

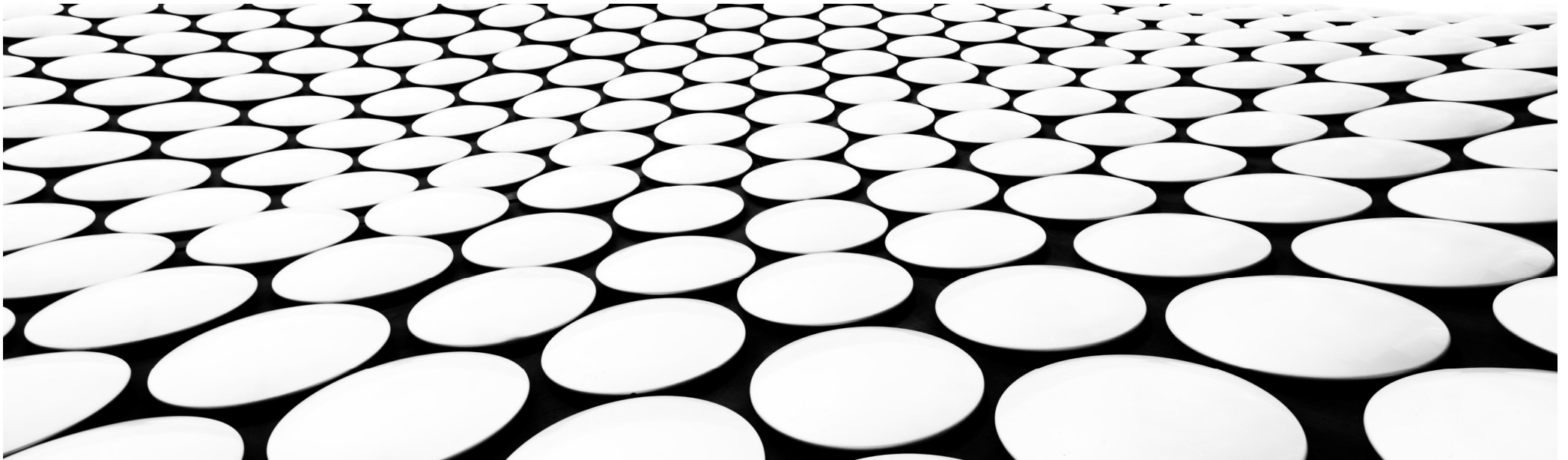
B조

정대욱  
홍상표  
조성민

# COVID19 질병 증상 분석

## CLINICAL CHARACTERIZATION OF COVID-19 BY EHR

SEP592 소프트웨어특강<데이터사이언스> 정유채 교수님



## 연구 대상 주제

COVID-19의 EHR 데이터 중 CONSORTIUM FOR CLINICAL CHARACTERIZATION OF COVID-19 데이터 분석  
환자의 ICD 코드 분석을 통한 코로나 증상의 통계적 분석

## 목표

ICD코드로 실제 증상에 대한 통계적 발현 증세를 확인

국내 한국질병분류 중심으로 질병증상을 분석하여 발현 증세 분석

## 분석 내용 COVID-19

질병코드 분류에 의한 순위 분석

질병코드 분류에 의한 증상 분석

시각화

## 분석 방법 PREPROCESSING

1. COVID-19 진단자료
2. ICD <-> KOICD
3. 질환 증상 Mapping

지역: 미국(45개), 프랑스(42개), 이탈리아(5개), 독일(3개), 싱가포르(1) 등  
총 96개 병원

인원 : 27,584명의 COVID-19 진단 환자

기간 : 2020년 1월 1일 ~ 2020년 4월 11일

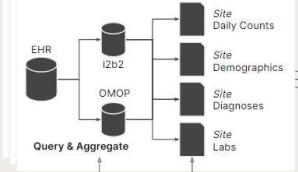
[https://covidclinical.net/data/index.html#publication\\_paper-01](https://covidclinical.net/data/index.html#publication_paper-01)

Consortium for Clinical Characterization of COVID-19 by EHR

International Electronic Health Record Consortium  
Course Profile: The 4CE Consortium

