Covid-19 Cases and Vaccines

Lindsay Brenner and Ari Sherman

https://github.com/arisherm817/SI-Final-Project

1 Goals and Goals Achieved

Goals:

- To compare COVID-19 cases to the number of people vaccinated in each per country.
- Use one API and one website.
- Join and make a calculation using the gathered data.
- Create at least two visual representations of the gathered data.

Goals Achieved:

- Gathered data from the COVID API and the wikipedia table on COVID vaccinations per country.
- Joined data and made multiple calculations.
    - Found the top ten COVID-19 death rates per country, top ten fewest number of confirmed cases per vaccinated person per country, top ten countries with the highest COVID-19 death rates, world average percent vaccinated, and the top ten highest vaccination percentages in the world.
- Displayed the data visually using the python package, Plotly.

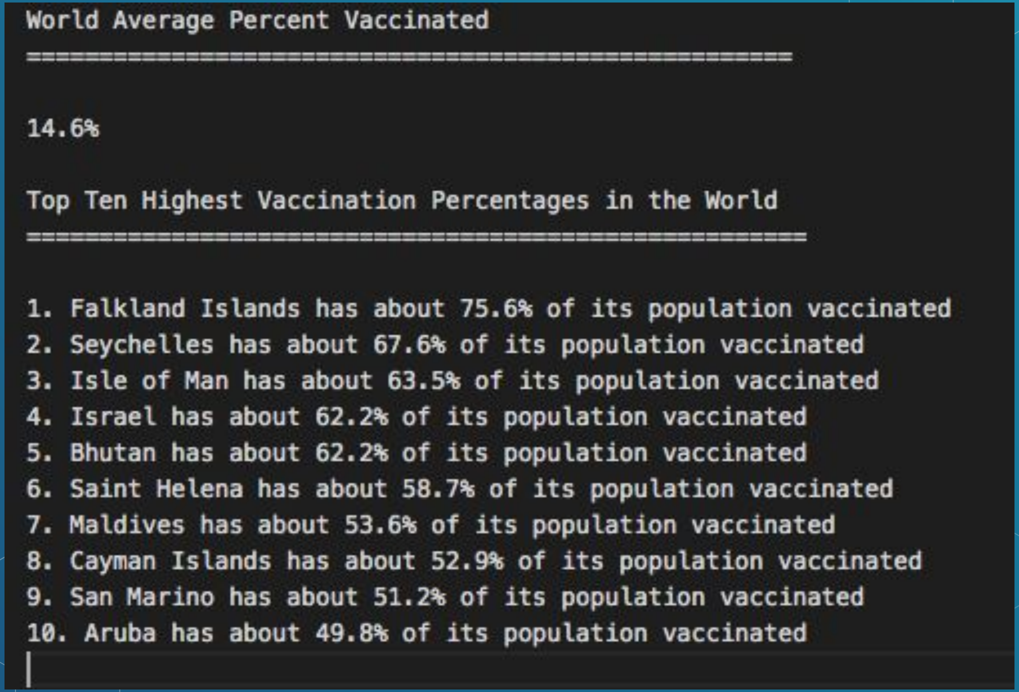
2 Problems

Problems:

- We originally planned to use two APIs, however when trying to access the data from our second API (Uber), we were unable to access the data because of the requirement for an OAuth key.
- It was very difficult to find a website with a good vaccine data because complex JavaScript concepts were needed to inspect the ‘Show more’ buttons.
- We experienced a ‘503 Service Temporarily Unavailable’ error when trying to gather data from the COVID API.
  - Unable to use the API for multiple hours.

