SI 206

FINAL PROJECT

https://github.com/emmiewells/SI206-FINAL

By: Emmie Wells, Lionel Robert, and Michelle Xie

MAIN FOCUS: THE COVID-19 PANDEMIC

Are the numbers in deaths of COVID-19 comparable to those of Pneumonia and Influenza? What age group is the most susceptible? Which gender has the most cases?

DISCLAIMER

All of our data does not account for non-binary or transgender people because the World Health Organization has not included it in their data sets. In addition, not all countries have made public their death count to the World Health Organization and thus is absent.

GOALS

- Use at least three APIs/websites
- Create at least three visualizations; two from calculations
- Initially, we were going to focus on the hospitalizations and use of ventilators for those that were infected
- We thought it could be interesting to measure cases by state, city, and county in the United States
- We thought about the possibility of using testing locations and coordinates

GOALS ACHIEVED

- Found four APIs, but used three (not covid-tracking API)
- Created five visualizations; three from calculations
- We were able to find the connection between pre-existing health conditions and COVID-19 deaths
- We collected data from countries all over the world and states within the US

PROBLEMS

- We had to find new APIs multiple times because they either had too little data or too much data- this made it challenging for making specifications within our code
- Not all countries (ex: China and Russia) provided updated information in regards to death and gender
- On multiple accounts we merged the wrong branches within git and had a few technical difficulties

CALCULATIONS

— 07

Global COVID Calculations
The country with the highest COVID case count is United States of America, with a total case count of 15,165,159, accounting for about 22% of all cases in the world!

The country with the highest COVID death count is United States of America, with a total death count of 286,249, accounting for 18% of all death.s

The country with the highest COVID recovery count is India, with a total recovery count of 9,215,581, accounting for 21% of all recoveries in the world!

US State COVID Calculations

Out of all of the US States, the state with the highest COVID cases is California with 1,389,707 cases, which make up about 9% of all cases countrywide.

Thus far, the deadliest US state is New York with 27,307 deaths, which account for 10% of all deaths.

But on the lighter side, the state with the most amount of recoveries is Texas with 1,050,416 total recovered and 18% of all recoveries!

USA Gender Data

In the US alone, men are dying at alarmingly higher rates than women, accounting for roughly 54% of all deaths right now. Women only make up about 46%%.
In terms of age group between the genders, the overall most vulnerable group to COVID are ages between 85 years and over.
Men however typically die a bit younger with their most vulnerable age group of 75–84 years taking the most amount of deaths while women are typically more vulnerable around 85 years and over.
Furthermore, we decided to compare COVID deaths with deaths of other and similiar conditions that it is consistently compared to, for example Pneumonia and Influenza. For all of America, COVID19 and Pneumonia deaths are roughly equal and account for about 10% of all deaths while the flu accounts for only about 0.25% of all *reported* related deaths thus far this year. Men tend to die from COVID 10% of the time compared to pneumonia and the flu while women tend to die from COVID 9% of the time.

Global Gender Data

Unfortunately, for a lot of countries, the gender distribution data is not provided. So take this data with a grain of salt.
In terms of gender data, the highest COVID count providing this data is Mexico, besides the United States. There, men account for 74% of all COVID deaths so far with females accounting for only 42%. That's quite alarming.
However, strangely enough, women make up the majority of the population at 51%, outnumbering men who make up only 49%

What is perhaps even more surprising, we looked at the age groups between men and women to see if there was a gender difference here.
Age Group: 0–4, Female Deaths: 95, Male Deaths: 102, Male Death Percentage 0.5177664974619289

