8-3. 데이터 분석 시각화 2

러닝스푼즈 나노디그리
데이터 이해를 위한 SQL

홍형경
chariehong@gmail.com
2021.06

Copyright © 2020, All rights reserved.

1. 코로나 데이터 생성 (2020.01 ~ 2021.02)

covid19_country 테이블

컬럼	데이터 형	Null	PK	설명
countrycode	VARCHAR2(10)	N	Y	국가 코드 (기본 키)
countryname	VARCHAR2(80)	N		국가 명
continent	VARCHAR2(50)	Y		대륙 명
population	NUMBER	Y		인구
population_density	NUMBER	Y		인구 밀도
median_age	NUMBER	Y		평균 연령
aged_65_older	NUMBER	Y		65세 이상 인구 비율
aged_70_older	NUMBER	Y		70세 이상 인구 비율
hospital_beds_per_thousand	NUMBER	Y		1000명당 병실 침대 수

Copyright © 2020, All rights reserved.

1. 코로나 데이터 생성 (2020.01 ~ 2021.02)

covid19_data 테이블

컬럼	데이터 형	Null	PK	설명
countrycode	VARCHAR2(10)	N	Y	국가 코드 (기본 키)
issue_date	DATE	N	Y	발생 일 (기본 키)
cases	NUMBER	Y		확진자 수
new_cases_per_million	NUMBER	Y		100만명당 확진자 수
deaths	NUMBER	Y		사망자 수
icu_patients	NUMBER	Y		중환자 수
hosp_patients	NUMBER	Y		병원 환자 수
tests	NUMBER	Y		검사자 수
reproduction_rate	NUMBER	Y		재생산 감염율
new_vaccinations	NUMBER	Y		백신 접종자 수
stringency_index	NUMBER	Y		방역 지수로 0에서 100사이의 값이 저장되어 있으며, 100이 가장 방역지수가 높음

Copyright © 2020, All rights reserved.

1. 코로나 데이터 생성 (2020.01 ~ 2021.02)

· covid19_country 테이블 생성

```
CREATE TABLE covid19_country (
  countrycode             VARCHAR2(10) NOT NULL,
  countryname             VARCHAR2(80) NOT NULL,
  continent               VARCHAR2(50),
  population              NUMBER,
  population_density      NUMBER,
  median_age              NUMBER,
  aged_65_older           NUMBER,
  aged_70_older           NUMBER,
  hospital_beds_per_thousand  NUMBER,
  PRIMARY KEY (countrycode)    );
```

러닝스푼즈 나노디그리 데이터 이해를 위한 SQL

Copyright © 2020, All rights reserved.

1. 코로나 데이터 생성 (2020.01 ~ 2021.02)

· covid19_data 테이블 생성

```
CREATE TABLE covid19_data (
  countrycode             VARCHAR2(10) NOT NULL,
  issue_date              DATE       NOT NULL,
  cases                   NUMBER,
  new_cases_per_million   NUMBER,
  deaths                  NUMBER,
  icu_patients            NUMBER,
  hosp_patients           NUMBER,
  tests                   NUMBER,
  reproduction_rate       NUMBER,
  new_vaccinations        NUMBER,
  stringency_index        NUMBER,
  PRIMARY KEY (countrycode, issue_date)   );
```

파이썬 나노디그리 데이터 이해를 위한 SQL

Copyright © 2020, All rights reserved.

1. 코로나 데이터 생성 (2020.01 ~ 2021.02)

- **04.covid19_country_insert.sql** 파일 실행

- **05.covid19_data_insert.sql** 파일 실행

러닝스푼즈 나노디그리
데이터 이해를 위한 SQL

Copyright © 2020, All rights reserved.

2. 코로나 데이터 정제

· covid19_data 테이블의 countrycode 값에 OWID로 시작되는 데이터 삭제

```
SELECT countrycode, COUNT(*)
    FROM covid19_data
  WHERE countrycode LIKE 'OWID%'
  GROUP BY countrycode ;
```

	COUNTRY...	COUNT(*)
1	OWID_AFR	382
2	OWID_ASI	404
3	OWID_EUN	403
4	OWID_EUR	403
5	OWID_INT	388
6	OWID_KOS	352
7	OWID_NAM	404
8	OWID_NCY	9
9	OWID_OCE	401
10	OWID_SAM	373
11	OWID_WRL	404

러닝스푼즈 나노디그리 데이터 이해를 위한 SQL

Copyright © 2020, All rights reserved.

2. 코로나 데이터 정제

· covid19_data 테이블의 countrycode 값에 OWID로 시작되는 데이터 삭제

```
DELETE  covid19_data
   WHERE countrycode LIKE 'OWID%' ;

 SELECT COUNT(*)
      FROM covid19_data
    WHERE countrycode LIKE 'OWID%' ;

COMMIT;
```

	COUNT(*)
1	0

나노디그리 데이터 이해를 위한 SQL 러닝 스쿨

Copyright © 2020, All rights reserved.

2. 코로나 데이터 정제

· covid19_data 테이블에서 숫자형 컬럼 NULL을 0으로 수정

· 숫자형 컬럼의 경우 NULL이 있으면 사칙연산시 결과가 NULL

· cases, new_cases_per_million, ~ stringency_index까지 총 9개 컬럼이 대상

	COUNTRYCODE	ISSUE_DATE	CASES	NEW_CASES_PER_MILLION	DEATHS	ICU_PATIENTS	HOSP_PATIENTS	TESTS	REPRODUCTION_RATE	NEW_VACCINATIONS	STRINGENCY_INDEX
1	ALB	2020-02-25 00:00:00	(null)	(null)	(null)	(null)	(null)	8	(null)	(null)	8.33
2	ALB	2020-02-26 00:00:00	(null)	(null)	(null)	(null)	(null)	5	(null)	(null)	8.33
3	ALB	2020-02-27 00:00:00	(null)	(null)	(null)	(null)	(null)	4	(null)	(null)	8.33
4	ALB	2020-02-28 00:00:00	(null)	(null)	(null)	(null)	(null)	1	(null)	(null)	8.33
5	ALB	2020-02-29 00:00:00	(null)	(null)	(null)	(null)	(null)	8	(null)	(null)	8.33
6	ALB	2020-03-01 00:00:00	(null)	(null)	(null)	(null)	(null)	3	(null)	(null)	8.33
7	ALB	2020-03-02 00:00:00	(null)	(null)	(null)	(null)	(null)	2	(null)	(null)	8.33
8	ALB	2020-03-03 00:00:00	(null)	(null)	(null)	(null)	(null)	5	(null)	(null)	8.33
9	ALB	2020-03-04 00:00:00	(null)	(null)	(null)	(null)	(null)	6	(null)	(null)	8.33
10	ALB	2020-03-05 00:00:00	(null)	(null)	(null)	(null)	(null)	8	(null)	(null)	8.33
11	ALB	2020-03-06 00:00:00	(null)	(null)	(null)	(null)	(null)	3	(null)	(null)	8.33
12	ALB	2020-03-07 00:00:00	(null)	(null)	(null)	(null)	(null)	5	(null)	(null)	8.33
13	ALB	2020-03-08 00:00:00	(null)	(null)	(null)	(null)	(null)	1	(null)	(null)	8.33
14	ALB	2020-06-12 00:00:00	31	10.772	1	(null)	(null)	300	1.6	(null)	71.3
15	ALB	2020-06-13 00:00:00	48	16.679	0	(null)	(null)	270	1.62	(null)	71.3
16	ALB	2020-06-14 00:00:00	57	19.807	0	(null)	(null)	362	1.63	(null)	71.3
17	ALB	2020-06-15 00:00:00	69	23.977	0	(null)	(null)	319	1.61	(null)	71.3

Copyright © 2020, All rights reserved.

