Statistical Analysis For Dataset Cities

Xinxin Xu USC ID 4205459366

2023-02-20

```
# import data
library(readxl)
df <- read_excel("cities1.xlsx")

library(cluster)  # silohuette()
library(factoextra)
```

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

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

```
#
str(df)
```

```
## tibble [325 x 21] (S3: tbl_df/tbl/data.frame)
##  $ Metropolitan_Area      : chr [1:325] "Abilene, TX" "Akron, OH" "Albany, GA" "Albany-Schenectady-T
##  $ Cost_Living            : num [1:325] 96.3 47.3 86.1 25.2 44.5 ...
##  $ Transportation         : num [1:325] 36.5 69.7 28 82.7 84.1 ...
##  $ Jobs                   : num [1:325] 17.3 86.1 32 53 90.7 ...
##  $ Education              : num [1:325] 49.3 72 26.6 99.4 71.7 ...
##  $ Climate                : num [1:325] 55.52 22.66 75.63 8.78 78.18 ...
##  $ Crime                  : num [1:325] 49.58 54.11 15.59 73.94 2.84 ...
##  $ Arts                   : num [1:325] 27.2 81.6 33.1 79.6 75.4 ...
##  $ Health_Care            : num [1:325] 45 24.1 20.1 77.3 77.9 ...
##  $ Recreation             : num [1:325] 2.83 77.33 6.79 77.62 70.25 ...
##  $ Population_2000        : num [1:325] 123711 689538 120838 885782 734255 ...
##  $ Total_Violent          : num [1:325] 582 518 761 365 1133 ...
##  $ Total_Property         : num [1:325] 4396 4527 7036 3531 7261 ...
##  $ Crime_Trend            : chr [1:325] "Constant" "Constant" "Constant" "Constant" ...
##  $ Unemployment_Threat    : chr [1:325] "Average" "Average" "Average" "Below average" ...
##  $ Past_Job_Growth        : num [1:325] 6.1 11.6 6.6 5.2 21.4 6.1 4.7 9.2 16.6 14.5 ...
##  $ Fcast_Future_Job_Growth: num [1:325] 4.5 7.5 7 3.7 8.1 4.8 3.2 4.4 4.4 8 ...
##  $ Fcast_Blue_Collar_Jobs : num [1:325] 345 6363 460 88 5846 ...
##  $ Fcast_White_Collar_Jobs: num [1:325] 3385 24773 4852 20043 30750 ...
##  $ Fcast_High_Jobs        : num [1:325] 641 6728 473 1325 8134 ...
##  $ Fcast_Average_Jobs     : num [1:325] 2234 18638 4603 15233 20712 ...
```

```
head(df)
```

```
## # A tibble: 6 x 21
##   Metropolit~1 Cost_~2 Trans~3  Jobs Educa~4 Climate Crime  Arts Healt~5 Recre~6
##   <chr>          <dbl>   <dbl> <dbl>   <dbl>   <dbl> <dbl> <dbl>   <dbl>   <dbl>
## 1 Abilene, TX     96.3    36.5  17.3    49.3   55.5  49.6   27.2    45.0    2.83
## 2 Akron, OH       47.3    69.7  86.1    72.0   22.7  54.1   81.6    24.1   77.3
## 3 Albany, GA      86.1    28.0  32.0    26.6   75.6  15.6   33.2    20.1    6.79
```

```
## 4 Albany-Sche~    25.2    82.7  53.0    99.4    8.78 73.9   79.6    77.3   77.6
## 5 Albuquerque~    44.5    84.1  90.6    71.7   78.2   2.84  75.4    77.9   70.2
## 6 Alexandria,~    92.4    42.5  19.3    11.6   66     7.09  40.8    62.0   22.7
## # ... with 11 more variables: Population_2000 <dbl>, Total_Violent <dbl>,
## #   Total_Property <dbl>, Crime_Trend <chr>, Unemployment_Threat <chr>,
## #   Past_Job_Growth <dbl>, Fcast_Future_Job_Growth <dbl>,
## #   Fcast_Blue_Collar_Jobs <dbl>, Fcast_White_Collar_Jobs <dbl>,
## #   Fcast_High_Jobs <dbl>, Fcast_Average_Jobs <dbl>, and abbreviated variable
## #   names 1: Metropolitan_Area, 2: Cost_Living, 3: Transportation,
## #   4: Education, 5: Health_Care, 6: Recreation
```

```r
#drop character
df= as.data.frame(df)
df$Crime_Trend = NULL
df$Unemployment_Threat = NULL
row.names(df) = df$Metropolitan_Area
df$Metropolitan_Area = NULL
head(df,5)
```

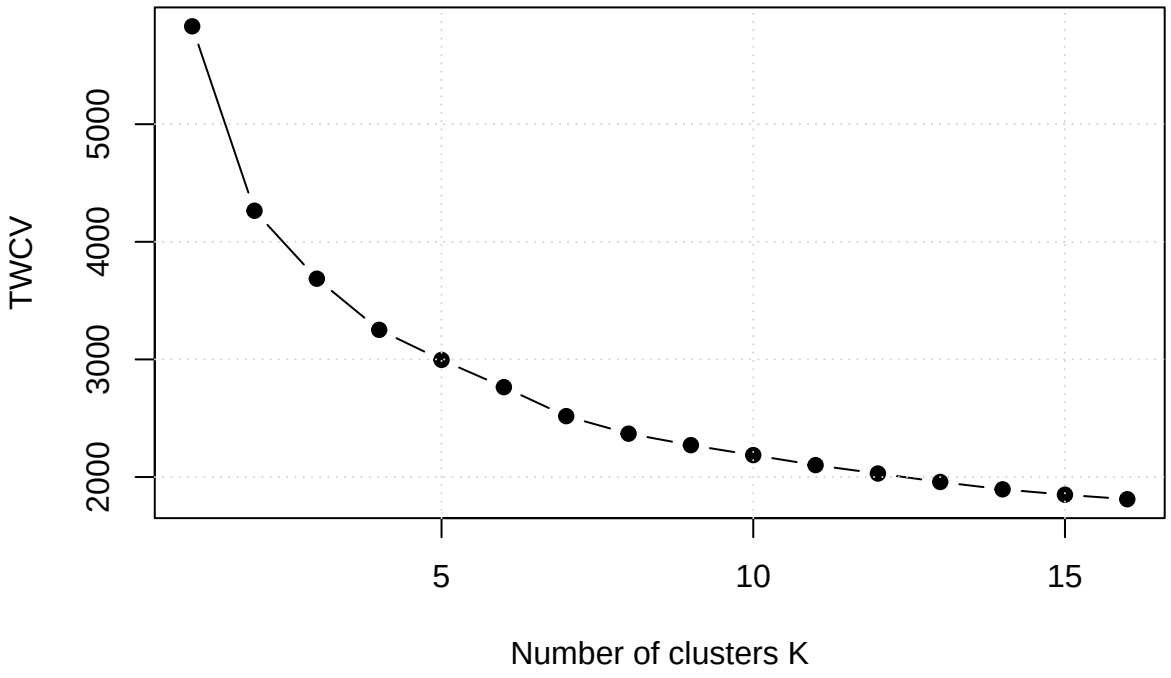
```
##                             Cost_Living Transportation  Jobs Education Climate
## Abilene, TX                       96.32          36.54 17.28     49.29   55.52
## Akron, OH                         47.31          69.68 86.11     71.95   22.66
## Albany, GA                        86.12          28.02 32.01     26.62   75.63
## Albany-Schenectady-Troy, NY       25.22          82.71 52.97     99.43    8.78
## Albuquerque, NM                   44.48          84.13 90.65     71.67   78.18
##                             Crime  Arts Health_Care Recreation Population_2000
## Abilene, TX                 49.58 27.20       45.04       2.83          123711
## Akron, OH                   54.11 81.59       24.07      77.33          689538
## Albany, GA                  15.59 33.15       20.11       6.79          120838
## Albany-Schenectady-Troy, NY 73.94 79.61       77.33      77.62          885782
## Albuquerque, NM              2.84 75.36       77.90      70.25          734255
##                             Total_Violent Total_Property Past_Job_Growth
## Abilene, TX                           582           4396             6.1
## Akron, OH                             518           4527            11.6
## Albany, GA                            761           7036             6.6
## Albany-Schenectady-Troy, NY           365           3531             5.2
## Albuquerque, NM                      1133           7261            21.4
##                             Fcast_Future_Job_Growth Fcast_Blue_Collar_Jobs
## Abilene, TX                                     4.5                    345
## Akron, OH                                       7.5                   6363
## Albany, GA                                      7.0                    460
## Albany-Schenectady-Troy, NY                     3.7                     88
## Albuquerque, NM                                 8.1                   5846
##                             Fcast_White_Collar_Jobs Fcast_High_Jobs
## Abilene, TX                                    3385             641
## Akron, OH                                     24773            6728
## Albany, GA                                     4852             473
## Albany-Schenectady-Troy, NY                   20043            1325
## Albuquerque, NM                               30750            8134
##                             Fcast_Average_Jobs
## Abilene, TX                               2234
## Akron, OH                                18638
## Albany, GA                                4603
## Albany-Schenectady-Troy, NY              15233
## Albuquerque, NM                          20712
```

```r
# scale data (always for distance-based methods)
df_scale = scale(df)
```