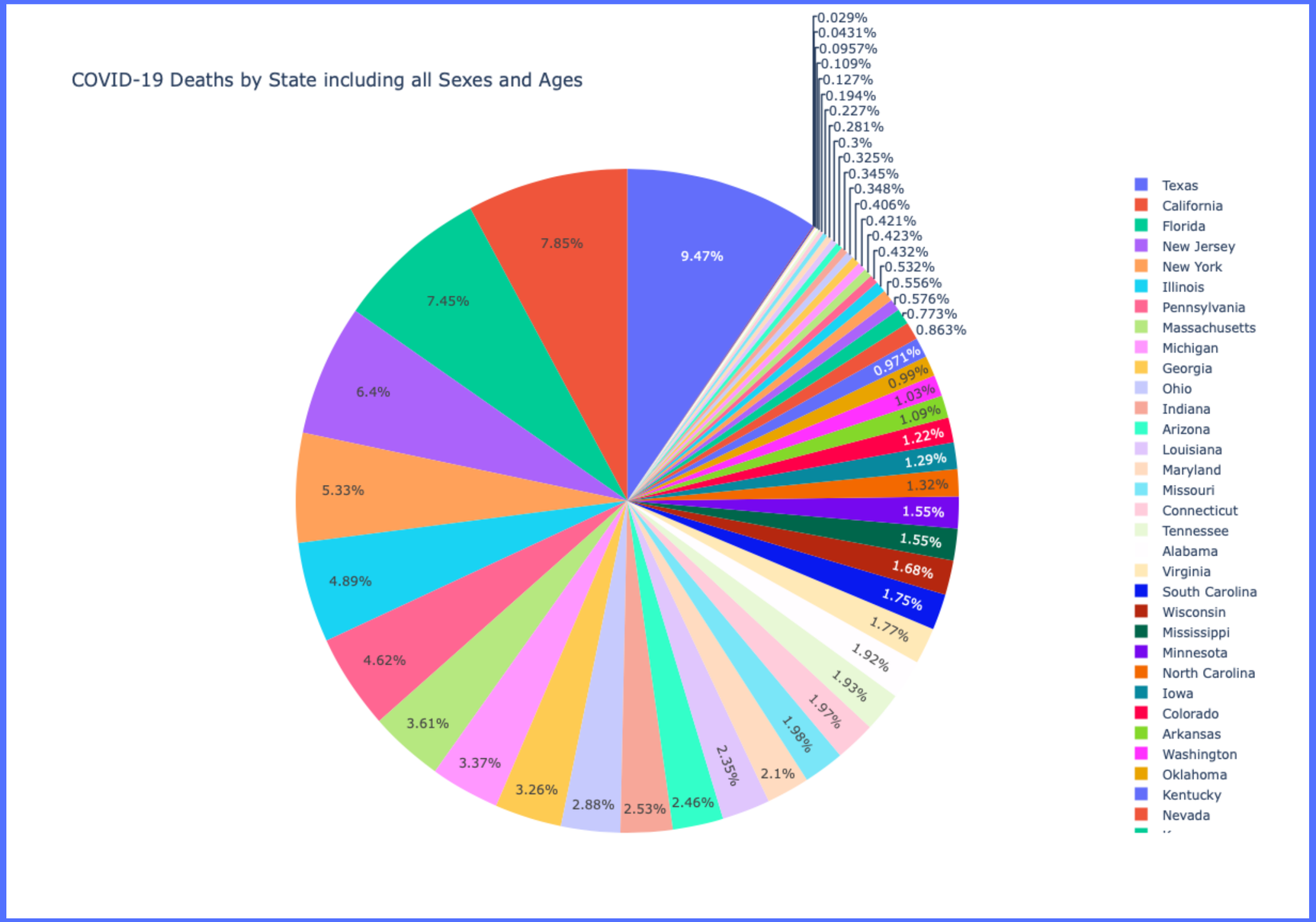
Global Gender Data

Unfortunately, for a lot of countries, the gender distribution data is not provided. So take this data with a grain of salt.
In terms of gender data, the highest COVID count providing this data is Mexico, besides the United States. There, men account for 74% of all COVID deaths so far with females accounting for only 42%. That's quite alarming.
However, strangely enough, women make up the majority of the population at 51%, outnumbering men who make up only 49%

