

# 조사방법론 팀 프로젝트 보고서

(국가별 COVID-19 발병률 및 사망률 현황)



6조

201512019 최호진

201711908 박상현

201811525 김승호

201812045 양우성

201911826 박종우

# 목 차

I. 서	론	.....	3
1. 문제 정의	.....		3
II. 본	론	.....	5
1. 데이터 선정	.....		5
2. 최적 군집 수 (K)	.....		5
3. K-means 군집분석	.....		5
4. Clusplot	.....		6
III. 결	론	.....	7
1. 결과 해석 및 시사점	.....		7
2. 한계점	.....		8
IV. 부	록	.....	9

# I. 서론

## 1. 문제 정의

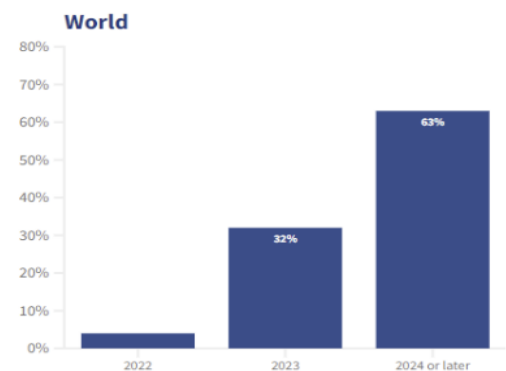
세계 각국에서는 위드코로나(with corona)가 시행되고 있다. 위드코로나란 코로나의 완전한 종식을 기대하는 것보다는 방역체계 등을 통해 코로나와의 공존을 준비한다는 뜻으로 확진자 수를 억제하는 것보다는 치사율을 줄이는 새로운 방역체계의 전환이라고 볼 수 있다. 사회적 거리두기 조치가 시행된 지 약 2년 1개월 뒤인 4월 18일부터 모든 조치가 해제되었다. 4월 25일부터 실내 취식이 가능해지게 되고 논문을 작성한 현 시점에서는 외부에서 마스크를 착용하지 않아도 된다. 한국도 세계의 추세를 맞추기 위해 노력하고 있는 것으로 판단된다.

사회적 거리두기가 해제되고 위드코로나 체제로 가면서 외부활동이 증가하고 있다. 배달에 의존하는 사람들이 최근 식당의 이용이 증가하였고, 재택근무를 선호하던 직장인들이 다시 회사에서 근무하는 등 코로나 이전의 일상적인 사회로 돌아가고 있다는 것을 알 수 있다.

세계 여행 시장을 다루는 각 기관들은 2022년부터 여행 시장의 회복이 본격적인 가속화가 시작하고 2024년부터는 완전히 회복할 것이라고 예상했다.

### a. 세계 관광 기구(UNWTO)

UNWTO는 올해 3분기부터 회복이 시작될 것이라고 예측했다. 코로나가 최초로 발생했던 2020년에 국제 관광객 수는 이전보다 73%나 감소하였다. 2021년에는 글로벌 관광산업이 2020년보다 4% 증가했지만 여전히 국제 관광객 수는 이전보다 72% 감소한 수치를 보여주어 여행 시장이 많은 어려움을 겪게 된다. 하지만 올해부터 여행 시장이 반등될 것이라 하였고 2024년부터는 코로나가 발생한 이전인 2019년의 수준으로 돌아올 것이라고 예상했다.



<출처 : UNWTO 2022-01-13 UN report Underscores Importance of Tourism for Economic Recovery in 2022 >

### b. 국제항공운송협회 (IATA)

IATA 역시 2022년부터 빠른 회복을 필두로 2024년에 완전히 회복될 것이라 예상했다. IATA는 대륙별로 통계치를 작성하여 각각의 전망을 자세하게 예측한 것이 특징이다.

