

Assingment 1

Jiseong Yang

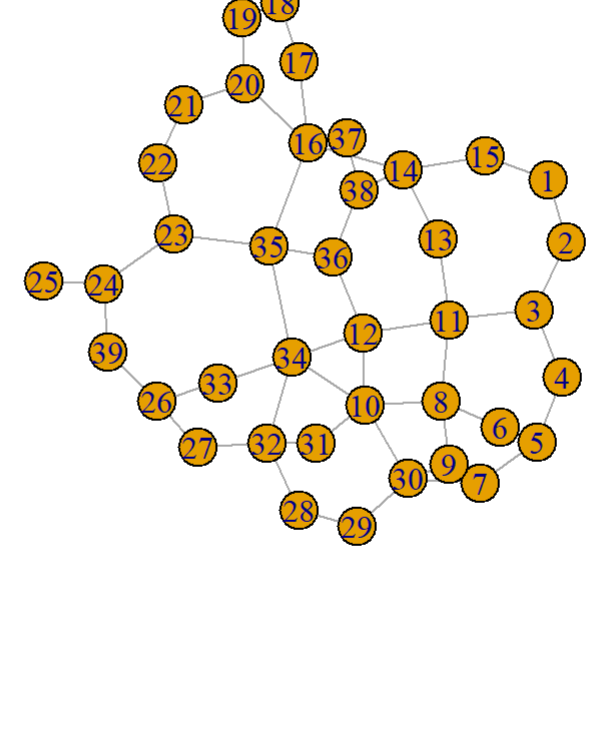
2018년 12월 13일

1. 사회연결망 분석

1. *igraph package*를 이용한 네트워크 그래프

- 그래프

Medieval Russian Trade Network



- 지역별 코드명

```
## [1] "1 : Novgorod"
## [1] "2 : Vitebsk"
## [1] "3 : Smolensk"
## [1] "4 : Kiev"
## [1] "5 : Chernikov"
## [1] "6 : Novgorod_Serverskiy"
## [1] "7 : Kursk"
## [1] "8 : Bryansk"
## [1] "9 : Kazachev"
## [1] "10 : Kozelsk"
## [1] "11 : Dorogobusch"
## [1] "12 : Vyazna"
## [1] "13 : A_region"
## [1] "14 : Tver"
## [1] "15 : Vishniy_Totochek"
## [1] "16 : Kayatyn"
## [1] "17 : Uglich"
## [1] "18 : Yaroslavl"
## [1] "19 : Rostov"
## [1] "20 : B_region"
## [1] "21 : C_region"
## [1] "22 : Suzdal"
## [1] "23 : Vladimir"
## [1] "24 : Nizhniy_Novgorod"
## [1] "25 : Bolgar"
## [1] "26 : Isad_Ryazan"
## [1] "27 : Pronsk"
## [1] "28 : Dubok"
## [1] "29 : Elest"
## [1] "30 : Mtsensk"
## [1] "31 : Tula"
## [1] "32 : Dedoslavl"
## [1] "33 : Pereslavl"
## [1] "34 : Kolonna"
## [1] "35 : Moscow"
## [1] "36 : Moshaysk"
## [1] "37 : Dmitrov"
## [1] "38 : Volok_Lamskiy"
## [1] "39 : Murom"
```

2. 연결중심성(degree centrality), 근접 중심성(closeness centrality), 매개 중심성(betweenness centrality), 아이겐벡터 중심성(Eigenvector centrality)산출을 통한 당시 교역의 중심지 제시

1. 연결 중심성

- 연결 중심성 최고 지역

```
##      region_name res
## 1      Kozelsk     5
## 2      Kayatyn     5
## 3      Kolonna     5
```

- 최고 연결중심성: 5

2. 근접 중심성

- 근접 중심성 최고 지역

```
##      region_name      res
## 1      Kolonna 0.3653846
```

- 최고 근접중심성: 0.3654 (소숫점 넷째자리까지 반올림)

3. 매개 중심성

- 매개 중심성 최고 지역

```
##      region_name      res
## 1      Kolonna 246.8524
```

- 최고 매개중심성: 246.8524 (소숫점 넷째자리까지 반올림)

4. 아이겐벡터 중심성

- 아이겐벡터 중심성 최고 지역

```
##      region_name vector
## 1      Kozelsk         1
```