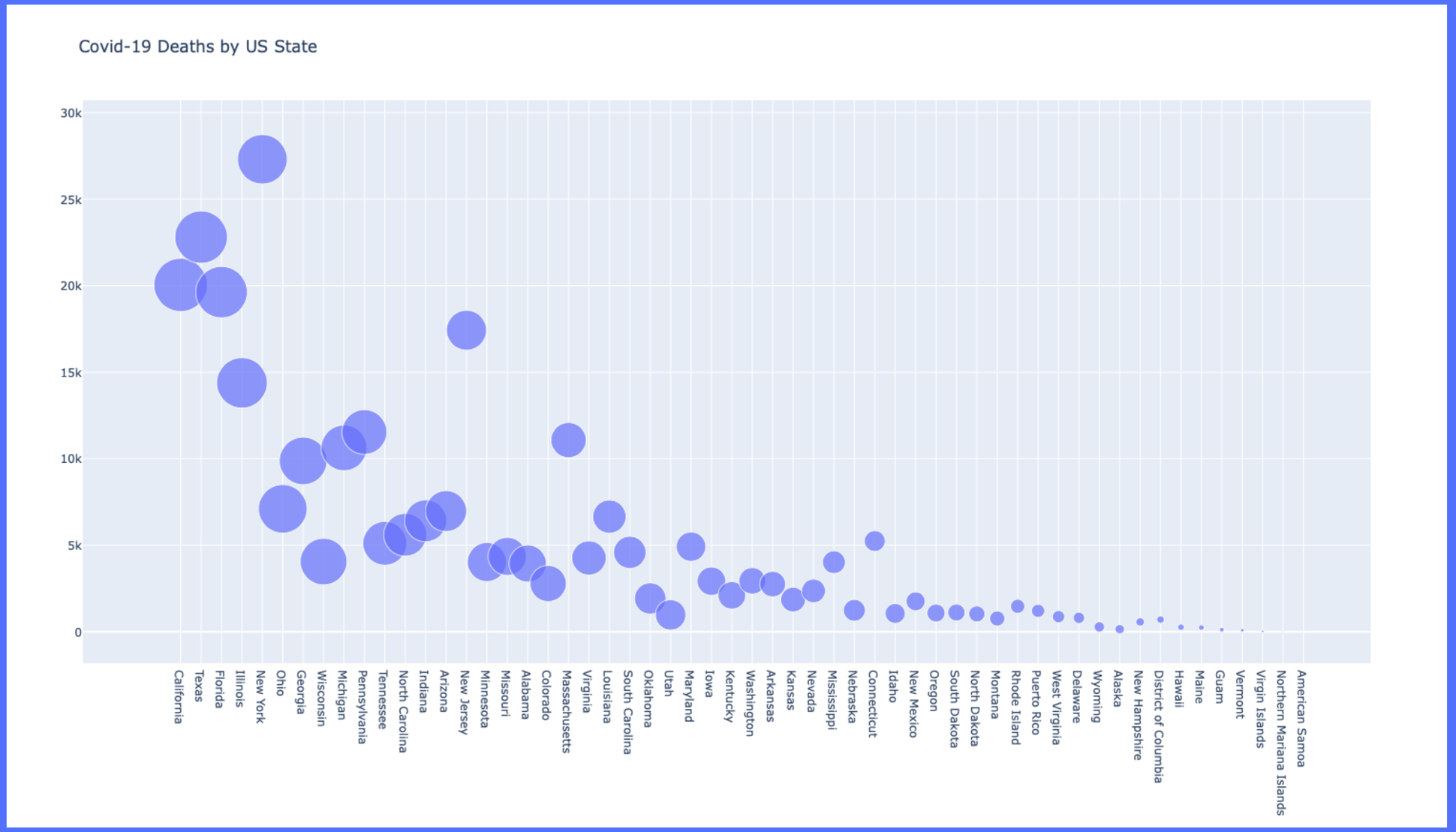
What is perhaps even more surprising, we looked at the age groups between men and women to see if there was a gender difference here.
Age Group: 0–4, Female Deaths: 95, Male Deaths: 102, Male Death Percentage: 52.0%, Female Death Percentage: 48.0
Age Group: 15–19, Female Deaths: 63, Male Deaths: 53, Male Death Percentage: 46.0%, Female Death Percentage: 54.0
Age Group: 20–24, Female Deaths: 156, Male Deaths: 232, Male Death Percentage: 60.0%, Female Death Percentage: 40.0
Age Group: 25–29, Female Deaths: 297, Male Deaths: 535, Male Death Percentage: 64.0%, Female Death Percentage: 36.0
Age Group: 30–34, Female Deaths: 528, Male Deaths: 1099, Male Death Percentage: 68.0%, Female Death Percentage: 32.0
Age Group: 35–39, Female Deaths: 777, Male Deaths: 1833, Male Death Percentage: 70.0%, Female Death Percentage: 30.0
Age Group: 40–44, Female Deaths: 1335, Male Deaths: 3108, Male Death Percentage: 70.0%, Female Death Percentghe: 30.0
Age Group: 45–49, Female Deaths: 2330, Male Deaths: 5047, Male Death Percentage: 68.0%, Female Death Percentage: 32.0
Age Group: 50–54, Female Deaths: 3226, Male Deaths: 6604, Male Death Percentage: 67.0%, Female Death Percentage: 33.0
Age Group: 55–59, Female Deaths: 4497, Male Deaths: 8282, Male Death Percentage: 65.0%, Female Death Percentage: 35.0
Age Group: 60–64, Female Deaths: 5426, Male Deaths: 9031, Male Death Percentage: 62.0%, Female Death Percentage: 38.0
Age Group: 65–69, Female Deaths: 5491, Male Deaths: 9071, Male Death Percentage: 62.0%, Female Death Percentage: 38.0
Age Group: 70–74, Female Deaths: 4867, Male Deaths: 8027, Male Death Percentage: 62.0%, Female Death Percentage: 38.0
Age Group: 75–79, Female Deaths: 3858, Male Deaths: 6082, Male Death Percentage: 61.0%, Female Death Percentage: 39.0
Age Group: 80+, Female Deaths: 4782, Male Deaths: 7247, Male Death Percentage: 60.0%, Female Death Percentage: 40.0
The differences in death are quite shocking.