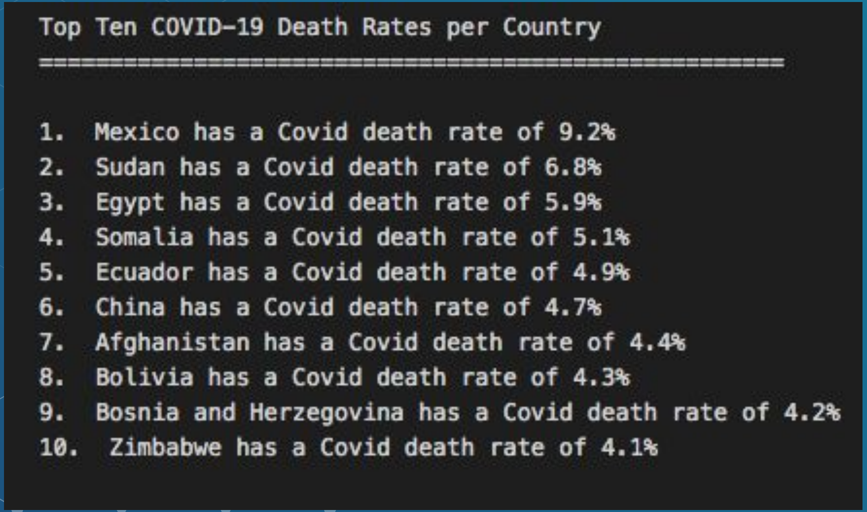
3 Calculations

Data calculated from gathering-vaccine-data.py

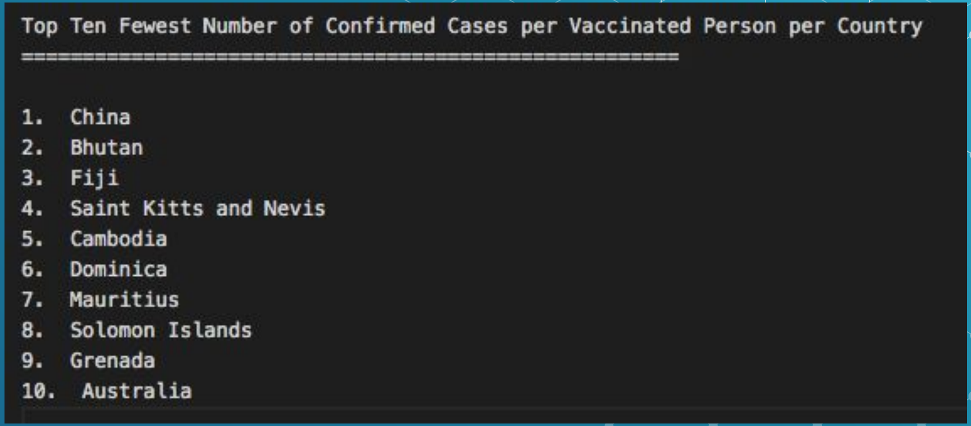


```
World Average Percent Vaccinated
=====================================================


14.6%

Top Ten Highest Vaccination Percentages in the World
======================================================


1. Falkland Islands has about 75.6% of its population vaccinated
2. Seychelles has about 67.6% of its population vaccinated
3. Isle of Man has about 63.5% of its population vaccinated
4. Israel has about 62.2% of its population vaccinated
5. Bhutan has about 62.2% of its population vaccinated
6. Saint Helena has about 58.7% of its population vaccinated
7. Maldives has about 53.6% of its population vaccinated
8. Cayman Islands has about 52.9% of its population vaccinated
9. San Marino has about 51.2% of its population vaccinated
10. Aruba has about 49.8% of its population vaccinated
```

