Comparing COVID Cases to Country Data

https://github.com/dhirajtadikamalla/SI206-Final-Project

Saya Takai, DJ Tadikamalla, Jamie Cha

Goals

Original Goals:

- Our original goal was to use Spotify, Soundcloud, and Genius APIs in order to collect data on streaming trends across the three platforms and compare them with each other.
- However, we realized that a lot of the data we wanted to collect from these APIs is restricted due to privacy reasons.
- Our new goal was to compare the number of COVID cases in each country to its government response and GDP.

Achieved Goals:

- Retrieved data using 2 APIs + 1 website.
- Used data retrieved from APIs + website to measure a variety of COVID-related calculations.
- Created 4 visualizations.
    - New daily US cases over 4 months
    - Comparing Average Cases of 10 countries and their Population
    - Average Daily Cases for 10 Countries
    - Comparing a Country's GDP with the Percent of Cases Recovered

API + Website Links

- https://covidtracker.bsg.ox.ac.uk
  - https://covidtrackerapi.bsg.ox.ac.uk/api/v2/stringency/date-range/2020-08-01/2020-12-01
- https://github.com/M-Media-Group/Covid-19-API
  - https://covid-api.mmediagroup.fr/v1/cases
- https://tradingeconomics.com/country-list/gdp

Problems We Faced & How We Solved Them

- Privacy restrictions on the Spotify API
  - We decided to switch topics and use different APIs
- Limiting the data to store 25 dates at a time
  - We realized we should select the last inputted data from the table and used a website (link in next slide) to figure out how to do so.
- Stringency API only had 4 data points over a span of 3 months
  - We decided to focus instead on the daily accumulating cases from the same API instead of stringency
- Trying to figure out Plotly and how Pandas dataframes work
  - We stuck with matplotlib instead

Resource Documentation

Date	Issue Description	Location of Resource	Result (did it solve the issue?)
12/4	Incorrect # of bindings error	https://stackoverflow.com/questions/16856647/sqlite3-programmingerror-incorrect-number-of-bindings-supplied-the-current-sta	Yes, we were missing a comma which made it a grouped expression rather than a tuple.
12/5	Difficulty figuring out how to add 25 data points at a time for date format	https://stackoverflow.com/questions/5191503/how-to-select-the-last-record-of-a-table-in-sql	Yes, we ended up using a SELECT statement to pull the last row in the table
12/6	Using plotly and understanding	https://stackoverflow.com/questions/62012194/plotly-how-to-make-a-line-plot-from-a-pandas-dataframe-with-a-long-or-wide-form	No, we ended up sticking to matplotlib.
12/8	Crowded data points on x-axis for our line plot using matplotlib	https://pythonpedia.com/en/knowledge-base/44863375/how-to-change-spacing-between-ticks-in-matplotlib-	Yes, we were able to space the labels apart to visualize it better

Instructions for running our code

1. Install country-converter 0.7.1, numpy, and any packages necessary
2. Run gdp.py four times.
3. Run cases.py five times and open COVID_CASES_USA.db
4. Run analysis.py once.
5. Open calculations.csv file using excel.

Functions in gdp.py

setUpDatabase: takes in a database name and sets up a database

setUpGDPTable: takes the database cursor and connection as inputs. Creates a table in the database with their countries and GDP

get_countries: retrieves data from both the website and an API and returns a list of countries with a country_id as a key

get_data: takes the database cursor and connection as inputs. Returns a list with 100 countries with their respective GDP's

main: calls all other functions and executes the program

Functions in cases.py

setUpDatabse: takes in a database name as input and sets up a database

setUpCasesTable: takes the database cursor and connection as inputs. Creates a table with 10 countries and their daily new cases for a 3-month period

setUpTotalCasesTable: takes the database cursor and connection as inputs. Creates a table of 100 countries and their total cases, recovered, and population

get_data: takes in the database cursor, connection, a start date, and an end date as inputs. Returns a data_list with all the cases confirmed over three months

country_data: takes in takes the database cursor and connection as inputs. Returns a list of countries with their total cases, recovered, and population

main: calls all other functions and executes the program

Functions in analysis.py

setUpDatabase: takes in database name as input and sets up database

cases_per_day: takes in database cursor and connection as inputs. Returns a list of number of new cases per day for USA over a span of three months

avg_new_cases: takes in database cursor and connection as inputs. Returns a list of average number of cases of 10 countries over a 3-month span

percentage_recovered: takes in database cursor and connection as inputs. Uses INNER JOIN to collect data from the GDP and CountryData tables. Returns a list of percentages recovered out of a country's infected population.