- 최고 아이겐벡터 중심성: 1

5. 교역의 중심지

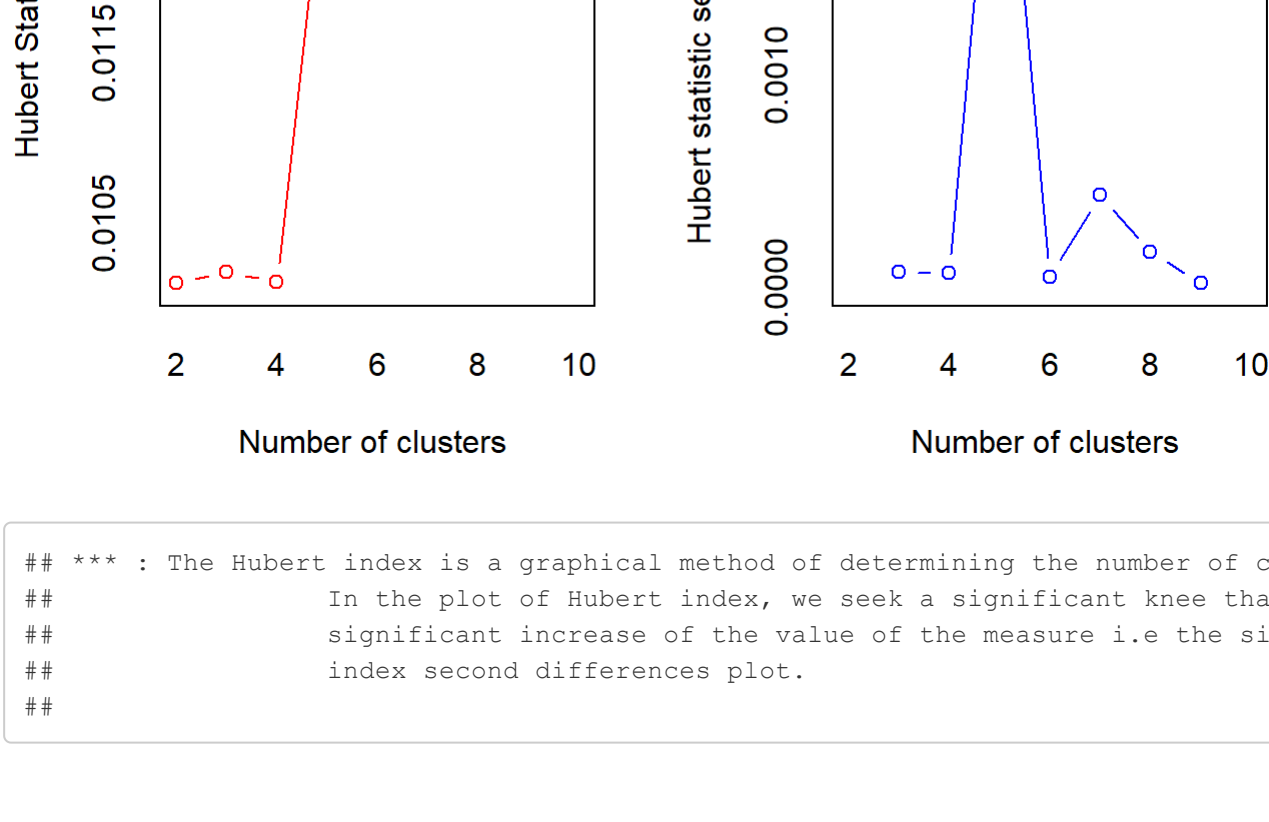
- 별도의 가중치는 부여되지 않았다. 주어진 정보 중 지도에 나타난 거리 차이를 가중치로 삼을 수도 있겠으나, 이는 거리가 짧은 교역량이 늘어난다는 것을 전제로 한다. 그러나 다른 변수들의 개입으로 인하여 꼭 그렇지 않을 수도 있기 때문에 가중치를 부여하지 않는 편이 bias의 개입을 최소화할 수 있다고 판단하였기 때문이다.
- 주어진 중심성을 모두 고려하였을 때, 연결중심성, 근접중심성, 매개중심성이 제일 높은 Kolonna 지역이 당시 무역의 중심지였던 것으로 보인다.
- 한편, Kozelsk 지역도 Kolonna 지역과 동일하게 높은 연결중심성과 가장 높은 아이겐벡터 중심성을 보이고 있는데, 이는 Kozelsk 지역과 연결된 지역들이 중요한 도시였음을 의미한다. 연결된 도시들은 Bryansk, Vyazna, Mtsensk, Kolonna인데, 이 중에서 Kolonna 지역이 포함되어 있음을 알 수 있다.