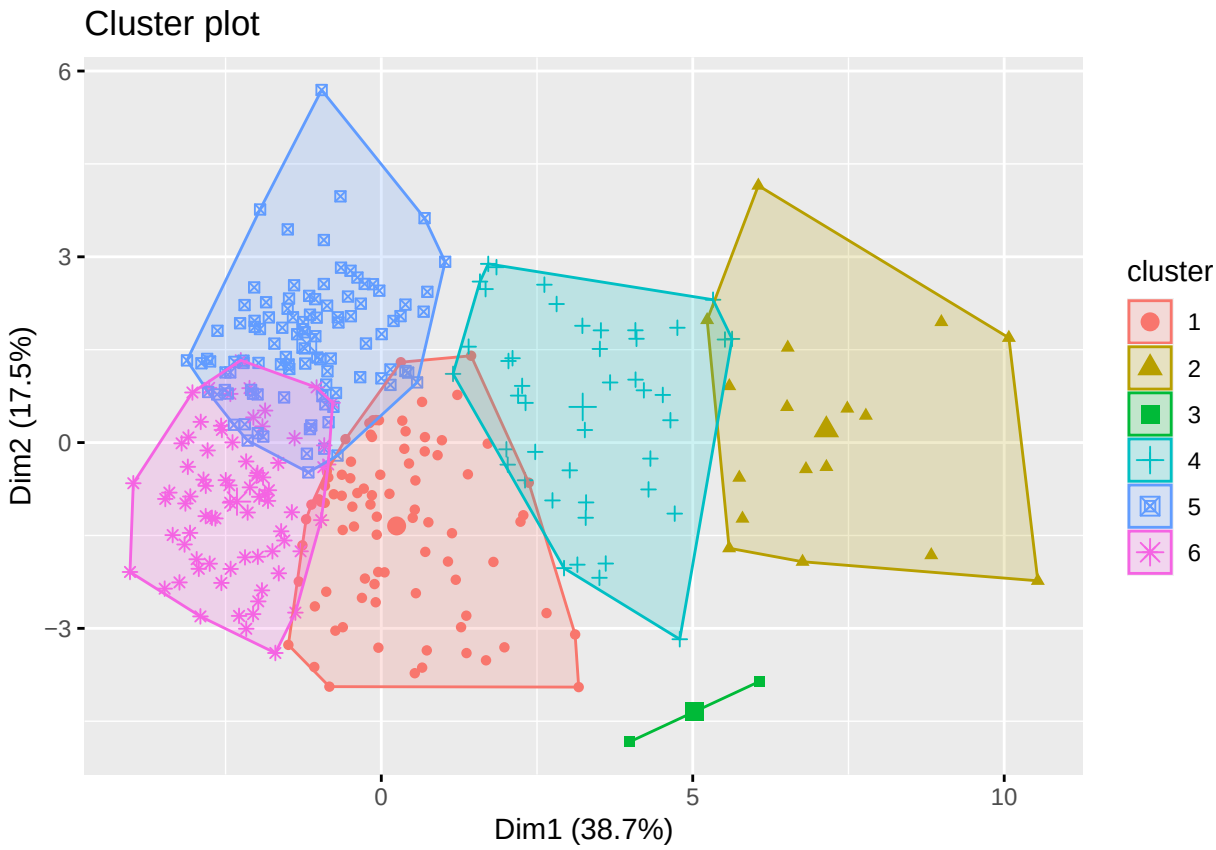
```r
#  Use set.seed(123) and the user function twcv to find TWCV values for k = 1 : 16. Use nstart = 25. Di

set.seed(123)
twcv = function(k) kmeans(df_scale, k, nstart = 25 )$tot.withinss
k <- 1:16
twcv_values <- sapply(k,twcv)
head(twcv_values)
```

```
## [1] 5832.000 4264.026 3686.381 3251.504 2996.052 2764.309
```

```r
plot(k, twcv_values,type="b",pch = 19,
     xlab="Number of clusters K",ylab="TWCV")
grid()
```

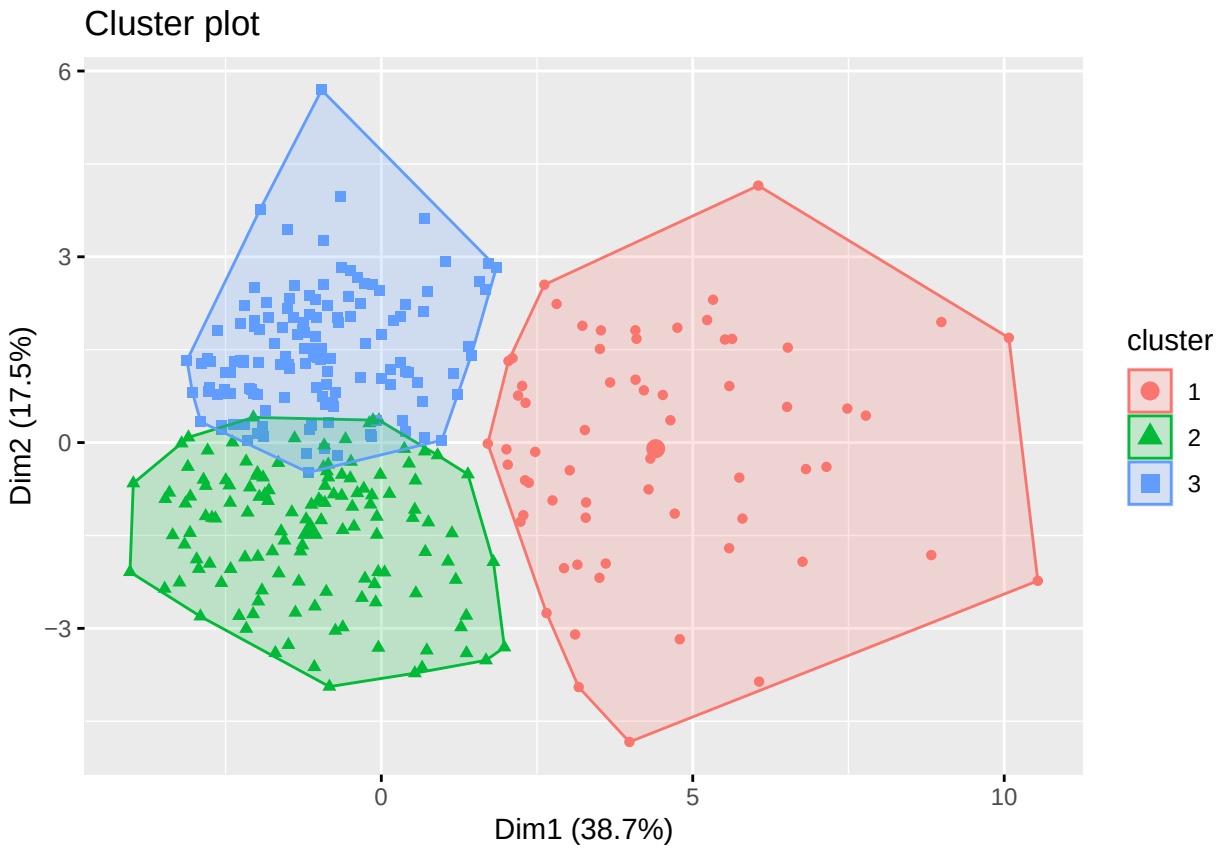



```r
#The best number of clusters is the smallest k such that the cluster plot shows the least amount of clu
final6 <- kmeans(df_scale, center = 6, nstart = 25)
fviz_cluster(final6, data = df_scale, geom = 'point')
```

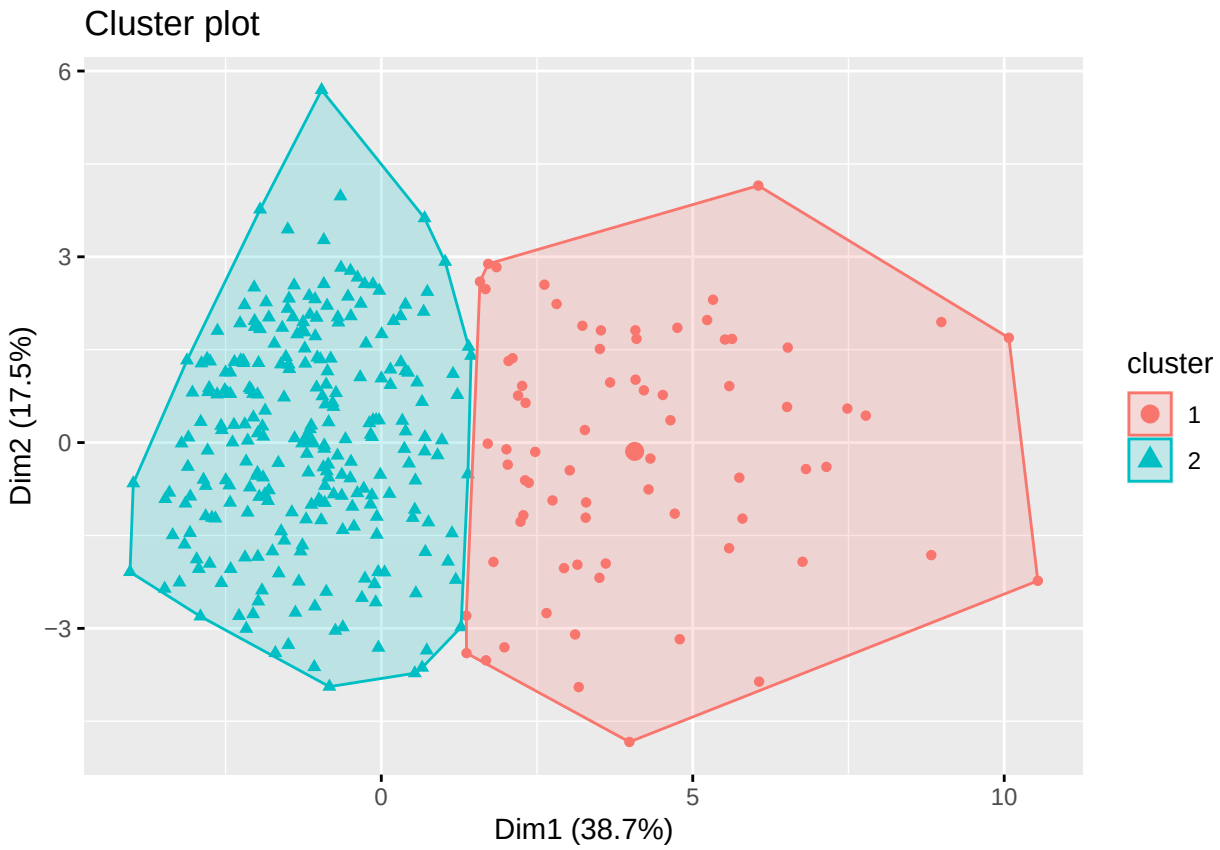



```
final4 <- kmeans(df_scale, center = 4, nstart = 25)
fviz_cluster(final4, data = df_scale, geom = 'point')
```


```
final3 <- kmeans(df_scale, center = 3, nstart = 25)
fviz_cluster(final3, data = df_scale, geom = 'point')
```




```
final2 <- kmeans(df_scale, center = 2, nstart = 25)
fviz_cluster(final2, data = df_scale, geom = 'point')
```