2. 코로나 데이터 정제

· covid19_data 테이블에서 숫자형 컬럼 NULL을 0으로 수정

```
UPDATE covid19_data
   SET cases = 0
 WHERE cases IS NULL;

UPDATE covid19_data
   SET deaths = 0
 WHERE deaths IS NULL;
```

· covid19_data 테이블의 총 9개 컬럼에 대해 UPDATE문 9개 실행

나노디그리 데이터 이해를 위한 SQL 비닝스퓨즈

Copyright © 2020, All rights reserved.

2. 코로나 데이터 정제

· covid19_data 테이블에서 숫자형 컬럼 NULL을 0으로 수정

```
UPDATE covid19_data
  SET cases                = NVL(cases, 0)
    ,new_cases_per_million = NVL(new_cases_per_million, 0)
    ,deaths                = NVL(deaths, 0)
    ,icu_patients          = NVL(icu_patients, 0)
    ,hosp_patients         = NVL(hosp_patients, 0)
    ,tests                 = NVL(tests, 0)
    ,reproduction_rate     = NVL(reproduction_rate, 0)
    ,new_vaccinations      = NVL(new_vaccinations, 0)
    ,stringency_index      = NVL(stringency_index, 0);

COMMIT;
```

리닝스푼즈 나노디그리
데이터 이해를 위한 SQL

Copyright © 2020, All rights reserved.

3. 코로나 데이터 분석

(1) 2020년 월별, 대륙별, 국가별 감염수

```
SELECT TO_CHAR(b.issue_date, 'YYYYMM') months,
       a.continent, a.countryname,
       SUM(b.cases) 감염수
FROM covid19_country a
INNER JOIN covid19_data b
  ON a.countrycode = b.countrycode
WHERE TO_CHAR(b.issue_date, 'YYYY') = '2020'
GROUP BY TO_CHAR(b.issue_date, 'YYYYMM'),
       a.continent, a.countryname
ORDER BY 1, 2, 3;
```

	MONTHS	CONTINENT	COUNTRYNAME	감염수
1	202001	Asia	Cambodia	1
2	202001	Asia	China	9254
3	202001	Asia	Hong Kong	0
4	202001	Asia	India	1
5	202001	Asia	Japan	13
6	202001	Asia	Malaysia	8
7	202001	Asia	Nepal	1
8	202001	Asia	Philippines	1
9	202001	Asia	Singapore	13
10	202001	Asia	South Korea	10
11	202001	Asia	Sri Lanka	1
12	202001	Asia	Taiwan	9
13	202001	Asia	Thailand	15
14	202001	Asia	United Arab Emirates	4
15	202001	Asia	Vietnam	2
16	202001	Europe	Finland	1
17	202001	Europe	France	5
18	202001	Europe	Germany	5

Copyright © 2020, All rights reserved.

3. 코로나 데이터 분석

(2-1) 2020년 월별, 대륙별, 국가별 감염수, 대륙 기준 감염수 비율

```
WITH covid1 AS (
SELECT TO_CHAR(b.issue_date, 'YYYYMM') months,
       a.continent, a.countryname,  SUM(b.cases) case_num
  FROM covid19_country a
 INNER JOIN covid19_data b
    ON a.countrycode = b.countrycode
 WHERE TO_CHAR(b.issue_date, 'YYYY') = '2020'
 GROUP BY TO_CHAR(b.issue_date, 'YYYYMM'),   a.continent, a.countryname ),
covid2 AS (
SELECT months, continent, countryname,  case_num,
       SUM(case_num) OVER (PARTITION BY months, continent) tot
  FROM covid1
)
SELECT months, continent, countryname,  case_num, tot,
       DECODE(tot, 0, 0, ROUND(case_num / tot * 100,2)) rates
  FROM covid2
ORDER BY 1, 2, 3 ;
```

	MO...	CONTINENT	COUNTRYNAME	CASE_NUM	TOT	RATES
1	202001	Asia	Cambodia	1	9333	0.01
2	202001	Asia	China	9254	9333	99.15
3	202001	Asia	Hong Kong	0	9333	0
4	202001	Asia	India	1	9333	0.01
5	202001	Asia	Japan	13	9333	0.14
6	202001	Asia	Malaysia	8	9333	0.09
7	202001	Asia	Nepal	1	9333	0.01
8	202001	Asia	Philippines	1	9333	0.01
9	202001	Asia	Singapore	13	9333	0.14
10	202001	Asia	South Korea	10	9333	0.11
11	202001	Asia	Sri Lanka	1	9333	0.01
12	202001	Asia	Taiwan	9	9333	0.1
13	202001	Asia	Thailand	15	9333	0.16
14	202001	Asia	United Arab Emirates	4	9333	0.04
15	202001	Asia	Vietnam	2	9333	0.02
16	202001	Europe	Finland	1	17	5.88
17	202001	Europe	France	5	17	29.41
18	202001	Europe	Germany	5	17	29.41
19	202001	Europe	Italy	2	17	11.76
20	202001	Europe	Russia	2	17	11.76
21	202001	Europe	Switzerland	0	17	0
22	202001	Europe	United Kingdom	2	17	11.76
23	202001	North America	Canada	4	11	36.36
24	202001	North America	Mexico	0	11	0
25	202001	North America	United States	7	11	63.64
26	202001	Oceania	Australia	9	9	100
27	202001	Oceania	Fiji	0	9	0
28	202001	South America	Argentina	0	0	0
29	202002	Africa	Algeria	1	3	33.33
30	202002	Africa	Egypt	1	3	33.33

Copyright © 2020, All rights reserved.

