DSCI 300 Mini Project 4 - Cluster Analysis

Nisi Mohan Kuniyil 300321388

2020-11-02

```
set.seed(42)
```

Tidying Data

Read only first 5 rows from BigTenexpand.csv, since there are two sets of data in single csv.

```
University_49 <- read.csv("BigTenexpand.csv", nrows = 5 )
```

```
head(University_49)
```

```
##   ï..Candidates.          School       Duke   Maryland North.Carolina.State
## 1             NA  StadiumCapacity -0.7451514  0.1113582            0.2643505
## 2             NA         Latitude -1.3000459 -0.1389516           -1.3610659
## 3             NA        Longitude  0.9159120  1.2319867            0.9568089
## 4             NA Endowment ($000)  2.5449230 -0.2651109           -0.3638772
## 5             NA       Enrollment -1.0502828  0.9118229            0.6626736
##   North.Carolina   Virginia Virginia.Tech Wake.Forest  Iowa.State      Kansas
## 1      0.4947998  0.4316045     0.6337013  -0.8493809   0.1540577 -0.05640823
## 2     -1.3203829 -0.5080417    -0.8178726  -1.2532009   1.0266317 -0.15173905
## 3      0.8958546  0.9844242     0.6743377   0.7007552  -1.4321853 -1.69232936
## 4      0.5679706  1.9853271    -0.3735079  -0.1140267  -0.3669104 -0.00504443
## 5      0.1640208 -0.2088616     0.3233594  -1.7655794   0.2060033  0.05791320
##   Kansas.State    Oklahoma Oklahoma.State West.Virginia  Cincinnati Louisville
## 1  -0.05943990  1.31172693      0.3768637     0.3906130 -0.69993262  0.1967572
## 2  -0.06424199 -1.49771144     -1.4977114     0.1073058 -0.08360839 -0.4359264
## 3  -1.90535282 -2.05276241     -2.0527624     0.7507380  0.02348501 -0.1736760
## 4  -0.52274758 -0.02683033     -0.3275228    -0.4918203 -0.14458022 -0.2762545
## 5  -0.30310079  0.30626531     -0.2705068     0.2065348  0.44098127 -0.5431273
##    Pittsburgh     Rutgers    Syracuse     Temple    Nebraska Penn.State
## 1   0.5831878  0.04534471 -0.09095214  0.7318675  1.26813073  2.3864737
## 2   0.4166961  0.43486666  1.41847570  0.2502149  0.56192510  0.3538765
## 3   0.7464394  1.62938202  1.35820015  1.5193176 -1.92102580  1.2400126
## 4   0.7190475 -0.31801740 -0.19601467 -0.5549155 -0.01007186  0.2641294
## 5   0.1311611  1.12173532 -0.57182415  0.8476091 -0.23844421  1.6246395
##    Charlotte East.Carolina   Marshall     Memphis Old.Dominion      Tulsa
## 1 -1.5411131  -0.059439896 -0.5711509  0.4691801   -1.3481967 -0.6767895
## 2 -1.6014758  -1.446589280 -0.3634646 -1.6368789   -0.9370565 -1.2433250
## 3  0.6095165   1.161816151  0.3542456 -0.8545836    1.3420984 -1.7978377
## 4 -0.6341987  -0.641208545 -0.6665787 -0.6034949   -0.6176053 -0.2599417
## 5 -0.1778619   0.008933626 -1.1796845 -0.4038942   -0.2242729 -2.0542312
##         Army        Navy Notre.Dame       Akron  Ball.State Bowling.Green
## 1 -0.4864350 -0.7426321  1.2554917 -0.91343020 -1.2336766    -1.1814124
```

```
## 2   0.7721271 -0.1483508  0.8926236  0.66357185  0.2922162      0.7768849
## 3   1.6960733  1.3009024 -0.2574030  0.49986072 -0.5090963      0.1598884
## 4  -0.6726008 -0.6631935  2.8353757 -0.61684713 -0.6345605     -0.6337026
## 5  -2.0071116 -2.0113630 -1.3534602  0.01637356 -0.4550880     -0.8598559
##        Buffalo Central.Michigan Eastern.Michigan  Kent.State Miami.University
## 1  -0.9555746      -0.90493300       -0.9048903 -0.89122645      -1.15314528
## 2   1.3602371       1.63235471        1.1116646  0.68924035       0.05725938
## 3   0.9247365      -0.01997824        0.1647239  0.52524203      -0.01502846
## 4  -0.4335337      -0.66575255       -0.6873944 -0.69298039      -0.48554371
## 5   0.1394868       0.08049871       -0.3493347  0.05047326      -0.87597577
##    Northern.Illinois        Ohio      Toledo Western.Michigan    Kentucky
## 1         -0.8708161 -1.16962729 -1.0736388       -0.9091602  0.692327767
## 2          0.9909318 -0.01261891  0.8882535        1.1234079 -0.506002614
## 3         -0.6542491  0.40815140  0.1710220       -0.1492752  0.030906808
## 4         -0.6858129 -0.52357547 -0.6021828       -0.6015806 -0.194732014
## 5         -0.3804230 -0.09602260 -0.4140798       -0.1947789 -0.005237678
##       Missouri  Tennessee Vanderbilt Arkansas.State Middle.Tennessee
## 1   0.83742072  2.1803632 -0.4954019     -0.8722679      -0.87073068
## 2  -0.15510802 -1.3023110 -1.2264609     -1.3611128      -1.34896027
## 3  -1.22485224  0.1128187 -0.3402659     -0.9629292      -0.27797671
## 4  -0.07956048 -0.2330615  1.2220829     -0.6917164      -0.67748870
## 5   0.57746865  0.2576400 -1.2797694     -1.1855302      -0.07467707
##    Western.Kentucky  X X.1 X.2 X.3 X.4 X.5 X.6 X.7 X.8 X.9 X.10 X.11 X.12 X.13
## 1        -1.2550263 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA   NA   NA
## 2        -0.9176115 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA   NA   NA
## 3        -0.2854229 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA   NA   NA
## 4        -0.6492243 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA   NA   NA
## 5        -0.5534900 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA   NA   NA
##    X.14 X.15 X.16 X.17
## 1    NA   NA   NA   NA
## 2    NA   NA   NA   NA
## 3    NA   NA   NA   NA
## 4    NA   NA   NA   NA
## 5    NA   NA   NA   NA
```