<2019년 대비 여행자 수 예측 통계>

	2021년	2022년	2023년	2024년	2025년
업계 전반	47%	83%	94%	103%	111%
아시아 태평양	40%	68%	84%	97%	109%
유럽	40%	86%	96%	105%	111%
북아메리카	56%	94%	102%	107%	112%
아프리카	46%	76%	85%	93%	101%
중동	42%	81%	90%	98%	105%
남아메리카	51%	88%	97%	103%	108%
중앙 아메리카	72%	96%	102%	109%	115%
카리브해	44%	72%	82%	92%	101%

<출처: IATA, 2022-03-01, Air Passenger Numbers to Recover in 2024 >

### c. 세계여행관광협회(WTTC)

WTTC에서 올해의 해외여행 지출이 작년에 비해 약 94% 증가할 것이라 예상했다. 2019년에 여행&관광 부문에서 약 9조 2000억 달러를 창출하였으나, 2020년에 코로나로 인하여 약 4조 5000억 달러의 손실이 발생한다. 하지만 2022년인 올해는 8조 6000억 정도로 회복될 것이라 전망하였고 이는 2019년보다 약 6.4% 정도 뒤쳐진 정도에 그친다고 예상하였고 이를 위해서는 백신과 부스터샷 도입에 집중하고 해외 여행에 대한 제한이 완화되어야 한다고 주장하였다.

### d. 전세계 여행 시장 정보 제공 업체 (CAPA)

CAPA는 전 세계 공항, 항공사, 공급업체 등에 관한 최신 데이터를 제공하는 업체이다. CAPA에서 임원진 400명에게 설문조사한 결과 2022년 말부터 회복을 시작하여 2023년부터 여행 재개에 대한 기대가 팬데믹 이전 수준으로 증가될 것이라 예측하였다.



(출처 : 아시아 태평양 2021 여행회복보고서)

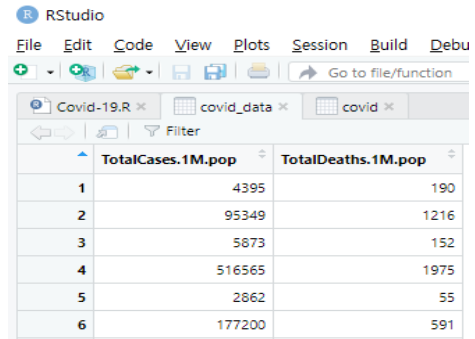
이처럼 코로나에 대한 규제가 완화되면서 세계 곳곳에서 여행자의 수가 늘어나고 있는 추세이다. 하지만 규제가 완화되었다고 해서 코로나가 완전히 종식된 것은 아니기 때문에 세계를 여행할 때 코로나에 관한 주의사항을 배제해서는 아니된다. 이번 프로젝트의 목적은 세계 각국의 코로나 발생자 수와 사망자 수가 기재되어 있는 통계치를 분석하여 안전한 세계 여행을 할 수 있도록 도움을 주기 위해서이다. R을 이용하여 진행하였고 분석방법은 K-Means 비계층적 군집분석을 채용하여 각 군집마다 어떠한 특징이 있는지 알아보았다.

※ 통계치 : <https://www.kaggle.com/datasets/imdevskp/corona-virus-report>

## II. 본론

### 1. 데이터 선정

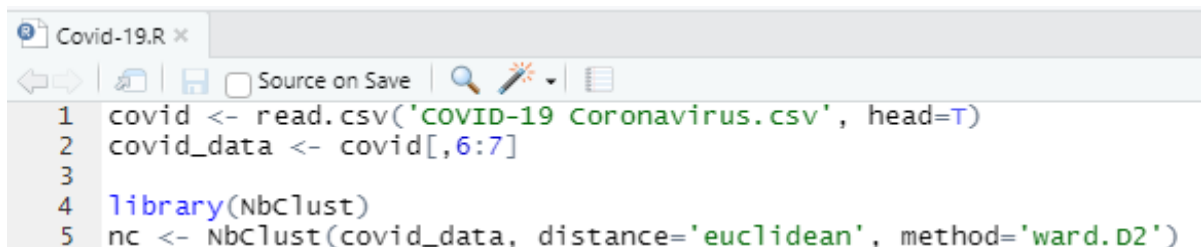
R studio를 실행하고 read.csv를 작성하여 통계치를 불러온다. 이후 사용할 열을 지정하여 사용할 데이터를 추출한다. 분석하고자 하는 데이터는 단위 당 발생자 수와 단위 당 사망자 수 (단위 : 백만 명 ) 이고 총 225개국 국가들을 조사하였다. 단위 당 발생자 수와 단위 당 사망자 수를 선정한 이유는 국가별 총 인구 수가 천차만별이고 발생자 수와 사망자 수를 비교하는 것이 국가의 안전도를 평가하는 데 가장 신빙성이 있을 것이라고 판단했기 때문이다.