3. 코로나 데이터 분석

(2-2) 2020년 월별, 대륙별, 국가별 감염수, 대륙 기준 감염수 비율

```
WITH covid1 AS (
SELECT TO_CHAR(b.issue_date, 'YYYYMM') months,
       a.continent, a.countryname,
       SUM(b.cases) case_num
  FROM covid19_country a
 INNER JOIN covid19_data b
    ON a.countrycode = b.countrycode
 WHERE TO_CHAR(b.issue_date, 'YYYY') = '2020'
 GROUP BY TO_CHAR(b.issue_date, 'YYYYMM'),
          a.continent, a.countryname
)
SELECT months, continent, countryname,
       case_num,
       ROUND(RATIO_TO_REPORT(case_num) OVER (PARTITION BY months, continent) * 100,2)  rates
  FROM covid1
 ORDER BY 1, 2, 3;
```

	MO...	CONTINENT	COUNTRYNAME	CASE_NUM	RATES
1	202001	Asia	Cambodia	1	0.01
2	202001	Asia	China	9254	99.15
3	202001	Asia	Hong Kong	0	0
4	202001	Asia	India	1	0.01
5	202001	Asia	Japan	13	0.14
6	202001	Asia	Malaysia	8	0.09
7	202001	Asia	Nepal	1	0.01
8	202001	Asia	Philippines	1	0.01
9	202001	Asia	Singapore	13	0.14
10	202001	Asia	South Korea	10	0.11
11	202001	Asia	Sri Lanka	1	0.01
12	202001	Asia	Taiwan	9	0.1
13	202001	Asia	Thailand	15	0.16
14	202001	Asia	United Arab Emirates	4	0.04
15	202001	Asia	Vietnam	2	0.02
16	202001	Europe	Finland	1	5.88
17	202001	Europe	France	5	29.41
18	202001	Europe	Germany	5	29.41
19	202001	Europe	Italy	2	11.76
20	202001	Europe	Russia	2	11.76
21	202001	Europe	Switzerland	0	0
22	202001	Europe	United Kingdom	2	11.76
23	202001	North America	Canada	4	36.36
24	202001	North America	Mexico	0	0
25	202001	North America	United States	7	63.64

Copyright © 2020, All rights reserved.

3. 코로나 데이터 분석

(3) 2020년 한국의 월별 검사 수, 확진자 수, 확진율

```
SELECT TO_CHAR(issue_date, 'MM') months,
       SUM(tests) 검사수,
       SUM(cases) 확진자수,
       CASE WHEN SUM(tests) = 0 THEN 0
            ELSE ROUND(SUM(cases) /SUM(tests) * 100,2)
       END 확진율
  FROM covid19_data
 WHERE countrycode = 'KOR'
   AND TO_CHAR(issue_date, 'YYYY') = '2020'
 GROUP BY TO_CHAR(issue_date, 'MM')
 ORDER BY 1;
```

	MONTHS	검사수	확진자수	확진율
1	01	144	10	6.94
2	02	85144	3139	3.69
3	03	228952	6636	2.9
4	04	167487	988	0.59
5	05	228833	729	0.32
6	06	365429	1347	0.37
7	07	260256	1486	0.57
8	08	334347	5846	1.75
9	09	382650	3707	0.97
10	10	296411	2746	0.93
11	11	398993	8017	2.01
12	12	959542	27117	2.83

Copyright © 2020, All rights reserved.

3. 코로나 데이터 분석

(4-1) 2020년 가장 많은 확진자가 나온 상위 5개 국가 정보

```
SELECT countryname, case_num
  FROM ( SELECT a.countryname,
                SUM(b.cases) case_num
           FROM covid19_country a
          INNER JOIN covid19_data b
             ON a.countrycode = b.countrycode
          WHERE TO_CHAR(b.issue_date, 'YYYY') = '2020'
          GROUP BY a.countryname
          ORDER BY 2 DESC )
 WHERE ROWNUM <= 5
 ORDER BY 2 DESC;
```

	COUNTRYNAME	CASE_NUM
1	United States	20061902
2	India	10266674
3	Brazil	7675973
4	Russia	3127347
5	France	2677666

Copyright © 2020, All rights reserved.

3. 코로나 데이터 분석

(4-2) 2020년 가장 많은 확진자가 나온 상위 5개 국가 정보

```
SELECT countryname, case_num, seq
  FROM ( SELECT a.countryname,
                SUM(b.cases) case_num,
                ROW_NUMBER() OVER (ORDER BY SUM(b.cases) DESC ) seq
           FROM covid19_country a
          INNER JOIN covid19_data b
             ON a.countrycode = b.countrycode
          WHERE TO_CHAR(b.issue_date, 'YYYY') = '2020'
          GROUP BY a.countryname
       )
 WHERE seq <= 5
 ORDER BY 3;
```

	COUNTRYNAME	CASE_NUM	SEQ
1	United States	20061902	1
2	India	10266674	2
3	Brazil	7675973	3
4	Russia	3127347	4
5	France	2677666	5

Copyright © 2020, All rights reserved.

3. 코로나 데이터 분석

(4-3) 2020년 가장 많은 확진자가 나온 상위 5개 국가 정보 (12c)

```
SELECT countryname, case_num
  FROM ( SELECT a.countryname,
                SUM(b.cases) case_num
           FROM covid19_country a
          INNER JOIN covid19_data b
             ON a.countrycode = b.countrycode
          WHERE TO_CHAR(b.issue_date, 'YYYY') = '2020'
          GROUP BY a.countryname
       )
 ORDER BY 2 DESC
 FETCH FIRST 5 ROWS ONLY;
```

	COUNTRYNAME	CASE_NUM
1	United States	20061902
2	India	10266674
3	Brazil	7675973
4	Russia	3127347
5	France	2677666

Copyright © 2020, All rights reserved.

3. 코로나 데이터 분석

(5) 2020년 인구 대비 사망률이 높은 20개 국가는?

```
SELECT *
  FROM ( SELECT a.countryname,
                MAX(a.population) popu,
                SUM(b.deaths) death_num,
                ROUND(DECODE(MAX(a.population),0,0,
                        SUM(b.deaths) / MAX(a.population))*100,4) death_rate
           FROM covid19_country a
          INNER JOIN covid19_data b
             ON a.countrycode = b.countrycode
          WHERE TO_CHAR(b.issue_date, 'YYYY') = '2020'
          GROUP BY a.countryname
          ORDER BY 4 DESC
       )
 WHERE ROWNUM <= 20 ;
```

	COUNTRYNAME	POPU	DEATH_NUM	DEATH_RATE
1	San Marino	33938	59	0.1738
2	Belgium	11589616	19528	0.1685
3	Slovenia	2078932	2697	0.1297
4	Bosnia and Herzegovina	3280815	4050	0.1234
5	Italy	60461828	74159	0.1227
6	North Macedonia	2083380	2503	0.1201
7	Peru	32971846	37680	0.1143
8	Bulgaria	6948445	7576	0.109
9	Spain	46754783	50837	0.1087
10	Andorra	77265	84	0.1087
11	Montenegro	628062	682	0.1086
12	United Kingdom	67886004	73622	0.1084
13	Czechia	10708982	11580	0.1081
14	United States	331002647	351817	0.1063
15	Liechtenstein	38137	39	0.1023
16	France	65273512	64759	0.0992
17	Hungary	9660350	9537	0.0987
18	Mexico	128932753	125807	0.0976
19	Argentina	45195777	43245	0.0957
20	Croatia	4105268	3920	0.0955

Copyright © 2020, All rights reserved.