2. 군집분석

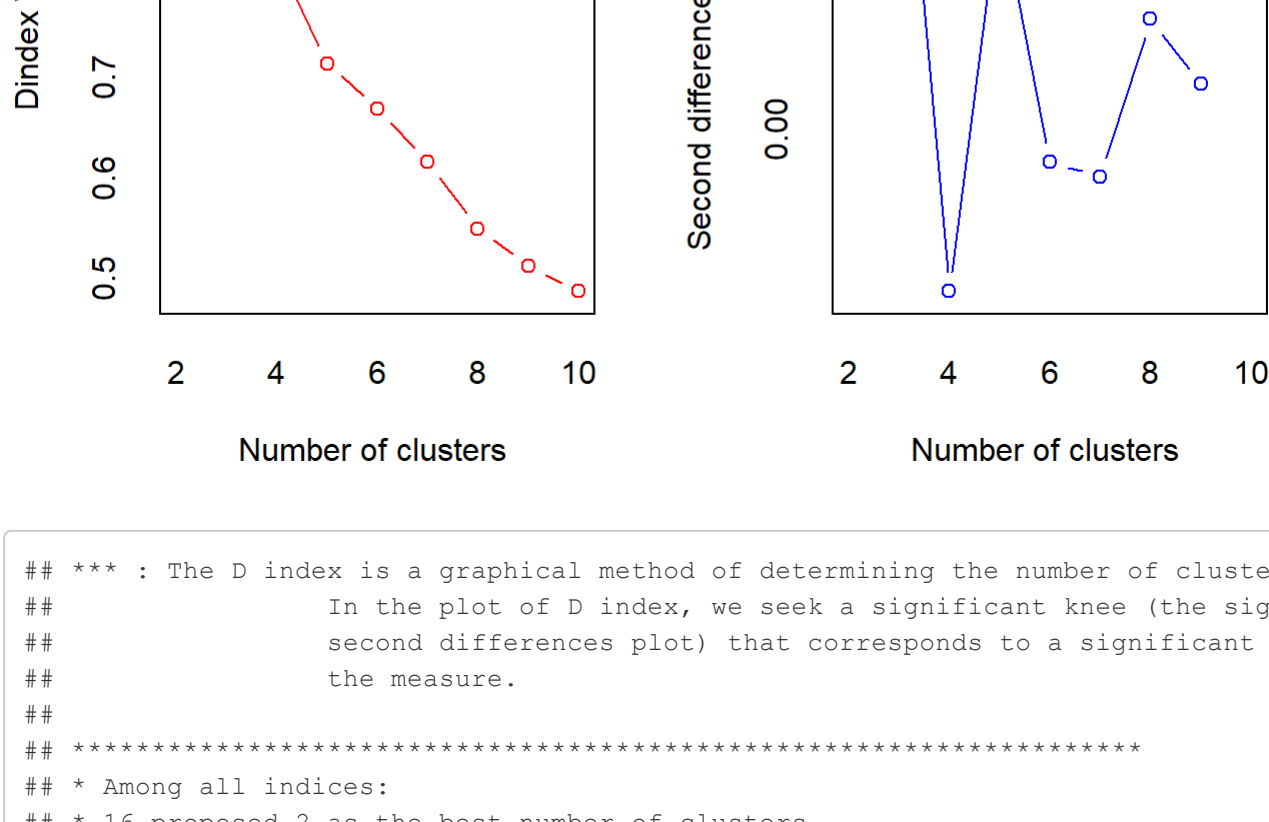
1. 연결망분석을 통해 도출된 4개의 지표를 변수로 하는 데이터 set을 만들고, K-means 방법을 이용한 군집분석. "NbClust Package"를 이용해 각자 최적의 군집의 수를 결정하고 그 이유 설명.