	TotalCases.1M.pop	TotalDeaths.1M.pop
1	4395	190
2	95349	1216
3	5873	152
4	516565	1975
5	2862	55
6	177200	591

### 2. 최적 군집 수 (K)

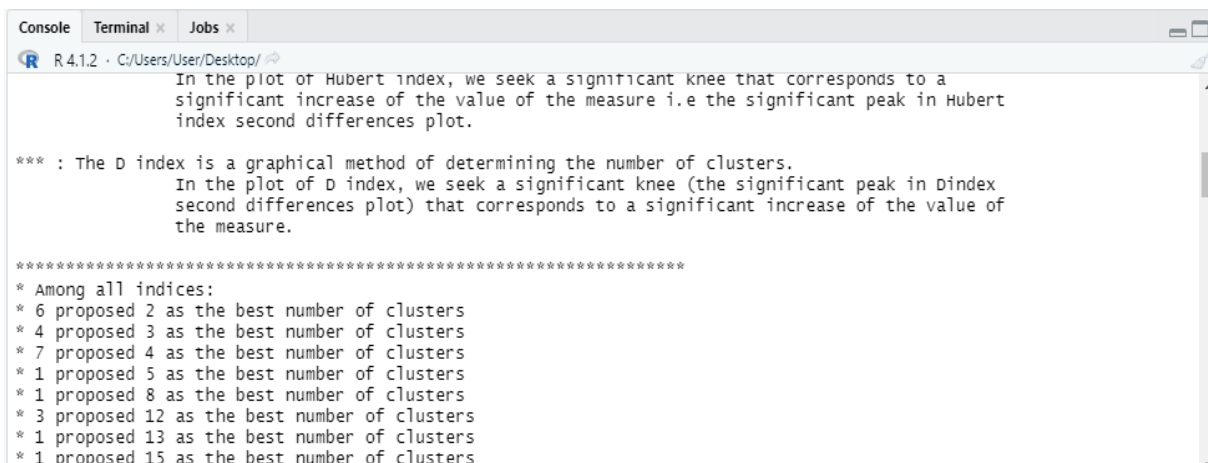
NbClust 패키지를 활용하여 분석하기 가장 이상적인 군집 수를 구해보았다. distance와



```
1 covid <- read.csv('COVID-19 Coronavirus.csv', head=T)
2 covid_data <- covid[,6:7]
3
4 library(NbClust)
5 nc <- NbClust(covid_data, distance='euclidean', method='ward.D2')
```

method는 가장 보편적으로 사용되는 euclidean과 ward.D2를 활용하였다.

가장 추천 수가 높은 군집 수 4개로 총 7표를 받았다. 따라서 최적 군집 수는 4로  $k = 4$ 로 두어 다음 단계로 진입하였다.