VISUALIZATIONS

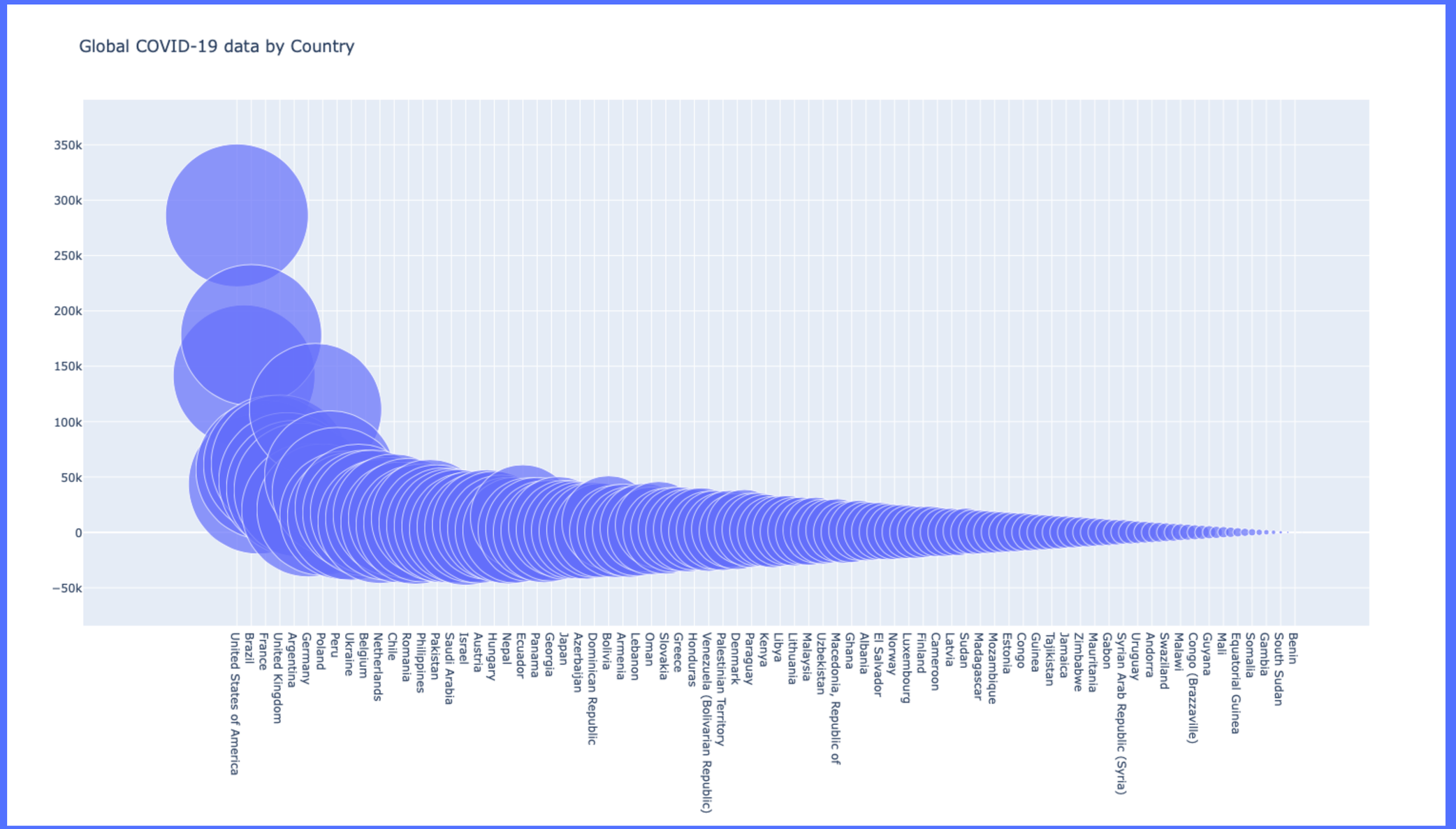
— 09