Remove the extra column "i..candidates" from the imported dataset University_49

```
University_49 <- University_49[, -1]
head(University_49)
```

```
##             School       Duke   Maryland North.Carolina.State North.Carolina
## 1  StadiumCapacity -0.7451514  0.1113582            0.2643505      0.4947998
## 2         Latitude -1.3000459 -0.1389516           -1.3610659     -1.3203829
## 3        Longitude  0.9159120  1.2319867            0.9568089      0.8958546
## 4 Endowment ($000)  2.5449230 -0.2651109           -0.3638772      0.5679706
## 5       Enrollment -1.0502828  0.9118229            0.6626736      0.1640208
##     Virginia Virginia.Tech Wake.Forest  Iowa.State      Kansas Kansas.State
## 1  0.4316045     0.6337013  -0.8493809  0.1540577 -0.05640823  -0.05943990
## 2 -0.5080417    -0.8178726  -1.2532009  1.0266317 -0.15173905  -0.06424199
## 3  0.9844242     0.6743377   0.7007552 -1.4321853 -1.69232936  -1.90535282
## 4  1.9853271    -0.3735079  -0.1140267 -0.3669104 -0.00504443  -0.52274758
## 5 -0.2088616     0.3233594  -1.7655794  0.2060033  0.05791320  -0.30310079
##      Oklahoma Oklahoma.State West.Virginia  Cincinnati Louisville Pittsburgh
## 1  1.31172693      0.3768637     0.3906130 -0.69993262  0.1967572  0.5831878
## 2 -1.49771144     -1.4977114     0.1073058 -0.08360839 -0.4359264  0.4166961
```

```
## 3 -2.05276241    -2.0527624     0.7507380  0.02348501 -0.1736760  0.7464394
## 4 -0.02683033    -0.3275228    -0.4918203 -0.14458022 -0.2762545  0.7190475
## 5  0.30626531    -0.2705068     0.2065348  0.44098127 -0.5431273  0.1311611
##        Rutgers    Syracuse      Temple   Nebraska Penn.State  Charlotte
## 1  0.04534471 -0.09095214  0.7318675  1.26813073  2.3864737 -1.5411131
## 2  0.43486666  1.41847570  0.2502149  0.56192510  0.3538765 -1.6014758
## 3  1.62938202  1.35820015  1.5193176 -1.92102580  1.2400126  0.6095165
## 4 -0.31801740 -0.19601467 -0.5549155 -0.01007186  0.2641294 -0.6341987
## 5  1.12173532 -0.57182415  0.8476091 -0.23844421  1.6246395 -0.1778619
##   East.Carolina    Marshall    Memphis Old.Dominion       Tulsa       Army
## 1  -0.059439896 -0.5711509  0.4691801   -1.3481967 -0.6767895 -0.4864350
## 2  -1.446589280 -0.3634646 -1.6368789   -0.9370565 -1.2433250  0.7721271
## 3   1.161816151  0.3542456 -0.8545836    1.3420984 -1.7978377  1.6960733
## 4  -0.641208545 -0.6665787 -0.6034949   -0.6176053 -0.2599417 -0.6726008
## 5   0.008933626 -1.1796845 -0.4038942   -0.2242729 -2.0542312 -2.0071116
##          Navy Notre.Dame       Akron Ball.State Bowling.Green     Buffalo
## 1 -0.7426321  1.2554917 -0.91343020 -1.2336766    -1.1814124 -0.9555746
## 2 -0.1483508  0.8926236  0.66357185  0.2922162     0.7768849  1.3602371
## 3  1.3009024 -0.2574030  0.49986072 -0.5090963     0.1598884  0.9247365
## 4 -0.6631935  2.8353757 -0.61684713 -0.6345605    -0.6337026 -0.4335337
## 5 -2.0113630 -1.3534602  0.01637356 -0.4550880    -0.8598559  0.1394868
##   Central.Michigan Eastern.Michigan Kent.State Miami.University
## 1       -0.90493300       -0.9048903 -0.89122645      -1.15314528
## 2        1.63235471        1.1116646  0.68924035       0.05725938
## 3       -0.01997824        0.1647239  0.52524203      -0.01502846
## 4       -0.66575255       -0.6873944 -0.69298039      -0.48554371
## 5        0.08049871       -0.3493347  0.05047326      -0.87597577
##   Northern.Illinois        Ohio      Toledo Western.Michigan     Kentucky
## 1        -0.8708161 -1.16962729 -1.0736388       -0.9091602  0.692327767
## 2         0.9909318 -0.01261891  0.8882535        1.1234079 -0.506002614
## 3        -0.6542491  0.40815140  0.1710220       -0.1492752  0.030906808
## 4        -0.6858129 -0.52357547 -0.6021828       -0.6015806 -0.194732014
## 5        -0.3804230 -0.09602260 -0.4140798       -0.1947789 -0.005237678
##      Missouri   Tennessee Vanderbilt Arkansas.State Middle.Tennessee
## 1  0.83742072  2.1803632 -0.4954019     -0.8722679      -0.87073068
## 2 -0.15510802 -1.3023110 -1.2264609     -1.3611128      -1.34896027
## 3 -1.22485224  0.1128187 -0.3402659     -0.9629292      -0.27797671
## 4 -0.07956048 -0.2330615  1.2220829     -0.6917164      -0.67748870
## 5  0.57746865  0.2576400 -1.2797694     -1.1855302      -0.07467707
##   Western.Kentucky  X X.1 X.2 X.3 X.4 X.5 X.6 X.7 X.8 X.9 X.10 X.11 X.12 X.13
## 1       -1.2550263 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA   NA   NA
## 2       -0.9176115 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA   NA   NA
## 3       -0.2854229 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA   NA   NA
## 4       -0.6492243 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA   NA   NA
## 5       -0.5534900 NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   NA   NA   NA   NA
##   X.14 X.15 X.16 X.17
## 1   NA   NA   NA   NA
## 2   NA   NA   NA   NA
## 3   NA   NA   NA   NA
## 4   NA   NA   NA   NA
## 5   NA   NA   NA   NA
```

Get the transpose of the dataframe "University_49"

```r
University_49 <- data.frame(t(University_49))
head(University_49)
```

```
##                                    X1         X2        X3                 X4
## School              StadiumCapacity   Latitude Longitude   Endowment ($000)
## Duke                     -0.7451514 -1.3000459 0.9159120          2.5449230
## Maryland                  0.1113582 -0.1389516 1.2319867         -0.2651109
## North.Carolina.State      0.2643505 -1.3610659 0.9568089         -0.3638772
## North.Carolina            0.4947998 -1.3203829 0.8958546          0.5679706
## Virginia                  0.4316045 -0.5080417 0.9844242          1.9853271
##                              X5
## School               Enrollment
## Duke                 -1.0502828
## Maryland              0.9118229
## North.Carolina.State  0.6626736
## North.Carolina        0.1640208
## Virginia             -0.2088616
```

Create a new column name and assign the first column values to this "newcol"

```r
University_49 <- cbind(newcol = rownames(University_49), University_49)
rownames(University_49) <- 1:nrow(University_49)

head(University_49)
```

```
##                 newcol              X1         X2        X3                 X4
## 1               School StadiumCapacity   Latitude Longitude   Endowment ($000)
## 2                 Duke     -0.7451514 -1.3000459 0.9159120          2.5449230
## 3             Maryland      0.1113582 -0.1389516 1.2319867         -0.2651109
## 4 North.Carolina.State      0.2643505 -1.3610659 0.9568089         -0.3638772
## 5       North.Carolina      0.4947998 -1.3203829 0.8958546          0.5679706
## 6             Virginia      0.4316045 -0.5080417 0.9844242          1.9853271
##           X5
## 1 Enrollment
## 2 -1.0502828
## 3  0.9118229
## 4  0.6626736
## 5  0.1640208
## 6 -0.2088616
```

