Machine Learning Assignment 4

Tom Scarberry

11/1/2020

Load and do an initial check of the data top and bottom six rows and summary information for the data file

```
All.rankings<-read.csv("Universities.csv")
head(All.rankings)
```

```
##                         College.Name State Public..1...Private..2.
## 1         Alaska Pacific University    AK                       2
## 2 University of Alaska at Fairbanks    AK                       1
## 3    University of Alaska Southeast    AK                       1
## 4 University of Alaska at Anchorage    AK                       1
## 5       Alabama Agri. & Mech. Univ.    AL                       1
## 6               Faulkner University    AL                       2
##   X..appli..rec.d X..appl..accepted X..new.stud..enrolled
## 1             193               146                    55
## 2            1852              1427                   928
## 3             146               117                    89
## 4            2065              1598                  1162
## 5            2817              1920                   984
## 6             345               320                   179
##   X..new.stud..from.top.10. X..new.stud..from.top.25. X..FT.undergrad
## 1                        16                        44             249
## 2                        NA                        NA            3885
## 3                         4                        24             492
## 4                        NA                        NA            6209
## 5                        NA                        NA            3958
## 6                        NA                        27            1367
##   X..PT.undergrad in.state.tuition out.of.state.tuition room board add..fees
## 1             869            7560                 7560 1620  2500       130
## 2            4519            1742                 5226 1800  1790       155
## 3            1849            1742                 5226 2514  2250        34
## 4           10537            1742                 5226 2600  2520       114
## 5             305            1700                 3400 1108  1442       155
## 6             578            5600                 5600 1550  1700       300
##   estim..book.costs estim..personal.. X..fac..w.PHD stud..fac..ratio
## 1               800              1500            76             11.9
## 2               650              2304            67             10.0
## 3               500              1162            39              9.5
## 4               580              1260            48             13.7
## 5               500               850            53             14.3
```

```
## 6               350                NA            52             32.8
##   Graduation.rate
## 1              15
## 2              NA
## 3              39
## 4              NA
## 5              40
## 6              55
```

```
tail(All.rankings)
```

```
##                               College.Name State Public..1...Private..2.
## 1297 West Virginia Institute of Technology    WV                       1
## 1298           West Virginia State College    WV                       1
## 1299              West Virginia University    WV                       1
## 1300        West Virginia Wesleyan College    WV                       2
## 1301               Wheeling Jesuit College    WV                       2
## 1302                 University of Wyoming    WY                       1
##      X..appli..rec.d X..appl..accepted X..new.stud..enrolled
## 1297            1594              1572                   675
## 1298            1869                NA                   957
## 1299            9630              7801                  2881
## 1300            1566              1400                   483
## 1301             903               755                   213
## 1302            2029              1516                  1073
##      X..new.stud..from.top.10. X..new.stud..from.top.25. X..FT.undergrad
## 1297                        NA                        NA            2432
## 1298                        NA                        NA            2817
## 1299                        23                        49           14524
## 1300                        28                        55            1509
## 1301                        15                        49             971
## 1302                        23                        46            7535
##      X..PT.undergrad in.state.tuition out.of.state.tuition room board add..fees
## 1297             616             3954                   NA 1800  1952        NA
## 1298            1939             1988                 4616 1500  1700        50
## 1299            1053             2128                 6370 2284  2026        NA
## 1300             170            14200                14200 1750  2025        NA
## 1301             305            10500                10500 2100  2445        NA
## 1302            1488             1908                 5988 1462  1960       300
##      estim..book.costs estim..personal.. X..fac..w.PHD stud..fac..ratio
## 1297               500                NA            NA             15.3
## 1298               750               750            38             19.2
## 1299                NA                NA            83             13.4
## 1300               450              1100            58             16.4
## 1301               600               600            66             14.1
## 1302               600              1500            91             15.1
##      Graduation.rate
## 1297              56
## 1298              NA
## 1299              57
## 1300              67
## 1301              72
## 1302              45
```

```
summary(All.rankings)
```

```
##  College.Name          State           Public..1...Private..2.
##  Length:1302        Length:1302        Min.   :1.000
##  Class :character   Class :character   1st Qu.:1.000
##  Mode  :character   Mode  :character   Median :2.000
##                                        Mean   :1.639
##                                        3rd Qu.:2.000
##                                        Max.   :2.000
##
##  X..appli..rec.d   X..appl..accepted X..new.stud..enrolled
##  Min.   :   35.0   Min.   :   35.0   Min.   :  18.0
##  1st Qu.:  695.8   1st Qu.:  554.5   1st Qu.: 236.0
##  Median : 1470.0   Median : 1095.0   Median : 447.0
##  Mean   : 2752.1   Mean   : 1870.7   Mean   : 778.9
##  3rd Qu.: 3314.2   3rd Qu.: 2303.0   3rd Qu.: 984.0
##  Max.   :48094.0   Max.   :26330.0   Max.   :7425.0
##  NA's   :10        NA's   :11        NA's   :5
##  X..new.stud..from.top.10. X..new.stud..from.top.25. X..FT.undergrad
##  Min.   : 1.00             Min.   :  6.00            Min.   :   59
##  1st Qu.:13.00             1st Qu.: 36.75            1st Qu.:  966
##  Median :21.00             Median : 50.00            Median : 1812
##  Mean   :25.67             Mean   : 52.35            Mean   : 3693
##  3rd Qu.:32.00             3rd Qu.: 66.00            3rd Qu.: 4540
##  Max.   :98.00             Max.   :100.00            Max.   :31643
##  NA's   :235               NA's   :202               NA's   :3
##  X..PT.undergrad   in.state.tuition out.of.state.tuition      room
##  Min.   :    1.0   Min.   :  480    Min.   : 1044        Min.   : 500
##  1st Qu.:  131.2   1st Qu.: 2580    1st Qu.: 6111        1st Qu.:1710
##  Median :  472.0   Median : 8050    Median : 8670        Median :2200
##  Mean   : 1081.5   Mean   : 7897    Mean   : 9277        Mean   :2515
##  3rd Qu.: 1313.0   3rd Qu.:11600    3rd Qu.:11659        3rd Qu.:3040
##  Max.   :21836.0   Max.   :25750    Max.   :25750        Max.   :7400
##  NA's   :32        NA's   :30       NA's   :20           NA's   :321
##      board        add..fees      estim..book.costs estim..personal..
##  Min.   : 531   Min.   :   9.0   Min.   :  90      Min.   :  75
##  1st Qu.:1619   1st Qu.: 130.0   1st Qu.: 480      1st Qu.: 900
##  Median :1980   Median : 264.5   Median : 502      Median :1250
##  Mean   :2061   Mean   : 392.0   Mean   : 550      Mean   :1389
##  3rd Qu.:2402   3rd Qu.: 480.0   3rd Qu.: 600      3rd Qu.:1794
##  Max.   :6250   Max.   :4374.0   Max.   :2340      Max.   :6900
##  NA's   :498    NA's   :274      NA's   :48        NA's   :181
##  X..fac..w.PHD    stud..fac..ratio Graduation.rate
##  Min.   :  8.00   Min.   : 2.30    Min.   :  8.00
##  1st Qu.: 57.00   1st Qu.:11.80    1st Qu.: 47.00
##  Median : 71.00   Median :14.30    Median : 60.00
##  Mean   : 68.65   Mean   :14.86    Mean   : 60.41
##  3rd Qu.: 82.00   3rd Qu.:17.60    3rd Qu.: 74.00
##  Max.   :105.00   Max.   :91.80    Max.   :118.00
##  NA's   :32       NA's   :2        NA's   :98
```