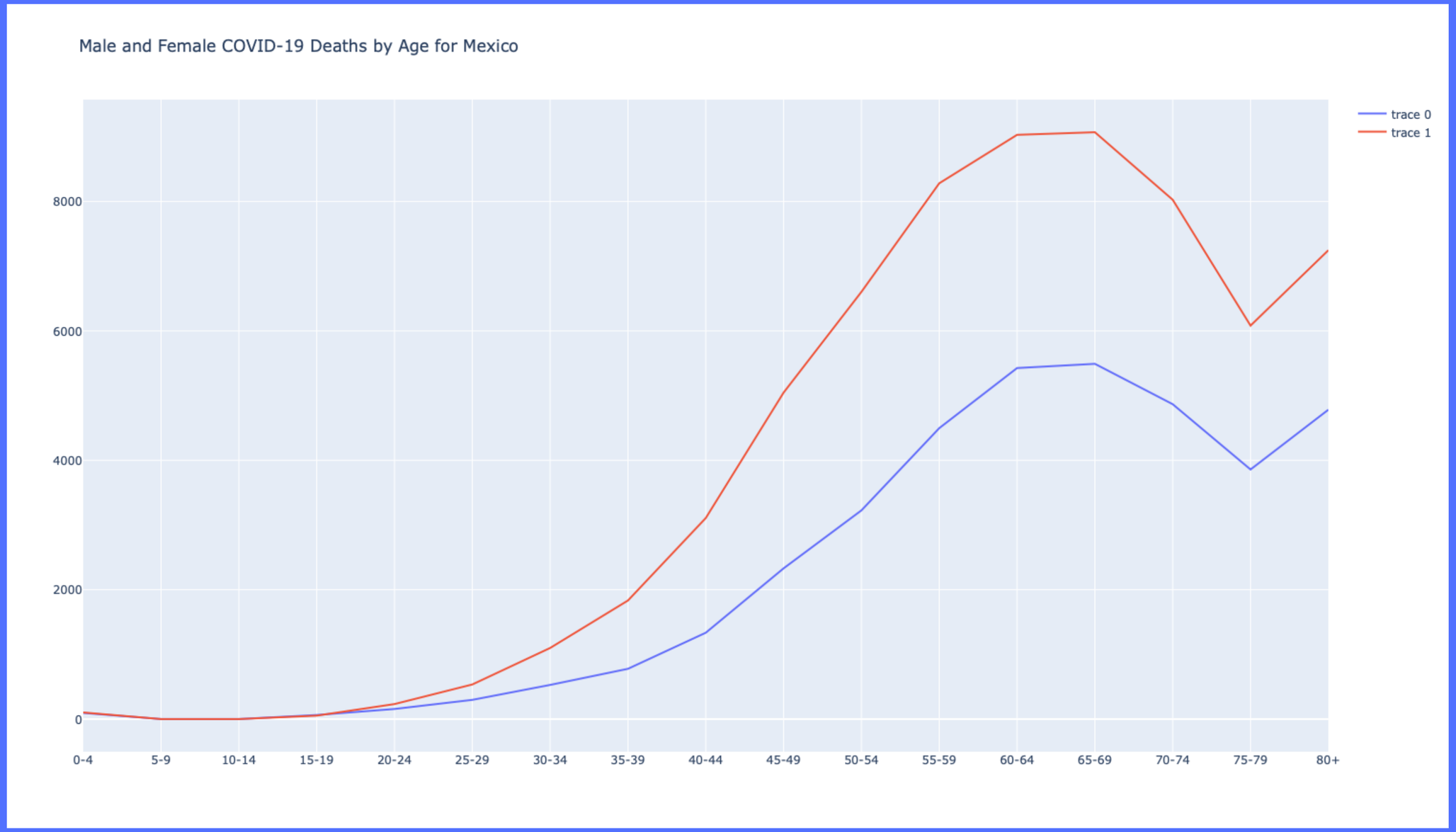
Calculations from "USAStateGender" Table