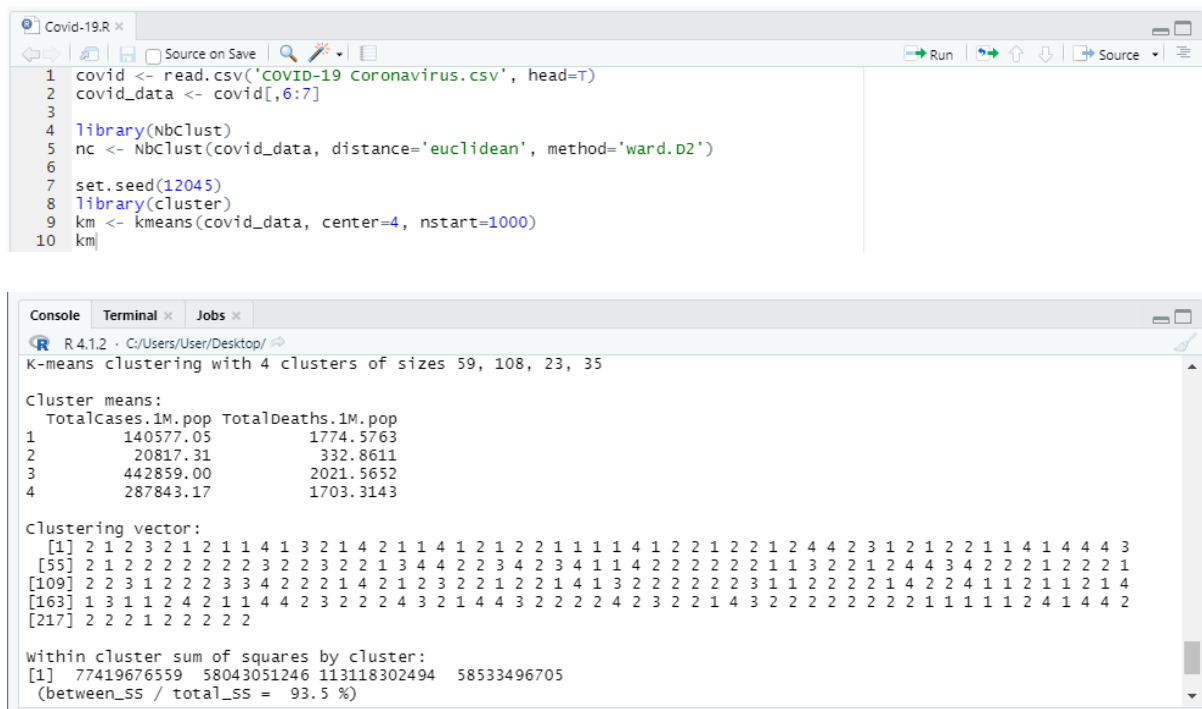
```
R 4.1.2 · C:/Users/User/Desktop/
In the plot of Hubert index, we seek a significant knee that corresponds to a
significant increase of the value of the measure i.e the significant peak in Hubert
index second differences plot.

*** : 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:
* 6 proposed 2 as the best number of clusters
* 4 proposed 3 as the best number of clusters
* 7 proposed 4 as the best number of clusters
* 1 proposed 5 as the best number of clusters
* 1 proposed 8 as the best number of clusters
* 3 proposed 12 as the best number of clusters
* 1 proposed 13 as the best number of clusters
* 1 proposed 15 as the best number of clusters
```

### 3. K-means 군집분석

k-means 함수를 사용하여 분석을 실시하였다. 함수를 적용하기 위해 cluster 패키지를 이용하였고 k = 4로 두고 iteration을 1000번 실행하였다.



```
1 covid <- read.csv('COVID-19 Coronavirus.csv', head=T)
2 covid_data <- covid[,6:7]
3
4 library(NbClust)
5 nc <- NbClust(covid_data, distance='euclidean', method='ward.D2')
6
7 set.seed(12045)
8 library(cluster)
9 km <- kmeans(covid_data, center=4, nstart=1000)
10 km
```

R 4.1.2 - C:/Users/User/Desktop/

K-means clustering with 4 clusters of sizes 59, 108, 23, 35

Cluster means:

	TotalCases.1M.pop	TotalDeaths.1M.pop
1	140577.05	1774.5763
2	20817.31	332.8611
3	442859.00	2021.5652
4	287843.17	1703.3143

Clustering vector:

```
[1] 2 1 2 3 2 1 2 1 1 4 1 3 2 1 4 2 1 1 4 1 2 1 2 2 1 1 1 1 4 1 2 2 1 2 2 1 2 4 4 2 3 1 2 1 2 2 1 1 4 1 4 4 4 3
[55] 2 1 2 2 2 2 2 2 2 3 2 2 2 2 1 3 4 4 2 2 3 4 2 3 4 1 1 1 4 2 2 2 2 2 1 1 3 2 2 1 2 4 4 3 4 2 2 2 1 2 2 2 1
[109] 2 2 3 1 2 2 2 3 3 4 2 2 2 1 4 2 1 2 3 2 2 1 2 2 1 4 1 3 2 2 2 2 2 3 1 1 2 2 2 2 1 4 2 2 4 1 1 2 1 1 2 1 4
[163] 1 3 1 1 2 4 2 1 1 4 4 2 3 2 2 2 4 3 2 1 4 4 3 2 2 2 2 4 2 3 2 2 1 4 3 2 2 2 2 2 2 2 2 1 1 1 1 1 2 4 1 4 4 2
[217] 2 2 2 1 2 2 2 2 2
```


