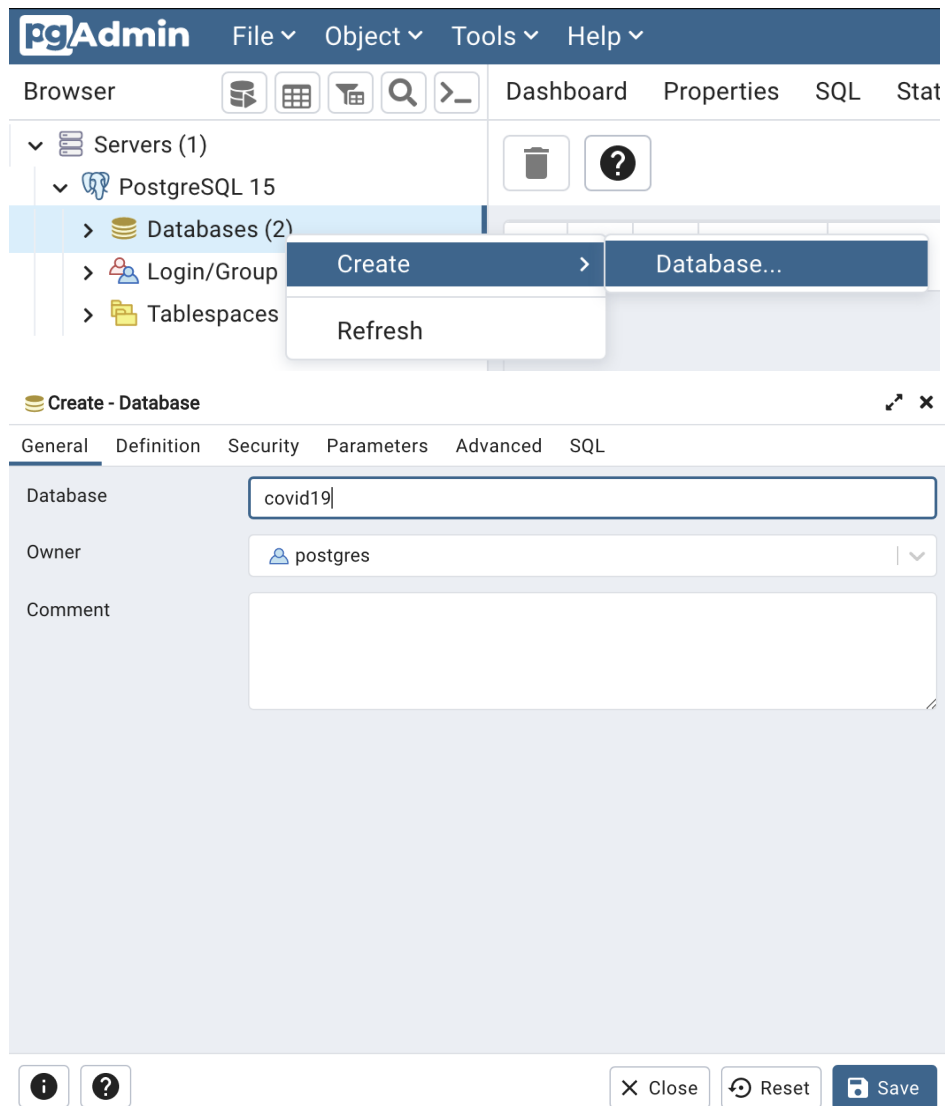# Database System HW1 by 吳承瑀 110550091

1. The process of creating the "covid19" databases.

I downloaded the required files given in PostgreSQL engines and opened pgAdmin4.
Inside the created servers, I clicked on 'Databases' and created a new database and named it 'covid19'.

2. The process of importing three required .csv files into covid19 database.

I downloaded the files given in HW1 requests by downloading the csv files. Then, I followed the tutorial of importing csv files into postgreSQL tables. The tutorial was quite clear, so there wasn't much problesm encountered here.

I opened the Query Tool in pgAdmin4 and created the following three tables: FirstOfOct, EleventhOfOct, and CodeList (I used all lowercase letters since SQL is case insensitive).

FirstOfOct and EleventhOfOct are basically the same with different dates. It contains multiple integer, varchar, real attributes and a timestamp. 50 and 60 was the estimated suitable number for those varchars. The primary key was chosen to be 'Combined_Key'.

```
1   create table firstofoct(
2       FIPS int,
3       Admin2 varchar(50),
4       Province_State varchar(50),
5       Country_Region varchar(50),
6       Last_Update timestamp,
7       Lat real,
8       Long real,
9       Confirmed int,
10      Deaths int,
11      Recovered int,
12      Active int,
13      Combined_Key varchar(60),
14      Incident_Rate real,
15      Case_Fatality_Ratio real,
16      primary key (Combined_Key)
17  );
18
19  copy firstofoct
20  from '/private/tmp/10-01-2022.csv'
21  delimiter ','
22  csv header;
```

```
24  create table eleventhofoct(
25      FIPS int,
26      Admin2 varchar(50),
27      Province_State varchar(50),
28      Country_Region varchar(50),
29      Last_Update timestamp,
30      Lat real,
31      Long real,
32      Confirmed int,
33      Deaths int,
34      Recovered int,
35      Active int,
36      Combined_Key varchar(60),
37      Incident_Rate real,
38      Case_Fatality_Ratio real,
39      primary key (Combined_Key)
40  );
41
42  copy eleventhofoct
43  from '/private/tmp/10-11-2022.csv'
44  delimiter ','
45  csv header;
```