3. 코로나 데이터 분석

(6) 2020년 국가별 확진자와 사망자의 월별 추이

```
CREATE OR REPLACE VIEW covid19_mon_v AS
WITH covid AS (
SELECT b.countryname,
      TO_CHAR(a.issue_date, 'MM') months,
      SUM(a.cases) case_num,
      SUM(a.deaths) death_num
 FROM covid19_data a
 INNER JOIN covid19_country b
   ON a.countrycode = b.countrycode
 GROUP BY b.countryname, TO_CHAR(a.issue_date, 'MM')
)
```

```
SELECT countryname,
      '1.확진' gubun,
      SUM(CASE WHEN months = 01 THEN case_num ELSE 0 END) "01",
      SUM(CASE WHEN months = 02 THEN case_num ELSE 0 END) "02",
      SUM(CASE WHEN months = 03 THEN case_num ELSE 0 END) "03",
      SUM(CASE WHEN months = 04 THEN case_num ELSE 0 END) "04",
      SUM(CASE WHEN months = 05 THEN case_num ELSE 0 END) "05",
      SUM(CASE WHEN months = 06 THEN case_num ELSE 0 END) "06",
      SUM(CASE WHEN months = 07 THEN case_num ELSE 0 END) "07",
      SUM(CASE WHEN months = 08 THEN case_num ELSE 0 END) "08",
      SUM(CASE WHEN months = 09 THEN case_num ELSE 0 END) "09",
      SUM(CASE WHEN months = 10 THEN case_num ELSE 0 END) "10",
      SUM(CASE WHEN months = 11 THEN case_num ELSE 0 END) "11",
      SUM(CASE WHEN months = 12 THEN case_num ELSE 0 END) "12"
FROM covid
 GROUP BY countryname
 UNION ALL
```

러닝스푼즈 나노디그리 데이터 이해를 위한 SQL

Copyright © 2020, All rights reserved.

3. 코로나 데이터 분석

(6) 2020년 국가별 확진자와 사망자의 월별 추이

```
SELECT countryname,
       '2.사망' gubun,
       SUM(CASE WHEN months = 01 THEN death_num ELSE 0 END) "01",
       SUM(CASE WHEN months = 02 THEN death_num ELSE 0 END) "02",
       SUM(CASE WHEN months = 03 THEN death_num ELSE 0 END) "03",
       SUM(CASE WHEN months = 04 THEN death_num ELSE 0 END) "04",
       SUM(CASE WHEN months = 05 THEN death_num ELSE 0 END) "05",
       SUM(CASE WHEN months = 06 THEN death_num ELSE 0 END) "06",
       SUM(CASE WHEN months = 07 THEN death_num ELSE 0 END) "07",
       SUM(CASE WHEN months = 08 THEN death_num ELSE 0 END) "08",
       SUM(CASE WHEN months = 09 THEN death_num ELSE 0 END) "09",
       SUM(CASE WHEN months = 10 THEN death_num ELSE 0 END) "10",
       SUM(CASE WHEN months = 11 THEN death_num ELSE 0 END) "11",
       SUM(CASE WHEN months = 12 THEN death_num ELSE 0 END) "12"
FROM covid
 GROUP BY countryname
 ORDER BY 1, 2 ;
```

러닝스푼즈 데이터 이해를 위한 SQL 나눔바그리

Copyright © 2020, All rights reserved.

3. 코로나 데이터 분석

(6) 2020년 국가별 확진자와 사망자의 월별 추이

```
SELECT *
FROM covid19_mon_v;
```

	COUNTRYNAME	GUBUN	01	02	03	04	05	06	07	08	09	10	11	12
1	Afghanistan	1.확진	3497	692	174	1952	13081	16299	5158	1494	1109	2157	4849	5252
2	Afghanistan	2.사망	209	43	4	60	194	494	532	119	57	78	257	396
3	Albania	1.확진	19811	29040	243	530	364	1398	2741	4237	4136	7226	17307	20134
4	Albania	2.사망	199	416	15	16	2	29	95	127	103	122	301	371
5	Algeria	1.확진	7729	5754	715	3290	5388	4513	16487	14100	7036	6412	25257	16411
6	Algeria	2.사망	135	92	44	406	203	259	298	300	226	228	467	325
7	Andorra	1.확진	1888	929	376	369	19	91	70	251	874	2706	1989	1304
8	Andorra	2.사망	17	9	12	30	9	1	0	1	0	22	1	8
9	Angola	1.확진	2243	1011	7	20	59	198	864	1506	2318	5833	4334	2414
10	Angola	2.사망	61	42	2	0	2	9	39	56	75	101	64	57
11	Anguilla	1.확진	0	0	0	0	0	0	0	0	0	0	0	0
12	Anguilla	2.사망	0	0	0	0	0	0	0	0	0	0	0	0
13	Antigua and Barbuda	1.확진	75	496	7	17	2	43	22	3	7	27	13	18
14	Antigua and Barbuda	2.사망	2	7	0	3	0	0	0	0	0	0	1	1
15	Argentina	1.확진	301725	180126	1054	3374	12423	47679	126772	226433	333266	415923	257609	200981
16	Argentina	2.사망	4729	3991	27	191	321	768	2236	5117	8277	14065	7728	4515
17	Armenia	1.확진	7617	5032	532	1534	7216	16260	13008	5231	6578	39454	45311	24285
18	Armenia	2.사망	257	112	3	29	99	312	295	141	80	382	823	659
19	Australia	1.확진	402	176	4534	2207	436	718	9360	8539	1277	499	317	513
20	Australia	2.사망	0	0	18	75	10	1	97	456	231	19	1	1
21	Austria	1.확진	53583	45051	10171	5272	1279	1035	3364	6308	17375	60112	177531	78359

Copyright © 2020, All rights reserved.

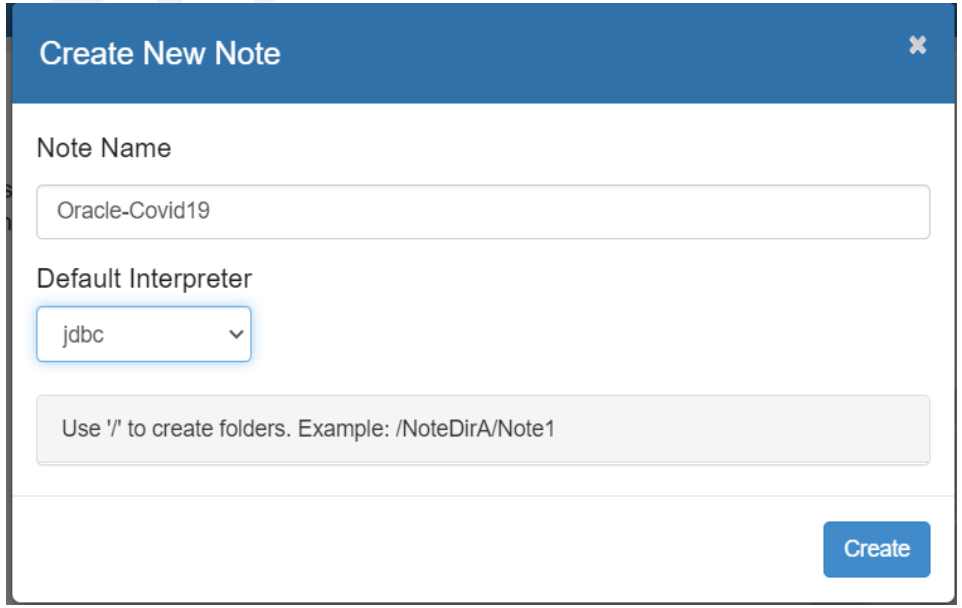
4. 코로나 데이터 분석 시각화

- Apache Zeppelin 실행

- 제플린 홈 창에서 Notebook – Create New Note 버튼 클릭

- Create New Note 창
  - Note Name : Oracle-Covid19
  - Default Interpreter : jdbc 선택
  - Create 버튼 클릭