← Data Apr 13, 2020 • 4CE Consortium

Participating Sites



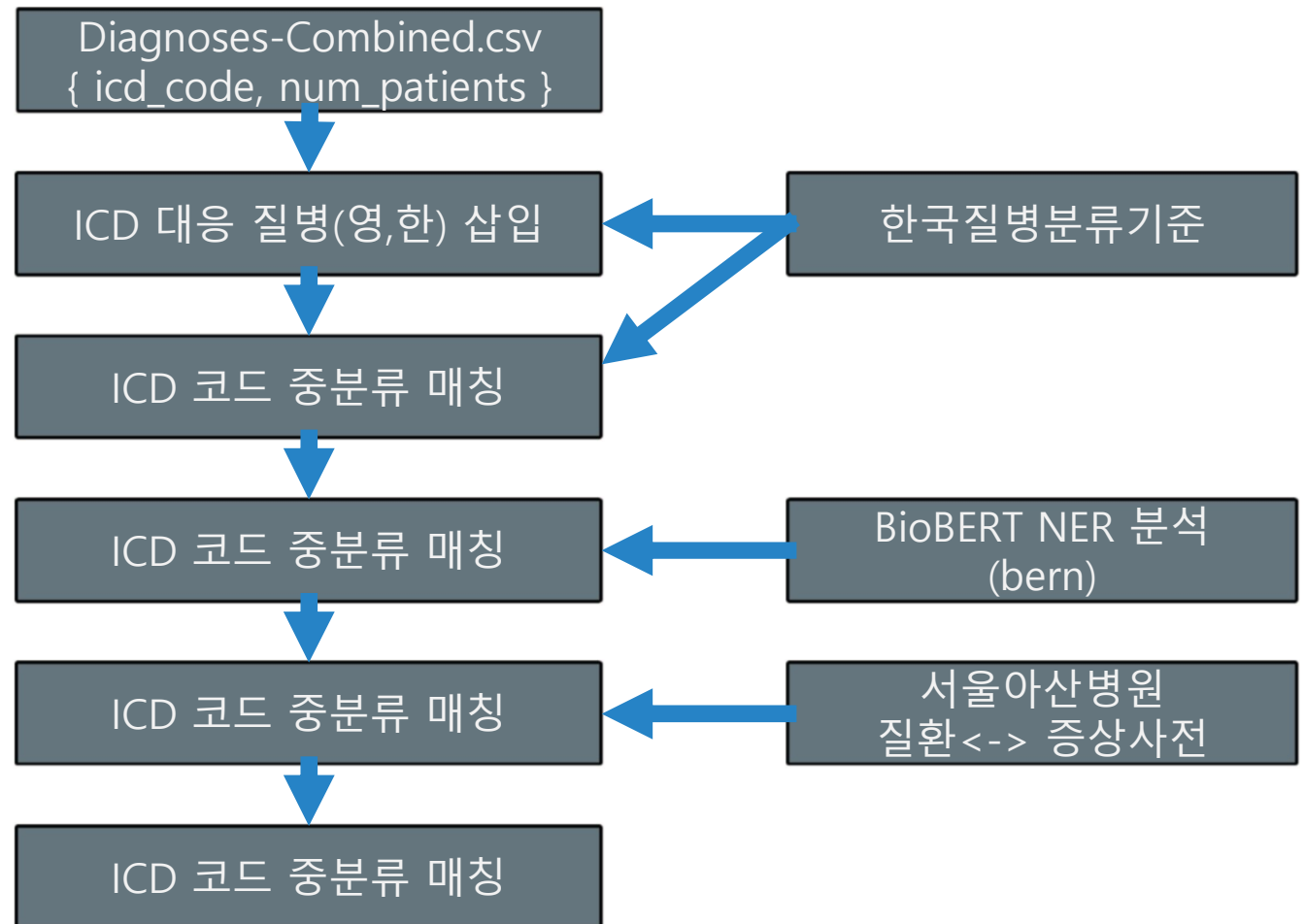
Download options:

- ↓ Daily Counts Data  
Download daily patient counts data
- ↓ Demographics Data  
Download demographics data
- ↓ **Diagnosis Data**  
Download diagnosis data
- ↓ Labs Data  
Download labs data
- ↓ Participating Healthcare Systems Table  
Download information about participating sites
- ↓ JHU CSSE Table  
Download global confirmed cases as of Apr 11, 2020