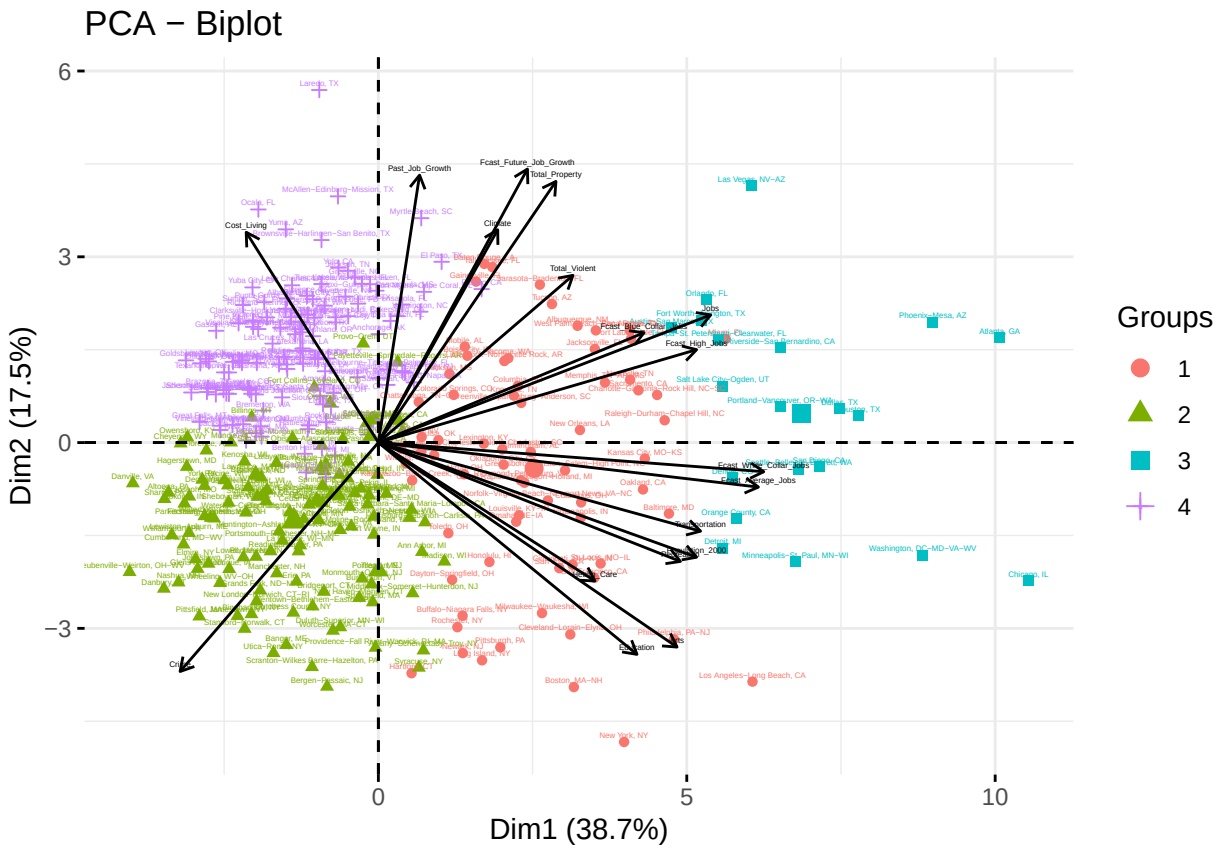

```
## I find the best value of K (the optimal number of clusters) is 4
cluster_number = final4$cluster
table(cluster_number)
```

```
## cluster_number
##   1   2   3   4
##  69 120  20 116
```

```
length(cluster_number)
```

```
## [1] 325
```

```
m1 = prcomp(df_scale, scale = T)
fviz_pca_biplot(m1, labelsize = 1, col.var = 'black', habillage = cluster_number)
```




```
# I chose clustering with k = 4
#
set.seed(123)
final <- kmeans(df_scale,centers = 4,nstart = 25)
final
```

```
## K-means clustering with 4 clusters of sizes 20, 69, 120, 116
##
## Cluster means:
##   Cost_Living Transportation       Jobs   Education    Climate      Crime
## 1  -0.6740665      1.4621261  1.5526656  0.97734506  0.5185996 -0.7371093
## 2  -0.2673169      1.0806001  0.8892767  0.91721059  0.2446817 -0.6455594
## 3  -0.2447100     -0.2886216 -0.5221906  0.06123641 -0.6006986  0.9511231
## 4   0.5283741     -0.5962874 -0.2564701 -0.77743795  0.3864552 -0.4728361
##           Arts Health_Care Recreation Population_2000 Total_Violent
## 1  1.327744006  0.58005311  1.3250248       2.0446160     0.5949124
## 2  1.000324555  0.86814166  0.9423201       0.6662064     0.6950926
## 3 -0.003721281 -0.05793164 -0.1369811      -0.3437348    -0.8042153
## 4 -0.820092420 -0.55647448 -0.6472660      -0.3932102     0.3159156
##   Total_Property Past_Job_Growth Fcast_Future_Job_Growth Fcast_Blue_Collar_Jobs
## 1      0.6669952       0.4531166              1.14970085             2.92891722
## 2      0.5551272      -0.1024415              0.06664932             0.03339108
## 3     -0.8494671      -0.2456713             -0.40351775            -0.29339548
## 4      0.4335549       0.2369543              0.17956301            -0.22133509
##   Fcast_White_Collar_Jobs Fcast_High_Jobs Fcast_Average_Jobs
## 1               3.0868595       3.0882617          3.0072150
## 2               0.5712124       0.1976550          0.5853708
```

```
## 3              -0.4161721      -0.3578688         -0.4016412
## 4              -0.4414672      -0.2798204         -0.4511891
##
## Clustering vector:
##                                Abilene, TX
##                                         4
##                                  Akron, OH
##                                         2
##                                 Albany, GA
##                                         4
##                Albany-Schenectady-Troy, NY
##                                         3
##                             Albuquerque, NM
##                                         2
##                              Alexandria, LA
##                                         4
##             Allentown-Bethlehem-Easton, PA
##                                         3
##                                Altoona, PA
##                                         3
##                               Amarillo, TX
##                                         4
##                              Anchorage, AK
##                                         4
##                              Ann Arbor, MI
##                                         3
##                               Anniston, AL
##                                         4
##                Appleton-Oshkosh-Neenah, WI
##                                         3
##                              Asheville, NC
##                                         3
##                                 Athens, GA
##                                         4
##                                Atlanta, GA
##                                         1
##                 Atlantic City-Cape May, NJ
##                                         4
##                       Augusta-Aiken, GA-SC
##                                         4
##                      Austin-San Marcos, TX
##                                         1
##                            Bakersfield, CA
##                                         4
##                              Baltimore, MD
##                                         2
##                                 Bangor, ME
##                                         3
##                    Barnstable-Yarmouth, MA
##                                         3
##                            Baton Rouge, LA
##                                         2
##                   Beaumont-Port Arthur, TX
##                                         4
```

```
##                            Bellingham, WA
##                                      4
##                          Benton Harbor, MI
##                                      4
##                         Bergen-Passaic, NJ
##                                      3
##                              Billings, MT
##                                      3
##             Biloxi-Gulfport-Pascagoula, MS
##                                      4
##                              Binghamton, NY
##                                      3
##                              Birmingham, AL
##                                      2
##                               Bismarck, ND
##                                      3
##                            Bloomington, IN
##                                      3
##                     Bloomington-Normal, IL
##                                      3
##                              Boise City, ID
##                                      2
##                               Boston, MA-NH
##                                      2
##                        Boulder-Longmont, CO
##                                      3
##                               Brazoria, TX
##                                      4
##                              Bremerton, WA
##                                      4
##                              Bridgeport, CT
##                                      3
##                               Brockton, MA
##                                      3
##       Brownsville-Harlingen-San Benito, TX
##                                      4
##                  Bryan-College Station, TX
##                                      4
##                  Buffalo-Niagara Falls, NY
##                                      2
##                              Burlington, VT
##                                      3
##                       Canton-Massillon, OH
##                                      3
##                                 Casper, WY
##                                      4
##                           Cedar Rapids, IA
##                                      3
##                       Champaign-Urbana, IL
##                                      3
##                              Charleston, WV
##                                      3
##             Charleston-North Charleston, SC
##                                      2
```

```
##        Charlotte-Gastonia-Rock Hill, NC-SC
##                                         2
##                        Charlottesville, VA
##                                         3
##                         Chattanooga, TN-GA
##                                         2
##                               Cheyenne, WY
##                                         3
##                                Chicago, IL
##                                         1
##                         Chico-Paradise, CA
##                                         4
##                        Cincinnati, OH-KY-IN
##                                         2
##            Clarksville-Hopkinsville, TN-KY
##                                         4
##                Cleveland-Lorain-Elyria, OH
##                                         2
##                        Colorado Springs, CO
##                                         2
##                               Columbia, MO
##                                         3
##                               Columbia, SC
##                                         2
##                             Columbus, GA-AL
##                                         4
##                               Columbus, OH
##                                         2
##                          Corpus Christi, TX
##                                         4
##                           Cumberland, MD-WV
##                                         3
##                                 Dallas, TX
##                                         1
##                                Danbury, CT
##                                         3
##                               Danville, VA
##                                         3
##        Davenport-Moline-Rock Island, IA-IL
##                                         3
##                      Dayton-Springfield, OH
##                                         2
##                           Daytona Beach, FL
##                                         4
##                                Decatur, IL
##                                         3
##                                 Denver, CO
##                                         1
##                             Des Moines, IA
##                                         3
##                                Detroit, MI
##                                         1
##                                 Dothan, AL
##                                         4
```

```
##                                Dover, DE
##                                       4
##                              Dubuque, IA
##                                       3
##                    Duluth-Superior, MN-WI
##                                       3
##                       Dutchess County, NY
##                                       3
##                            Eau Claire, WI
##                                       3
##                              El Paso, TX
##                                       4
##                          Elkhart-Goshen, IN
##                                       4
##                               Elmira, NY
##                                       3
##                                 Enid, OK
##                                       4
##                                 Erie, PA
##                                       3
##                    Eugene-Springfield, OR
##                                       3
##                 Evansville-Henderson, IN-KY
##                                       3
##                      Fargo-Moorhead, ND-MN
##                                       3
##                           Fayetteville, NC
##                                       4
##         Fayetteville-Springdale-Rogers, AR
##                                       3
##                   Fitchburg-Leominster, MA
##                                       3
##                           Flagstaff, AZ-UT
##                                       4
##                                Flint, MI
##                                       4
##                              Florence, AL
##                                       3
##                              Florence, SC
##                                       4
##                  Fort Collins-Loveland, CO
##                                       3
##                        Fort Lauderdale, FL
##                                       2
##                  Fort Myers-Cape Coral, FL
##                                       4
##              Fort Pierce-Port St. Lucie, FL
##                                       4
##                          Fort Smith, AR-OK
##                                       4
##                       Fort Walton Beach, FL
##                                       4
##                              Fort Wayne, IN
##                                       3
```

```
##                   Fort Worth-Arlington, TX
##                                         1
##                                Fresno, CA
##                                         4
##                                Gasden, AL
##                                         4
##                           Gainesville, FL
##                                         2
##                  Galveston-Texas City, TX
##                                         4
##                                Gary, IN
##                                         4
##                           Glens Falls, NY
##                                         3
##                             Goldsboro, NC
##                                         4
##                         Grands Fork, ND-MN
##                                         3
##                         Grand Junction, CO
##                                         4
##          Grand Rapids-Muskegon-Holland, MI
##                                         2
##                             Great Falls, MT
##                                         4
##                               Greeley, CO
##                                         4
##                               Green Bay, WI
##                                         3
##    Greensboro-Winston-Salem-High Point, NC
##                                         2
##                              Greenville, NC
##                                         4
##        Greenville-Spartanburg-Anderson, SC
##                                         2
##                              Hagerstown, MD
##                                         3
##                    Hamilton-Middletown, OH
##                                         4
##            Harrisburg-Lebanon-Carlisle, PA
##                                         3
##                               Hartford, CT
##                                         2
##                             Hattiesburg, MS
##                                         4
##               Hickory-Morganton-Lenoir, NC
##                                         3
##                               Honolulu, HI
##                                         2
##                                 Houma, LA
##                                         4
##                               Houston, TX
##                                         1
##               Huntington-Ashland, WV-KY-OH
##                                         3
```

```
##                             Huntsville, AL
##                                        2
##                           Indianapolis, IN
##                                        2
##                              Iowa City, IA
##                                        3
##                                Jackson, MI
##                                        4
##                                Jackson, MS
##                                        2
##                                Jackson, TN
##                                        4
##                           Jacksonville, FL
##                                        2
##                           Jacksonville, NC
##                                        4
##                              Jamestown, NY
##                                        3
##                      Janesville-Beloit, WI
##                                        3
##                             Jersey City, NJ
##                                        4
##      Johnson City-Kingsport-Bristol, TN-VA
##                                        3
##                                Johnstown, PA
##                                        3
##                                  Joplin, MO
##                                        4
##                  Kalamazoo-Battle Creek, MI
##                                        2
##                                Kankakee, IL
##                                        4
##                          Kansas City, MO-KS
##                                        2
##                                 Kenosha, WI
##                                        3
##                          Killeen-Temple, TX
##                                        4
##                              Knoxville, TN
##                                        2
##                                 Kokomo, IN
##                                        3
##                            La Crosse, WI-MN
##                                        3
##                              Lafayette, IN
##                                        3
##                              Lafayette, LA
##                                        4
##                            Lake Charles, LA
##                                        4
##                   Lakeland-Winter Haven, FL
##                                        4
##                              Lancaster, PA
##                                        3
```

```
##                   Lansing-East Lansing, MI
##                                        3
##                                Laredo, TX
##                                        4
##                             Las Cruces, NM
##                                        4
##                           Las Vegas, NV-AZ
##                                        1
##                              Lawrence, KS
##                                        4
##                            Lawrence, MA-NH
##                                        3
##                                Lawton, OK
##                                        4
##                         Lewiston-Auburn, ME
##                                        3
##                               Lexington, KY
##                                        2
##                                  Lima, OH
##                                        4
##                                Lincoln, NE
##                                        2
##          Little Rock-North Little Rock, AR
##                                        2
##                             Long Island, NY
##                                        2
##                        Longview-Marshall, TX
##                                        4
##                  Los Angeles-Long Beach, CA
##                                        2
##                            Louisville, KY-IN
##                                        2
##                               Lowell, MA-NH
##                                        3
##                                Lubbock, TX
##                                        4
##                               Lynchburg, VA
##                                        3
##                                   Macon, GA
##                                        4
##                                Madison, WI
##                                        3
##                             Manchester, NH
##                                        3
##                               Mansfield, OH
##                                        4
##                 McAllen-Edinburg-Mission, TX
##                                        4
##                         Medford-Ashland, OR
##                                        4
##          Melbourne-Titusville-Palm Bay, FL
##                                        4
##                           Memphis, TN-AR-MS
##                                        2
```

```
##                                Merced, CA
##                                       4
##                                 Miami, FL
##                                       2
##           Middlesex-Somerset-Hunterdon, NJ
##                                       3
##                     Milwaukee-Waukesha, WI
##                                       2
##                Minneapolis-St. Paul, MN-WI
##                                       1
##                                Mobile, AL
##                                       2
##                               Modesto, CA
##                                       4
##                          Monmouth-Ocean, NJ
##                                       3
##                                Monroe, LA
##                                       4
##                             Montgomery, AL
##                                       4
##                                Muncie, AL
##                                       3
##                            Myrtle Beach, SC
##                                       4
##                                Naples, FL
##                                       4
##                                Nashua, NH
##                                       3
##                              Nashville, TN
##                                       2
##                             New Bedford, MA
##                                       4
##                       New Haven-Meriden, CT
##                                       3
##                   New London-Norwich, CT-RI
##                                       3
##                             New Orleans, LA
##                                       2
##                                New York, NY
##                                       2
##                                Newark, NJ
##                                       2
##                             Newburgh, NY-PA
##                                       3
## Norfolk-Virginia Beach-Newport News, VA-NC
##                                       2
##                                Oakland, CA
##                                       2
##                                  Ocala, FL
##                                       4
##                           Odessa-Midland, TX
##                                       4
##                            Oklahoma City, OK
##                                       2
```

```
##                               Olympia, WA
##                                       3
##                              Omaha, NE-IA
##                                       2
##                           Orange County, CA
##                                       1
##                                Orlando, FL
##                                       1
##                              Owensboro, KY
##                                       3
##                             Panama City, FL
##                                       4
##                  Parkersburg-Marietta, WV-OH
##                                       3
##                                Pensacola, FL
##                                       4
##                             Peoria-Pekin, IL
##                                       3
##                          Philadelphia, PA-NJ
##                                       2
##                             Phoenix-Mesa, AZ
##                                       1
##                               Pine Bluff, AR
##                                       4
##                               Pittsburgh, PA
##                                       2
##                               Pittsfield, MA
##                                       3
##                                 Portland, ME
##                                       3
##                    Portland-Vancouver, OR-WA
##                                       1
##                  Portsmouth-Rochester, NH-ME
##                                       3
##       Providence-Fall River-Warwick, RI-MA
##                                       3
##                               Provo-Orem, UT
##                                       3
##                                  Pueblo, CO
##                                       4
##                             Punta Gorda, FL
##                                       4
##                                  Racine, WI
##                                       3
##               Raleigh-Durham-Chapel Hill, NC
##                                       2
##                             Rapid City, SD
##                                       4
##                                Reading, PA
##                                       3
##                                Redding, CA
##                                       4
##                                   Reno, NV
##                                       2
```