Data calculated from gathering-covid-data.py

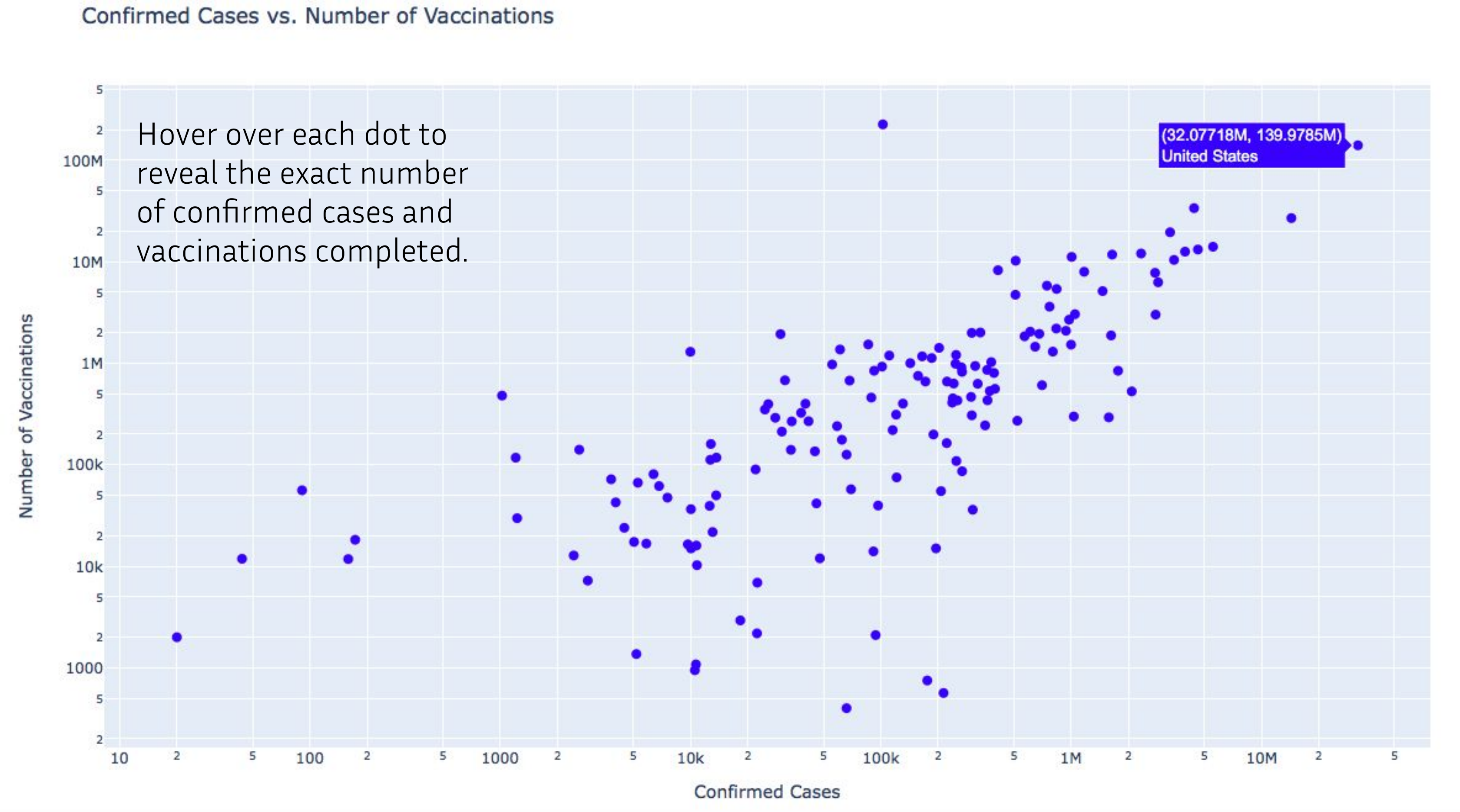


```
Top Ten COVID-19 Death Rates per Country
=====================================================

1.  Mexico has a Covid death rate of 9.2%
2.  Sudan has a Covid death rate of 6.8%
3.  Egypt has a Covid death rate of 5.9%
4.  Somalia has a Covid death rate of 5.1%
5.  Ecuador has a Covid death rate of 4.9%
6.  China has a Covid death rate of 4.7%
7.  Afghanistan has a Covid death rate of 4.4%
8.  Bolivia has a Covid death rate of 4.3%
9.  Bosnia and Herzegovina has a Covid death rate of 4.2%
10.  Zimbabwe has a Covid death rate of 4.1%
```