within cluster sum of squares by cluster:

```
[1] 77419676559 58043051246 113118302494 58533496705
(between_SS / total_SS = 93.5 %)
```

실행 결과 다음과 같은 표가 도출되었고 이는 각 나라마다 속해 있는 군집을 번호로 나타내어 표시한 결과물이다. 하지만 225개의 많은 데이터를 표로 분석하기에는 어려움이 있었다. 따라서 clusplot으로 나타낸 이후에 분석하기로 하였다.

### 4. Clusplot

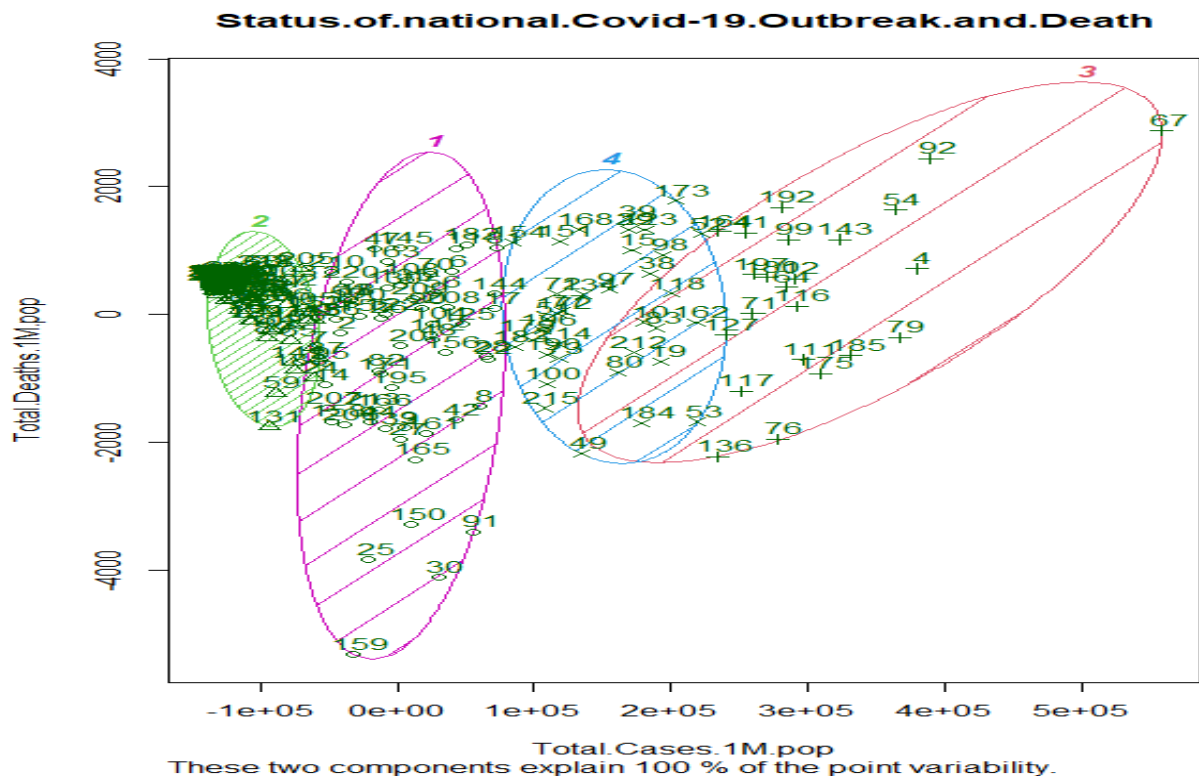
clusplot 함수를 사용하여 분석 결과를 그림으로 알아보기 쉽게 나타낸다. x축은 단위 당 발병자 수이고 y축은 단위 당 사망자 수를 적용시켜 나타내었다. clusplot을 통해 최종적으로 군집 분석을 완료하였고 분석 결과 다음과 같은 결론을 내리게 되었다.



```
1 covid <- read.csv('COVID-19 Coronavirus.csv', head=T)
2 covid_data <- covid[,8:9]
3
4 library(NbClust)
5 nc <- NbClust(covid_data, distance='euclidean', method='ward.D2')
6
7 set.seed(12045)
8 library(cluster)
9 km <- kmeans(covid_data, center=4, nstart=1000)
10 km
11
12 clusplot(x=covid_data, clus=km$cluster, lines=0, shade=T, color=T, labels=2,
13         main="Status.of.national.Covid-19.Outbreak.and.Death", xlab="Total.Cases.1M.pop", ylab="Total.Deaths.1M.p
14
15 covid$cluster <- km$cluster
```