write_csv: takes in database cursor, connection, and filename as inputs. Returns a csv file with the 3 tables: GDP/Percent Recovered, Average Cases, and New Cases

Functions in analysis.py (cont.)

new_cases_US: takes in database cursor and connection as inputs. Creates a line graph of the number of new cases in USA over a span of three months

cases_vs_population: takes in database cursor and connection as inputs. Creates a scatter plot of the average cases for 10 countries and their respective population

recovered_vs_gdp: takes in database cursor and connection as inputs. Creates a scatter plot of the percent of people recovered from 100 countries and their respective GDP

zoomed_in: takes in database cursor and connection as inputs. Creates the same scatter plot from recovered_vs_gdp, but excluding data from USA to visualize better

Functions in analysis.py (cont.)

case_vs_country: takes in database cursor and connection as inputs. Creates a bar graph of the average cases for a 3 month period for 10 different countries

case_vs_country_zoomed: takes in database cursor and connection as inputs. Creates the same bar graph for the function above, excluding the data from USA

main: calls all other functions and executes the program

CSV Calculation File

https://drive.google.com/file/d/1_x2WvcidtturZqmX78YvY22dTmXjvGTj/view?usp=sharing

PERCENT RECOVERED		
Country	Percent (%)	GDP (USD Billion)
Algeria	65.2855518	170
Angola	54.4992151	94.64
Argentina	88.9010603	450
Australia	91.6571429	1393
Austria	85.9659917	446
Azerbaijan	61.3685232	48.05
Bahrain	97.9535567	38.57
Bangladesh	84.4612266	303
Belarus	85.5517125	63.08
Belgium	0	530
Bolivia	85.9167465	40.9
Brazil	88.6699942	1840
Bulgaria	42.0297038	67.93
Cambodia	86.2359551	27.09
Cameroon	93.5144013	38.76
Canada	80.6497287	1736
Chile	95.3830944	282
China	93.2868844	14343
Colombia	92.4909473	324
Costa Rica	72.3824384	61.77
Croatia	83.3705303	60.42
Cuba	89.4455942	100
Cyprus	15.4824627	24.56
Denmark	75.4105338	348
Dominican Republic	78.2818463	88.94
Ecuador	86.929269	107
Egypt	86.9442449	303
El Salvador	91.239783	27.02
Estonia	59.9168575	31.39
Ethiopia	77.1281207	96.11
Finland	67.6315434	269
France	7.54114853	2716
Germany	73.378032	3846
Ghana	97.6866775	66.98
Greece	19.0296323	210
Guatemala	91.1226582	85.3
Honduras	45.1166749	25.1
Hungary	29.1928771	161
India	94.7369154	2875
Indonesia	82.141909	1119
Iran	71.8513827	445
Iraq	87.8712234	234
Ireland	31.0679095	389
Israel	94.6387032	395
Italy	57.5215133	2001
Japan	82.8711654	5082
Jordan	81.6360555	43.74
Kazakhstan	84.5475399	180
Kenya	78.9037152	95.5
Kuwait	97.0885597	135

AVERAGE DAILY CASES	
Country	Average Cases Per Day
United_States	74084.78862
Russia	11855.05691
Poland	7752.666667
Norway	222.2601626
Egypt	178.7560976
New_Zealand	4.024390244
Cuba	46.73170732
Ghana	119.1300813
Lebanon	1014.02439
Uganda	161.4552846