Remove the first row from the new dataset "Univeristy_49"

```r
University_49 <- University_49%>%
  row_to_names(row_number = 1)
head(University_49)
```

```
##                 School StadiumCapacity   Latitude Longitude Endowment ($000)
## 2                 Duke     -0.7451514 -1.3000459 0.9159120        2.5449230
## 3             Maryland      0.1113582 -0.1389516 1.2319867       -0.2651109
## 4 North.Carolina.State      0.2643505 -1.3610659 0.9568089       -0.3638772
## 5       North.Carolina      0.4947998 -1.3203829 0.8958546        0.5679706
## 6             Virginia      0.4316045 -0.5080417 0.9844242        1.9853271
## 7         Virginia.Tech      0.6337013 -0.8178726 0.6743377       -0.3735079
##    Enrollment
## 2 -1.0502828
## 3  0.9118229
```

```
## 4  0.6626736
## 5  0.1640208
## 6 -0.2088616
## 7  0.3233594
```

There are lot of extra rows got added after clean up, remove those extra rows.

```
# removing rextra rows

University_49 <- head(University_49, 49)
head(University_49)
```

```
##                    School StadiumCapacity   Latitude   Longitude Endowment ($000)
## 2                    Duke      -0.7451514 -1.3000459  0.9159120        2.5449230
## 3                Maryland       0.1113582 -0.1389516  1.2319867       -0.2651109
## 4 North.Carolina.State       0.2643505 -1.3610659  0.9568089       -0.3638772
## 5          North.Carolina       0.4947998 -1.3203829  0.8958546        0.5679706
## 6                Virginia       0.4316045 -0.5080417  0.9844242        1.9853271
## 7            Virginia.Tech       0.6337013 -0.8178726  0.6743377       -0.3735079
##    Enrollment
## 2 -1.0502828
## 3  0.9118229
## 4  0.6626736
## 5  0.1640208
## 6 -0.2088616
## 7  0.3233594
```

Big Ten Schools importing.

```
BigTenUniversity <- read.csv("BigTenExpand.csv", skip = 7)[,1:2]
head(BigTenUniversity)
```

```
##             Big.Ten            X
## 1            School      Illinois
## 2   StadiumCapacity    0.39616393
## 3          Latitude   0.291098113
## 4         Longitude  -0.576520912
## 5 Endowment ($000)   0.193512318
## 6        Enrollment   1.516494718
```

The dataset needs to be changed to wide format, in order to do this we need a unique id for each group. For this creating a new column to use in reshaping.

```
#adding a dummy vector to get a unique column for long to wide
vector_col <-  vector(mode = "integer", length = 60)
for(i in seq_len(length(vector_col)-1)){

  vector_col[i+1] <- floor(i/6)

}

vector_col
```

```
##  [1] 0 0 0 0 0 0 1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3 3 4 4 4 4 4 4 5 5 5 5 5 5 6 6
## [39] 6 6 6 6 7 7 7 7 7 7 8 8 8 8 8 8 9 9 9 9 9 9
```

combining this vector with our datset

```r
#cbind

BigTenUniversity <- cbind(BigTenUniversity, vector_col)

head(BigTenUniversity)
```

```
##              Big.Ten           X vector_col
## 1            School    Illinois           0
## 2    StadiumCapacity  0.39616393          0
## 3           Latitude 0.291098113          0
## 4          Longitude -0.576520912         0
## 5 Endowment ($000)  0.193512318          0
## 6         Enrollment 1.516494718          0
```

Using reshape, changing to wide format from long.

```r
#changing to wide format from long
BigTenUniversity <- reshape(BigTenUniversity, timevar = "Big.Ten",
                            idvar = "vector_col", direction = "wide")
head(BigTenUniversity)
```

```
##    vector_col      X.School X.StadiumCapacity   X.Latitude  X.Longitude
## 1           0      Illinois        0.39616393  0.291098113 -0.576520912
## 7           1       Indiana       0.065626984 -0.07379062 -0.299015746
## 13          2          Iowa       0.819529623  0.886058875 -1.098502436
## 19          3 Michigan State       1.00911547  1.301000932  0.026938684
## 25          4      Michigan       2.707659483  1.123658154  0.147208124
## 31          5      Minnesota      -0.025066787  2.158029349 -1.374918363
##    X.Endowment ($000) X.Enrollment
## 1          0.193512318  1.516494718
## 7          0.178889342  1.368050303
## 13        -0.122052057   0.22362889
## 19        -0.533940497  1.830654826
## 25         3.728560154  1.366721743
## 31         0.705385671  2.238345543
```

Rename the column names same as the dataset "University_49".

```r
ColumnNames <- c("vector_col","School", "StadiumCapacity","Latitude",
                 "Longitude","Endowment ($000)","Enrollment")

colnames(BigTenUniversity) <- ColumnNames
```

```r
BigTenUniversity
```

```
##    vector_col           School StadiumCapacity     Latitude    Longitude
## 1           0         Illinois      0.39616393  0.291098113 -0.576520912
## 7           1          Indiana     0.065626984 -0.07379062 -0.299015746
## 13          2             Iowa     0.819529623  0.886058875 -1.098502436
## 19          3    Michigan State     1.00911547  1.301000932  0.026938684
## 25          4         Michigan     2.707659483  1.123658154  0.147208124
## 31          5         Minnesota   -0.025066787  2.158029349 -1.374918363
## 37          6     Northwestern   -0.091208336  1.035487299 -0.485474329
## 43          7        Ohio State     2.174983032  0.243329145  0.265878343
## 49          8           Purdue     0.474304043  0.418798388 -0.360553436
## 55          9        Wisconsin     1.235252102  1.433391974 -0.755704228
##    Endowment ($000)    Enrollment
```

```
## 1        0.193512318  1.516494718
## 7        0.178889342  1.368050303
## 13      -0.122052057   0.22362889
## 19      -0.533940497  1.830654826
## 25       3.728560154  1.366721743
## 31       0.705385671  2.238345543
## 37       3.358842341 -0.648083481
## 43       0.488439062   2.62008506
## 49       0.420896539  1.094012701
## 55       0.347935893  1.342364813
```