- Notebook – Oracle-Covid19 클릭

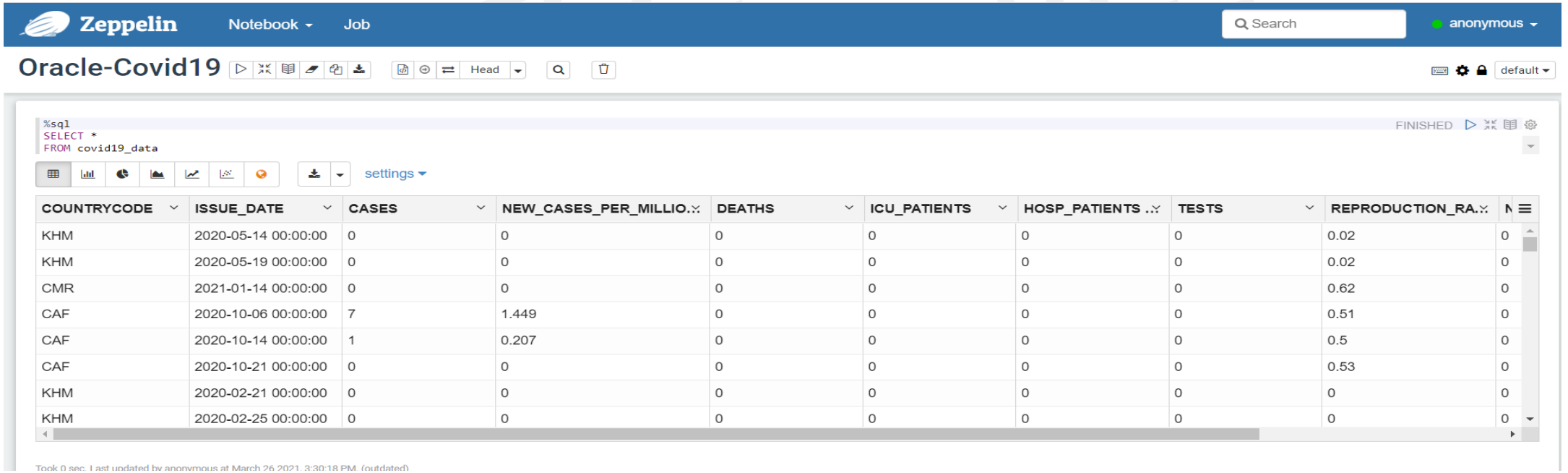


Copyright © 2020, All rights reserved.

4. 코로나 데이터 분석 시각화

- **%sql** (소문자로) 을 입력 후 SELECT 문장 입력 (세미콜론 뺄것) → shift + Enter 키

```
%sql
 SELECT *
     FROM covid19_data
```



COUNTRYCODE	ISSUE_DATE	CASES	NEW_CASES_PER_MILLIO..	DEATHS	ICU_PATIENTS	HOSP_PATIENTS ..	TESTS	REPRODUCTION_RA..	N
KHM	2020-05-14 00:00:00	0	0	0	0	0	0	0.02	0
KHM	2020-05-19 00:00:00	0	0	0	0	0	0	0.02	0
CMR	2021-01-14 00:00:00	0	0	0	0	0	0	0.62	0
CAF	2020-10-06 00:00:00	7	1.449	0	0	0	0	0.51	0
CAF	2020-10-14 00:00:00	1	0.207	0	0	0	0	0.5	0
CAF	2020-10-21 00:00:00	0	0	0	0	0	0	0.53	0
KHM	2020-02-21 00:00:00	0	0	0	0	0	0	0	0
KHM	2020-02-25 00:00:00	0	0	0	0	0	0	0	0

Copyright © 2020, All rights reserved.

4. 코로나 데이터 분석 시각화

(1) 전 기간 가장 많은 사망자를 낸 상위 10개 국가 현황 조회

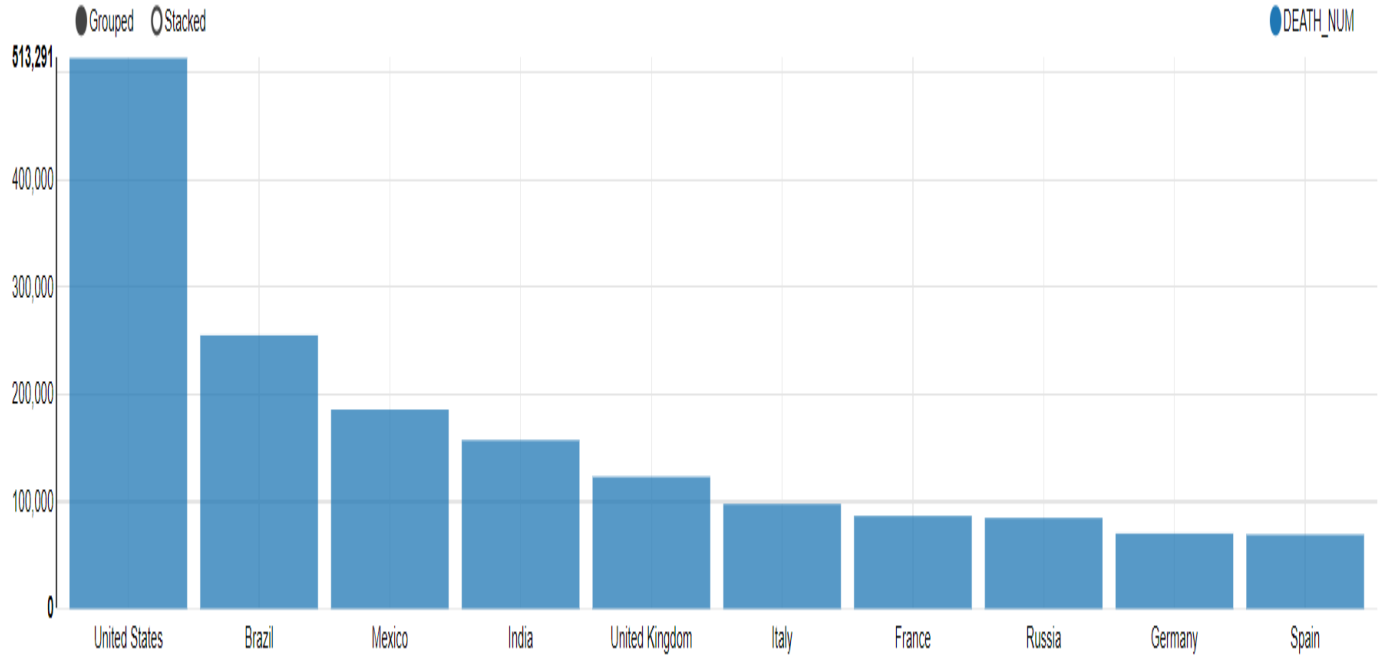
```
    %sql
SELECT *
  FROM ( SELECT b.countryname,
                SUM(a.deaths) death_num
           FROM covid19_data a
          INNER JOIN covid19_country b
             ON a.countrycode = b.countrycode
          GROUP BY b.countryname
          ORDER BY 2 DESC
       )
 WHERE ROWNUM <= 10
```

Copyright © 2020, All rights reserved.