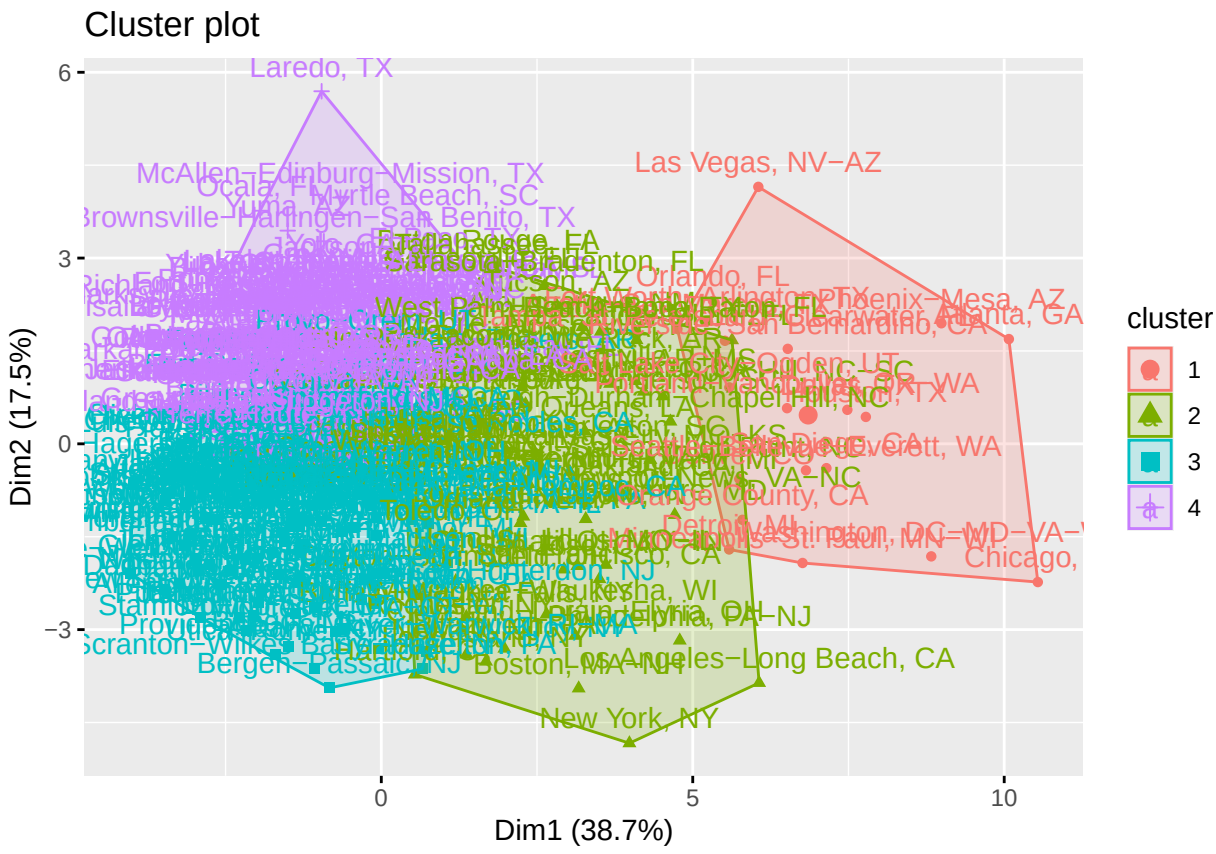
```
##               Richland-Kennewick-Pasco, WA
##                                         4
##                    Richmond-Petersburg, VA
##                                         2
##               Riverside-San Bernardino, CA
##                                         1
##                                Roanoke, VA
##                                         3
##                              Rochester, MN
##                                         3
##                              Rochester, NY
##                                         2
##                               Rockford, IL
##                                         4
##                             Rocky Mount, NC
##                                         4
##                              Sacramento, CA
##                                         2
##               Saginaw-Bay City-Midland, MI
##                                         3
##                               St. Cloud, MN
##                                         3
##                              St. Joseph, MO
##                                         3
##                            St. Louis, MO-IL
##                                         2
##                                 Salem, OR
##                                         4
##                                Salinas, CA
##                                         3
##                   Salt Lake City-Ogden, UT
##                                         1
##                              San Angelo, TX
##                                         4
##                             San Antonio, TX
##                                         2
##                               San Diego, CA
##                                         1
##                           San Francisco, CA
##                                         2
##                                San Jose, CA
##                                         2
## San Luis Obispo-Atascadero-Paso Robles, CA
##                                         3
##       Santa Barbara-Santa Maria-Lompoc, CA
##                                         3
##                 Santa Cruz-Watsonville, CA
##                                         4
##                                Santa Fe, NM
##                                         4
##                              Santa Rosa, CA
##                                         3
##                     Sarasota-Bradenton, FL
##                                         2
```

```
##                              Savannah, GA
##                                       4
##         Scranton-Wilkes Barre-Hazelton, PA
##                                       3
##               Seattle-Bellevue-Everett, WA
##                                       1
##                                Sharon, PA
##                                       3
##                              Sheboygan, WI
##                                       3
##                         Sherman-Denison, TX
##                                       4
##                 Shreveport-Bossier City, LA
##                                       4
##                            Sioux City, IA-NE
##                                       4
##                              Sioux Falls, SD
##                                       3
##                               South Bend, IN
##                                       3
##                                  Spokane, WA
##                                       2
##                              Springfield, IL
##                                       3
##                              Springfield, MA
##                                       3
##                              Springfield, MO
##                                       3
##                         Stamford-Norwalk, CT
##                                       3
##                              State College, PA
##                                       3
##                 Steubenville-Weirton, OH-WV
##                                       3
##                            Stockton-Lodi, CA
##                                       4
##                                 Sumter, SC
##                                       4
##                                Syracuse, NY
##                                       3
##                                 Tacoma, WA
##                                       2
##                              Tallahassee, FL
##                                       2
##         Tampa-St. Petersburg-Clearwater, FL
##                                       1
##                              Terre Haute, IN
##                                       4
##                 Texarkana, TX-Texarkana, AR
##                                       4
##                                 Toledo, OH
##                                       2
##                                 Topeka, KS
##                                       4
```

```
##                               Trenton, NJ
##                                      3
##                                Tucson, AZ
##                                      2
##                                 Tulsa, OK
##                                      2
##                             Tuscaloosa, AL
##                                      4
##                                 Tyler, TX
##                                      4
##                             Utica-Rome, NY
##                                      3
##                  Vallejo-Fairfield-Napa, CA
##                                      4
##                                Ventura, CA
##                                      3
##                               Victoria, TX
##                                      4
##            Vineland-Millville-Bridgeton, NJ
##                                      4
##              Visalia-Tulare-Porterville, CA
##                                      4
##                                  Waco, TX
##                                      4
##                   Washington, DC-MD-VA-WV
##                                      1
##                               Waterbury, CT
##                                      3
##                   Waterloo-Cedar Falls, IA
##                                      3
##                                Wausau, WI
##                                      3
##              West Palm Beach-Boca Raton, FL
##                                      2
##                             Wheeling, WV-OH
##                                      3
##                                Wichita, KS
##                                      2
##                           Wichita Falls, TX
##                                      4
##                            Williamsport, PA
##                                      3
##                              Wilmington, NC
##                                      4
##                   Wilmington-Newark, DE-MD
##                                      3
##                             Worcester, MA-CT
##                                      3
##                                Yakima, WA
##                                      4
##                                  Yolo, CA
##                                      4
##                                  York, PA
##                                      3
```

```
##                         Youngstown-Warren, OH
##                                             3
##                                  Yuba City, CA
##                                             4
##                                      Yuma, AZ
##                                             4
##
## Within cluster sum of squares by cluster:
## [1]  276.3662  915.4033 1071.4458  988.2893
##  (between_SS / total_SS =  44.2 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
fviz_cluster(final,data = df_scale)
```