Creating index as school names

```
BigTenUniversity <- BigTenUniversity[, -1]
rownames(BigTenUniversity) <- NULL
BigTenUniversity <- column_to_rownames(BigTenUniversity, "School")
BigTenUniversity
```

```
##                 StadiumCapacity     Latitude     Longitude Endowment ($000)
## Illinois             0.39616393 0.291098113 -0.576520912        0.193512318
## Indiana             0.065626984 -0.07379062 -0.299015746        0.178889342
## Iowa                0.819529623 0.886058875 -1.098502436       -0.122052057
## Michigan State       1.00911547 1.301000932  0.026938684       -0.533940497
## Michigan            2.707659483 1.123658154  0.147208124        3.728560154
## Minnesota          -0.025066787 2.158029349 -1.374918363        0.705385671
## Northwestern       -0.091208336 1.035487299 -0.485474329        3.358842341
## Ohio State          2.174983032 0.243329145  0.265878343        0.488439062
## Purdue              0.474304043 0.418798388 -0.360553436        0.420896539
## Wisconsin           1.235252102 1.433391974 -0.755704228        0.347935893
##                  Enrollment
## Illinois        1.516494718
## Indiana         1.368050303
## Iowa             0.22362889
## Michigan State  1.830654826
## Michigan        1.366721743
## Minnesota       2.238345543
## Northwestern   -0.648083481
## Ohio State       2.62008506
## Purdue          1.094012701
## Wisconsin       1.342364813
```

1.Using Euclidean distance to measure dissimilarity between observations, determine which school (in its own cluster of one) that hierarchical clustering with complete linkage would recommend integrating into the Big Ten Conference. That is, which school is the most similar with respect to complete linkage to the cluster of ten schools that were members of the Big Ten from 1946 to 1990?

Take average of BigTen universities and add this as a fake college in the dataset "University_49"

```
ColumnNames <- c("StadiumCapacity","Latitude","Longitude",
                 "Endowment ($000)","Enrollment")

BigTenUniversity[ColumnNames] <- sapply(BigTenUniversity[ColumnNames],
                                        as.numeric)
sapply(BigTenUniversity, class)
```

```
##  StadiumCapacity          Latitude         Longitude Endowment ($000)
##        "numeric"         "numeric"         "numeric"        "numeric"
##       Enrollment
##        "numeric"
```

Taking the mean of BigTenUniversity

```
BigTen_mean <- colMeans(BigTenUniversity)
BigTen_mean
```

```
##  StadiumCapacity          Latitude         Longitude Endowment ($000)
##        0.8766360         0.8817062        -0.4510664        0.8766469
##       Enrollment
##        1.2952275
```

Appending name to it inorder to add that in University_49

```
BigTen_mean <- append(BigTen_mean, "BigTen", 0)
BigTen_mean
```

```
##                   StadiumCapacity          Latitude         Longitude
##        "BigTen"   "0.8766359544"   "0.8817061609"   "-0.4510664299"
## Endowment ($000)        Enrollment
##   "0.8766468766"   "1.2952275116"
```

adding a new row BigTen

```
University_50 <- rbind(University_49, BigTen_mean)
tail(University_50)
```

```
##                 School StadiumCapacity       Latitude        Longitude
## 46            Tennessee       2.1803632     -1.3023110        0.1128187
## 47           Vanderbilt      -0.4954019     -1.2264609       -0.3402659
## 48        Arkansas.State      -0.8722679     -1.3611128       -0.9629292
## 49    Middle.Tennessee     -0.87073068    -1.34896027      -0.27797671
## 50    Western.Kentucky      -1.2550263     -0.9176115       -0.2854229
## 501             BigTen    0.8766359544   0.8817061609     -0.4510664299
##      Endowment ($000)    Enrollment
## 46         -0.2330615     0.2576400
## 47          1.2220829    -1.2797694
## 48         -0.6917164    -1.1855302
## 49         -0.67748870   -0.07467707
## 50         -0.6492243    -0.5534900
## 501      0.8766468766   1.2952275116
```

converting columns from char to double

```
##                 School StadiumCapacity   Latitude  Longitude Endowment ($000)
## 46           Tennessee       2.1803632 -1.3023110  0.1128187       -0.2330615
## 47          Vanderbilt      -0.4954019 -1.2264609 -0.3402659        1.2220829
## 48      Arkansas.State      -0.8722679 -1.3611128 -0.9629292       -0.6917164
## 49   Middle.Tennessee      -0.8707307 -1.3489603 -0.2779767       -0.6774887
## 50   Western.Kentucky      -1.2550263 -0.9176115 -0.2854229       -0.6492243
## 501            BigTen       0.8766360  0.8817062 -0.4510664        0.8766469
##        Enrollment
## 46    0.25764000
## 47   -1.27976940
## 48   -1.18553020
## 49   -0.07467707
## 50   -0.55349000
## 501   1.29522751
```

Converting column "School" to rownames

```r
rownames(University_50) <- NULL
University_50 <- column_to_rownames(University_50, "School")
tail(University_50)
```

```
##                  StadiumCapacity   Latitude  Longitude Endowment ($000)
## Tennessee              2.1803632 -1.3023110  0.1128187       -0.2330615
## Vanderbilt            -0.4954019 -1.2264609 -0.3402659        1.2220829
## Arkansas.State        -0.8722679 -1.3611128 -0.9629292       -0.6917164
## Middle.Tennessee      -0.8707307 -1.3489603 -0.2779767       -0.6774887
## Western.Kentucky      -1.2550263 -0.9176115 -0.2854229       -0.6492243
## BigTen                 0.8766360  0.8817062 -0.4510664        0.8766469
##                   Enrollment
## Tennessee         0.25764000
## Vanderbilt       -1.27976940
## Arkansas.State   -1.18553020
## Middle.Tennessee -0.07467707
## Western.Kentucky -0.55349000
## BigTen            1.29522751
```

clustering methord to figure out the distance b/w each college and cluster based on the farthest distance.

```r
maximum_dist <- dist(University_50, method = "maximum")
#maximum_dist
```

School closer to the BigTen

```r
maximum_dist <- as.matrix(maximum_dist)
which.min(maximum_dist["BigTen", -50])
```

```
## Missouri
##       44
```

```r
min(maximum_dist["BigTen",-50])
```

```
## [1] 1.036814
```

We got Missouri as our first school which is similar to Big Ten

2.Add the single school identified in (1) to create a cluster of 11 schools representing a hypothetical BigTen Conference. Repeat the calculations to identify the school most similar with respect to complete linkage to this new cluster of 11 schools.

Adding Missouri to BigTenUniversity dataframe

```
BigElevenUniversity <- rbind(BigTenUniversity, University_50["Missouri",])
BigElevenUniversity
```

```
##                 StadiumCapacity    Latitude   Longitude Endowment ($000)
## Illinois            0.39616393  0.29109811 -0.57652091        0.19351232
## Indiana             0.06562698 -0.07379062 -0.29901575        0.17888934
## Iowa                0.81952962  0.88605888 -1.09850244       -0.12205206
## Michigan State      1.00911547  1.30100093  0.02693868       -0.53394050
## Michigan            2.70765948  1.12365815  0.14720812        3.72856015
## Minnesota          -0.02506679  2.15802935 -1.37491836        0.70538567
## Northwestern       -0.09120834  1.03548730 -0.48547433        3.35884234
## Ohio State          2.17498303  0.24332914  0.26587834        0.48843906
## Purdue              0.47430404  0.41879839 -0.36055344        0.42089654
## Wisconsin           1.23525210  1.43339197 -0.75570423        0.34793589
## Missouri            0.83742072 -0.15510802 -1.22485224       -0.07956048
##                 Enrollment
## Illinois         1.5164947
## Indiana          1.3680503
## Iowa             0.2236289
## Michigan State   1.8306548
## Michigan         1.3667217
## Minnesota        2.2383455
## Northwestern    -0.6480835
## Ohio State       2.6200851
## Purdue           1.0940127
## Wisconsin        1.3423648
## Missouri         0.5774687
```