- NbCluster를 이용한 최적의 군집개수 찾기

```
set.seed(123)
NbClust(set_centr_scaled, min.nc=2, max.nc=10, index="alllong", method="kmeans")
```



```
## *** : The Hubert index is a graphical method of determining the number of clusters.
##
##      In the plot of Hubert index, we seek a significant knee that corresponds to a
##      significant increase of the value of the measure i.e the significant peak in Hubert
##      index second differences plot.
##
```



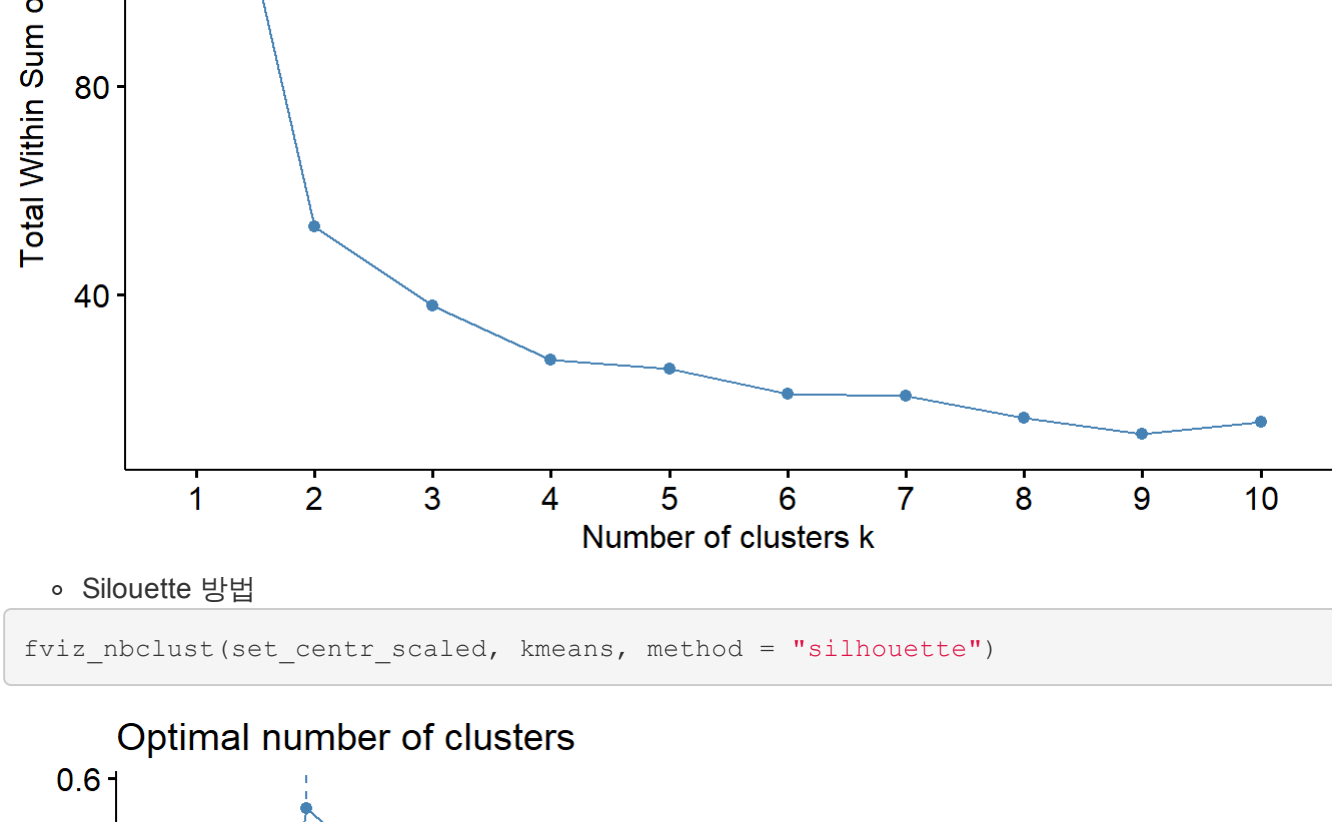
```
## *** : The D index is a graphical method of determining the number of clusters.
##
##      In the plot of D index, we seek a significant knee (the significant peak in Dindex
##      second differences plot) that corresponds to a significant increase of the value of
##      the measure.
##
## *****
## * Among all indices:
## * 16 proposed 2 as the best number of clusters
## * 3 proposed 3 as the best number of clusters
## * 3 proposed 4 as the best number of clusters
## * 1 proposed 5 as the best number of clusters
## * 1 proposed 7 as the best number of clusters
## * 4 proposed 9 as the best number of clusters
##
## ***** Conclusion *****
##
## * According to the majority rule, the best number of clusters is  2
##
## *****
```

```
## $All.index
##      KL      CH Hartigan      CCC      Scott      Marriot      TrCovW      TraceW
## 2  16.0604 68.8647  11.3990 -0.0932  58.2099 28094.244  73.4563 53.1244
## 3  1.9994 49.3684   4.2011 -1.7312  92.6547 26135.527  73.7457 40.6125
## 4  0.1535 32.0536  14.8113 -3.4421 125.9621 19779.045 59.0417 36.3684
## 5 10.1691 42.0590   4.2723 -1.8671 147.2039 17925.976 52.4591 25.5543
## 6  0.2187 37.5905   7.7884 -2.2762 163.3861 17046.845 38.1714 22.7017
## 7  1.4982 38.8040   5.8563 -1.7279 197.5436  9664.242 25.1086 18.3669
## 8  2.4051 38.9284   3.4054 -1.4846 214.5164  8168.587 12.1264 15.5256
## 9  3.0601 36.9966  1.1507 -1.7215 233.1109  6417.827 11.2791 13.9889
## 10 0.3633 33.1326  2.7190 -2.4607 238.3085  6934.639  9.8447 13.4722
##      Friedman      Rubin Cindex      DB Silhouette      Duda Pseudo2      Beale
## 2  15.8679  2.8612  0.3149  0.7169   0.5762  1.0736  -1.7146 -0.1490
## 3  22.8310  3.7427  0.2881  1.0246   0.3669  0.7347   3.9721  0.7925
## 4  30.5245  4.1795  0.2685  1.1257   0.3552  0.5618   7.0211  1.6741
## 5  34.3687  5.9481  0.3627  1.2797   0.2559  0.5929   4.8073  1.4507
## 6  37.9702  6.6955  0.3341  1.1535   0.2801  1.1636  -0.9844 -0.2263
## 7  47.1470  8.2758  0.3179  1.0927   0.2799  0.6465   6.0158  1.1736
## 8  48.6253  9.7903  0.2716  1.0195   0.3202  1.0809  -0.6735 -0.1355
## 9  58.2304 10.8658  0.2670  0.9026   0.3321  1.6543  -3.5597 -0.8185
## 10 59.5091 11.2825  0.2763  0.8674   0.3172  1.3758  -2.1852 -0.3297
##      