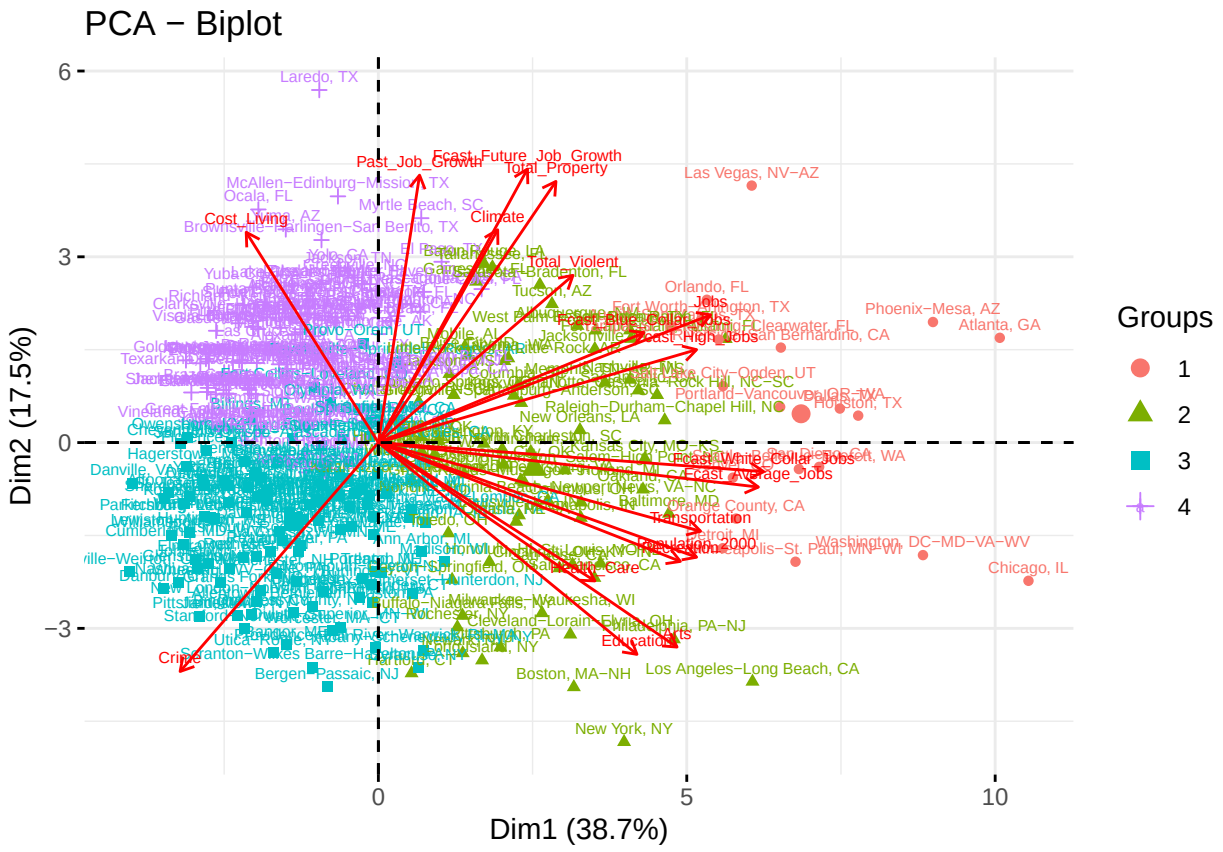



```
# add cluster number to dataframe
#
cluster_number = as.factor(final$cluster)
df$cluster = cluster_number
head(df)
```

```
##                          Cost_Living Transportation  Jobs Education Climate
## Abilene, TX                    96.32          36.54 17.28     49.29   55.52
## Akron, OH                      47.31          69.68 86.11     71.95   22.66
## Albany, GA                     86.12          28.02 32.01     26.62   75.63
```

```
## Albany-Schenectady-Troy, NY       25.22          82.71 52.97     99.43    8.78
## Albuquerque, NM                   44.48          84.13 90.65     71.67   78.18
## Alexandria, LA                    92.36          42.49 19.26     11.61   66.00
##                             Crime  Arts Health_Care Recreation Population_2000
## Abilene, TX                 49.58 27.20       45.04       2.83          123711
## Akron, OH                   54.11 81.59       24.07      77.33          689538
## Albany, GA                  15.59 33.15       20.11       6.79          120838
## Albany-Schenectady-Troy, NY 73.94 79.61       77.33      77.62          885782
## Albuquerque, NM              2.84 75.36       77.90      70.25          734255
## Alexandria, LA               7.09 40.80       62.03      22.66          127635
##                             Total_Violent Total_Property Past_Job_Growth
## Abilene, TX                           582           4396             6.1
## Akron, OH                             518           4527            11.6
## Albany, GA                            761           7036             6.6
## Albany-Schenectady-Troy, NY           365           3531             5.2
## Albuquerque, NM                      1133           7261            21.4
## Alexandria, LA                       1100           5581             6.1
##                             Fcast_Future_Job_Growth Fcast_Blue_Collar_Jobs
## Abilene, TX                                     4.5                    345
## Akron, OH                                       7.5                   6363
## Albany, GA                                      7.0                    460
## Albany-Schenectady-Troy, NY                     3.7                     88
## Albuquerque, NM                                 8.1                   5846
## Alexandria, LA                                  4.8                    627
##                             Fcast_White_Collar_Jobs Fcast_High_Jobs
## Abilene, TX                                    3385             641
## Akron, OH                                     24773            6728
## Albany, GA                                     4852             473
## Albany-Schenectady-Troy, NY                   20043            1325
## Albuquerque, NM                               30750            8134
## Alexandria, LA                                 2687             768
##                             Fcast_Average_Jobs cluster
## Abilene, TX                               2234       4
## Akron, OH                                18638       2
## Albany, GA                                4603       4
## Albany-Schenectady-Troy, NY              15233       3
## Albuquerque, NM                          20712       2
## Alexandria, LA                            1879       4
```

```r
# biplot with clusters
#
m3 = prcomp(df_scale, scale=T)
fviz_pca_biplot(m3,labelsize = 2,col.var = "red",
                habillage = cluster_number)
```