Change Public and Private variable from integer to a factor with two selection options 1 or 2. Normalize continuous variables and check the variable type loaded from the csv file.

```
str(All.rankings)
```

```
## 'data.frame':    1302 obs. of  20 variables:
##  $ College.Name             : chr  "Alaska Pacific University" "University of Alaska at Fairbanks" "U
##  $ State                    : chr  "AK" "AK" "AK" "AK" ...
##  $ 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 ...
```

```
All.rankings[,3]<-as.factor(All.rankings[,3])
str(All.rankings)
```

```
## 'data.frame':    1302 obs. of  20 variables:
##  $ College.Name             : chr  "Alaska Pacific University" "University of Alaska at Fairbanks" "U
##  $ State                    : chr  "AK" "AK" "AK" "AK" ...
##  $ Public..1...Private..2.  : Factor w/ 2 levels "1","2": 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 ...
```

```r
library(caret)
```

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

```
## Loading required package: ggplot2
```

```r
z.norm<-preProcess(All.rankings[ ,4:20], method= c("center","scale"))
rankings.norm<-All.rankings
rankings.norm[,4:20]<-predict(z.norm,All.rankings[,4:20])
summary(rankings.norm)
```

```
##  College.Name          State           Public..1...Private..2.
##  Length:1302        Length:1302        1:470
##  Class :character   Class :character   2:832
##  Mode  :character   Mode  :character
##
##
##
##
##  X..appli..rec.d   X..appl..accepted X..new.stud..enrolled
##  Min.   :-0.7671   Min.   :-0.8155   Min.   :-0.8602
##  1st Qu.:-0.5806   1st Qu.:-0.5847   1st Qu.:-0.6137
##  Median :-0.3620   Median :-0.3446   Median :-0.3752
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.1587   3rd Qu.: 0.1921   3rd Qu.: 0.2319
##  Max.   :12.8013   Max.   :10.8666   Max.   : 7.5133
##  NA's   :10        NA's   :11        NA's   :5
##  X..new.stud..from.top.10. X..new.stud..from.top.25. X..FT.undergrad
##  Min.   :-1.3473           Min.   :-2.2197           Min.   :-0.7995
##  1st Qu.:-0.6920           1st Qu.:-0.7471           1st Qu.:-0.5999
##  Median :-0.2551           Median :-0.1125           Median :-0.4138
##  Mean   : 0.0000           Mean   : 0.0000           Mean   : 0.0000
##  3rd Qu.: 0.3456           3rd Qu.: 0.6537           3rd Qu.: 0.1863
##  Max.   : 3.9496           Max.   : 2.2819           Max.   : 6.1499
##  NA's   :235               NA's   :202               NA's   :3
##  X..PT.undergrad   in.state.tuition   out.of.state.tuition      room
##  Min.   :-0.6462   Min.   :-1.38688   Min.   :-1.9740      Min.   :-1.7506
##  1st Qu.:-0.5683   1st Qu.:-0.99423   1st Qu.:-0.7591      1st Qu.:-0.6992
##  Median :-0.3645   Median : 0.02856   Median :-0.1455      Median :-0.2734
##  Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.0000      Mean   : 0.0000
##  3rd Qu.: 0.1384   3rd Qu.: 0.69234   3rd Qu.: 0.5711      3rd Qu.: 0.4565
##  Max.   :12.4115   Max.   : 3.33810   Max.   : 3.9497      Max.   : 4.2450
##  NA's   :32        NA's   :30         NA's   :20           NA's   :321
##      board           add..fees       estim..book.costs estim..personal..
##  Min.   :-2.3121   Min.   :-0.8160   Min.   :-2.7485   Min.   :-1.8401
##  1st Qu.:-0.6675   1st Qu.:-0.5582   1st Qu.:-0.4181   1st Qu.:-0.6850
##  Median :-0.1224   Median :-0.2717   Median :-0.2867   Median :-0.1950
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.5146   3rd Qu.: 0.1875   3rd Qu.: 0.2989   3rd Qu.: 0.5666
##  Max.   : 6.3303   Max.   : 8.4835   Max.   :10.6960   Max.   : 7.7154
##  NA's   :498       NA's   :274       NA's   :48        NA's   :181
##  X..fac..w.PHD     stud..fac..ratio  Graduation.rate
```

```
##  Min.   :-3.4022   Min.   :-2.4215   Min.   :-2.77437
##  1st Qu.:-0.6533   1st Qu.:-0.5898   1st Qu.:-0.70969
##  Median : 0.1321   Median :-0.1077   Median :-0.02146
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.00000
##  3rd Qu.: 0.7492   3rd Qu.: 0.5285   3rd Qu.: 0.71971
##  Max.   : 2.0394   Max.   :14.8352   Max.   : 3.04910
##  NA's   :32        NA's   :2         NA's   :98
```