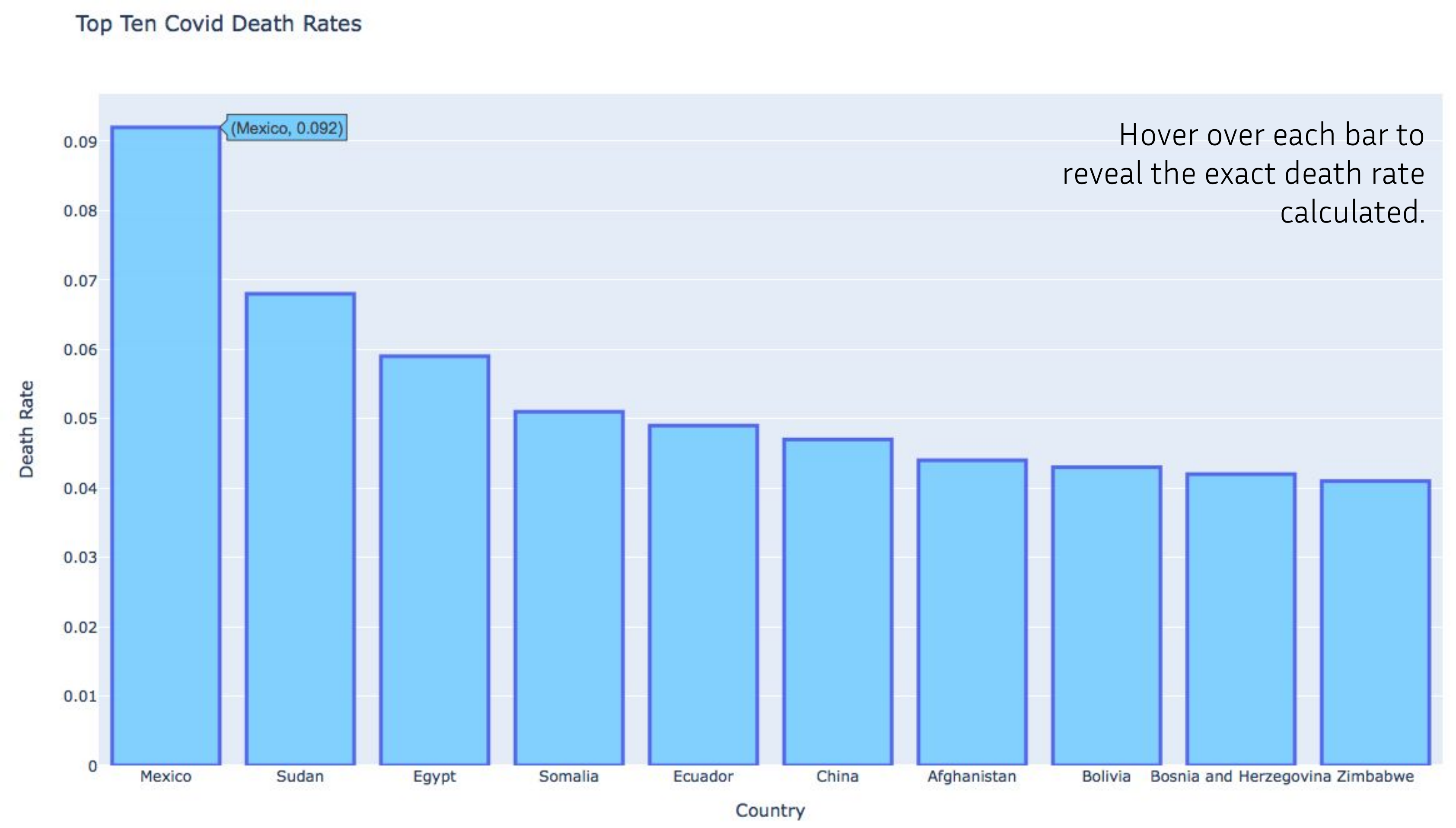
```
Top Ten Fewest Number of Confirmed Cases per Vaccinated Person per Country
=====================================================================

1.  China
2.  Bhutan
3.  Fiji
4.  Saint Kitts and Nevis
5.  Cambodia
6.  Dominica
7.  Mauritius
8.  Solomon Islands
9.  Grenada
10.  Australia
```

4 Visualizations



Confirmed Cases vs. Number of Vaccinations

Hover over each dot to reveal the exact number of confirmed cases and vaccinations completed.