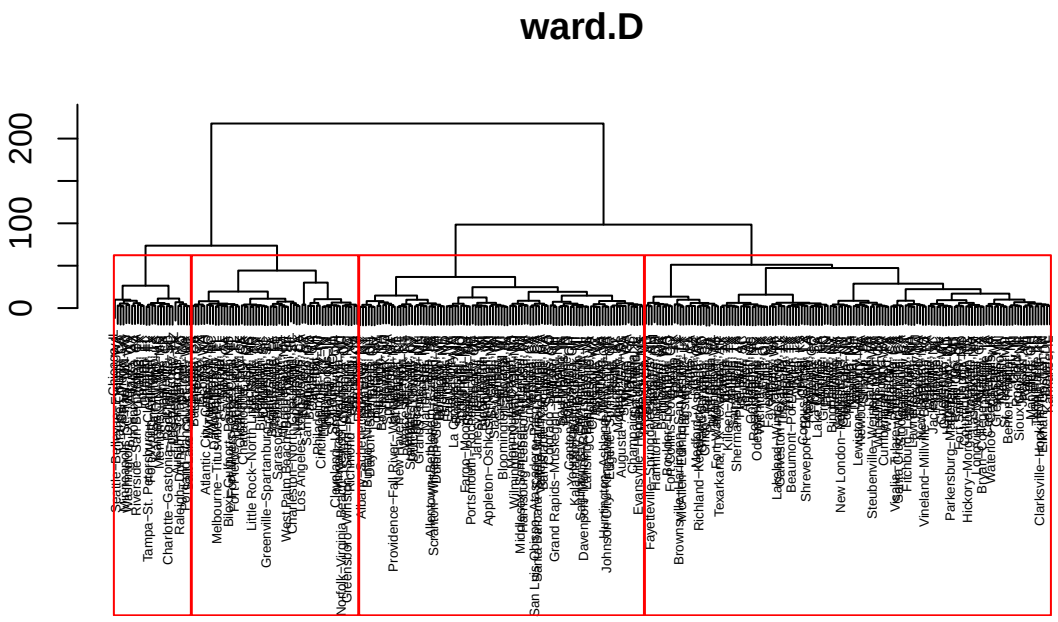

```
# For my choice of K clusters, find the median (or mean, if you prefer) of each numerical column (on th

df = lapply(df,as.numeric)
aggregate(df, list(cluster_number), median)
```

```
##   Group.1 Cost_Living Transportation   Jobs Education Climate  Crime  Arts
## 1       1      26.920         92.065 97.305    82.005   70.82 22.665 91.65
## 2       2      47.310         80.730 81.010    80.730   64.58 27.200 80.46
## 3       3      45.615         40.785 30.450    52.830   32.15 80.315 51.28
## 4       4      76.070         30.730 43.760    24.215   67.13 31.305 23.94
##   Health_Care Recreation Population_2000 Total_Violent Total_Property
## 1      65.290     90.365       2818808.5           753         5878.5
## 2      76.480     78.750       1059044.0           696         5436.0
## 3      43.055     46.880        227733.5           273         3645.0
## 4      29.175     23.790        179977.5           653         5472.0
##   Past_Job_Growth Fcast_Future_Job_Growth Fcast_Blue_Collar_Jobs
## 1            15.6                    8.3                20447.5
## 2            10.9                    5.9                 3388.0
## 3             8.3                    4.8                  436.0
## 4            11.9                    6.0                  797.5
##   Fcast_White_Collar_Jobs Fcast_High_Jobs Fcast_Average_Jobs cluster
## 1                119533.5         23248.0             83826.0       1
## 2                 33198.0          4976.0             23990.0       2
## 3                  6518.5           796.5              4489.5       3
## 4                  6020.0          1367.5              3721.0       4
```

```
## Group (cluster) 1 & 2 have high rate on every category except cost of living and crime
## Group (cluster) 3 has high crime
```

```
## Group (cluster) 4 has high rates on cost of living
```

```
# Use function hclust with linkage ward.D to create object h1 and display the four clusters on the dend

# Find distances
distance = dist(df_scale)

# Dendrogram - Ward
h1 = hclust(distance, method = 'ward.D')
plot(h1, cex = 0.4, xlab = '', main = 'ward.D', ylab = '')
rect.hclust(h1, k = 4, border = 'red')
```