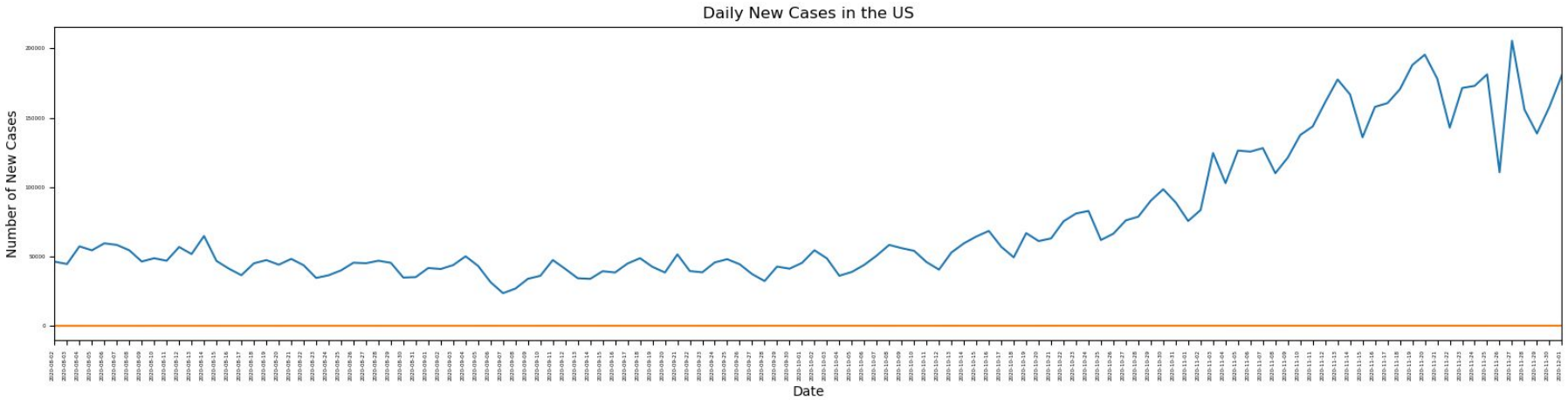
NEW DAILY CASES	
Date	New Cases
8/2/2020	46269
8/3/2020	44567
8/4/2020	57283
8/5/2020	54433
8/6/2020	59481
8/7/2020	58322
8/8/2020	54449
8/9/2020	46369
8/10/2020	48733
8/11/2020	46910
8/12/2020	56811
8/13/2020	51742
8/14/2020	64764
8/15/2020	46790
8/16/2020	41181
8/17/2020	36484
8/18/2020	45034
8/19/2020	47393
8/20/2020	44087
8/21/2020	48253
8/22/2020	43653
8/23/2020	34510
8/24/2020	36432
8/25/2020	39979
8/26/2020	45520
8/27/2020	45120
8/28/2020	46929
8/29/2020	45414
8/30/2020	34787
8/31/2020	35138
9/1/2020	41677
9/2/2020	40951
9/3/2020	43781
9/4/2020	50129
9/5/2020	43192
9/6/2020	31511
9/7/2020	23545
9/8/2020	26845
9/9/2020	33895
9/10/2020	36066
9/11/2020	47439
9/12/2020	41003
9/13/2020	34305
9/14/2020	33842
9/15/2020	39385
9/16/2020	38460
9/17/2020	44922
9/18/2020	48750
9/19/2020	42512
9/20/2020	38444

Latvia	7.87142496	34.12
Lebanon	67.3317533	53.37
Libya	66.4050323	52.08
Lithuania	41.6949799	54.22
Luxembourg	80.3046055	71.1
Malaysia	84.3781449	365
Mexico	73.775855	1200
Morocco	88.3546302	119
Nepal	94.2808874	30.64
Netherlands	1.26978155	909
New Zealand	96.1685824	207
Nigeria	92.3205366	448
Norway	45.2574935	403
Oman	93.361927	76.98
Pakistan	87.1927413	278
Panama	85.7320865	66.8
Papua New Guinea	87.8654971	24.97
Paraguay	70.875025	38.15
Peru	93.4345053	227
Philippines	91.8117341	377
Poland	71.8738658	592
Portugal	77.4291706	238
Qatar	98.2329414	183
Romania	80.6294483	250
Russia	79.1456232	1700
Saudi Arabia	97.3448521	793
Serbia	0	51.41
Singapore	99.8130264	372
Slovakia	73.2102689	105
Slovenia	75.2485803	53.74
South Africa	90.428245	351
Spain	8.74250606	1394
Sri Lanka	72.717473	84.01
Sweden	0	531
Switzerland	78.2218365	703
Taiwan	81.4917127	605
Tanzania	35.9528487	63.18
Thailand	93.2597745	544
Tunisia	76.2359248	38.8
Turkey	47.3539548	754
Uganda	37.329965	34.39
Ukraine	54.8415779	154
United Arab Emirates	89.1601665	421
United Kingdom	0.21120102	2827
United States	37.9251414	21428
Uruguay	68.6574531	56.05
Uzbekistan	96.2978872	57.92
Venezuela	95.3708952	482
Vietnam	88.4476534	262
Yemen	66.5223665	27.59