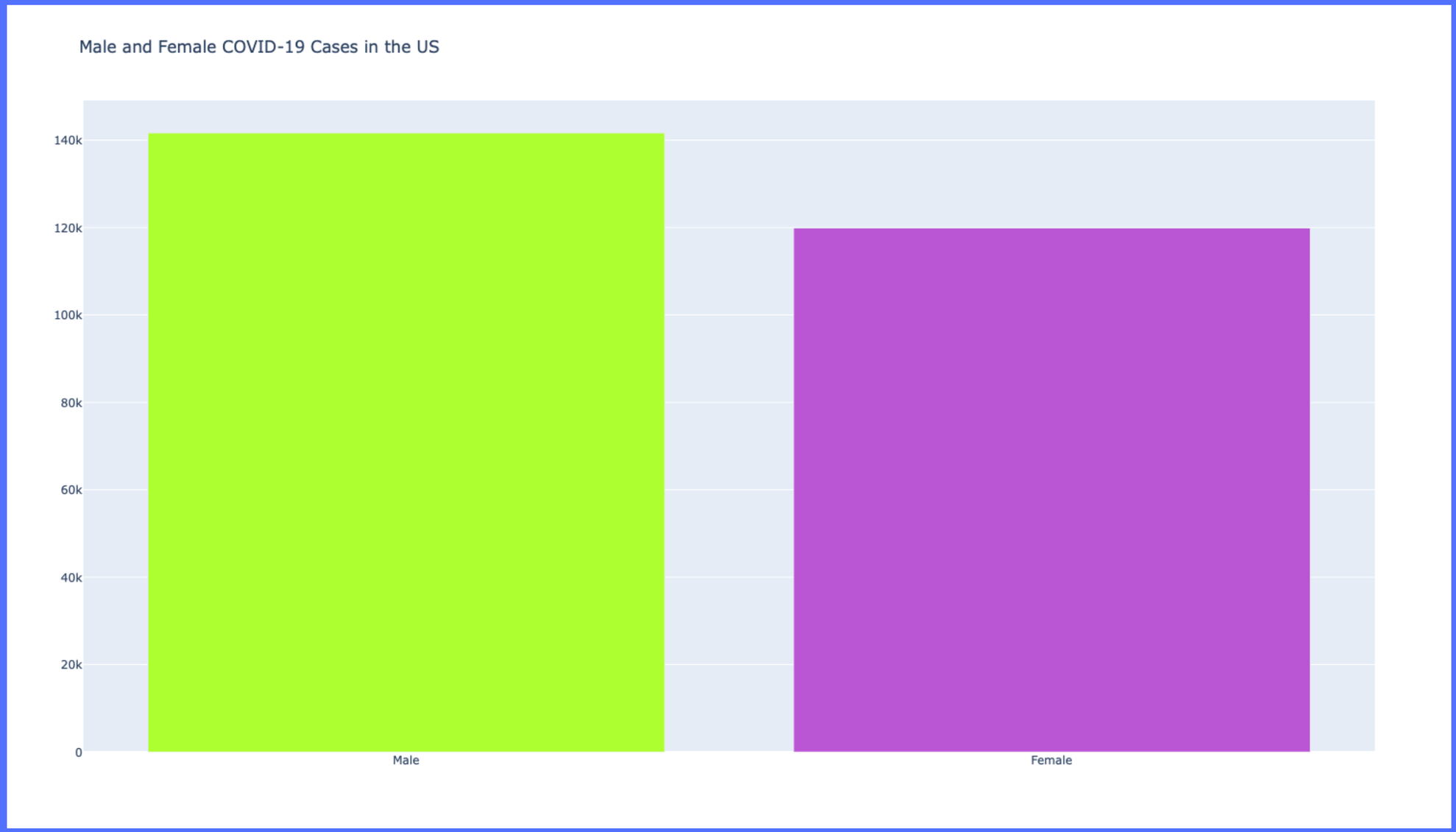
Calculations from "States" Table



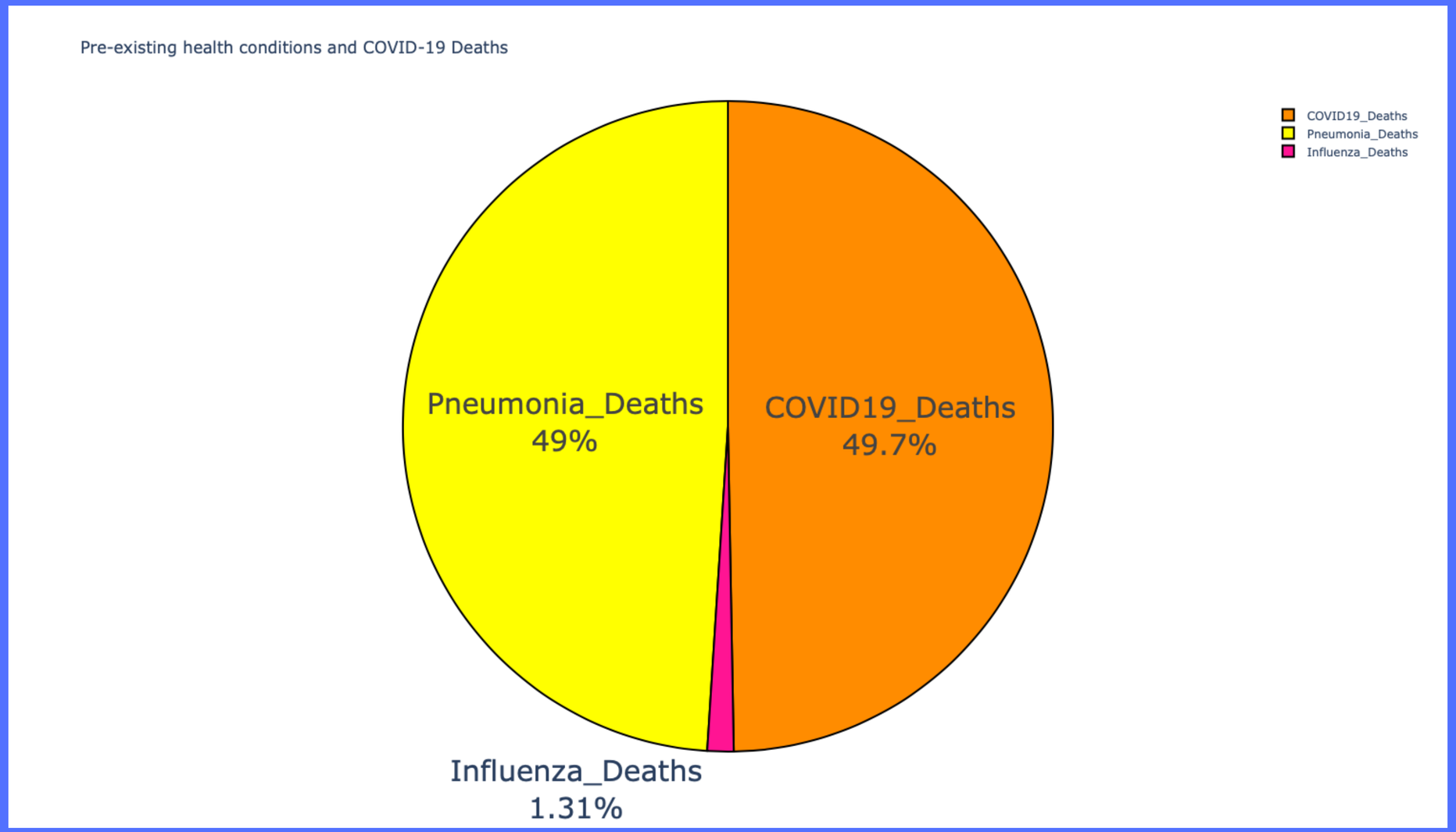
Calculations from "Countries" Table



Calculations from "CountriesAgeSex" Table



Extra Chart –
comparison of male and female deaths in the US



Extra Chart –

comparison of pre-existing health condition deaths and Covid-19 Deaths

INSTRUCTIONS

Instructions for Running Code

1. The first thing to do after the code is unzipped is to open the main.py file and run it.
2. You will see a prompt in your terminal asking what you would like to do. Press 1 and then Enter to 'Create Database'.
3. The program should loop and again prompt you for what you would like to do. Next, press 2 and then Enter to 'Update database'.

4. Then you will enter into a new prompt interface asking which tables to update. There is a specific order to insert the data into the databases.

    1. Press 1 and then Enter to update the Global and USA tables. Only press this once
    2. Then, press 2 and then Enter to update the USA State table. Run this 3 and only 3 times
    3. Finally, after you've filled the States table above and see the prompt, press 3 and then Enter to update every other table. You will need to do this at least 14 times in a row
    4. After doing all of this, press 0 and then Enter to exit out of that prompt to return to the main prompt interface. All of the tables should be ready do go now.

5. Now that the tables are ready, and you are on the main prompt, we can calculate the data. Press 3 and then Enter to Calculate results and it will generate a file called "covid.txt" which will display the calculated results.

6. Finally, press 4 and then Enter to Plot data, which will plot all of our visualizations into the browser to be seen.

CODE DOCUMENTATION

This is the code documentation for every single function that we used in this project.

main.py

def main(): """this is the main function for running everything in the project. Follow the instructions in the readme for how to accomplish this. """

database.py

def create_databases():
"""this creates the tables for the entire database
"""
def retrieve_data(url):
"""retrieves data from API
Args:
url (string): the url for the API

Returns:
dictionary: JSON data as a dictionary
"""
def save_usa_data():
"""this function specifically saves USA COVID data into a single row table
"""
def save_usa_state_data():
"""this function saves specific USA state COVID data into a multi-row table