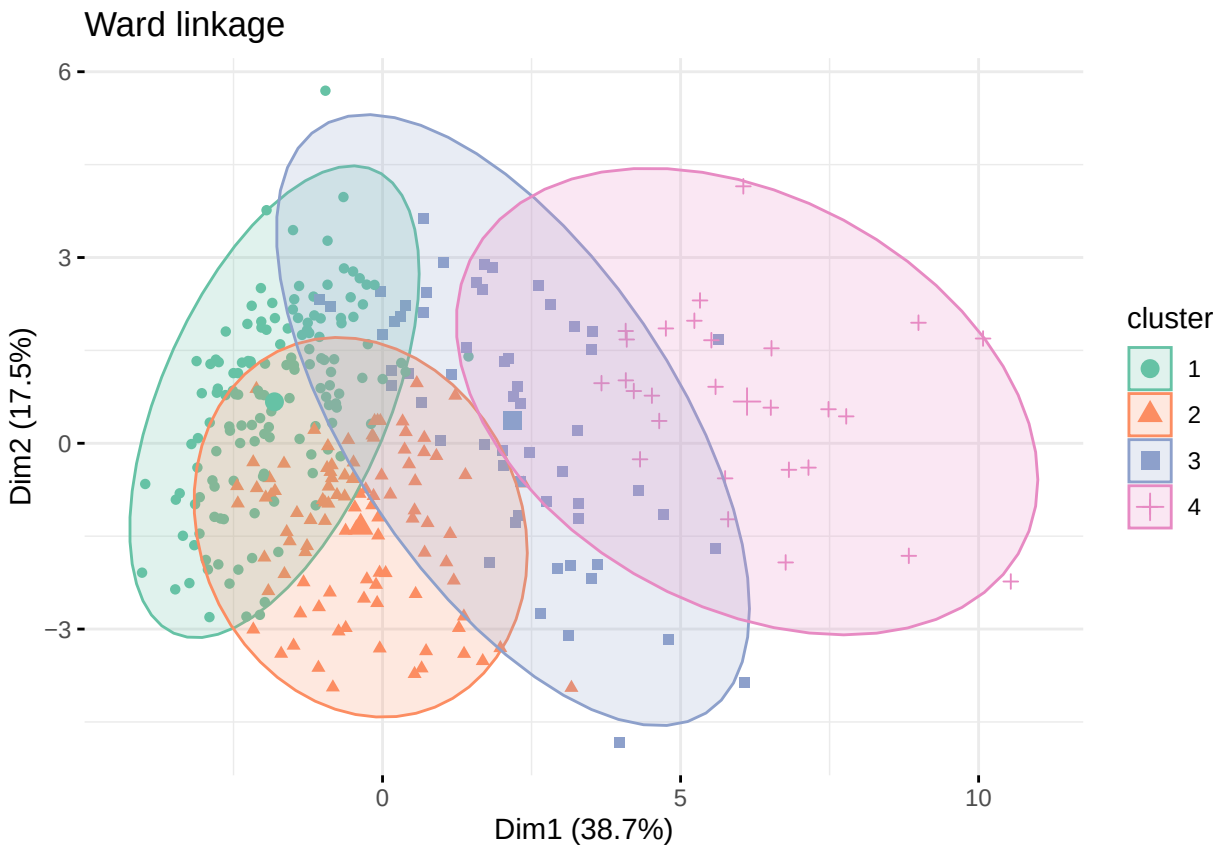



```
# CUT the dendrograms to 4 clusters
cut1 = cutree(h1,k=4)
# dataframe with cluster numbers
df1 = data.frame(df,cluster = cut1)
# number of members per cluster
table(cut1)
```

```
## cut1
##   1   2   3   4
## 141  99  58  27
```

```
# library factoextra
library(factoextra)

# cluster plots for Ward linkage
fviz_cluster(list(data = df_scale, cluster = cut1),geom = 'point', main="Ward linkage", ellipse.type =
             palette = "Set2",
```

```
            ggtheme = theme_minimal()
            )
```




```
# CPCC = correlation (euclidean distances, cophenetic distances)
c1 = cophenetic(h1)
cor(distance, c1)
```

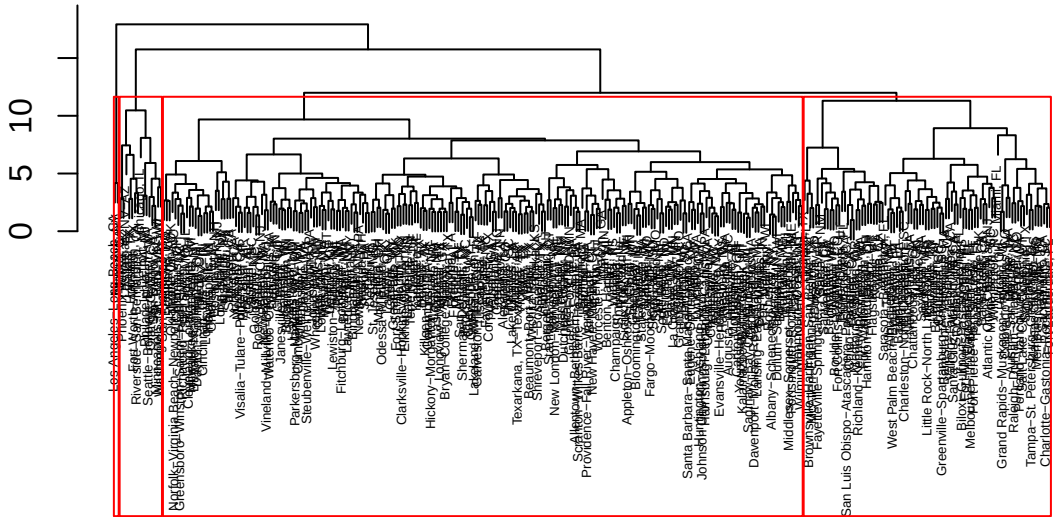
```
## [1] 0.5079247
```

```
# Use function hclust with linkage complete to create object h2 and display the four clusters on the de

h2 = hclust(distance, method ='complete')
cut2 = cutree(h2,k=4)
# number of members per cluster
table(cut2)
```

```
## cut2
##   1   2   3   4
## 222  86  15   2
```

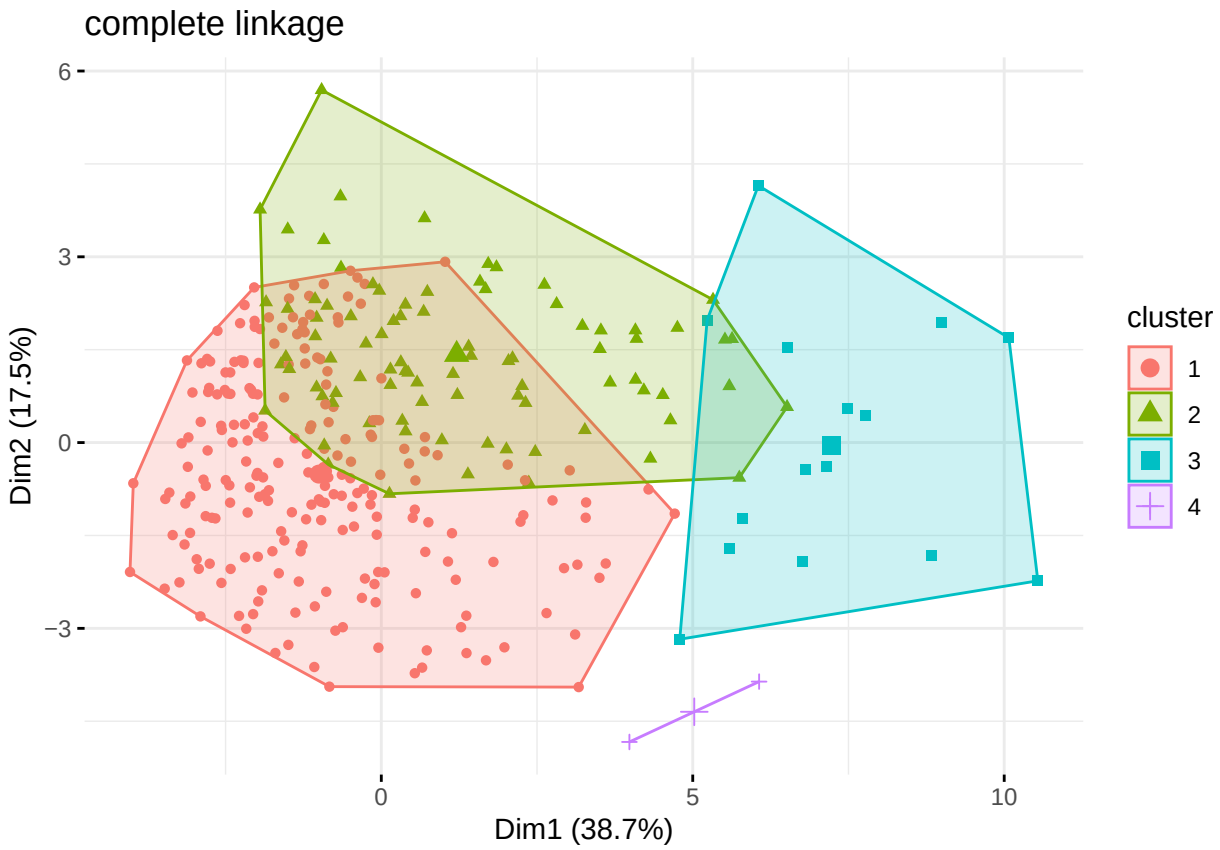
```
# dendrogram - complete linkage
plot(h2, cex = 0.4, xlab = '', main = 'Complete', ylab = '')
rect.hclust(h2, k = 4, border = 'red')
```