### III. 결 론

#### 1. 분석 결과



##### a. 1번 군집

total case/1m : 140577.05      total death/1m : 1774.5763

Case Rate : 14.058 %      Death Rate : 1.262 %

발병률이 상대적으로 적지만 사망자 수는 상대적으로 많다. 이는 코로나 백신의 부족 등 의료 체계가 상대적으로 부족하다는 것을 알 수 있다. 따라서 코로나에 감염되지 않도록 각별한 주의가 필요하다.

##### b. 2번 군집

total case/1m : 20817.31      total death/1m : 332.8611

Case Rate : 2.082 %      Death Rate : 1.599 %

백만 명당 발병률이 다른 군집 대비 현저히 적다. 걸린 인구가 적어 사망률은 다른 군집보다 조금 높다고 나타난다. 방문을 하였을 때 질병 감염률이 통계상으로는 전세계 나라 중 안전한 국가들이라 말할 수 있지만 사망률이 높아 고위험층 환자는 주의가 필요하다.

c. 3번 군집

total case/1m : 442859.00      total death/1m : 2021.5652

Case Rate : 44.286 %      Death percentage : 0.457 %

사망률은 낮지만 발병률이 높은 국가로 구성된 군집이다. 감염률이 높기 때문에 백신 등 예방 활동이 매우 중요시되는 국가이다.

d. 4번 군집

total case/1m: 287843.17      total death /1m : 1703.3143

Case Rate : 28.784 %      Death percentage : 0.592 %

발병률과 사망률이 상대적으로 낮은 국가이다. 타 군집들에 비해 상대적으로 안전하다고 볼 수 있다. 이 군집에는 주로 선진국 나라들이 위치되어있고, 전세계 나라 중 방역 체계가 잘 되어 있는 국가들이다.

## 2. 한계점

2번 군집이 수치상으로는 발병률이 가장 낮고 사망률이 가장 높으나 정치적으로 불안정하거나 의료 체계가 체계적으로 구축되어 있지 않은 나라들이 다수 포함되어 있어 발병자 집계에 대한 통계의 신뢰도가 낮아 정확한 판단을 하기엔 다소 어려움이 따른다.



## IV. 부 록

### 1번 군집

[1] "Albania"	"Anguilla"	"Argentina"
[4] "Armenia"	"Australia"	"Bahamas"
[7] "Barbados"	"Belarus"	"Belize"
[10] "Bermuda"	"Bosnia and Herzegovina"	"Botswana"
[13] "Brazil"	"British Virgin Islands"	"Bulgaria"
[16] "Cabo Verde"	"Canada"	"Chile"
[19] "Colombia"	"Cook Islands"	"Costa Rica"
[22] "Cuba"	"Dominica"	"Finland"
[25] "Greenland"	"Grenada"	"Hong Kong"
[28] "Hungary"	"Iran"	"Jordan"
[31] "Kuwait"	"Lebanon"	"Malaysia"
[34] "Malta"	"Mayotte"	"Moldova"
[37] "Mongolia"	"New Caledonia"	"New Zealand"
[40] "North Macedonia"	"Palestine"	"Panama"
[43] "Paraguay"	"Peru"	"Poland"
[46] "Qatar"	"Romania"	"Russia"
[49] "Saint Kitts and Nevis"	"Saint Lucia"	"Singapore"
[52] "Suriname"	"Trinidad and Tobago"	"Tunisia"
[55] "Turkey"	"Turks and Caicos"	"UAE"
[58] "Ukraine"	"Vietnam"	