Ratkowsky      Ball Ptbiserial      Gap      Frey McClain      Gamma      Gplus
## 2  0.5699 26.5622   0.7572 -0.1991  3.8847  0.2582  0.9146  7.6829
## 3  0.4948 18.5375   0.5449 -0.5607  1.6895  0.7829  0.7568 20.0256
## 4  0.4355  9.0921   0.4993 -1.1978  1.4147  1.0278  0.7562 17.8570
## 5  0.4078  5.1109   0.4337 -1.1795  0.5619  1.5147  0.7336 15.4534
## 6  0.3764  3.7836   0.4168 -1.4631  0.1045  1.7097  0.7606 12.3576
## 7  0.3543  2.6238   0.4217 -1.7316  0.4526  1.7004  0.8049  9.5439
## 8  0.3350  1.9407   0.3957 -1.8236  1.1682  1.9638  0.8410  6.4049
## 9  0.3176  1.5543   0.3859 -1.8398  1.9718  2.0711  0.8415  6.0850
## 10 0.3019  1.3472   0.3424 -1.8068  0.3089  2.6871  0.8022  6.4291
##      Tau      Dunn      Hubert      SIndex      Dindex      SDbw
## 2 164.6127 0.2500 0.0101  3.0154 1.0460 0.4757
## 3 124.6154 0.2094 0.0102  3.3940 0.8796 0.3772
## 4 110.7638 0.2239 0.0102  3.5044 0.8176 0.2807
## 5  85.1093 0.1072 0.0128  4.5272 0.7134 0.2973
## 6  78.5344 0.1072 0.0128  4.5989 0.6686 0.2642
## 7  78.7530 0.1120 0.0132  3.9646 0.6151 0.1904
## 8  67.7679 0.1024 0.0134  4.1355 0.5493 0.1695
## 9  64.6154 0.1024 0.0134  4.1582 0.5122 0.1589
## 10 52.1457 0.1024 0.0135  5.7593 0.4870 0.1626
##
## $All.CriticalValues
##      CritValue_Duda CritValue_Pseudo2 Pvalue_Beale CritValue_Gap
## 2      0.2316      82.9272      1.0000      0.3951
## 3      0.2576      31.7063      0.5370      0.6704
## 4      0.2019      35.5759      0.1802      0.0220
## 5      0.1673      34.8308      0.2436      0.3343
## 6     -0.1694     -48.3196      1.0000      0.3341
## 7      0.2019      43.4816      0.3409      0.1609
## 8     -0.0628     -152.4134  1.0000      0.0879
## 9      0.1265      62.1541  1.0000      0.0386
## 10     -0.3257     -32.5626  1.0000      0.1978
##
## $Best.nc
##
##      Value_Index      KL      CH Hartigan      CCC      Scott      Marriot      TrCovW
## 2 16.0604 68.8647  10.6102 -0.0932 34.4448 5886.948 14.704
##      Value_Index      TraceW      Friedman      Rubin Cindex      DB Silhouette      Duda
## 2 3.0000 3.0000  9.0000 5.0000  9.000 2.0000  2.0000 2.0000
##      Value_Index      Pseudo2      Beale      Ratkowsky      Ball Ptbiserial      Gap
## 2 8.2678  9.6051 -1.0212  0.267 0.7169   0.5762  1.0736
##      Value_Index      Dunn      Hubert      SIndex      Dindex      SDbw
## 2 2.0000 2.0000  2.0000 3.0000 3.0000 2.0000 2.0000
##      Value_Index      Frey McClain      Gamma      Gplus      Tau      Dunn      Hubert      SIndex
## 2 1.4147  0.2582 0.9146 6.085 164.6127 0.25  0  3.0154
##      Value_Index      Dindex      SDbw
## 2 0.0 0.1589
##
## $Best.partition
##
## [1] 2 2 2 2 2 2 2 1 1 1 1 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 1 1
## [36] 1 2 2 2
```