Hover over each bar to reveal the exact death rate calculated.

5 Instructions

Instructions:

1. Ensure that the database, **‘vaccine.db’** does not exist in your files. If it does exist, delete it.
2. Install the python plotly package by typing **‘pip install plotly==4.14.3’** into your terminal window.

Instructions (continued):

3. Open the project folder. Within this folder, there should be four files. After opening this folder, open and run the file entitled **‘gathering-vaccine-data.py’** eight times. This file creates the **‘vaccine.db’** database which will appear in the project folder after running the program. This database creates two tables both containing 190 rows of data. This program also creates the **‘vaccine_data.txt’** file which contains the formatted results from calculations made in the program.

Instructions (continued):

4. The next file that should be run is the **‘gathering-covid-data.py’** file. Run this program seven times. This program adds a third table to the vaccine database. The new table contains information regarding covid cases and deaths in 151 countries worldwide. This program also creates the **‘covid_data.txt’** file which contains formatted calculations made in the program.

Instructions (continued):

5. The final file that should be run is the **‘visuals.py’** file. This program creates two visualizations of the data in the database and opens them in an internet window on your computer.

6. After running the three python programs, there should be three new files in the project folder: 1 database containing 3 tables and two text files containing formatted calculations.

6 Code Documentation

‘Gathering-vaccine-data.py’:

```
def set_up_database(name):
   """Takes in the name of a database an input and returns the cursor and connection to the database."""
def get_vaccine_number_data():
""" Takes in a website url and uses BeautifulSoup to locate and  read the countries name and total vaccinations. Returns a list of tuples in the format (Country, Total Vaccinations)."""
def get_vaccine_percent_data():
"""Returns a list of tuples in the format (Country, Percent Vaccinated). Uses BeautifulSoup to read the countries name and percent vaccinated"""
```

‘Gathering-vaccine-data.py’ continued:

```
def fill_vaccine_number_table(cur, conn):
 """ Fills the Vaccine Number Table table with the country names and the total
number of vaccinations in the country"""
def fill_vaccine_percent_table(cur, conn):
   """ Fills the Vaccine Table table with the country names and the percent of the
country population that is vaccinated"""
def set_up_vaccine_tables(cur, conn):
  """Creates two tables. One table that contain the country vaccination rankings and
another table that has vaccine data for each country """
def calculate_average_percent_vaccinated(cur):
  """Calculates the world average percent vaccinated"""
```

‘Gathering-vaccine-data.py’ continued:

```
def top_ten_percentages(cur):
    """Sorts the countries by percent vaccinated and returns a list of tuples with the country
id and percent vaccinated """
def write_data_file(filename, cur, conn):
""”Writes the world average percent vaccinated and the top ten high vaccination
percentages to a filename that is given in the input"""
def main():
 """Calls the functions set_up_database(), set_up_vaccine_tables(),
fill_vaccine_number_table(), fill_vaccine_percent_table(), and write_data_file(). Closes the
database connection. """
```

‘Gathering-covid-data.py’:

```
def set_up_covid_table(cur, conn):
   """Pulls data from Covid API and sets up table with covid data """
def calculate_death_rate(cur):
   """Uses information form Covid table to calculate the Covid death rate.
Returns country names and death rates in descending order"""
def join_table(cur, conn):
   """Joins the Vaccine and Covid table. Returns a list of tuples with country
names, total number vaccinated, and total number of confirmed cases """
```

‘Gathering-covid-data.py’ continued:

```
def calculate_cases_per_vaccine(cur, conn):
  """Uses the joined table information to calculate the number of cases per vaccine in
each country. Returns a list of tuples with the country name and the cases per
vaccine"""
def write_data_file(filename, cur, conn):
  """Writes the top ten Covid-19 death rates per country and the top ten fewest number
of confirmed cases per vaccinated person to the filename given in the input """
def main():
  """Calls the functions set_up_covid_table(), calculate_death_rate(),
calculate_cases_per_vaccine(), write_data_file(). Closes connection to database"""
```