4. 코로나 데이터 분석 시각화

(1) 전 기간 가장 많은 사망자를 낸 상위 10개 국가 현황 조회

```
%sql
SELECT *
 FROM ( SELECT b.countryname,
             SUM(a.deaths) death_num
        FROM covid19_data a
       INNER JOIN covid19_country b
          ON a.countrycode = b.countrycode
       GROUP BY b.countryname
       ORDER BY 2 DESC
     )
WHERE ROWNUM <= 10
```

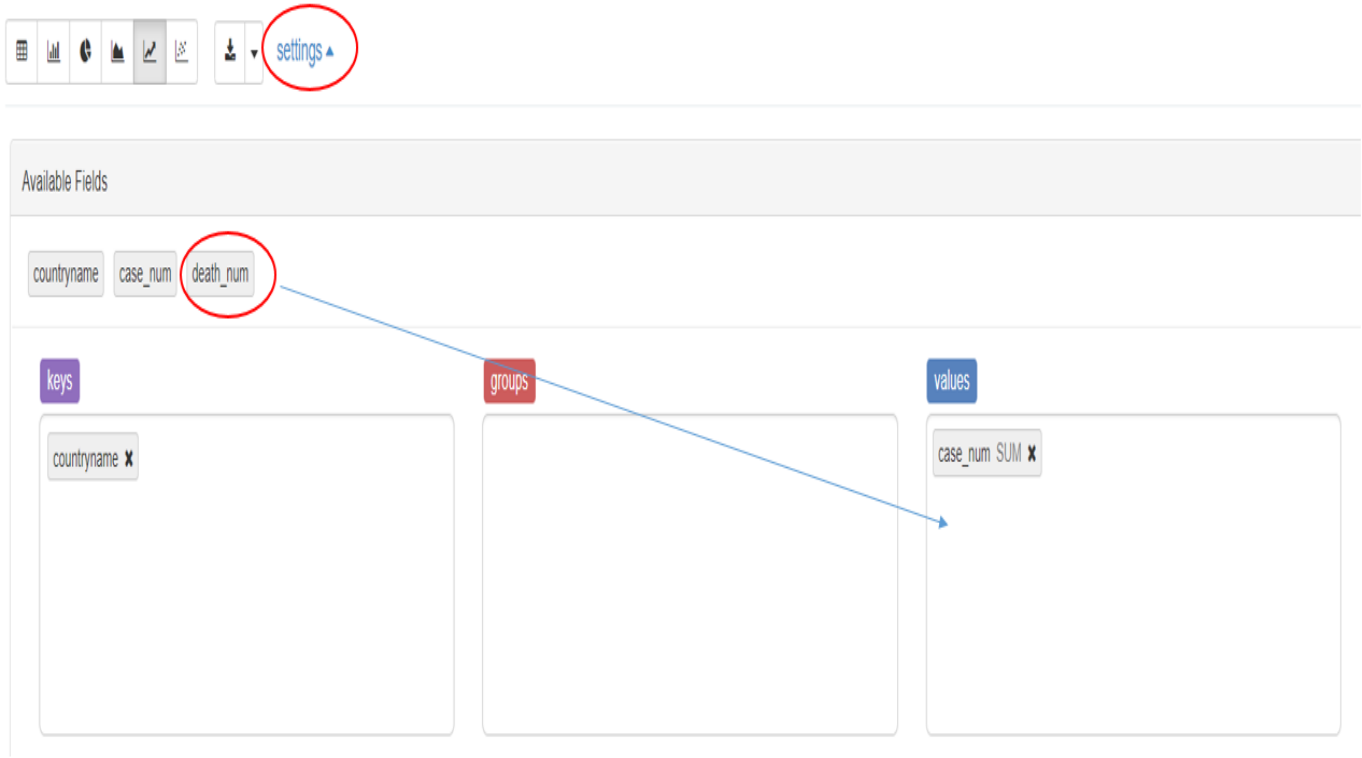


Copyright © 2020, All rights reserved.

4. 코로나 데이터 분석 시각화

(2) 전 기간 사망자 상위 10개 국가의 확진자와 사망자 현황

```
%sql
SELECT *
  FROM ( SELECT b.countryname,
                SUM(a.deaths) death_num,
                SUM(a.cases) case_num
           FROM covid19_data a
          INNER JOIN covid19_country b
             ON a.countrycode = b.countrycode
          GROUP BY b.countryname
          ORDER BY 2 DESC
       )
 WHERE ROWNUM <= 10
```