Remove missing records from the both original and normalized data frames.

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------------------------------------------------ tidy
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x purrr::lift()   masks caret::lift()
```

```
rankings<-rankings.norm%>%drop_na()
rankings.values<-All.rankings%>%drop_na()
```

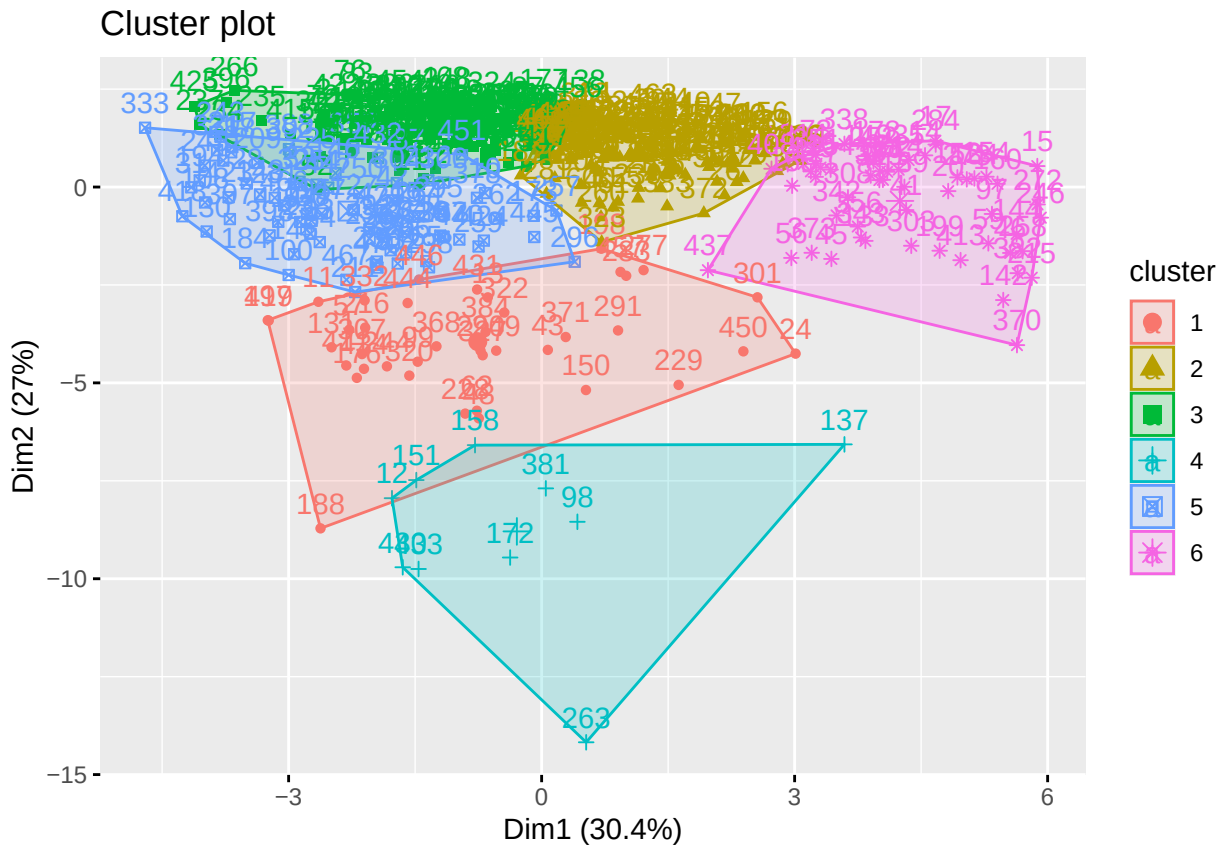
Initial k means cluster of 6, then use elbow method to find the turning point using total within-cluster sum of square (WSS) value. The best choice is where the elbow turns to the right for the optimal number of clusters - 3 clusters.