Labs Data – 국가별 코로나 진단환자 .CSV/DataSet

## 분석 방법 PREPROCESSING

1. COVID-19 진단자료
2. ICD <-> KOICD
3. 질환 증상 Mapping



# 분석 방법 PREPROCESSING

1. COVID-19 진단자료
2. ICD <-> KOICD
3. 질환 증상 Mapping

Labs Data – 국가별 코로나  
- 국가명 + ICD코드 + 환자수 -> R/Python으로 데이터 수집

	A	B	C	D
1	siteid	icd_code	icd_version	num_patients
2	France	Z29.0	10	1232
3	France	J12.8	10	1123
4	France	J96.0	10	620
5	France	I10	10	585
6	France	R06.0	10	309
7	France	J80	10	306
8	France	R50.9	10	255
9	France	E11.9	10	248
10	France	R53	10	223
11	France	R05	10	214
12	France	J12.9	10	170
13	France	Z71.1	10	149
14	France	E66.0	10	148
15	France	Z51.5	10	115
16	France	J18.9	10	112
17	France	N17.9	10	103
18	France	R65.0	10	95

```

13 # [steve] (https://github.com/steve-lemon)
14
15
16 ##[notebook help]##
17
18 - "cmd-shift+enter" to run
19 - "cmd-option+enter" to insert code chunk
20 - "cmd-shift+enter" to preview in html (no code run)
21
22
23 ---[r]
24 # Below are the required packages.
25 # install.packages("curl")
26 # install.packages("httr")
27 # install.packages("assertthat")
28 # install.packages("dplyr")
29
30
31 ---[r]
32 ## Commons [tabset]
33
34
35
36
37 ---[r]
38 ## Run Environment
39
40 # current directory
41 cat("current-directory =", getwd(), "\n")
42 # current time in local
43 cat("current-timestamp =", format(Sys.time(), "%a %b %d %X %Y"), "\n")
44
45
46
47 ---[r]
48 ## ICD 코드 구하기
49
50 # Fetching ICD code information from REST APIs
51 library(jsonlite)
52 mysearchICD <- function(code, all=1) {
53   req = httr::POST("https://www.kotcd.kr/2016/easyfind/search.json",
54     httr::add_headers(
55       "Authorization" = paste("n", "", secret)),
56       "content-type" = "application/x-www-form-urlencoded; charset=utf-8",
57       "Cookie" = "SESSIONID=89384048886C81290A86E4A42773; dft_popupdone="
58     ),
59     body = paste("searchtype=01&code=", code, sep="")
60   )
61   json = httr::content(req, as = "text")
62   data <- fromJSON(json)
63   # print(length(data$list))
64   if (length(data$list) == 0) return(list(code_title="", eng_text="", kor_text=""))
65   dd = data$list[1, c("code_title", "eng_text", "kor_text")]
66   if (all) return(dd)
67   return(dd[code_title == code,])
68 }
69 # sample for J20
70 mysearchICD("J20")
71
72

```

# 분석 방법 PREPROCESSING

1. COVID-19 진단자료
2. ICD <-> KOICD
3. 질환 증상 Mapping

## KOICD (한국질병분류) – 질병분류 수집

KOICD 질병분류 정보센터

통합검색 | 질병코드, 한글, 영문명을 입력하세요. | 인기검색어 | M5422 | 메모장 ? | 개선의견

검색어를 입력하세요. | 검색 | 상세검색 | 질병분류목록(1권)(3차정보 2016.11.30.) > | 기호 및 약어 설명 | KCD7 | 과거버전 | 부가서비스

HOME | MENU | LIST

- ▶ A00-B99 I. 특정 감염성 및 기생충성 질환
- ▶ C00-D48 II. 신장질환
- ▶ D50-D89 III. 철액 및 조절기관의 질환과 면역 메커니즘을 침범한 특정 장애
- ▶ E00-E90 IV. 내분비, 영양 및 대사 질환
- ▶ F00-F99 V. 정신 및 행동 장애
- ▶ G00-G99 VI. 신경계통의 질환
- ▶ H00-H59 VII. 눈 및 눈 부속기의 질환
- ▶ H60-H95 VIII. 귀 및 유도의 질환
- ▶ I00-I99 IX. 순환계통의 질환

### I. 특정 감염성 및 기생충성 질환(Certain infectious and parasitic diseases)(A00-B99)

항균제 내성의 분류를 원한다면 부가분류번호(U82-U84)를 사용 할 것.

포함 : 일반적으로 전염 또는 전파하는 것으로 인정된 질환

제외 : 감염성 질환의 보균자 또는 의심되는 보균자(Z22-)

특정 국소감염 - 신체계통에 관련된 장(章) 참조

임신, 출산, 산후기에 합병된 감염성 및 기생충성 질환[산과적 파상풍 제외](O98-)

출생전후기에 특이한 감염성 및 기생충성 질환[신생아파상풍, 선천매]

Includes : Diseases generally recognized as communicable or transmissible

Excludes : Carrier or suspected carrier of infectious diseases

Certain localized infections-see body system-related chapters

Infectious and parasitic diseases complicating pregnancy, childbirth and the puerperium[except obstetrical tetanus]

Infectious and parasitic diseases specific to the perinatal

## 서울아산병원(질환백과) – 질환명 기준 증상 수집

### 질환백과

질환명 검색 | 부위별 | 질환별

질환명 검색 | 코로나 | 검색

중동 호흡기 증후군(MERS)(Middle East Respiratory Syndrome)

증상 : 기침, 호흡곤란, 발열, 식욕부진, 오심, 구토, 설사, 복통, 근육통, 관절통, 두통, 목통, 인후통, 폐렴, 신장질환, 흉막염, 심근염, 뇌염, 뇌수막염, 뇌실막염, 폐색전증, 심근경색, 뇌졸중, 말초혈관질환, 당뇨병, 고혈압, 만성질환 악화, 임신, 출산, 산후기 합병증, 기타

【호흡곤란】

성인 호흡곤란 증후군(Adult respiratory distress syndrome)

증상 : 호흡곤란, 호흡부전, 저산소혈증, 폐성심

【신장 질환】

남성 신장 질환(Cystic kidney disease)