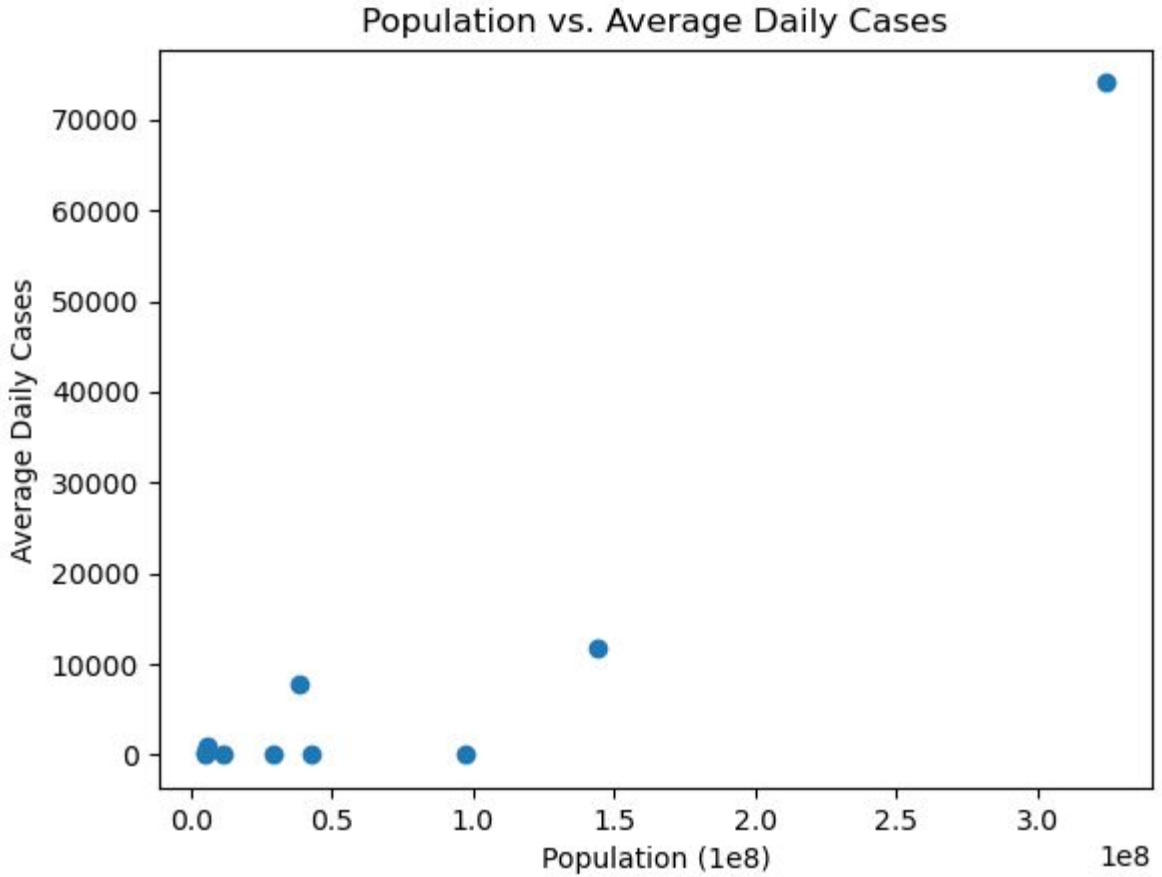
9/21/2020	51495
9/22/2020	39482
9/23/2020	38581
9/24/2020	45710
9/25/2020	48054
9/26/2020	44394
9/27/2020	37280
9/28/2020	32284
9/29/2020	42692
9/30/2020	41196
10/1/2020	45357
10/2/2020	54469
10/3/2020	48627
10/4/2020	36092
10/5/2020	38835
10/6/2020	43949
10/7/2020	50641
10/8/2020	58329
10/9/2020	55994
10/10/2020	54074
10/11/2020	45948
10/12/2020	40523
10/13/2020	52879
10/14/2020	59473
10/15/2020	64358
10/16/2020	68449
10/17/2020	56954
10/18/2020	49354
10/19/2020	66869
10/20/2020	61088
10/21/2020	63026
10/22/2020	75335
10/23/2020	80987
10/24/2020	82821
10/25/2020	61802
10/26/2020	66504
10/27/2020	76001
10/28/2020	78684
10/29/2020	90168
10/30/2020	98519
10/31/2020	88937
11/1/2020	75571
11/2/2020	83487
11/3/2020	124553
11/4/2020	102873
11/5/2020	126190
11/6/2020	125708
11/7/2020	128036
11/8/2020	110026
11/9/2020	121210
11/10/2020	137694
11/11/2020	143649