"""
def save_usa_gender_data():
"""this function populates gender COVID data for the entire United States along with comparisons to other illnesses
"""
def save_usa_state_gender_data():
"""this function saves Gender COVID data for every individual US State as well as the territories
"""

```
def save_countries_age_gender():
"""this function saves gender data on multiple countries around the world, in particular COVID deaths by country, and also the age groups of the deaths. Note that not every country has offered data on this
"""
def save_global_data():
"""this function inserts current global COVID data in total
"""
def save_country_data():
"""
this function saves current COVID data on individual countries that offer that data
"""
def save_countries_gender():
"""this function saves saves general COVID gender deaths for every country, rather than age groups
"""
```

calculate.py

```
def calculate_global_data():
"""this in general calculates global data
"""
def calculate_usa_data():
"""this calculates general COVID data in the US
"""
def calculate_usa_gender_data():
"""this calculates the gender distribution of COVID deaths in the US.
"""
def calculate_countries_gender_data():
"""this generates the information on the country besides the US with the most COVID cases according to the data and finds out whether men or women die more
"""
```

visualizations.py

```
def main():
"""this will generate visualizations of the data we fetched from the database
"""
```

DOCUMENTATION

Date	Issue	Source	Solution
11/11/20	Started Final Project Plan	Final Project Rubric	Have general idea of what the final project will be like.
11/15/20	Had to finish the Final Project Plan	Final Project Rubric	Finished Final Project Plan
11/25/20	Tried to find better API as our third API.	https://covid-19-apis.postman.com	Used Postman to finalize APIs.
11/29/20	Started to discuss how we were going to do calculations and visualizations; decided to use plotly	https://plotly.com	Figured out our next step and continued working
12/2/20	In the fourth API, we were specifically looking for 'DeathsbyAgeSex' and found that multiple countries did not include that data. This then created an error when running the other json files.	https://api.globalhealth5050.org/api/v1/agesex	Filtered out the countries that had data and ignored the ones that did not.
12/9/20	Had some trouble creating the visualizations	https://plotly.com	Finalized visualizations and rest of the project