증상 : 혈뇨, 단백뇨, 신장기능 저하, 고혈압, 만성질환 악화, 임신, 출산, 산후기 합병증, 기타

## 분석 방법 PREPROCESSING

1. COVID-19 진단자료
2. ICD <-> KOICD
3. 질환 증상 Mapping

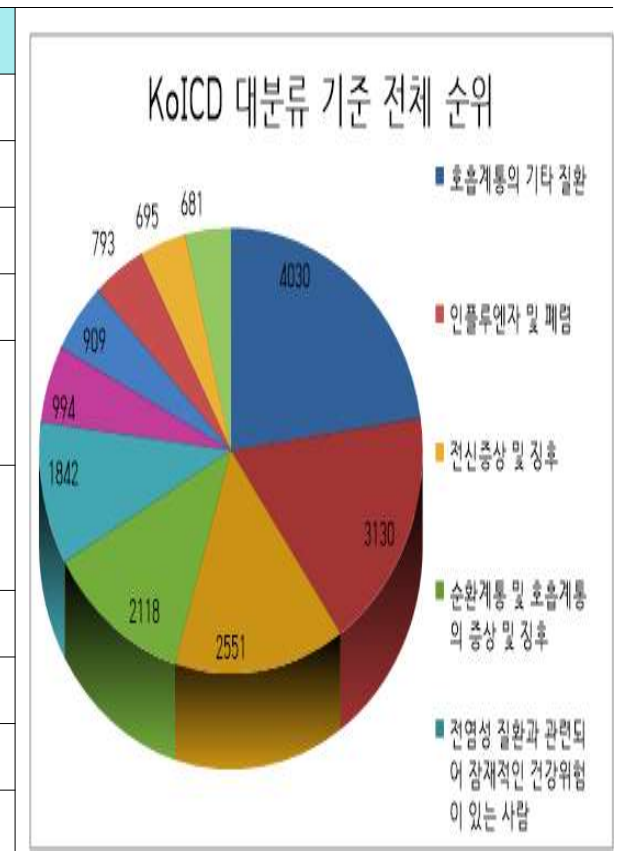
환자수별 ICD ⇒ KOICD ⇒ 질환명 ⇒ 증상

ICD No	영문 질환	한글 질환	증세	대분류
A04	6 Other bacterial intestinal infections	기타 세균성 장감염	설사  구토  두통  복부경련  빈혈  혈소판감소  혈변  복부 통증  요독증	장감염질환
A08	4 Viral and other specified intestinal infections	바이러스성 및 명시된 장감염	설사  구토  콧물  탈수  열  기침	장감염질환
A09	32 Other gastroenteritis and colitis of infectious and unspecified origin	감염성 및 상세불명 기원의 기타 위장염 및 결장염	설사  구토  두통  탈수  복부팽만감  열  복부 통증	장감염질환
A40	3 Streptococcal sepsis	연쇄구균패혈증	빈호흡  설사  물집  의식 저하  심부전  오한  저혈압  괴사  열  기운없음  환부 부종  반상출혈  소변량 감소  복부 통증  빈맥	기타 세균성 질환
A41	95 Other sepsis	기타 패혈증	빈호흡  설사  물집  의식 저하  심부전  오한  저혈압  괴사  열  기운없음  환부 부종  반상출혈  소변량 감소  복부 통증  빈맥	기타 세균성 질환
A46	9 Erysipelas	단독(丹毒)	림프 부종  오한  발진  열감  환부 통증	기타 세균성 질환
A49	6 Bacterial infection of unspecified site	상세불명 부위의 세균감염	목의 통증  설사  구토  두통  근육통  콧물  오심  오한  열  기침  권태감	기타 세균성 질환

## 분석 결과 VISUALIZATION

1. 환자 질환 순위
2. COVID 증상 20순위
3. Visualization
  - WordCount
  - 증상 빈도
  - 증상 연관도

KoICD 대분류 기준	인원
호흡계통의 기타 질환	4030
인플루엔자 및 폐렴	3130
전신증상 및 징후	2551
순환계통 및 호흡계통의 증상 및 징후	2118
전염성 질환과 관련되어 잠재적인 건강위험이 있는 사람	1842
병인이 불확실한 신종질환의 임시적 지정이나 응급사용	994
고혈압성 질환	909
급성 상기도감염	793
당뇨병	695
세균, 바이러스 및 기타 감염체	681





# 분석 결과 VISUALIZATION

1. 환자 질환 순위
2. COVID 증상 20순위
3. Visualization
  - WordCount
  - 증상 빈도
  - 증상 연관도

순위	증상20순위(가중치)	가중치점수
1	열	218
2	호흡곤란	193
3	기침	158
4	피로감	122
5	가래	111
6	두통	111
7	근육통	106
8	복부 통증	99
9	구토	95
10	오한	91
11	의식 저하	87
12	가슴 통증	78
13	설사	76
14	청색증	69
15	식욕부진	69
16	천명음	66
17	목의 통증	64
18	환부 부종	59
19	오심	53
20	체중감소	48

순위	증상 20순위	점수
1	열	108
2	호흡곤란	81
3	복부 통증	72
4	구토	69
5	피로감	59
6	기침	58
7	두통	58
8	식욕부진	53
9	설사	45
10	오심	41
11	체중감소	39
12	근육통	38
13	가래	36
14	환부 부종	34
15	무증상	33
16	오한	32
17	피부 긴장도 저하	29
18	어지러움	28
19	가슴 통증	27
20	황달	26