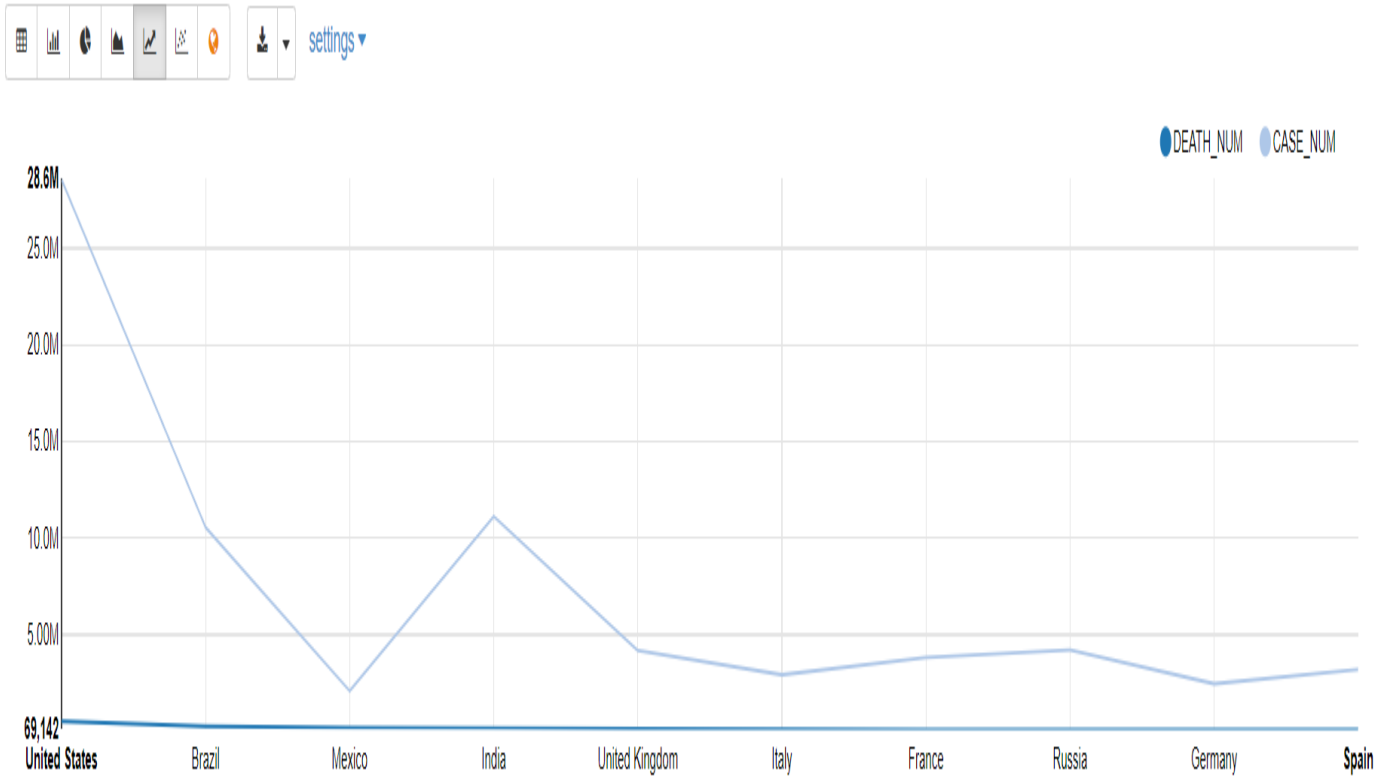


Copyright © 2020, All rights reserved.

4. 코로나 데이터 분석 시각화

(2) 전 기간 사망자 상위 10개 국가의 확진자와 사망자 현황

```
%sql
SELECT *
  FROM ( SELECT b.countryname,
                SUM(a.deaths) death_num,
                SUM(a.cases) case_num
           FROM covid19_data a
          INNER JOIN covid19_country b
             ON a.countrycode = b.countrycode
          GROUP BY b.countryname
          ORDER BY 2 DESC
       )
 WHERE ROWNUM <= 10
```

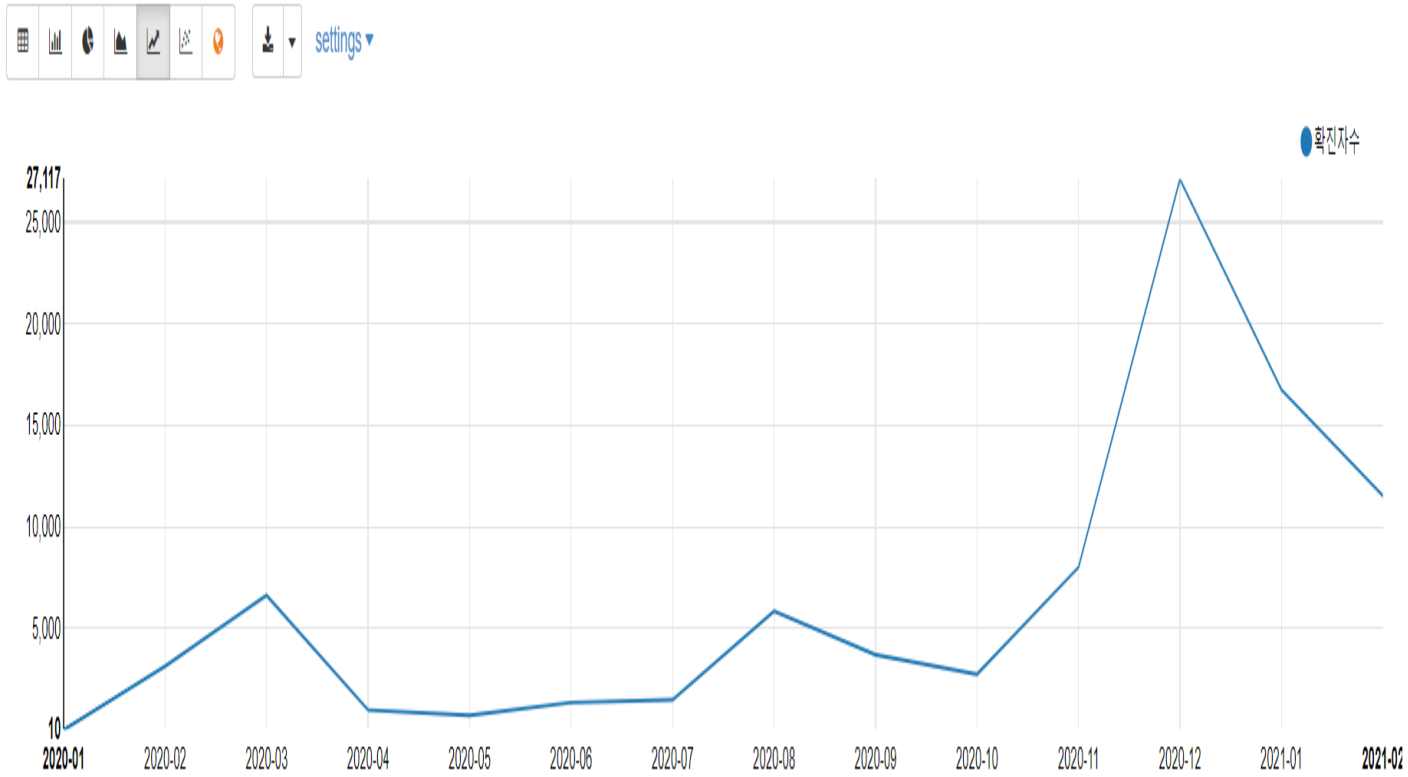


Copyright © 2020, All rights reserved.

4. 코로나 데이터 분석 시각화

(3) 한국의 월별 확진자 현황

```
%sql
SELECT TO_CHAR(issue_date, 'YYYY-MM') months,
       SUM(cases) 확진자수
  FROM covid19_data
 WHERE countrycode = 'KOR'
 GROUP BY TO_CHAR(issue_date, 'YYYY-MM')
 ORDER BY 1
```



Copyright © 2020, All rights reserved.