- factoextra 패키지를 활용한 군집 찾기

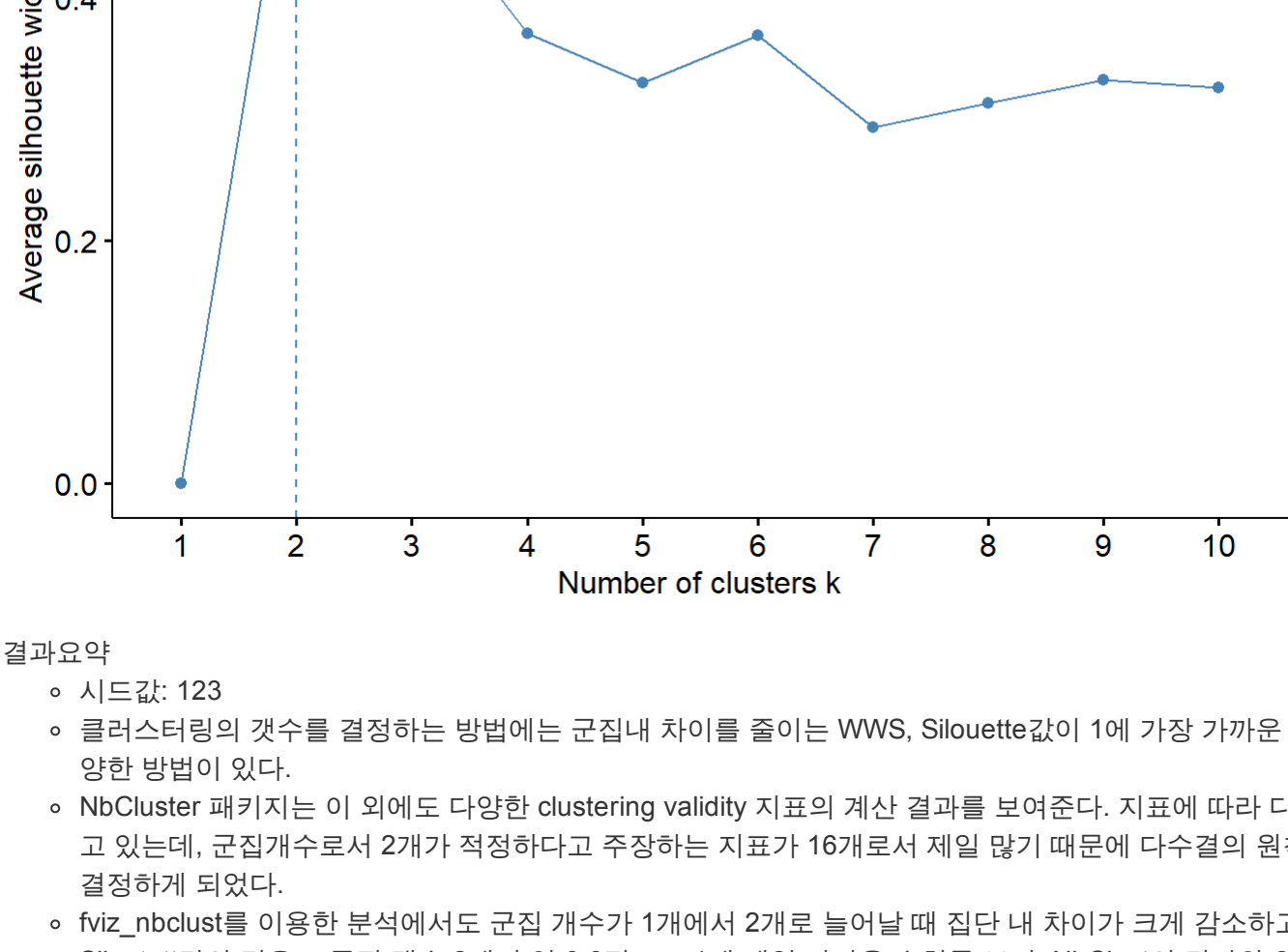
- WSS 방법

```
fvsiz_nbclust(set_centr_scaled, kmeans, method = "wss")
```