```
library(factoextra)
```

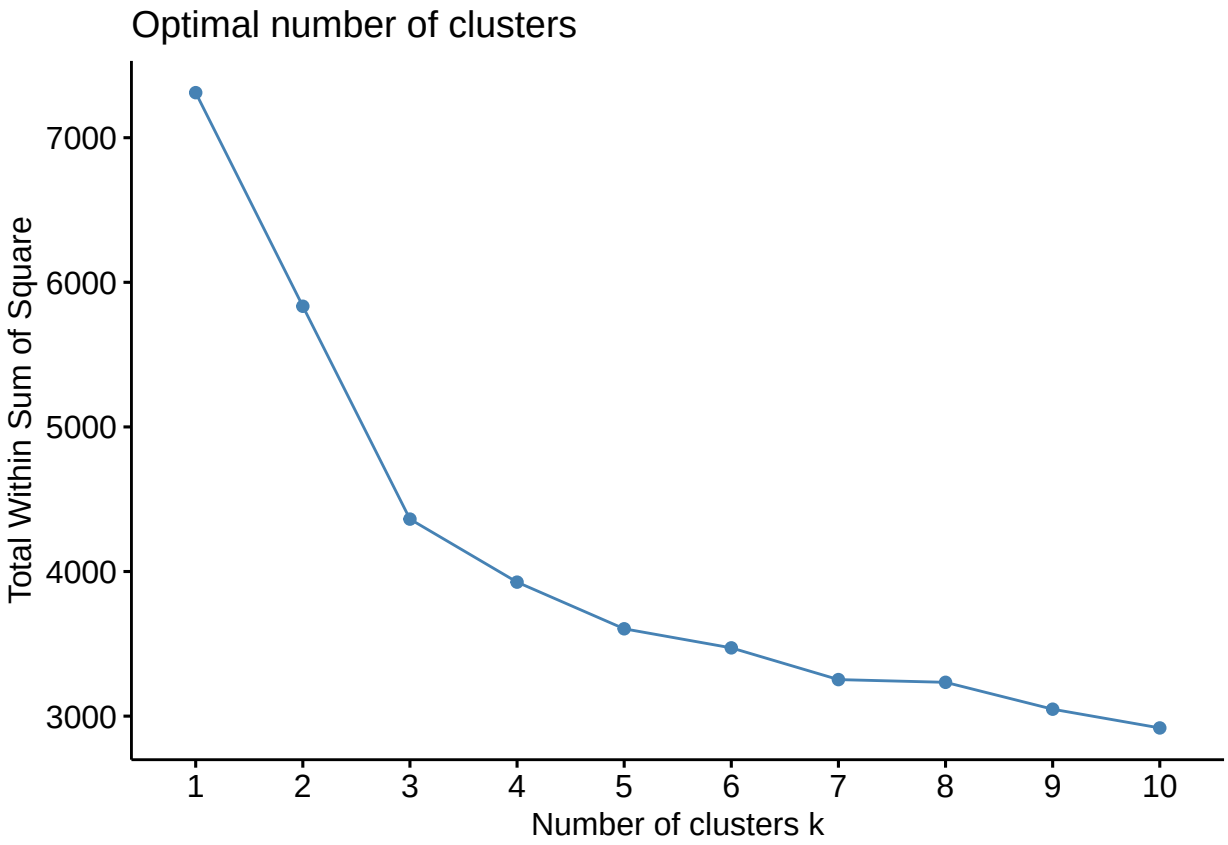
```
## Warning: package 'factoextra' was built under R version 4.0.3
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
set.seed(20)
k.rankings<-kmeans(rankings[4:20],centers=6,nstart = 25)
fviz_cluster(k.rankings, data = rankings[4:20])
```

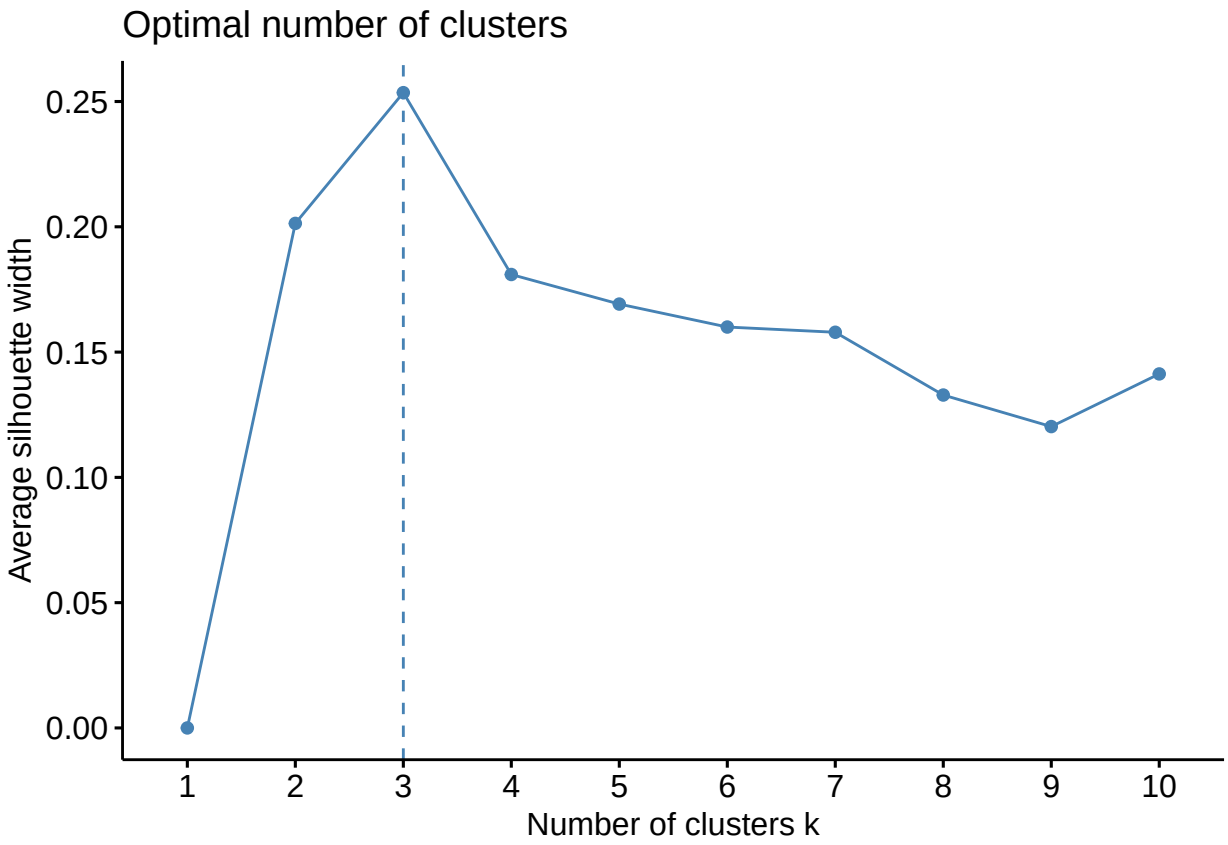