11/10/2020	137694
11/11/2020	143649
11/12/2020	161147
11/13/2020	177568
11/14/2020	166750
11/15/2020	135941
11/16/2020	157820
11/17/2020	160574
11/18/2020	170513
11/19/2020	188033
11/20/2020	195518
11/21/2020	178097
11/22/2020	142807
11/23/2020	171515
11/24/2020	172988
11/25/2020	181287
11/26/2020	110654
11/27/2020	205514
11/28/2020	155880
11/29/2020	138633
11/30/2020	157851
12/1/2020	180637

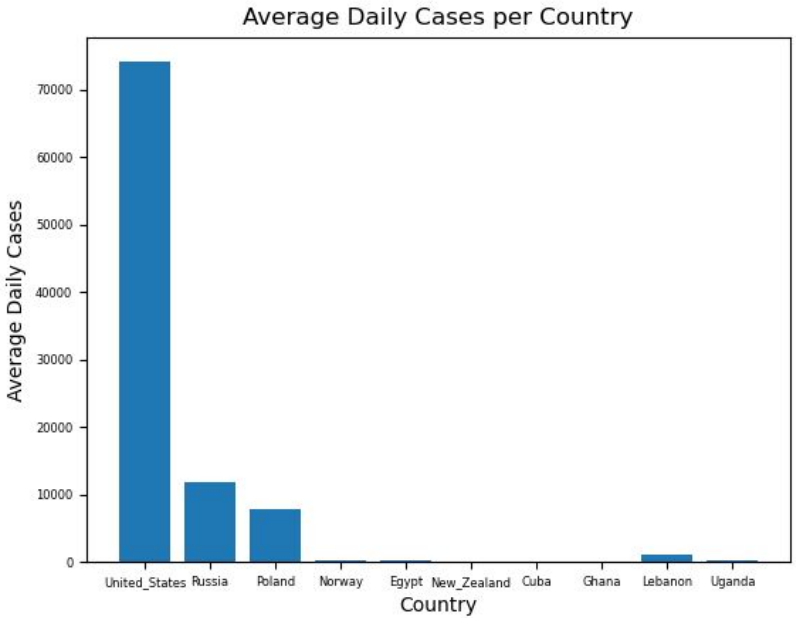
New Daily Cases in USA (4 months)

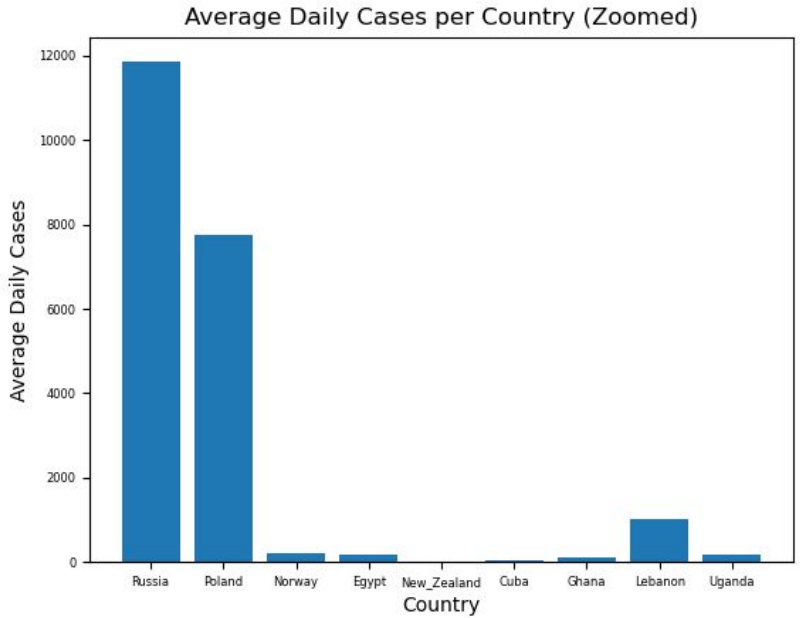


Avg. Cases of 4 months vs. Country Population



Avg. Cases of 4 months for 10 Countries





Country GDP vs. Percent Recovered