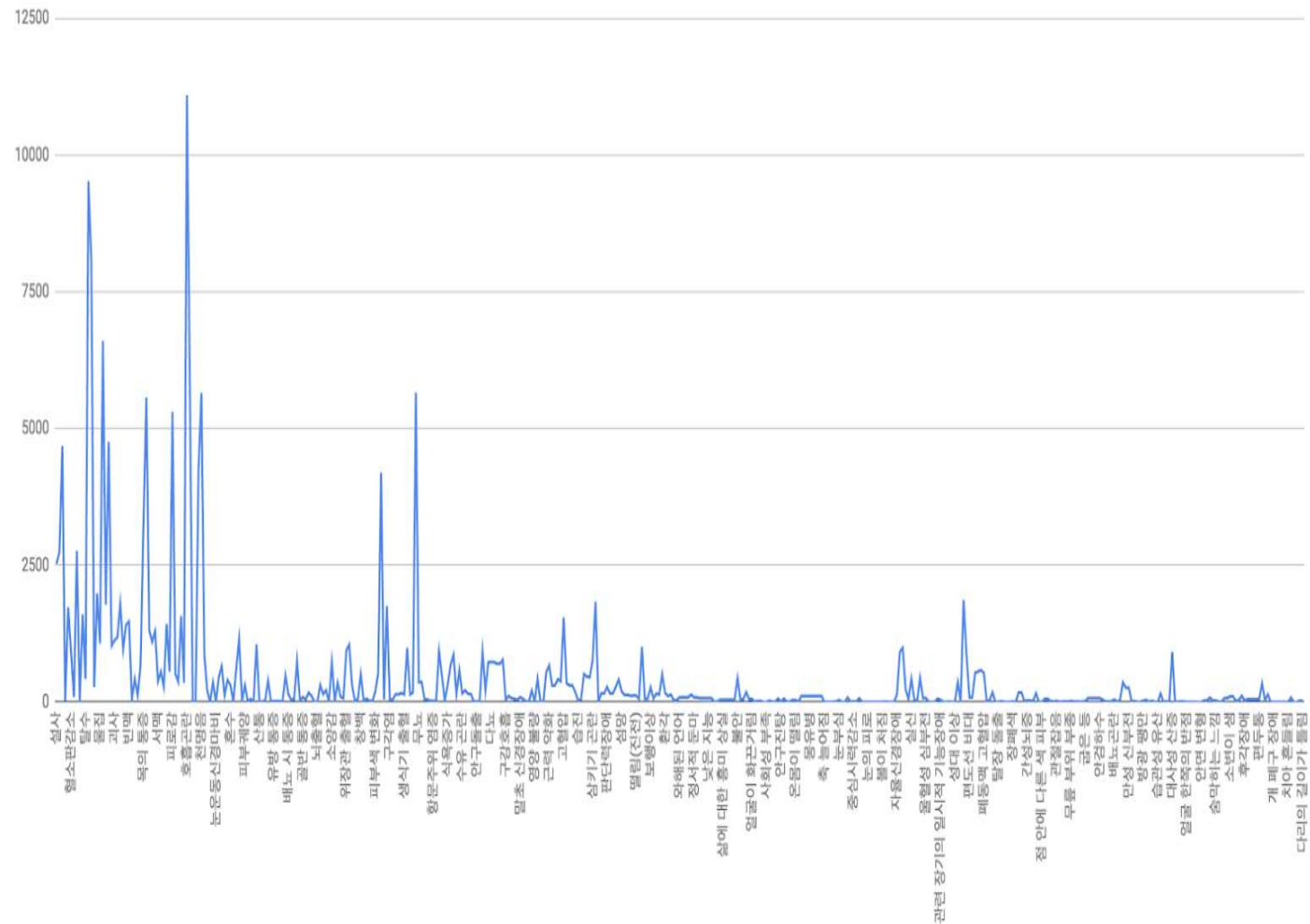
## 분석 결과 VISUALIZATION

1. 환자 질환 순위
2. COVID 증상 20순위
3. Visualization
  - WordCount