Take average of BigElevenUniveristy

```
BigEleven_mean <- colMeans(BigElevenUniversity)
BigEleven_mean
```

```
##  StadiumCapacity          Latitude         Longitude Endowment ($000)
##        0.8730709         0.7874503        -0.5214106        0.7897189
##       Enrollment
##        1.2299767
```

```
BigEleven_mean <- append(BigEleven_mean, "BigEleven", 0)
BigEleven_mean
```

```
##                            StadiumCapacity                   Latitude
##         "BigEleven"  "0.873070933090909"  "0.787450326272727"
##           Longitude       Endowment ($000)                 Enrollment
## "-0.521410594454545"  "0.789718935090909"        "1.229976706"
```

Removing Missouri from University_49 and adding BigEleven

```
University_48 <- University_49 %>% filter(School != "Missouri")
head(University_48)
```

```
##                 School StadiumCapacity   Latitude  Longitude Endowment ($000)
```

10

```
## 1            Duke       -0.7451514 -1.3000459  0.9159120       2.5449230
## 2         Maryland       0.1113582 -0.1389516  1.2319867      -0.2651109
## 3 North.Carolina.State   0.2643505 -1.3610659  0.9568089      -0.3638772
## 4      North.Carolina    0.4947998 -1.3203829  0.8958546       0.5679706
## 5           Virginia     0.4316045 -0.5080417  0.9844242       1.9853271
## 6       Virginia.Tech    0.6337013 -0.8178726  0.6743377      -0.3735079
##   Enrollment
## 1 -1.0502828
## 2  0.9118229
## 3  0.6626736
## 4  0.1640208
## 5 -0.2088616
## 6  0.3233594
```

Adding BigElevenMean to University_48

```
University_49 <- rbind(University_48, BigEleven_mean)
rownames(University_49) <- NULL
University_49 <- column_to_rownames(University_49, "School")
tail(University_49)
```

```
##                   StadiumCapacity            Latitude            Longitude
## Tennessee               2.1803632          -1.3023110            0.1128187
## Vanderbilt             -0.4954019          -1.2264609           -0.3402659
## Arkansas.State         -0.8722679          -1.3611128           -0.9629292
## Middle.Tennessee      -0.87073068         -1.34896027          -0.27797671
## Western.Kentucky       -1.2550263          -0.9176115           -0.2854229
## BigEleven        0.873070933090909 0.787450326272727 -0.521410594454545
##                  Endowment ($000)  Enrollment
## Tennessee              -0.2330615   0.2576400
## Vanderbilt              1.2220829  -1.2797694
## Arkansas.State         -0.6917164  -1.1855302
## Middle.Tennessee      -0.67748870  -0.07467707
## Western.Kentucky       -0.6492243  -0.5534900
## BigEleven       0.789718935090909 1.229976706
```

```
University_49[ColumnNames] <- sapply(University_49[ColumnNames], as.numeric)
tail(University_49)
```

```
##                  StadiumCapacity  Latitude  Longitude Endowment ($000)
## Tennessee              2.1803632 -1.3023110  0.1128187      -0.2330615
## Vanderbilt            -0.4954019 -1.2264609 -0.3402659       1.2220829
## Arkansas.State        -0.8722679 -1.3611128 -0.9629292      -0.6917164
## Middle.Tennessee      -0.8707307 -1.3489603 -0.2779767      -0.6774887
## Western.Kentucky      -1.2550263 -0.9176115 -0.2854229      -0.6492243
## BigEleven              0.8730709  0.7874503 -0.5214106       0.7897189
##                   Enrollment
## Tennessee          0.25764000
## Vanderbilt        -1.27976940
## Arkansas.State    -1.18553020
## Middle.Tennessee  -0.07467707
## Western.Kentucky  -0.55349000
## BigEleven          1.22997671
```

```
maximum_dist <- dist(University_49, method = "maximum")
#maximum_dist
```

School closer to the BigTen

```
maximum_dist <- as.matrix(maximum_dist)

#colnames(maximum_dist)

which.min(maximum_dist["BigEleven", -49])
```

```
## Iowa.State
##          8
```

```
min(maximum_dist["BigEleven",-49])
```

```
## [1] 1.156629
```

We got Iowa.State as our second school which is similar to Big Ten

3.Add the school identified in (2) to create a cluster of 12 schools representing a hypothetical Big Ten Conference. Repeat the calculations to identify the school most similar with respect to complete linkage to this new cluster of 12 schools.

Adding Iowa.State to BigElevenUniversity dataframe

```
BigTwelveUniversity <- rbind(BigElevenUniversity, University_49["Iowa.State",])
BigTwelveUniversity
```

```
##                 StadiumCapacity     Latitude    Longitude Endowment ($000)
## Illinois            0.39616393   0.29109811  -0.57652091       0.19351232
## Indiana             0.06562698  -0.07379062  -0.29901575       0.17888934
## Iowa                0.81952962   0.88605888  -1.09850244      -0.12205206
## Michigan State      1.00911547   1.30100093   0.02693868      -0.53394050
## Michigan            2.70765948   1.12365815   0.14720812       3.72856015
## Minnesota          -0.02506679   2.15802935  -1.37491836       0.70538567
## Northwestern       -0.09120834   1.03548730  -0.48547433       3.35884234
## Ohio State          2.17498303   0.24332914   0.26587834       0.48843906
## Purdue              0.47430404   0.41879839  -0.36055344       0.42089654
## Wisconsin           1.23525210   1.43339197  -0.75570423       0.34793589
## Missouri            0.83742072  -0.15510802  -1.22485224      -0.07956048
## Iowa.State          0.15405770   1.02663170  -1.43218530      -0.36691040
##                 Enrollment
## Illinois         1.5164947
## Indiana          1.3680503
## Iowa             0.2236289
## Michigan State   1.8306548
## Michigan         1.3667217
## Minnesota        2.2383455
## Northwestern    -0.6480835
## Ohio State       2.6200851
## Purdue           1.0940127
## Wisconsin        1.3423648
## Missouri         0.5774687
## Iowa.State       0.2060033
```

Finding mean of BigTwelveUniverisity

```
BigTwelve_mean <- colMeans(BigTwelveUniversity)
BigTwelve_mean
```

```
##  StadiumCapacity          Latitude         Longitude Endowment ($000)
##        0.8131532         0.8073821        -0.5973085        0.6933332
##       Enrollment
##        1.1446456
```

```
BigTwelve_mean <- append(BigTwelve_mean, "BigTwelve", 0)
BigTwelve_mean
```

```
##                       StadiumCapacity                Latitude
##         "BigTwelve"  "0.813153163666667"  "0.807382107416667"
##            Longitude     Endowment ($000)              Enrollment
## "-0.597308486583333"  "0.693333157166667"    "1.14464558883333"
```