```
fviz_nbclust(rankings[4:20],kmeans, method = "wss")
```



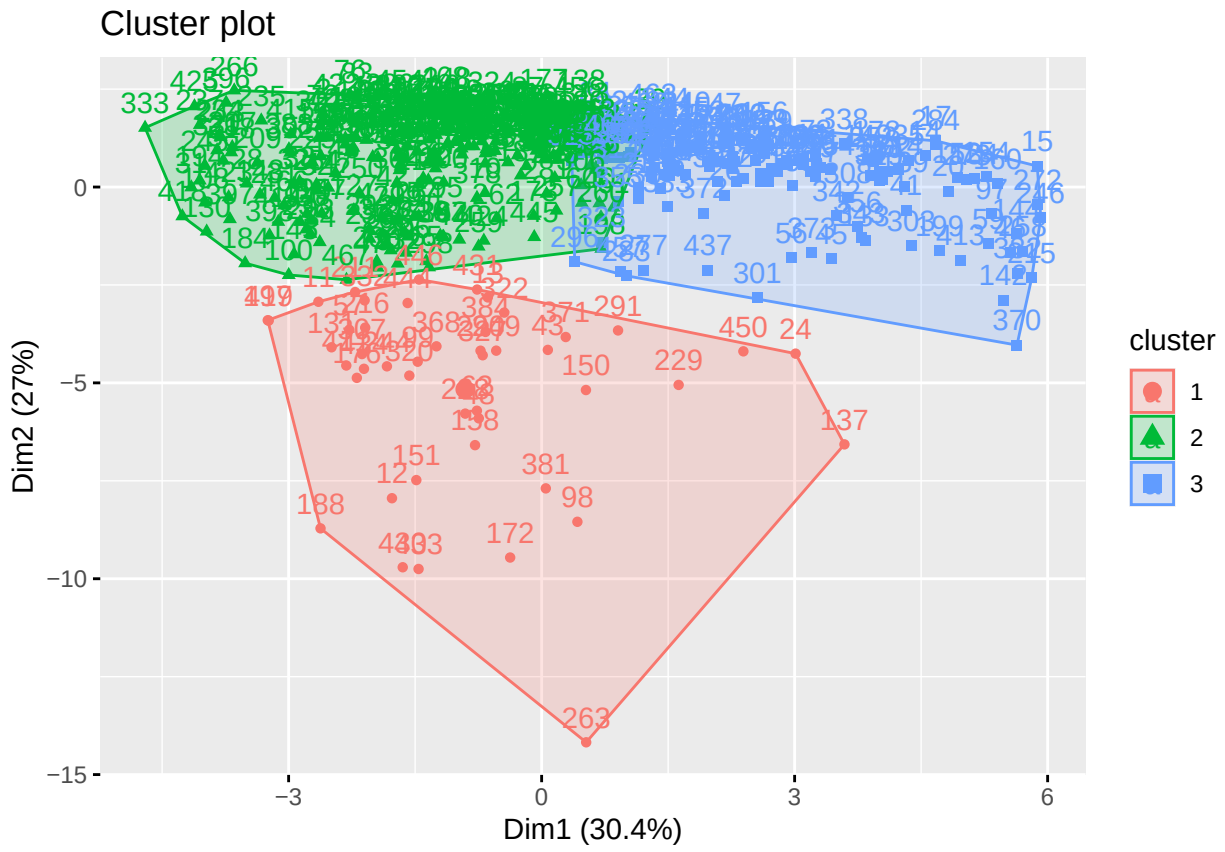
The Silhouette method confirms the optimal # of clusters should be 3

```
fviz_nbclust(rankings[4:20],kmeans, method = "silhouette")
```



The Elbow and Silhouette method both confirm the optimal number of clusters is 3, below is the k-means cluster with 3 clusters. Additionally the center values and size of the clusters are displayed.