4. 코로나 데이터 분석 시각화

(4) 사망자 상위 5개 국가의 월별 사망자 현황

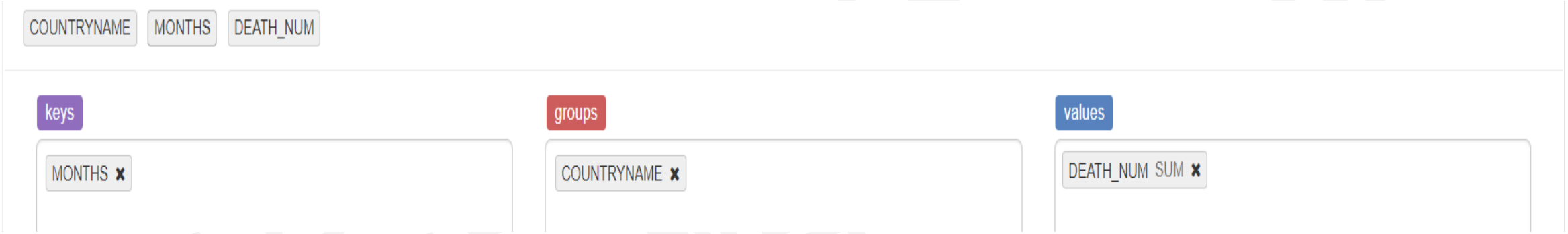
```
 %sql
WITH covid1 AS (
SELECT a.countrycode,   SUM(a.deaths) death_num
  FROM covid19_data a
 GROUP BY a.countrycode
 ORDER BY 2 DESC
),
covid2 AS (
SELECT *
  FROM covid1
 WHERE ROWNUM <= 5 )
```

```
SELECT b.countryname,
      TO_CHAR(a.issue_date, 'YYYY-MM') months,
      SUM(a.deaths) death_num
 FROM covid19_data a
INNER JOIN covid19_country b
   ON a.countrycode = b.countrycode
WHERE EXISTS ( SELECT 1
                 FROM covid2 c
                WHERE a.countrycode = c.countrycode)
GROUP BY b.countryname, TO_CHAR(a.issue_date, 'YYYY-MM')
ORDER BY 2
```

Copyright © 2020, All rights reserved.

4. 코로나 데이터 분석 시각화

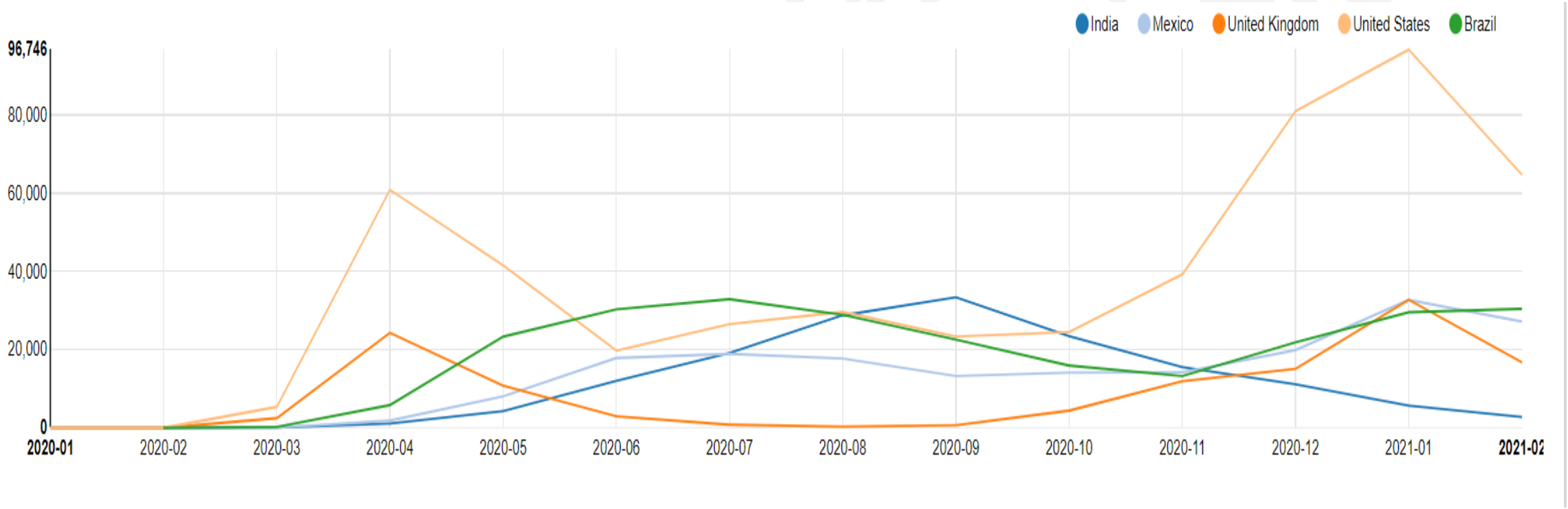
(4) 사망자 상위 5개 국가의 월별 사망자 현황



Copyright © 2020, All rights reserved.

4. 코로나 데이터 분석 시각화

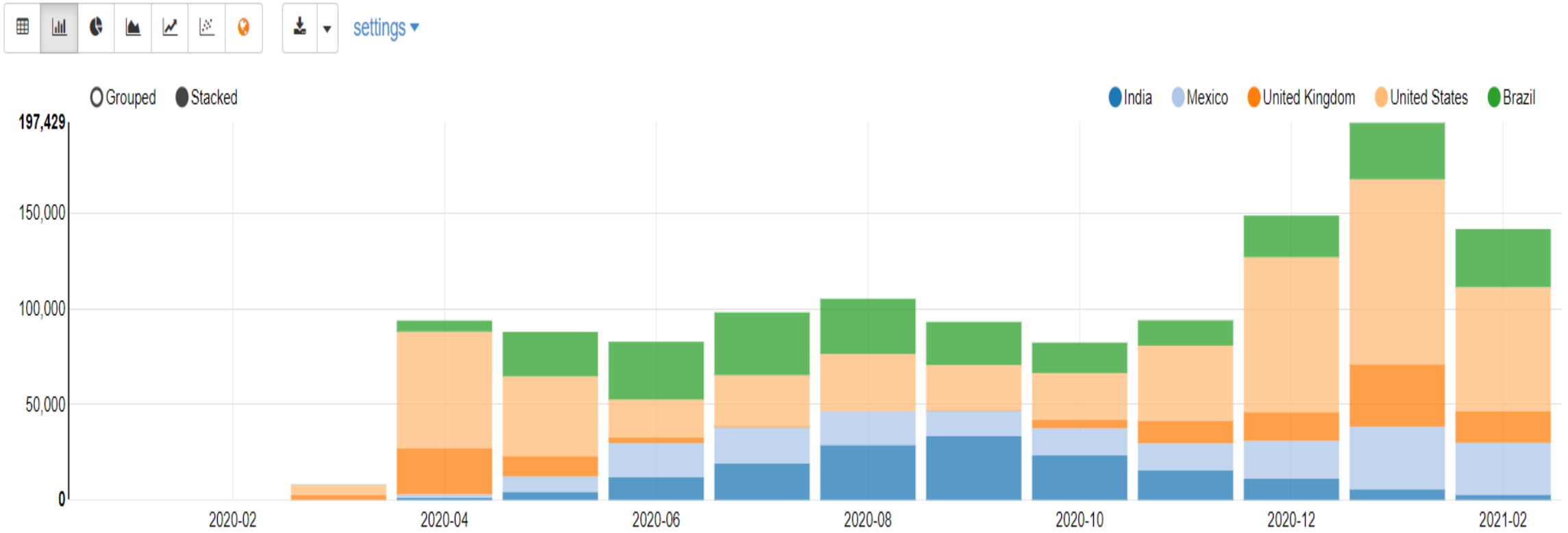
(4) 사망자 상위 5개 국가의 월별 사망자 현황



Copyright © 2020, All rights reserved.

4. 코로나 데이터 분석 시각화

(4) 사망자 상위 5개 국가의 월별 사망자 현황



Copyright © 2020, All rights reserved.