Removing Iowa.State from University__48

```
University_47 <- University_48 %>% filter(School != "Iowa.State")
head(University_47)
```

13

```
##                  School StadiumCapacity    Latitude   Longitude Endowment ($000)
## 1                  Duke      -0.7451514 -1.3000459   0.9159120        2.5449230
## 2              Maryland       0.1113582 -0.1389516   1.2319867       -0.2651109
## 3 North.Carolina.State       0.2643505 -1.3610659   0.9568089       -0.3638772
## 4        North.Carolina       0.4947998 -1.3203829   0.8958546        0.5679706
## 5              Virginia       0.4316045 -0.5080417   0.9844242        1.9853271
## 6          Virginia.Tech       0.6337013 -0.8178726   0.6743377       -0.3735079
##   Enrollment
## 1 -1.0502828
## 2  0.9118229
## 3  0.6626736
## 4  0.1640208
## 5 -0.2088616
## 6  0.3233594
```

Adding BigTwelveMean to University_47

```
University_48 <- rbind(University_47, BigTwelve_mean)
rownames(University_48) <- NULL
University_48 <- column_to_rownames(University_48, "School")
tail(University_48)
```

```
##                      StadiumCapacity            Latitude           Longitude
## Tennessee                  2.1803632          -1.3023110           0.1128187
## Vanderbilt                -0.4954019          -1.2264609          -0.3402659
## Arkansas.State            -0.8722679          -1.3611128          -0.9629292
## Middle.Tennessee         -0.87073068         -1.34896027         -0.27797671
## Western.Kentucky          -1.2550263          -0.9176115          -0.2854229
## BigTwelve         0.813153163666667 0.807382107416667 -0.597308486583333
##                   Endowment ($000)           Enrollment
## Tennessee               -0.2330615            0.2576400
## Vanderbilt               1.2220829           -1.2797694
## Arkansas.State          -0.6917164           -1.1855302
## Middle.Tennessee        -0.67748870          -0.07467707
## Western.Kentucky        -0.6492243           -0.5534900
## BigTwelve        0.693333157166667    1.14464558883333
```

Again typecasting to double

```
University_48[ColumnNames] <- sapply(University_48[ColumnNames], as.numeric)
head(University_48)
```

```
##                      StadiumCapacity    Latitude Longitude Endowment ($000)
## Duke                      -0.7451514 -1.3000459 0.9159120        2.5449230
## Maryland                   0.1113582 -0.1389516 1.2319867       -0.2651109
## North.Carolina.State       0.2643505 -1.3610659 0.9568089       -0.3638772
## North.Carolina             0.4947998 -1.3203829 0.8958546        0.5679706
## Virginia                   0.4316045 -0.5080417 0.9844242        1.9853271
## Virginia.Tech              0.6337013 -0.8178726 0.6743377       -0.3735079
##                      Enrollment
## Duke                 -1.0502828
## Maryland              0.9118229
## North.Carolina.State  0.6626736
## North.Carolina        0.1640208
## Virginia             -0.2088616
## Virginia.Tech         0.3233594
```

```r
maximum_dist <- dist(University_48, method = "maximum")
#maximum_dist
```

School closer to the BigTen

```r
maximum_dist <- as.matrix(maximum_dist)

#colnames(maximum_dist)

which.min(maximum_dist["BigTwelve", -48])
```

```
## Kansas
##      8
```

```r
min(maximum_dist["BigTwelve",-48])
```

```
## [1] 1.095021
```

We got Kansas as our third school.

4.Add the school identified in (3) to create a cluster of 13 schools representing a hypothetical Big Ten Conference. Repeat the calculations to identify the school most similar with respect to complete linkage to this new cluster of 13 schools. Add this school to create a 14-school cluster.

Adding Kansas to BigTwelveUniversity dataframe

```
BigThirteenUniversity <- rbind(BigTwelveUniversity, University_48["Kansas",])
tail(BigThirteenUniversity)
```

```
##              StadiumCapacity   Latitude  Longitude Endowment ($000) Enrollment
## Ohio State        2.17498303  0.2433291  0.2658783       0.48843906  2.6200851
## Purdue            0.47430404  0.4187984 -0.3605534       0.42089654  1.0940127
## Wisconsin         1.23525210  1.4333920 -0.7557042       0.34793589  1.3423648
## Missouri          0.83742072 -0.1551080 -1.2248522      -0.07956048  0.5774687
## Iowa.State        0.15405770  1.0266317 -1.4321853      -0.36691040  0.2060033
## Kansas           -0.05640823 -0.1517390 -1.6923294      -0.00504443  0.0579132
```

Finding mean of BigThirteenUniverisity

```
BigThirteen_mean <- colMeans(BigThirteenUniversity)
BigThirteen_mean
```

```
##  StadiumCapacity         Latitude        Longitude Endowment ($000)
##        0.7462638        0.7336036       -0.6815409        0.6396118
##       Enrollment
##        1.0610508
```

```
BigThirteen_mean <- append(BigThirteen_mean, "BigThirteen", 0)
BigThirteen_mean
```

```
##                          StadiumCapacity              Latitude
##      "BigThirteen"  "0.746263825692308"  "0.733603556846154"
##          Longitude     Endowment ($000)            Enrollment
## "-0.681540861461538"  "0.639611804307692"   "1.06105078969231"
```

Removing Kansas from University_48

```
University_46 <- University_47 %>% filter(School != "Kansas")
head(University_46)
```

```
##                     School StadiumCapacity   Latitude  Longitude Endowment ($000)
## 1                    Duke      -0.7451514 -1.3000459  0.9159120        2.5449230
## 2                Maryland       0.1113582 -0.1389516  1.2319867       -0.2651109
## 3 North.Carolina.State       0.2643505 -1.3610659  0.9568089       -0.3638772
## 4         North.Carolina       0.4947998 -1.3203829  0.8958546        0.5679706
## 5                Virginia       0.4316045 -0.5080417  0.9844242        1.9853271
## 6           Virginia.Tech       0.6337013 -0.8178726  0.6743377       -0.3735079
##    Enrollment
## 1 -1.0502828
## 2  0.9118229
## 3  0.6626736
## 4  0.1640208
## 5 -0.2088616
## 6  0.3233594
```

Adding BigTwelveMean to University_47

```
University_47 <- rbind(University_46, BigThirteen_mean)
rownames(University_47) <- NULL
```