```
k.rankings.3<-kmeans(rankings[4:20],centers=3,nstart = 25)
fviz_cluster(k.rankings.3, data = rankings[4:20])
```




```
k.rankings.3$centers
```

```
##   X..appli..rec.d X..appl..accepted X..new.stud..enrolled
## 1       2.3905697        2.56587741             2.5326119
## 2      -0.2974318       -0.29626363            -0.3215895
## 3       0.1656253        0.02665701            -0.1796647
##   X..new.stud..from.top.10. X..new.stud..from.top.25. X..FT.undergrad
## 1                 0.2624575                0.4059173       2.5633440
## 2                -0.3910478               -0.3573682      -0.3278163
## 3                 1.0473856                1.0364907      -0.2743664
##   X..PT.undergrad in.state.tuition out.of.state.tuition       room      board
## 1       1.4466178       -0.8009155           -0.1971571 -0.2791611 -0.0574326
## 2      -0.2732909       -0.1293629           -0.2311175 -0.4534200 -0.2094851
## 3      -0.4773773        1.3790073            1.4729401  0.1196992  0.6969544
##     add..fees estim..book.costs estim..personal.. X..fac..w.PHD
## 1  0.34784232        0.15240247       0.787711624     0.8955099
## 2 -0.06683581       -0.06037046      -0.006399304    -0.2421210
## 3 -0.07107609        0.04230999      -0.573668320     0.9810908
##   stud..fac..ratio Graduation.rate
## 1       0.28878469      0.02918124
## 2       0.01738081     -0.13155901
## 3      -0.66753932      1.09775850
```

```
k.rankings.3$size
```

```
## [1]  46 276 149
```

Cluster 2 is the largest with 276 Universities, followed by Cluster 3 with 149 and Cluster 1 with 46.

Pull cluster results into original data values in order to compare summary statistics for each cluster and convert the variable to a factor.

I will use mean value of several of the data variables to describe each cluster of universities.

```
rankings.clusters<-cbind(rankings.values,k.rankings.3$cluster)
rankings.clusters$`k.rankings.3$cluster`<-as.factor(rankings.clusters$`k.rankings.3$cluster`)

Cluster.means<-rankings.clusters%>%
  group_by(`k.rankings.3$cluster`)%>%
  summarise(Mean_Out_Tuition=mean(out.of.state.tuition),
            Mean_In_Tuition=mean(in.state.tuition),
            Mean_Applications=mean(X..appli..rec.d),
            Mean_Accepted=mean(X..appl..accepted),
            Mean_Enrolled=mean(X..new.stud..enrolled),
            Mean_Top_10=mean(X..new.stud..from.top.10.),
            Mean_Top_25=mean(X..new.stud..from.top.25.),
            Mean_FT_Undergrad=mean(X..FT.undergrad),
            Mean_PT_Undergrad=mean(X..PT.undergrad),
            Mean_Room=mean(room),
            Mean_Board=mean(board),
            Mean_Additional_Fee=mean(add..fees),
            Mean_Book_Cost=mean(estim..book.costs),
            Mean_Personal_Cost=mean(estim..personal..),
            Mean_Faculty_Phd=mean(X..fac..w.PHD),
            Mean_Stud_Fac_Ratio=mean(stud..fac..ratio),
            Mean_Grad_Rate=mean(Graduation.rate))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
Cluster.means
```