‘Visuals.py’:

```
def main():
   """ Takes in no inputs and returns nothing. Fetches data from the vaccine
database and loads the information into lists. Usings the imputed python
package, plotly to create a scatterplot of confirmed cases vs. number of
vaccinations per country and a barplot of the ten countries with the highest
death rates. """
```

7 Resources

Date	Issue Description	Location of Resource	Result (How we solved the issue).
4/14/21	Need for JavaScript to overcome the ‘Show More’ button on websites.	Office hours and https://en.wikipedia.org/wiki/Deployment_of_COVID-19_vaccines	Went to office hours and found a website that didn’t require JavaScript.
4/19/21	Unsure how to limit data sent to the database to 25.	https://realpython.com/python-mysql/	Attended office hours and researched methods of limiting data.
4/19/21	Country names were slightly different on the website and API.	https://stackoverflow.com/questions/60100484/consolidate-different-alternative-names-of-countries-under-the-official-name	Only matched countries that had the same name in both sources.

Date	Issue Description	Location of Resource	Result (How we solved the issue)
4/22/21	Our calculations resulted in long sequences of decimals.	https://www.w3schools.com/python/ref_func_round.asp	Researched methods of rounding numbers in python.
4/22/21	503 Service Temporarily Unavailable error	https://ubiq.co/tech-blog/fix-503-service-temporarily-unavailable-error-nginx/	We waited and checked the API until the service was back up.
4/24/21	Unable to label points on the scatter plot.	https://plotly.com/python/text-and-annotations/	Added a list of the country names to the points on the scatter plot.

8

Tables

Shared key: Country_id

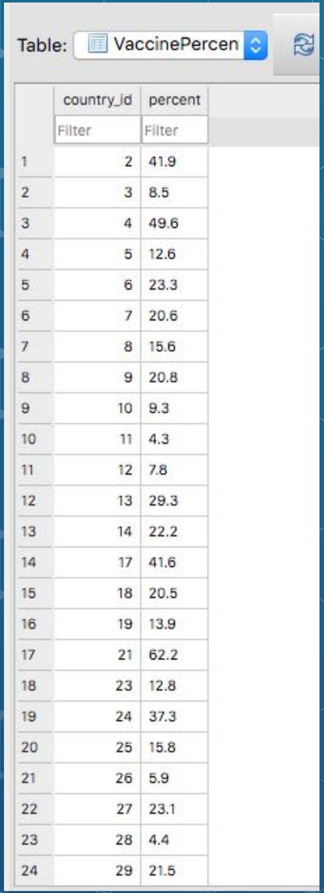


Table: VaccinePercen

	country_id	percent
	Filter	Filter
1	2	41.9
2	3	8.5
3	4	49.6
4	5	12.6
5	6	23.3
6	7	20.6
7	8	15.6
8	9	20.8
9	10	9.3
10	11	4.3
11	12	7.8
12	13	29.3
13	14	22.2
14	17	41.6
15	18	20.5
16	19	13.9
17	21	62.2
18	23	12.8
19	24	37.3
20	25	15.8
21	26	5.9
22	27	23.1
23	28	4.4
24	29	21.5

