medium
## 567	Medium
## 571	Medium
## 572	Medium
## 608	Medium
## 614	Medium
## 654	Medium
## 655	Medium
## 677	Medium
## 702	Medium
## 716	Medium
## 732	Medium
## 742	Medium
## 745	Medium
## 769	Medium
## 777	Medium
## 778	Medium
## 782	Medium
## 783	Medium
## 792	Medium
## 793	Medium
## 794	Medium
## 801	Medium
## 803	Medium
## 815	Medium
## 823	Medium
## 824	Medium
## 825	Medium
## 831	Medium
## 846	Medium
## 848	Medium
## 851	Medium
## 874	Medium
## 882	Medium
## 889	Medium
## 892	Medium
## 896	Medium
## 904	Medium
## 911	Medium
## 945	Medium
## 950	Medium
## 952	Medium
## 955	Medium
## 958	Medium
## 959	Medium
## 963	Medium
## 965	Medium
## 967	Medium
## 969	Medium
## 971	Medium
## 975	Medium

```
## 979      Medium
## 986      Medium
## 988      Medium
## 989      Medium
## 991      Medium
## 996      Medium
## 997      Medium
## 1021     Medium
## 1025     Medium
## 1027     Medium
## 1029     Medium
## 1030     Medium
## 1041     Medium
## 1043     Medium
## 1060     Medium
## 1065     Medium
## 1107     Medium
## 1117     Medium
## 1118     Medium
## 1163     Medium
## 1164     Medium
## 1189     Medium
## 1194     Medium
## 1196     Medium
## 1222     Medium
## 1237     Medium
## 1239     Medium
## 1246     Medium
## 1253     Medium
## 1257     Medium
## 1262     Medium
## 1268     Medium
## 1269     Medium
```

```
# Cluster 2 shows more Public Universities when compared with private
```

```
MyData2[MyData2$k3$cluster=="High",]
```

```

##          College.Name State
## 49      Northern Arizona University AZ
## 50          University of Arizona  AZ
## 63      California Polytechnic-San Luis CA
## 110     University of Southern California CA
## 156     University of Connecticut at Storrs CT
## 168          University of Delaware DE
## 194          University of Central Florida FL
## 230          University of Georgia GA
## 322     University of Illinois - Urbana IL
## 323     University of Illinois at Chicago IL
## 418     Louisiana State University at Baton Rouge LA
## 441          Boston University MA
## 467     Northeastern University MA
## 477     University of Massachusetts at Amherst MA
## 505     University of Maryland at College Park MD
## 543     Michigan State University MI
## 556     Western Michigan University MI
## 584     University of Minnesota Twin Cities MN
## 607     Southwest Missouri State University MO
## 646     Appalachian State University NC
## 656     East Carolina University NC
## 681     North Carolina State University at Raleigh NC
## 682     University of North Carolina at Chapel Hill NC
## 712     University of Nebraska at Lincoln NE
## 723     University of New Hampshire NH
## 751     Rutgers at New Brunswick NJ
## 833     SUNY at Albany NY
## 834     SUNY at Binghamton NY
## 872     Bowling Green State University OH
## 898     Ohio University OH
## 903     University of Cincinnati OH
## 916     Miami University at Oxford OH
## 932     Oklahoma State University OK
## 1023     Temple University PA
## 1026     University of Pittsburgh-Main Campus PA
## 1039     Pennsylvania State Univ. Main Campus PA
## 1047     University of Rhode Island RI
## 1115     University of Tennessee at Knoxville TN
## 1138     University of North Texas TX
## 1168     University of Texas at Austin TX
## 1172     Baylor University TX
## 1177     Texas A&M Univ. at College Station TX
## 1181     University of Utah UT
## 1206     Virginia Commonwealth University VA
## 1214     George Mason University VA
## 1223     University of Virginia VA
##      Public..1...Private..2. k3$cluster
## 49      public      High
## 50      public      High
## 63      public      High
## 110     private     High
## 156     public      High

```

## 168	private	High
## 194	public	High
## 230	public	High
## 322	public	High
## 323	public	High
## 418	public	High
## 441	private	High
## 467	private	High
## 477	public	High
## 505	public	High
## 543	public	High
## 556	public	High
## 584	public	High
## 607	public	High
## 646	public	High
## 656	public	High
## 681	public	High
## 682	public	High
## 712	public	High
## 723	public	High
## 751	public	High
## 833	public	High
## 834	public	High
## 872	public	High
## 898	public	High
## 903	public	High
## 916	public	High
## 932	public	High
## 1023	public	High
## 1026	public	High
## 1039	public	High
## 1047	public	High
## 1115	public	High
## 1138	public	High
## 1168	public	High
## 1172	private	High
## 1177	public	High
## 1181	public	High
## 1206	public	High
## 1214	public	High
## 1223	public	High

Cluster 3 shows more Private Universities when compared with public

k3\$withinss

[1] 2562.342 1424.892 1044.680

k3\$iter


```
## [1] 3
```

f. 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.

```
MyData3 <- read.csv("Universities.csv")
# Considering Tufts University
MyData4 <- MyData3[ MyData3$College.Name== "Tufts University",]
View(MyData4)
# Compute Euclidian Distance
dist(rbind(MyData4[, -c(1, 2, 3,10)], k3$centers[1,]))
```

```
##          476
## 2 29819.09
```

```
dist(rbind(MyData4[, -c(1, 2, 3,10)], k3$centers[2,]))
```

```
##          476
## 2 29817.8
```

```
dist(rbind(MyData4[, -c(1, 2, 3,10)], k3$centers[3,]))
```

```
##          476
## 2 29816.76
```

Based on the above distances it is closest to Cluster 3

Impute the missing values for Tufts by taking the average of the cluster on those measurements

```
MyData5 <- read.csv("Universities.csv")
MyData5 <- na.omit(MyData5)
MyData5 <- cbind(MyData5,k3$cluster)
View(MyData5)
#Filter the values which falls under cluster3
MyData6 <- filter(MyData5, MyData5`k3$cluster` == 3)
MyData7 <- mean(MyData6[,c(10)])
MyData7
```

```
## [1] 3500.565
```

```
MyData4$X..PT.undergrad <- MyData7
# After filling missing value with mean
MyData4
```

```
##      College.Name State Public..1...Private..2. X..appli..rec.d
## 476 Tufts University    MA                2                7614
##      X..appl..accepted X..new.stud..enrolled X..new.stud..from.top.10.
## 476                3605                1205                60
##      X..new.stud..from.top.25. X..FT.undergrad X..PT.undergrad
## 476                90                4598                3500.565
##      in.state.tuition out.of.state.tuition room board add..fees
## 476                19701                19701 3038 2930                503
##      estim..book.costs estim..personal.. X..fac..w.PHD stud..fac..ratio
## 476                600                928                99                10.3
##      Graduation.rate
## 476                92
```