  - 증상 빈도
  - 증상 연관도



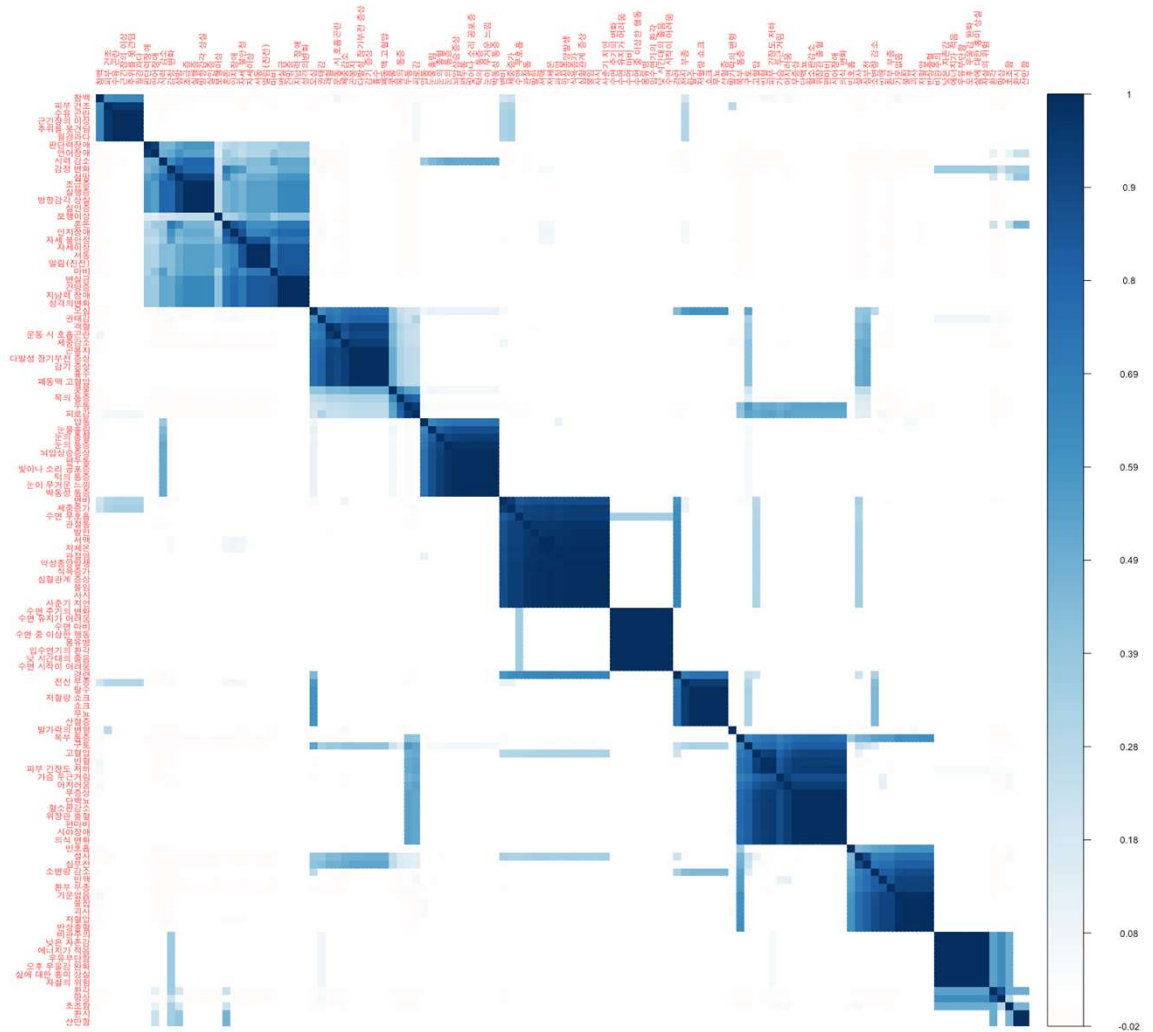
1. 환자 질환 순위
2. COVID 증상 20순위
3. Visualization