Complete



hclust (*, "complete")

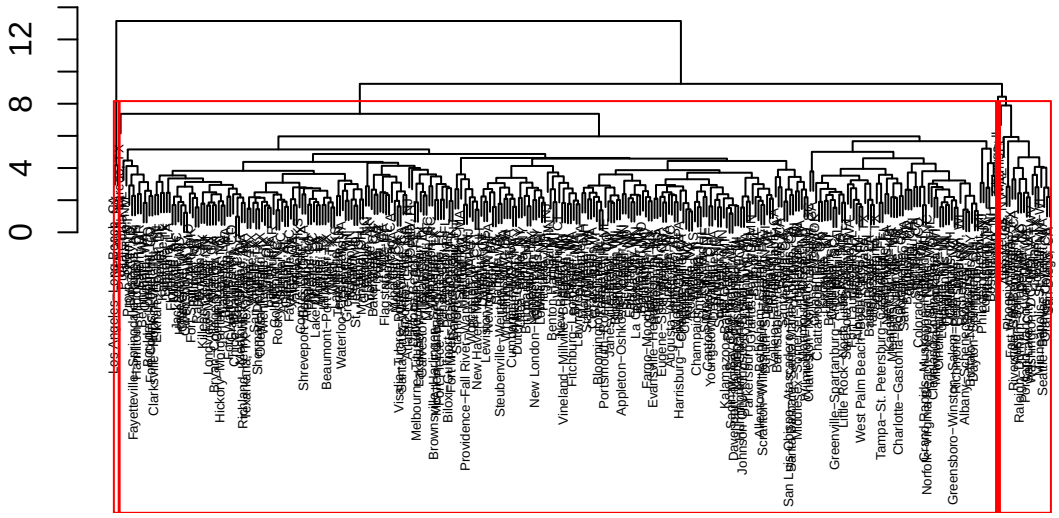
```
# cluster plots for complete linkage
fviz_cluster(list(data = df_scale, cluster = cut2), geom = 'point',
             main = 'complete linkage',
             palatte = 'Set2', ggtheme = theme_minimal())
```




```
# COPHENETIC distances
c2 = cophenetic(h2)
# CPCC = correlation (euclidean distances, cophenetic distances)
cor(distance,c2)
```

```
## [1] 0.6848473
```

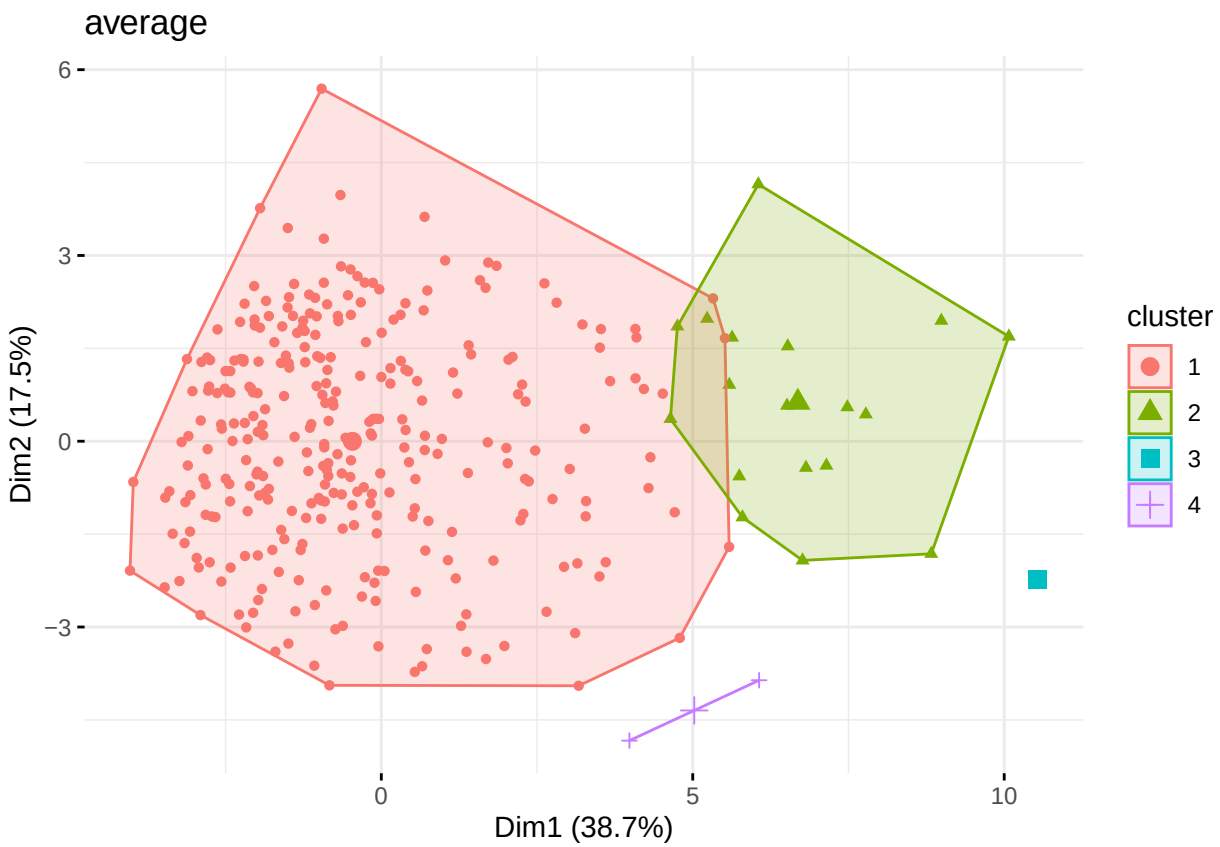
```
# Use function hclust with linkage average to create object h3 and display the four clusters on the den

# dendrogram - complete linkage
h3 = hclust(distance, method = 'average')
plot(h3, cex = 0.4, xlab = '', main = 'average', ylab = '')
rect.hclust(h3, k = 4, border = 'red')
```




```
# number of members per cluster
cut3 = cutree(h3, k = 4)
table(cut3)
```

```
## cut3
##   1   2   3   4
## 304  18   1   2
```

```
# cluster plots for average linkage
fviz_cluster(list(data = df_scale, cluster = cut3), geom = 'point',
             main = 'average', palatte = 'Set2', ggtheme = theme_minimal())
```




```
# CPCC = correlation (euclidean distances, cophenetic distances)
c3 = cophenetic(h3)
cor(distance, c3)
```

```
## [1] 0.8047003
```

```
# What linkage prefer? For the clusters found for this linkage find the median (or mean, if you prefer)

print('I perfer ward.D because the corrlation is the smallest.')
```

```
## [1] "I perfer ward.D because the corrlation is the smallest."
```

```
aggregate(df, list(cut1), median)
```

```
##   Group.1 Cost_Living Transportation   Jobs Education Climate Crime   Arts
## 1       1       69.13         27.190 32.010    26.060  52.690 51.00 24.930
## 2       2       45.90         55.800 49.570    67.130  32.570 70.26 61.480
## 3       3       51.14         75.635 79.595    70.105  74.925 20.40 73.235
## 4       4       33.43         91.500 96.030    82.710  70.820 20.68 88.390
##   Health_Care Recreation Population_2000 Total_Violent Total_Property
## 1      29.740      22.37          164563         533.0         4937.0
## 2      63.450      56.94          333180         344.0         4095.0
## 3      67.275      82.29          759293         775.5         5808.5
## 4      67.700      86.68         1888819         763.0         6102.0
##   Past_Job_Growth Fcast_Future_Job_Growth Fcast_Blue_Collar_Jobs
## 1            10.1                    4.90                   538
## 2             9.1                    5.40                   729
## 3            11.4                    6.35                  2809
## 4            15.2                    8.10                 17876
```

```
##   Fcast_White_Collar_Jobs Fcast_High_Jobs Fcast_Average_Jobs cluster
## 1                    4874             939             3105.0       4
## 2                    9733            1464             7302.0       3
## 3                   29210            4397            20133.5       2
## 4                   94192           19962            70187.0       1
```

Group 1 have high rate on cost of living, but low on Health_Care and Recreation, maybe not a good choice if people want to relocate to area in this group.

Group 2 have low rate on Cost_Living, Total_Property, but high rate on crime, indicating thus area is not safe, and very poor.

Group 3 has high rate on climate, health care and total violent, but low on crime.

Group 4 has high rates on education, arts, health care and recreation, population_2000, Past_Job_Growth, and Fcast_Future_Job_Growth, indicating group 4 as very suited for living. People there should live happily.