- Silhouette 방법

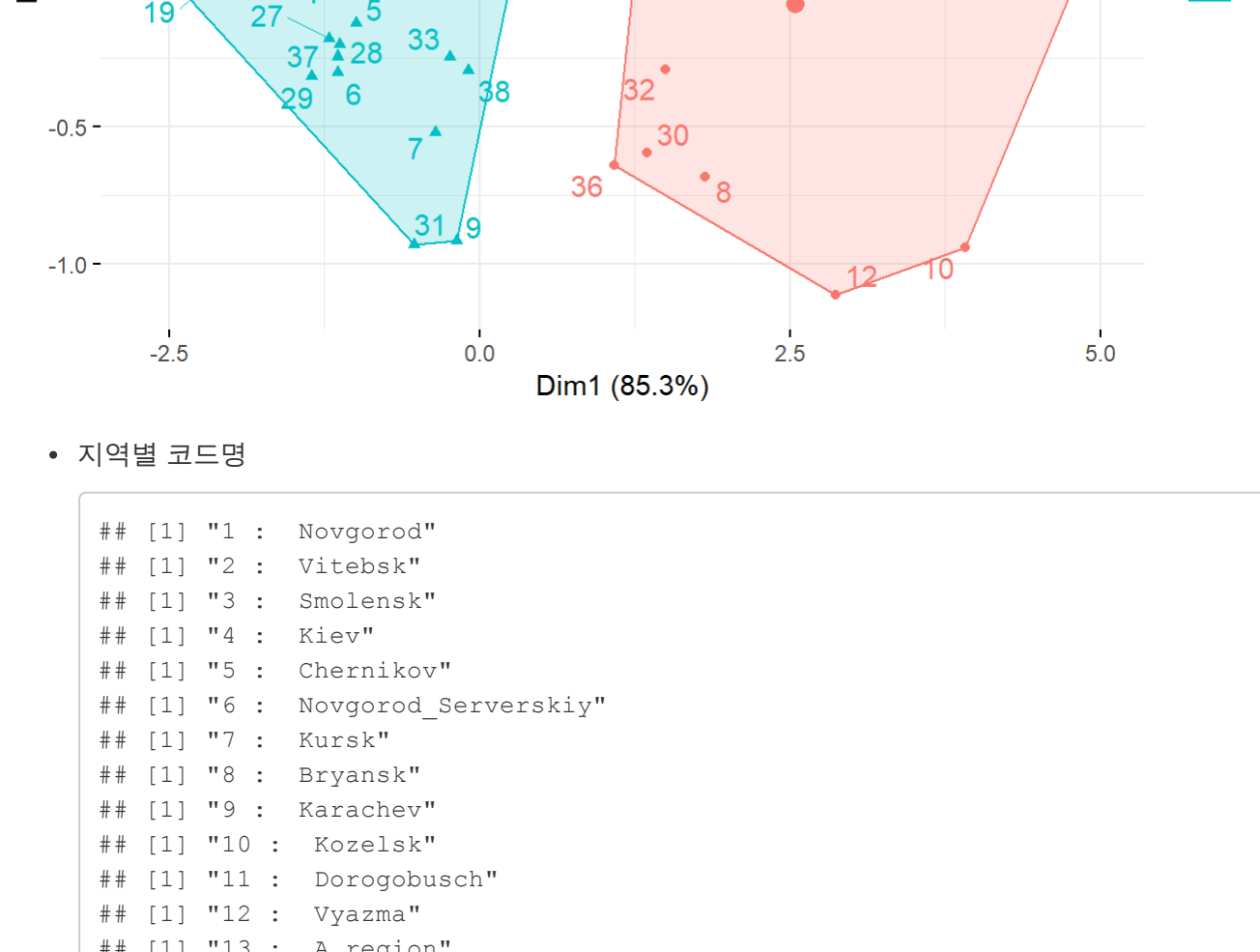
```
fvsiz_nbclust(set_centr_scaled, kmeans, method = "silhouette")
```