CodeList has multiple varchar, char, and int attributes. The primary key is set to be 'Continent_Code' and 'Two_Letter_Country_Code'.

```
47  create table codelist(
48      Continent_Name varchar(20),
49      Continent_Code char(2),
50      Country_Name varchar(60),
51      Two_Letter_Country_Code char(2),
52      Three_Letter_Country_Code char(3),
53      Country_Number int,
54      primary key (Continent_Code, Two_Letter_Country_Code)
55  );
56
57  copy codelist
58  from '/private/tmp/country-and-continent-codes-list.csv'
59  delimiter ','
60  csv header;
```

Using '\dt' in the PSQL Tool, we can see the list of relations we have created.

```
covid19=# \dt
            List of relations
 Schema |     Name      | Type  |  Owner
--------+---------------+-------+----------
 public | codelist      | table | postgres
 public | eleventhofoct | table | postgres
 public | firstofoct    | table | postgres
(3 rows)
```

3. The SQL statements and output results of 4a.

The sum of confirmed cases in US, California on October 11th.

There is a total of 11,310,690 cases on October 11th in California.

```
covid19=# select sum(confirmed)
covid19-# from eleventhofoct
covid19-# where country_region = 'US' and province_state = 'California';
   sum
----------
 11310690
(1 row)
```

4. The SQL statements and output results of 4b.

The sum of confirmed cases in US, California on October 1st.

There is a total of 11,268,292 cases on October 1st in California.

```
covid19=# select sum(confirmed)
covid19-# from firstofoct
covid19-# where country_region = 'US' and province_state = 'California';
   sum
----------
 11268292
(1 row)
```

5. The SQL statements and output results of 4c.

The difference between confirmed cases in US, California on October 11th and October 1st.

The difference is 42,398 cases.

```
covid19=# select sum(a.confirmed - b.confirmed)
covid19-# from eleventhofoct as a, firstofoct as b
covid19-# where a.country_region = 'US' and a.province_state = 'California'
covid19-# and b.country_region = 'US' and b. province_state = 'California'
covid19-# and a.admin2 = b.admin2;
  sum
-------
 42398
(1 row)
```

6. The SQL statements and output results of 4d.

Countries that have more than 20 million confirmed cases on October 11th.

There is a total of 10 countries that have more than 20 million cases on October 11th and the countries are the following in descending order: US, India, France, Brazil, Germany, South Korea, UK, Italy, Japan, and Russia.

```
covid19=# with countrycases(country_region, casetotal) as
covid19-# (select country_region, sum(confirmed)
covid19(# from eleventhofoct
covid19(# group by country_region)
covid19-#
covid19-# select country_region, casetotal
covid19-# from countrycases
covid19-# where casetotal > 20000000
covid19-# order by casetotal desc;
 country_region | casetotal
----------------+-----------
 US             |  96785810
 India          |  44616235
 France         |  36187658
 Brazil         |  34731539
 Germany        |  34257916
 Korea, South   |  25025749
 United Kingdom |  23957457
 Italy          |  22896742
 Japan          |  21593704
 Russia         |  20929929
(10 rows)
```

7. The SQL statements and output results of 4e.

Countries in Asia that have more than 20 million confirmed cases on October 11th.

The Asian countries are the following in descending order: India, South Korea, Japan and Russia. (*Note: I counted Russia as an Asian country, not an European country)

```
covid19=# with countrycases(country_region, casetotal) as
covid19-# (select country_region, sum(confirmed)
covid19(# from eleventhofoct
covid19(# group by country_region),
covid19-# twentymil(country_region, casetotal) as
covid19-# (select country_region, casetotal
covid19(# from countrycases
covid19(# where casetotal > 20000000)
covid19-#
covid19-# select t.country_region, t.casetotal
covid19-# from twentymil as t, codelist as c
covid19-# where t.country_region = c.country_name
covid19-# and c.continent_code = 'AS'
covid19-# order by t.casetotal desc;
 country_region | casetotal
----------------+-----------
 India          |  44616235
 Korea, South   |  25025749
 Japan          |  21593704
 Russia         |  20929929
(4 rows)
```

8. The SQL statements and output results of 4f.

Countries that have more than 100,000 new cases since October 1st to October 11th.

There are a total of 10 countries that fits the requirements. The countries are as the following in descending order: Germany, France, Taiwan, Italy, US, Japan, Russia, South Korea, Austria, and Greece.

```
covid19=# with countrycases11(country_region, casetotal) as
covid19-# (select country_region, sum(confirmed)
covid19(# from eleventhofoct
covid19(# group by country_region),
covid19-# countrycases1(country_region, casetotal) as
covid19-# (select country_region, sum(confirmed)
covid19(# from firstofoct
covid19(# group by country_region),
covid19-# newcases(country_region, newcase) as
covid19-# (select a.country_region, sum(a.casetotal - b.casetotal)
covid19(# from countrycases11 as a, countrycases1 as b
covid19(# where a.country_region =  b.country_region
covid19(# group by a.country_region)
covid19-#

covid19-# select *
covid19-# from newcases
covid19-# where newcase > 100000
covid19-# order by newcase desc;
 country_region | newcase
----------------+---------
 Germany        |  871687
 France         |  579373
 Taiwan*        |  440596
 Italy          |  396396
 US             |  388817
 Japan          |  264185
 Russia         |  212106
 Korea, South   |  206138
 Austria        |  129544
 Greece         |  106302
(10 rows)
```