16

```
University_47 <- column_to_rownames(University_47, "School")
tail(University_47)
```

```
##                     StadiumCapacity              Latitude              Longitude
## Tennessee                 2.1803632            -1.3023110              0.1128187
## Vanderbilt               -0.4954019            -1.2264609             -0.3402659
## Arkansas.State           -0.8722679            -1.3611128             -0.9629292
## Middle.Tennessee         -0.87073068           -1.34896027            -0.27797671
## Western.Kentucky         -1.2550263            -0.9176115             -0.2854229
## BigThirteen       0.746263825692308 0.733603556846154 -0.681540861461538
##                     Endowment ($000)          Enrollment
## Tennessee                 -0.2330615            0.2576400
## Vanderbilt                 1.2220829           -1.2797694
## Arkansas.State            -0.6917164           -1.1855302
## Middle.Tennessee          -0.67748870          -0.07467707
## Western.Kentucky          -0.6492243           -0.5534900
## BigThirteen        0.639611804307692 1.06105078969231
```

Typecasting to *double

```
University_47[ColumnNames] <- sapply(University_47[ColumnNames], as.numeric)
head(University_47)
```

```
##                     StadiumCapacity  Latitude Longitude Endowment ($000)
## Duke                     -0.7451514 -1.3000459 0.9159120        2.5449230
## Maryland                  0.1113582 -0.1389516 1.2319867       -0.2651109
## North.Carolina.State      0.2643505 -1.3610659 0.9568089       -0.3638772
## North.Carolina            0.4947998 -1.3203829 0.8958546        0.5679706
## Virginia                  0.4316045 -0.5080417 0.9844242        1.9853271
## Virginia.Tech             0.6337013 -0.8178726 0.6743377       -0.3735079
##                     Enrollment
## Duke                 -1.0502828
## Maryland              0.9118229
## North.Carolina.State  0.6626736
## North.Carolina        0.1640208
## Virginia             -0.2088616
## Virginia.Tech         0.3233594
```

```
maximum_dist <- dist(University_47, method = "maximum")
#maximum_dist
```

School closer to the BigTen

```
maximum_dist <- as.matrix(maximum_dist)

#colnames(maximum_dist)

which.min(maximum_dist["BigThirteen", -47])
```

```
## Kentucky
##       41
```

```
min(maximum_dist["BigThirteen",-47])
```

```
## [1] 1.239606
```

Adding Kentucky to BigThirteenUniveristy dataframe

17

```
Big14University <- rbind(BigThirteenUniversity, University_47["Kentucky",])
tail(Big14University)
```

```
##               StadiumCapacity    Latitude   Longitude Endowment ($000)    Enrollment
## Purdue             0.47430404   0.4187984 -0.36055344       0.42089654   1.094012701
## Wisconsin          1.23525210   1.4333920 -0.75570423       0.34793589   1.342364813
## Missouri           0.83742072  -0.1551080 -1.22485224      -0.07956048   0.577468650
## Iowa.State         0.15405770   1.0266317 -1.43218530      -0.36691040   0.206003300
## Kansas            -0.05640823  -0.1517390 -1.69232936      -0.00504443   0.057913200
## Kentucky           0.69232777  -0.5060026  0.03090681      -0.19473201  -0.005237678
```

We got our 4th school, kentucky

5. How does the hypothetical 14-team cluster created in (4) compare to the actual 14-team Big Ten Conference? For both the hypothetical 14-team Big Ten Conference and the actual 14-team Big Ten Conference, compute the cluster centroid, the distance from each cluster member to the cluster centroid, and average distance between the observations in the cluster. What do you observe when comparing these two clusters? Which cluster has the smaller average distance between observations? Is this surprising? Explain.

First take Penn.State,Nebraska,Maryland, and Rutgers to BigTenUniversity and make a actual BigTenUniveristy.

```
Actual_Big14Univ <- rbind(BigTenUniversity,
                    University_47[c("Penn.State","Nebraska",
                                    "Maryland","Rutgers"),])
Actual_Big14Univ
```

```
##                StadiumCapacity    Latitude    Longitude Endowment ($000)
## Illinois           0.39616393  0.29109811 -0.57652091       0.19351232
## Indiana            0.06562698 -0.07379062 -0.29901575       0.17888934
## Iowa               0.81952962  0.88605888 -1.09850244      -0.12205206
## Michigan State     1.00911547  1.30100093  0.02693868      -0.53394050
## Michigan           2.70765948  1.12365815  0.14720812       3.72856015
## Minnesota         -0.02506679  2.15802935 -1.37491836       0.70538567
## Northwestern      -0.09120834  1.03548730 -0.48547433       3.35884234
## Ohio State         2.17498303  0.24332914  0.26587834       0.48843906
## Purdue             0.47430404  0.41879839 -0.36055344       0.42089654
## Wisconsin          1.23525210  1.43339197 -0.75570423       0.34793589
## Penn.State         2.38647370  0.35387650  1.24001260       0.26412940
## Nebraska           1.26813073  0.56192510 -1.92102580      -0.01007186
## Maryland           0.11135820 -0.13895160  1.23198670      -0.26511090
## Rutgers            0.04534471  0.43486666  1.62938202      -0.31801740
##                Enrollment
## Illinois          1.5164947
## Indiana           1.3680503
## Iowa              0.2236289
## Michigan State    1.8306548
## Michigan          1.3667217
## Minnesota         2.2383455
## Northwestern     -0.6480835
## Ohio State        2.6200851
## Purdue            1.0940127
## Wisconsin         1.3423648
## Penn.State        1.6246395
## Nebraska         -0.2384442
## Maryland          0.9118229
## Rutgers           1.1217353
```

Get the mean of hypothetical and actual Big 14 Universities.

```
ActualBig14_mean <- colMeans(Actual_Big14Univ)
ActualBig14_mean
```

```
##  StadiumCapacity         Latitude        Longitude Endowment ($000)
##        0.8984048        0.7163413       -0.1664506        0.6026713
##       Enrollment
##        1.1694306
```

```r
class(ActualBig14_mean)
```

```
## [1] "numeric"
```

```r
HypotheticalBig14_mean <- colMeans(Big14University)
HypotheticalBig14_mean
```

```
##   StadiumCapacity          Latitude         Longitude Endowment ($000)
##        0.7424113         0.6450603        -0.6306517        0.5800158
##       Enrollment
##        0.9848873
```

```r
class(HypotheticalBig14_mean)
```

```
## [1] "numeric"
```