```
## # A tibble: 3 x 18
##   `k.rankings.3$c~ Mean_Out_Tuition Mean_In_Tuition Mean_Applicatio~
##   <fct>                       <dbl>           <dbl>            <dbl>
## 1 1                           8455.           3614.           11219.
## 2 2                           8313.           7205.            1699.
## 3 3                          15420.          15272.            3339.
## # ... with 14 more variables: Mean_Accepted <dbl>, Mean_Enrolled <dbl>,
## #   Mean_Top_10 <dbl>, Mean_Top_25 <dbl>, Mean_FT_Undergrad <dbl>,
## #   Mean_PT_Undergrad <dbl>, Mean_Room <dbl>, Mean_Board <dbl>,
## #   Mean_Additional_Fee <dbl>, Mean_Book_Cost <dbl>, Mean_Personal_Cost <dbl>,
## #   Mean_Faculty_Phd <dbl>, Mean_Stud_Fac_Ratio <dbl>, Mean_Grad_Rate <dbl>
```

Cluster 1 has the lowest in state tuition but average out of state tuition, an average acceptance rate, the largest number of students enrolling every year, and an average number of top 10 & 25 students (lowest from a % of students enrolled perspective).

Cluster 2 has mid priced out and in state tuition, a better than average acceptance rate, the smallest students enrolling every year, and the smallest number of top 10 & 25 students (average from a % of students enrolled).

Cluster 3 has the highest tuition for both in state and out of state students, the lowest acceptance rate, an average number of students enrolling every year, and the largest number of top 10 & 25 students (highest % of students enrolled).

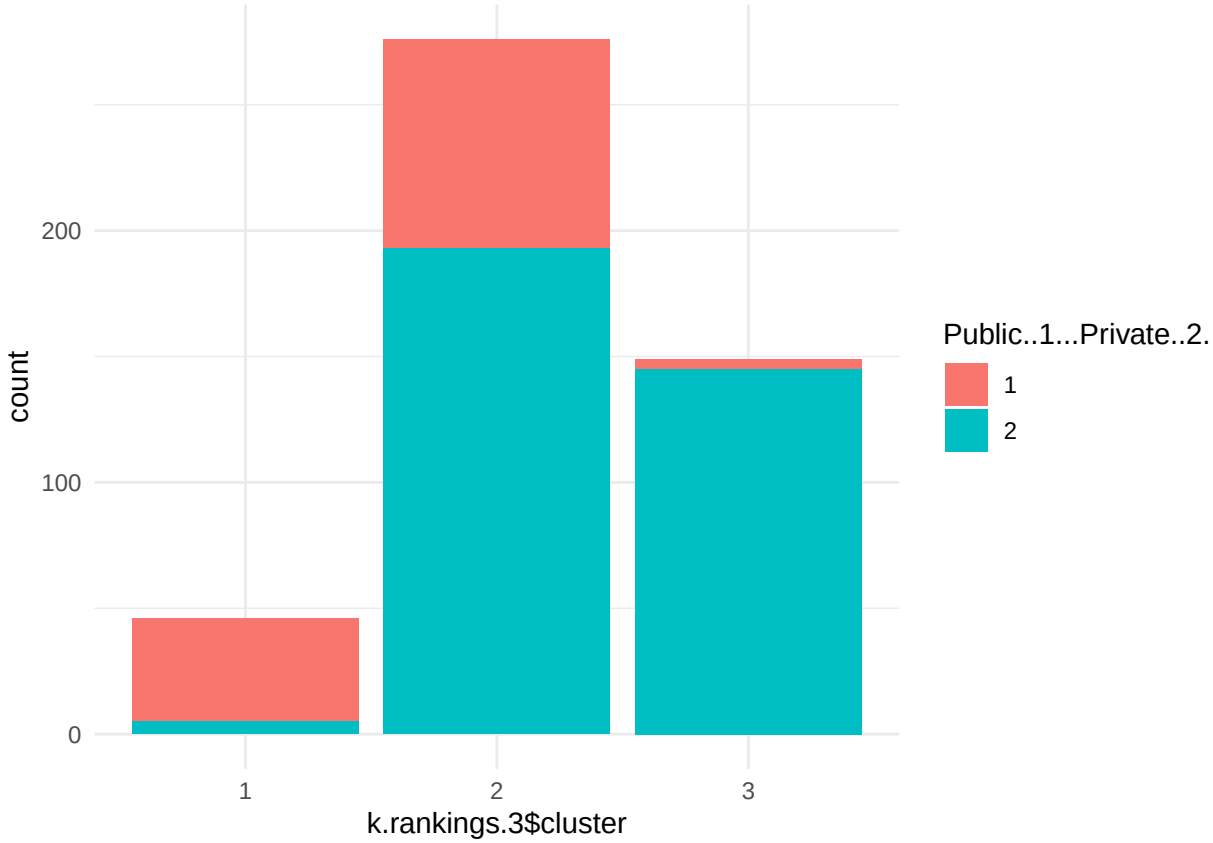
Create charts to assess the categorical measurements and relationships to the three clusters

```
library(esquisse)
```

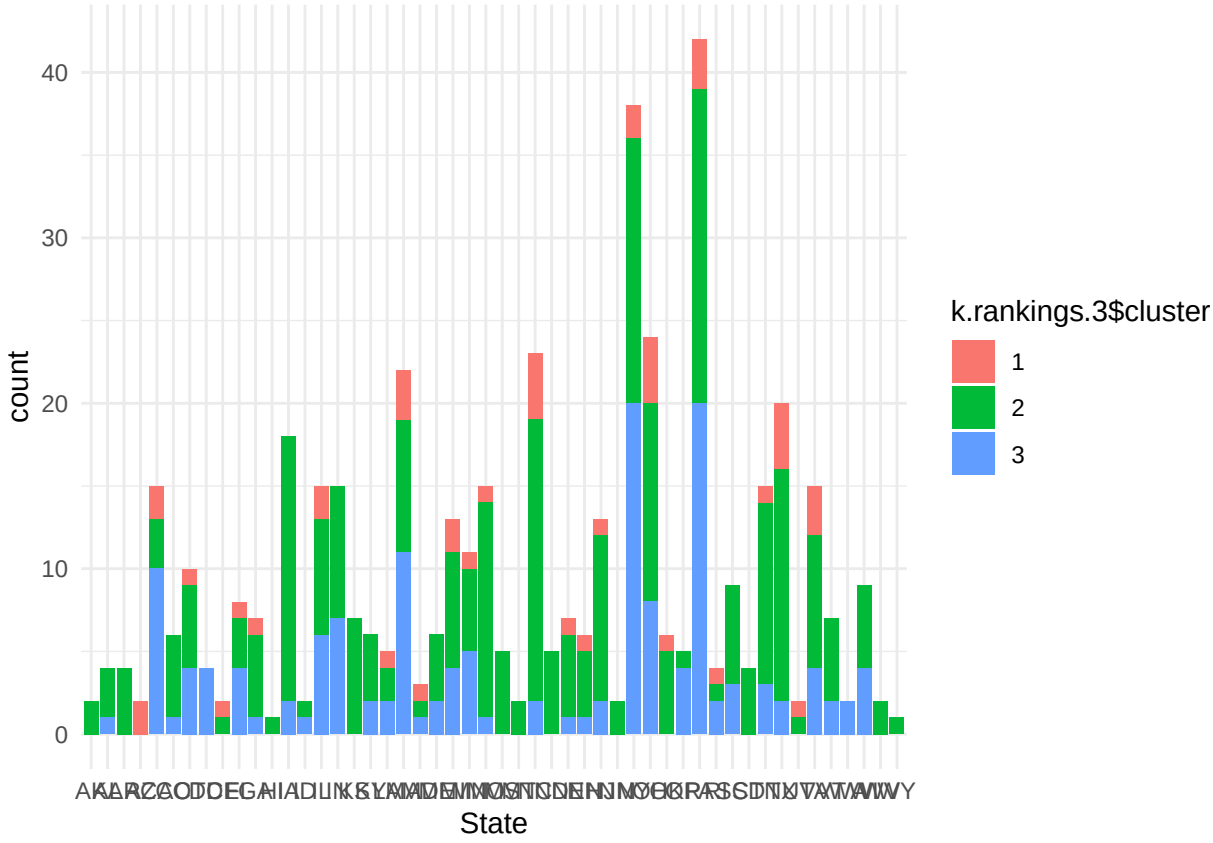
```
## Warning: package 'esquisse' was built under R version 4.0.3
```