- 결과요약

- 시드값: 123
- 클러스터링의 결과를 결정하는 방법에는 군집내 차이를 줄이는 WWS, Silhouette값이 1에 가장 가까운 클러스터 개수를 선택하는 등 다양한 방법이 있다.
- NbCluster 패키지는 이 외에도 다양한 clustering 지표의 계산 결과를 보여준다. 지표에 따라 다양한 최적 클러스터 값을 제시하고 있는데, 군집개수가 2개 이상 정정하고 주장하는 지표가 16개로서 제일 많기 때문에 다수결의 원칙에 따라 클러스터 수를 2개로 결정하게 되었다.
- fvsiz_nbclust를 이용한 분석에서도 군집 개수가 1개에서 2개로 늘어날 때 집단 내 차이가 크게 감소하고 이후부터는 차이가 크지 않고, Silouett값의 경우도 군집 개수 2에서 약 0.6정도로 10에 제일 가까운 수치를 보아, NbClust의 결과와 일치한다.

2. 위의 과정을 통해 도출된 군집분석 결과를 그래프로 시각화



- 지역별 코드명

```
## [1] "1 : Novgorod"
## [1] "2 : Vitebsk"
## [1] "3 : Smolensk"
## [1] "4 : Kiev"
## [1] "5 : Chernikov"
## [1] "6 : Novgorod_Serverskiy"
## [1] "7 : Kursk"
## [1] "8 : Bryansk"
## [1] "9 : Kazachev"
## [1] "10 : Kozelsk"
## [1] "11 : Dorogobusch"
## [1] "12 : Vyazna"
## [1] "13 : A_region"
## [1] "14 : Tver"
## [1] "15 : Vishniy_Totochek"
## [1] "16 : Kayatyn"
## [1] "17 : Uglich"
## [1] "18 : Yaroslavl"
## [1] "19 : Rostov"
## [1] "20 : B_region"
## [1] "21 : C_region"
## [1] "22 : Suzdal"
## [1] "23 : Vladimir"
## [1] "24 : Nizhniy_Novgorod"
## [1] "25 : Bolgar"
## [1] "26 : Isad_Ryazan"
## [1] "27 : Pronsk"
## [1] "28 : Dubok"
## [1] "29 : Elest"
## [1] "30 : Mtsensk"
## [1] "31 : Tula"
## [1] "32 : Dedoslavl"
## [1] "33 : Pereslavl"
## [1] "34 : Kolonna"
## [1] "35 : Moscow"
## [1] "36 : Moshaysk"
## [1] "37 : Dmitrov"
## [1] "38 : Volok_Lamskiy"
## [1] "39 : Murom"
```