### 2번 군집

[1] "Afghanistan"	"Algeria"	"Angola"
[4] "Antigua and Barbuda"	"Azerbaijan"	"Bangladesh"
[7] "Benin"	"Bhutan"	"Bolivia"
[10] "Burkina Faso"	"Burundi"	"Cambodia"
[13] "Cameroon"	"CAR"	"Chad"
[16] "China"	"Comoros"	"Congo"
[19] "Democratic Republic of the Congo"	"Djibouti"	"Dominican Republic"
[22] "Ecuador"	"Egypt"	"El Salvador"
[25] "Equatorial Guinea"	"Eritrea"	"Eswatini"
[28] "Ethiopia"	"Falkland Islands"	"Fiji"
[31] "Gabon"	"Gambia"	"Ghana"
[34] "Guatemala"	"Guinea"	"Guinea-Bissau"
[37] "Guyana"	"Haiti"	"Honduras"
[40] "India"	"Indonesia"	"Iraq"
[43] "Ivory Coast"	"Jamaica"	"Japan"
[46] "Kazakhstan"	"Kenya"	"Kiribati"
[49] "Kyrgyzstan"	"Laos"	"Lesotho"
[52] "Liberia"	"Libya"	"Macao"
[55] "Madagascar"	"Malawi"	"Mali"
[58] "Marshall Islands"	"Mauritania"	"Mauritius"
[61] "Mexico"	"Micronesia"	"Montserrat"
[64] "Morocco"	"Mozambique"	"Myanmar"
[67] "Namibia"	"Nepal"	"Nicaragua"
[70] "Niger"	"Nigeria"	"Niue"
[73] "Oman"	"Pakistan"	"Papua New Guinea"
[76] "Philippines"	"Rwanda"	"Saint Helena"
[79] "Samoa"	"Sao Tome and Principe"	"Saudi Arabia"
[82] "Senegal"	"Sierra Leone"	"Solomon Islands"
[85] "Somalia"	"South Africa"	"South Sudan"
[88] "Sri Lanka"	"St. Vincent Grenadines"	"Sudan"
[91] "Syria"	"Taiwan"	"Tajikistan"
[94] "Tanzania"	"Thailand"	"Timor-Leste"
[97] "Togo"	"Tonga"	"Uganda"
[100] "Uzbekistan"	"Vanuatu"	"Vatican City"
[103] "Venezuela"	"Wallis and Futuna"	"Western Sahara"
[106] "Yemen"	"Zambia"	"Zimbabwe"

### 3번 군집

[1] "Andorra"	"Austria"	"Channel Islands"	"Denmark"
[5] "Estonia"	"Faeroe Islands"	"France"	"Georgia"
[9] "Gibraltar"	"Iceland"	"Israel"	"Latvia"
[13] "Liechtenstein"	"Lithuania"	"Martinique"	"Montenegro"
[17] "Netherlands"	"R <del>책</del> union"	"San Marino"	"Seychelles"
[21] "Slovenia"	"St. Barth "	"Switzerland"	

### 4번 군집

[1] "Aruba"	"Bahrain"	"Belgium"
[4] "Brunei "	"Caribbean Netherlands"	"Cayman Islands"
[7] "Croatia"	"Cura <del>책</del> ao"	"Cyprus"
[10] "Czechia"	"French Guiana"	"French Polynesia"
[13] "Germany"	"Greece"	"Guadeloupe"
[16] "Ireland"	"Isle of Man"	"Italy"
[19] "Luxembourg"	"Maldives"	"Monaco"
[22] "Norway"	"Palau"	"Portugal"
[25] "S. Korea"	"Saint Martin"	"Saint Pierre Miquelon"
[28] "Serbia"	"Sint Maarten"	"Slovakia"
[31] "Spain"	"Sweden"	"UK"
[34] "Uruguay"	"USA"	