TABLES

Table: CountriesGender

Filter in any column

	id	Country	Date	country_id	MaleCases	FemaleCases	TotalCases	MaleDeaths	FemaleDeaths	TotalDeaths	MalePop	FemalePop	TotalPop
	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	1	China	2/28/2020	72	28577	27347	78959	1343	766	2791	735624.3	698159.4	1433784
2	2	France	10/27/2020	5	496959	550124	927193	13495	9378	33350	31524.49	33605.24	65129.73
3	3	Iran	3/17/2020	14	8545	6446	16169	503	350	988	41889.89	41024	82913.89
4	4	South Korea	11/25/2020	92	15079	16656	31735	269	244	513	25649.46	25575.86	51225.32
5	5	Germany	11/26/2020	11	480322	495571	942687	7951	6395	14361	41249.12	42267.93	83517.05
6	6	Italy	11/24/2020	6	604107	627236	1144552	26289	19268	44683	29461.04	31089.06	60550.09
7	7	Portugal	11/24/2020	34	118649	145191	260758	2095	1961	3897	4836.67	5389.5	10226.18
8	8	Netherlands	11/25/2020	21	208356	248555	452157	4738	3878	8519	8514.84	8582.29	17097.12
9	9	Spain	5/21/2020	8	321224	367750	232555	11604	8914	27888	22960.65	23776.13	46736.78
10	10	Sweden	11/24/2020	36	104477	121083	225934	3543	2957	6604	5025.6	5010.79	10036.39
11	11	Denmark	11/24/2020	71	35729	35925	70485	449	348	784	2869.67	2902.21	5771.88
12	12	Philippines	11/26/2020	27	230865	192050	421722	4949	3266	8185	54316.07	53800.55	108116.6
13	13	Switzerland	11/26/2020	32	151194	162706	308254	2291	1817	4016	4260.66	4330.7	8591.36
14	14	Belgium	11/24/2020	19	249941	308703	560299	7719	8009	15862	5711.48	5827.85	11539.33
15	15	Greece	11/25/2020	63	48259	42563	93006	1093	722	1714	5140.52	5332.93	10473.45
16	16	Canada	11/25/2020	28	155452	170523	337555	5380	6139	11521	18563.54	18847.5	37411.04
17	17	Norway	11/25/2020	93	17965	15749	33183	167	149	314	2716.81	2662.05	5378.86
18	18	India	4/6/2020	2	3089	976	4067	81	30	109	710129.6	656288.2	1366418
19	19	Peru	11/24/2020	15	494390	458049	954459	24591	11084	35727	16148.24	16362.22	32510.46
20	20	Australia	11/24/2020	98	13509	14213	27835	439	468	907	12551.31	12651.89	25203.2

id	Country	date_id	TotalConfirmed	TotalDeaths	TotalRecovered
1	United States of America	1	15165159	286249	5786915
2	India	1	9735850	141360	9215581
3	Brazil	1	6674999	178159	5965492
4	Russian Federation	1	2492713	43674	1963864
5	France	1	2363196	56453	177558
6	Italy	1	1757394	61240	958629
7	United Kingdom	1	1754911	62130	3765
8	Spain	1	1702328	46646	150376
9	Argentina	1	1469919	40009	1305587
10	Colombia	1	1384610	38158	1278326
11	Germany	1	1229269	20002	909058
12	Mexico	1	1193255	110874	881050
13	Poland	1	1076180	20592	738845
14	Iran	1	1062397	50917	754224
15	Peru	1	973912	36274	909389
16	Turkey	1	893630	15314	441515
17	Ukraine	1	855054	14413	454226
18	South Africa	1	821889	22432	753072
19	Belgium	1	594572	17507	0
20	Indonesia	1	586842	18000	483497
21	Netherlands	1	570437	9775	0
22	Iraq	1	568138	12477	498064
23	Chile	1	563534	15680	537727
24	Czech Republic	1	551070	9036	484781
25	Romania	1	524675	12660	416797
26	Bangladesh	1	481945	6906	401194
27	Philippines	1	442785	8670	408790
28	Canada	1	432717	12886	349629
29	Pakistan	1	426142	8547	372271

CONCLUSION

- The country with the most COVID-19 deaths was the United States. In the United States, men tended to die at a higher rate than women
- For other countries where gender data is available such as Mexico, men also tend to die at higher rates than women
- Amount of COVID-19 deaths were similar to amount of pneumonia deaths

RESOURCES

- Global COVID-19 API: https://documenter.getpostman.com/view/10808728/SzS8rjbc
- US COVID-19 API: https://api.covidtracking.com/v1/states/current.json
- Global Gender Data: https://globalhealth5050.org/the-sex-gender-and-covid-19-project/
- CDC COVID-19 for the US API: https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-Sex-Age-and-S/9bhg-hcku

RESOURCES (OLD)

- https://covidtracking.com/data/api
- https://documenter.getpostman.com/view/10808728/SzS8rjbc
- https://documenter.getpostman.com/view/11144369/Szf6Z9B3?version=latest