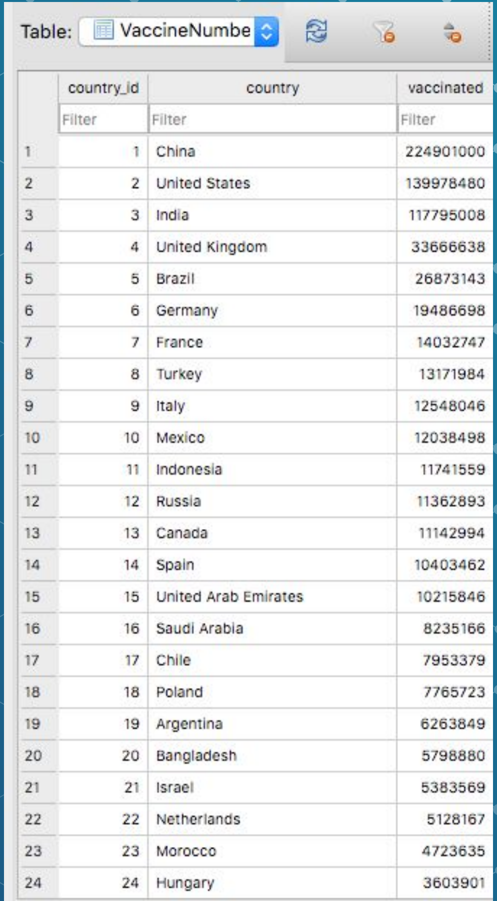
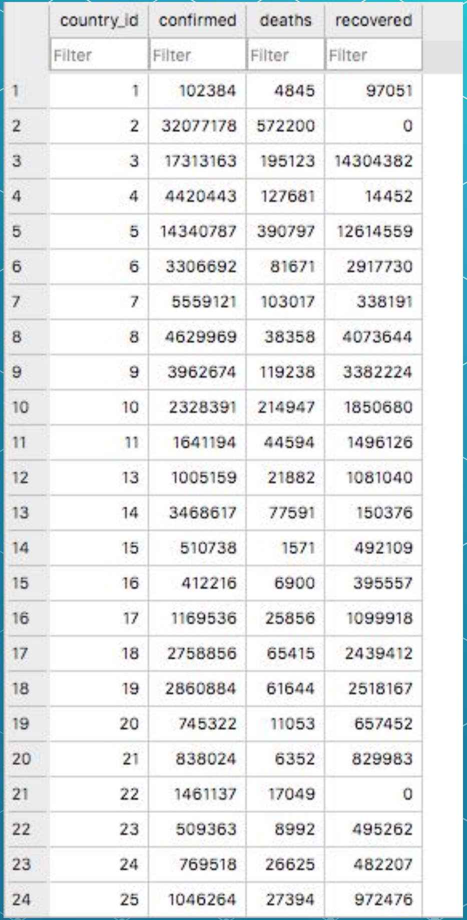


Table: VaccineNumbe

	country_id	country	vaccinated
	Filter	Filter	Filter
1	1	China	224901000
2	2	United States	139978480
3	3	India	117795008
4	4	United Kingdom	33666638
5	5	Brazil	26873143
6	6	Germany	19486698
7	7	France	14032747
8	8	Turkey	13171984
9	9	Italy	12548046
10	10	Mexico	12038498
11	11	Indonesia	11741559
12	12	Russia	11362893
13	13	Canada	11142994
14	14	Spain	10403462
15	15	United Arab Emirates	10215846
16	16	Saudi Arabia	8235166
17	17	Chile	7953379
18	18	Poland	7765723
19	19	Argentina	6263849
20	20	Bangladesh	5798880
21	21	Israel	5383569
22	22	Netherlands	5128167
23	23	Morocco	4723635
24	24	Hungary	3603901

Gathering-vaccine data.py



	country_id	confirmed	deaths	recovered
	Filter	Filter	Filter	Filter
1	1	102384	4845	97051
2	2	32077178	572200	0
3	3	17313163	195123	14304382
4	4	4420443	127681	14452
5	5	14340787	390797	12614559
6	6	3306692	81671	2917730
7	7	5559121	103017	338191
8	8	4629969	38358	4073644
9	9	3962674	119238	3382224
10	10	2328391	214947	1850680
11	11	1641194	44594	1496126
12	13	1005159	21882	1081040
13	14	3468617	77591	150376
14	15	510738	1571	492109
15	16	412216	6900	395557
16	17	1169536	25856	1099918
17	18	2758856	65415	2439412
18	19	2860884	61644	2518167
19	20	745322	11053	657452
20	21	838024	6352	829983
21	22	1461137	17049	0
22	23	509363	8992	495262
23	24	769518	26625	482207
24	25	1046264	27394	972476

Gathering-covid-data.py

THANKS!

ANY QUESTIONS?