```
library(ggplot2)

ggplot(rankings.clusters) +
 aes(x = `k.rankings.3$cluster`, fill = Public..1...Private..2.) +
 geom_bar() +
 scale_fill_hue() +
 theme_minimal()
```




```
ggplot(rankings.clusters) +
 aes(x = State, fill = `k.rankings.3$cluster`) +
 geom_bar() +
 scale_fill_hue() +
 theme_minimal()
```


40
30
20
10
0
count
State
k.rankings.3$cluster
1
2
3

Most States have a high % of cluster 2 schools, which is a blend of private and public universities with mid priced tuition and average acceptance and top student %.

The second most prevelant type of school is cluster 3, which is predominately private schools with high tuition and lower acceptance rates.

The least represented across all states is cluster 1, which is predominately public schools with low in state tuition.

While the cluster of shool types are represented fairly proportionally across all the states, it is clear that cluster 1 is generally a public school and cluster 3 is generally a private school. Cluster 2 has a blend of public and private schools included.

There are several other student factors that could be considered when clustering universities that may help improve the results. Those factors include: ACT scores, SAT scores, ethnicity, high school GPA, and high school type (e.g. public). Other factors that could be included on the universities include: % employment upon graduation, % employeed 3 months post graduation, avg. starting salary for employment, avg. class size, and avg. scholarship value.

Evaluate Tufts vs the center values for three clusters to determine closest cluster

```
Tufts<-rankings.norm[476,]
Cluster1.mean<-k.rankings$centers[1,]
Cluster2.mean<-k.rankings$centers[2,]
Cluster3.mean<-k.rankings$centers[3,]
Tufts.1<-rbind(Tufts[,4:20],Cluster1.mean)
Tufts.2<-rbind(Tufts[,4:20],Cluster2.mean)
Tufts.3<-rbind(Tufts[,4:20],Cluster3.mean)
dist(Tufts.1, method = "euclidean")
```

```
##        476
## 2 6.089343
```

```
dist(Tufts.2, method = "euclidean")
```

```
##        476
## 2 4.347511
```

```
dist(Tufts.3, method = "euclidean")
```

```
##        476
## 2 6.760563
```

Tufts is closest to cluster 2 with the lowest distance measurement vs. the cluster center values for each numerical variable

Replace missing value in Tufts data with cluster 2 mean value

```
Tufts.new<-Tufts
is.na(Tufts)
```

```
##     College.Name State Public..1...Private..2. X..appli..rec.d
## 476        FALSE FALSE                   FALSE           FALSE
##     X..appl..accepted X..new.stud..enrolled X..new.stud..from.top.10.
## 476             FALSE                 FALSE                     FALSE
##     X..new.stud..from.top.25. X..FT.undergrad X..PT.undergrad in.state.tuition
## 476                     FALSE           FALSE            TRUE            FALSE
##     out.of.state.tuition  room board add..fees estim..book.costs
## 476                FALSE FALSE FALSE     FALSE             FALSE
##     estim..personal.. X..fac..w.PHD stud..fac..ratio Graduation.rate
## 476             FALSE         FALSE            FALSE           FALSE
```

```
Tufts.new[,10]<-k.rankings$center[2,7]
Tufts.new
```

```
##         College.Name State Public..1...Private..2. X..appli..rec.d
## 476 Tufts University    MA                        2        1.372653
##     X..appl..accepted X..new.stud..enrolled X..new.stud..from.top.10.
## 476         0.7705108             0.4817205                  1.874556
##     X..new.stud..from.top.25. X..FT.undergrad X..PT.undergrad in.state.tuition
## 476                  1.803047       0.1992003       -0.460936         2.207062
##     out.of.state.tuition      room    board add..fees estim..book.costs
## 476             2.499321 0.4547283 1.313225 0.2364556         0.2989274
##     estim..personal.. X..fac..w.PHD stud..fac..ratio Graduation.rate
## 476        -0.6458426      1.702848       -0.8789855        1.672645
```

After identifying the missing value as PT undergrad, identified column and replaced with the Cluster 2 center value for the cluster