- WordCount
- 증상 빈도
- 증상 연관도



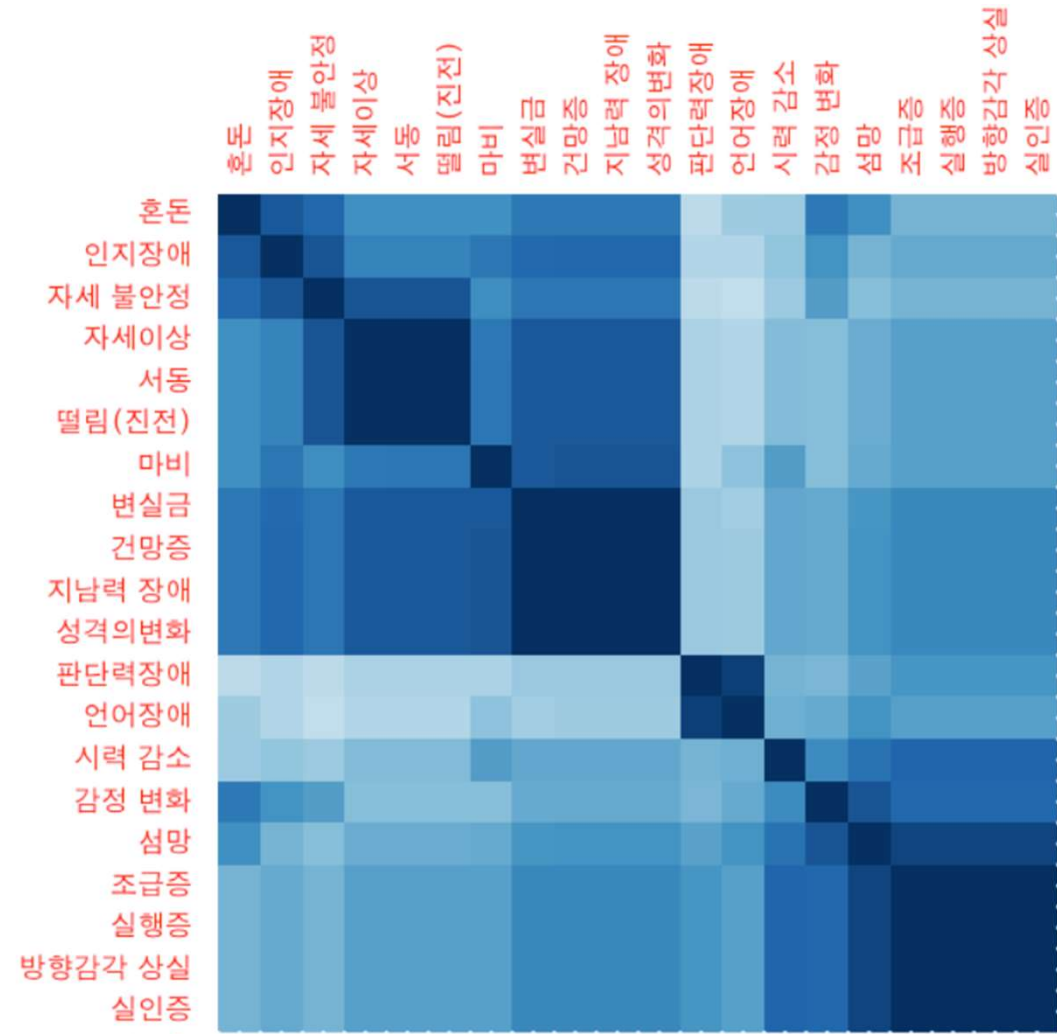
# 분석 결과 VISUALIZATION

1. 환자 질환 순위
2. COVID 증상 20순위
3. Visualization
  - WordCount
  - 증상 빈도
  - 증상 연관도



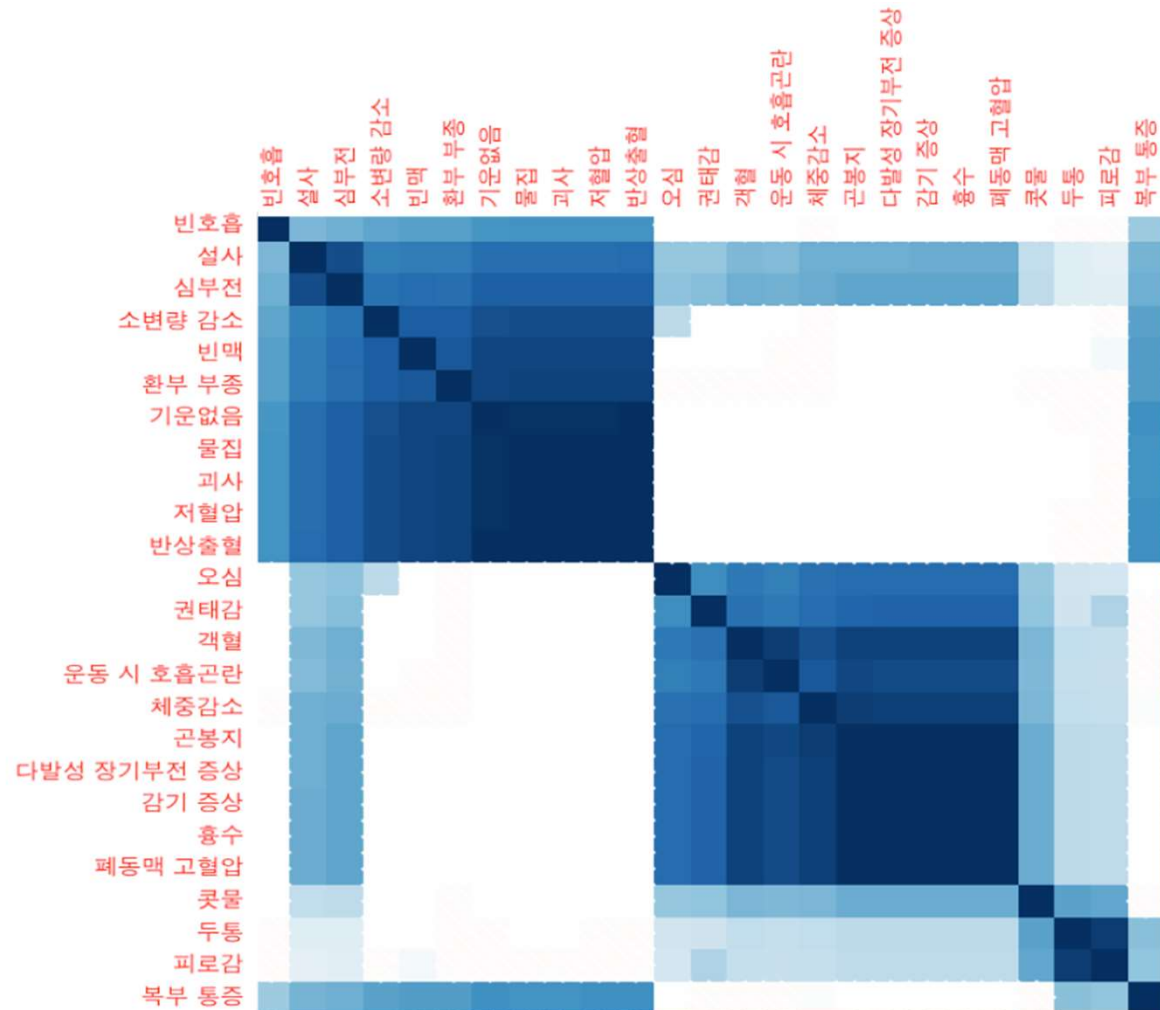
# 분석 결과 VISUALIZATION

1. 환자 질환 순위
2. COVID 증상 20순위
3. Visualization
  - WordCount
  - 증상 빈도
  - 증상 연관도



# 분석 결과 VISUALIZATION

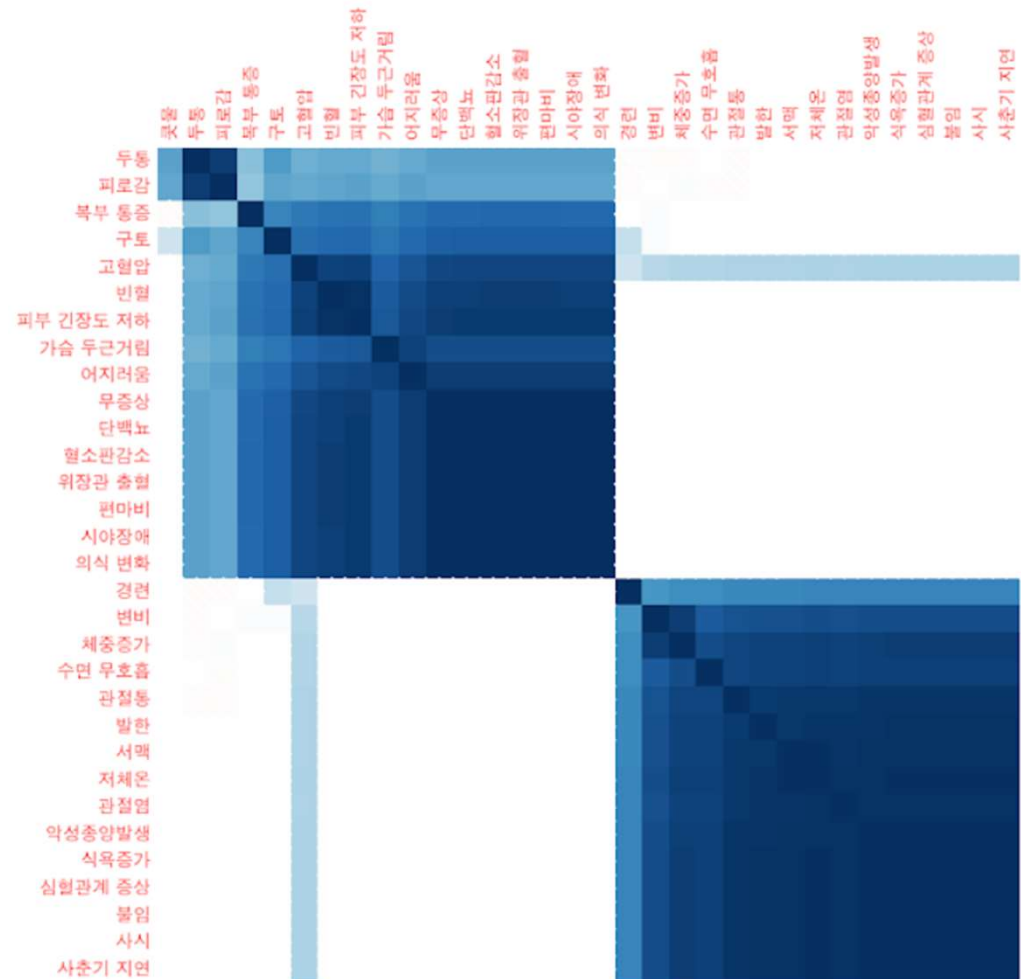
1. 환자 질환 순위
2. COVID 증상 20순위
3. Visualization
  - WordCount
  - 증상 빈도
  - 증상 연관도



# 분석 결과 VISUALIZATION

1. 환자 질환 순위
2. COVID 증상 20순위
3. Visualization

- WordCount
- 증상 빈도
- 증상 연관도



## 결론

코로나 증상의 대표적 증상 , 강도 확인

**분석에서 InSight** ICD코드로 실제 증상에 대한 통계적 발현 증세를 확인

주된 증상은 열, 호흡곤란이 가장 큰 빈도를 차지하여 코로나 증상⇔ 고열 임을 다시 확인

호흡기관, 열 이외 소수 증상으로 혈관계(고혈압,신장이상 단백질, 혈당) 증상이 다음 상관관계를 보임

보건,예방조치로 환자 질환코드를 등록한 경향이 보여 유럽/미국의 4월의 긴박한 상황을 추측가능