Append ActualBig14_mean and hypotheticalBig14_mean to respective data frames.

```r
Actual_Big14Univ <- rbind(Actual_Big14Univ, ActualBig14_mean)
```

```r
rownames(Actual_Big14Univ)[15]<-"Actual_centroid"
Actual_Big14Univ
```

```
##                 StadiumCapacity    Latitude   Longitude Endowment ($000)
## Illinois             0.39616393  0.29109811 -0.57652091       0.19351232
## Indiana              0.06562698 -0.07379062 -0.29901575       0.17888934
## Iowa                 0.81952962  0.88605888 -1.09850244      -0.12205206
## Michigan State       1.00911547  1.30100093  0.02693868      -0.53394050
## Michigan             2.70765948  1.12365815  0.14720812       3.72856015
## Minnesota           -0.02506679  2.15802935 -1.37491836       0.70538567
## Northwestern        -0.09120834  1.03548730 -0.48547433       3.35884234
## Ohio State           2.17498303  0.24332914  0.26587834       0.48843906
## Purdue               0.47430404  0.41879839 -0.36055344       0.42089654
## Wisconsin            1.23525210  1.43339197 -0.75570423       0.34793589
## Penn.State           2.38647370  0.35387650  1.24001260       0.26412940
## Nebraska             1.26813073  0.56192510 -1.92102580      -0.01007186
## Maryland             0.11135820 -0.13895160  1.23198670      -0.26511090
## Rutgers              0.04534471  0.43486666  1.62938202      -0.31801740
## Actual_centroid      0.89840478  0.71634130 -0.16645063       0.60267129
##                 Enrollment
## Illinois         1.5164947
## Indiana          1.3680503
## Iowa             0.2236289
## Michigan State   1.8306548
## Michigan         1.3667217
## Minnesota        2.2383455
## Northwestern    -0.6480835
## Ohio State       2.6200851
## Purdue           1.0940127
## Wisconsin        1.3423648
## Penn.State       1.6246395
## Nebraska        -0.2384442
## Maryland         0.9118229
## Rutgers          1.1217353
## Actual_centroid  1.1694306
```

```
Hypothetical_Big14Univ <- rbind(Big14University, HypotheticalBig14_mean)
rownames(Hypothetical_Big14Univ)[15]<-"Hypothetical_centroid"
Hypothetical_Big14Univ
```

```
##                      StadiumCapacity     Latitude   Longitude Endowment ($000)
## Illinois                  0.39616393   0.29109811 -0.57652091       0.19351232
## Indiana                   0.06562698  -0.07379062 -0.29901575       0.17888934
## Iowa                      0.81952962   0.88605888 -1.09850244      -0.12205206
## Michigan State            1.00911547   1.30100093  0.02693868      -0.53394050
## Michigan                  2.70765948   1.12365815  0.14720812       3.72856015
## Minnesota                -0.02506679   2.15802935 -1.37491836       0.70538567
## Northwestern             -0.09120834   1.03548730 -0.48547433       3.35884234
## Ohio State                2.17498303   0.24332914  0.26587834       0.48843906
## Purdue                    0.47430404   0.41879839 -0.36055344       0.42089654
## Wisconsin                 1.23525210   1.43339197 -0.75570423       0.34793589
## Missouri                  0.83742072  -0.15510802 -1.22485224      -0.07956048
## Iowa.State                0.15405770   1.02663170 -1.43218530      -0.36691040
## Kansas                   -0.05640823  -0.15173905 -1.69232936      -0.00504443
## Kentucky                  0.69232777  -0.50600261  0.03090681      -0.19473201
## Hypothetical_centroid     0.74241125   0.64506026 -0.63065174       0.58001582
##                        Enrollment
## Illinois              1.516494718
## Indiana               1.368050303
## Iowa                  0.223628890
## Michigan State        1.830654826
## Michigan              1.366721743
## Minnesota             2.238345543
## Northwestern         -0.648083481
## Ohio State            2.620085060
## Purdue                1.094012701
## Wisconsin             1.342364813
## Missouri              0.577468650
## Iowa.State            0.206003300
## Kansas                0.057913200
## Kentucky             -0.005237678
## Hypothetical_centroid 0.984887328
```

Computing the distance from each cluster member to the centroid of actual big 14 universities.

```
maximum_dist <- dist(Actual_Big14Univ, method = "euclidian")
#maximum_dist
```

```
maximum_dist <- as.matrix(maximum_dist)
maximum_dist["Actual_centroid",-15]
```

```
##      Illinois        Indiana           Iowa  Michigan State       Michigan
##     0.9429209      1.2467725      1.5243066       1.4562235      3.6534636
##     Minnesota   Northwestern      Ohio State          Purdue      Wisconsin
##     2.3547373      3.4760323      2.0390557       0.5871959      1.0342359
##    Penn.State       Nebraska        Maryland         Rutgers
##     2.1553873      2.3657222      2.0312565       2.2095015
```

```
AverageDistance_Actual <- mean(maximum_dist["Actual_centroid",-15])
AverageDistance_Actual
```

```
## [1] 1.934058
```

Computing the distance from each cluster member to the centroid of hypothetical big 14 universities.

```
maximum_dist <- dist(Hypothetical_Big14Univ, method = "euclidian")
# maximum_dist
```

```
maximum_dist <- as.matrix(maximum_dist)
maximum_dist["Hypothetical_centroid",-15]
```

```
##       Illinois       Indiana           Iowa  Michigan State       Michigan
##      0.8246805     1.1800349      1.1641845       1.7000105      3.8412825
##      Minnesota  Northwestern     Ohio State          Purdue      Wisconsin
##      2.2402915     3.3551307      2.3873990       0.4829661      1.0303632
##       Missouri     Iowa.State         Kansas        Kentucky
##      1.2662548     1.6240512      1.8978244       1.8291243
```

```
AverageDistance <-mean(maximum_dist["Hypothetical_centroid",-15])
AverageDistance
```

```
## [1] 1.773114
```

Observations about Actual 14-team cluster and Hypothetical 14-team cluster

Hypothetical Big Ten

- Hypothetical Big ten is formed by finding the nearest schools in vector space to the original Big Ten. Missouri, Iowa.State, kansas, and Kentucky are the schools added to the original Big Ten Schools.

- The centroid of the 14- team cluster is 0.74241125, 0.64506026, -0.63065174 0.58001582, 0.984887328 with respect to "StadiumCapacity","Latitude", "Longitude", "Endowment ($000)","Enrollment".

- The average distance between the observations in the cluster is 1.773114

Actual Big Ten

- Actual 14- team Big Ten is formed by adding "Penn.State","Nebraska", "Maryland", "Rutgers" to the original Big Ten schools during the years 1990-2014.

- The centroid of the actual 14 team school is 0.89840478, 0.71634130, -0.16645063, 0.60267129, 1.1694306 with respect to "StadiumCapacity", "Latitude", "Longitude", "Endowment ($000)","Enrollment".

- The average distance between the observatiobs in the cluster is 1.934058.

Stadium Capacity

The stadium capacity of hypothetical cluster is approximatly 0.74 standard deviation away from the mean, whereas the actual cluster is approximatly 0.90 standard deviation from the mean, which tells us that the stadium capacity for actual 14 team cluster is higher.

Endowment

The Endowment of the hypothetical cluster is around 0.58 standard deviation away from the mean and the actual cluster is 0.61 standard deviation away from the mean. This also conclude that the actual cluster has more funding than the hypothetical.

Enrollment

The hypothetical cluster has a lower enrollment(0.98 standard deviation away from the mean) compared to the actual 14-team cluster(1.30 standard deviation away from the mean).

Conclusion

The hypothetical cluster has the smaller average distance(1.77) between observations compared to the actual cluster(1.93). It is surprising to see the actual cluster more spread than the hypothetical cluster. It indicates that the four schools added between 1990 to 2014 were not the most similar schools to the original Big Ten as found in the cluster analysis.