

Analysing trends of COVID-19 globally

9-08

(<https://covidhunter.netlify.app/>)

Group Leader:

Ho Jun Yi (3i1)

Members:

Gareth Lau Jing Sheng (3i1)

Lim Yee Hin Cedric (3i1)

1 Introduction

a. Rationale

The 2019 novel coronavirus (COVID-19), is a severe widespread virus that has infected millions and been declared a global pandemic by the World Health Organisation. People now rely on many online sources for updates on COVID-19 statistics. However, data across an assortment of platforms are often replete with inaccuracies, becoming a breeding ground for fake news. Such misinformation stems from the lack of fact-checking, resulting in the spread of falsehoods.

Therefore, there is a need to provide the public with accurate statistics to enable the public to be better informed on the latest COVID-19 changes and to prevent the public from panicking over misinformation. This project collects data from multiple sources and compares them to ensure reliability and accuracy of our website's statistics.

b. Idea Description

Our proposed solution comprises 2 phases.

The first phase is data collection and cleaning, where we scrape several data sources. These data sources include Worldometer, CoronaTracker, ncov2019live, MOH's daily Situation Reports, and local media company Mothership's Telegram channel. Mothership was chosen due to their efficiency in releasing statistics as compared to other news companies, allowing for swift and reliable updates for our platform.

The second phase is data analysis and visualisation, where we presented myriads of graphs and tables for diverse groups of data. These data include the statistics worldwide and per continent, which are further categorised into total and daily numbers. These groups of statistics would then be presented in various forms of graphs where applicable, ranging from pie charts to bar graphs. We created a multitude of graphs with the intention of giving users a better experience visualising these data, so as to stand out from other similar websites.

c. Focus and Significance of Project

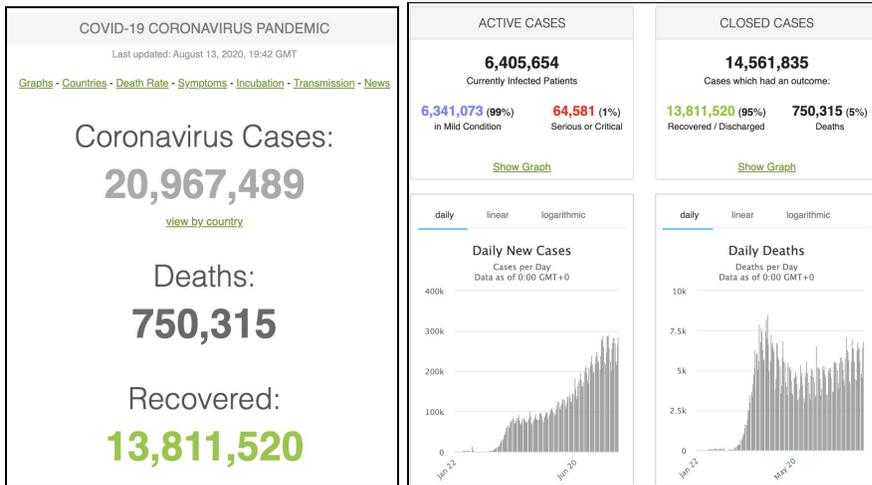
Our platform focuses on providing accurate statistics and detailed graphs for the public in hopes of allowing them to stay aware of COVID-19 statistics and trends worldwide.

d. Project Scope

Our platform collects, analyses and graphs global and local data daily for the statistics over the past 2 weeks.

2 Literature review

a. Worldometers:



#	Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	New Recovered	Active Cases	Serious, Critical	Tot Cases/ 1M pop	Deaths/ 1M pop	Total Tests	Tests/ 1M pop	Population	1 Case every X ppl	1 Death every X ppl	1 Test every X ppl
	World	20,967,489	+169,935	750,315	+3,909	13,811,520	+113,568	6,405,654	64,581	2,690	96.3						
1	USA	5,392,097	+31,795	169,806	+675	2,619,182	+6,579	2,403,109	17,342	16,279	513	68,062,560	205,484	331,229,890	61	1,951	5
2	Brazil	3,180,758	+10,284	104,528	+265	2,309,477		766,753	8,318	14,952	491	13,231,548	62,197	212,735,424	67	2,035	16
3	India	2,459,613	+64,142	48,144	+1,096	1,750,636	+54,776	660,833	8,944	1,780	35	26,845,688	19,431	1,381,567,238	562	28,697	51
4	Russia	907,758	+5,057	15,384	+124	716,396	+6,098	175,978	2,300	6,220	105	31,598,302	216,513	145,941,946	161	9,487	5
5	South Africa	568,919		11,010		432,029		125,880	539	9,579	185	3,295,434	55,484	59,393,824	104	5,395	18
6	Peru	498,555		21,713		341,938		134,904	1,501	15,097	657	2,643,464	80,047	33,023,850	66	1,521	12
7	Mexico	498,380	+5,858	54,666	+737	336,635	+3,835	107,079	3,775	3,861	423	1,127,115	8,731	129,088,353	259	2,361	115
8	Colombia	422,519		13,837		239,785		168,897	1,493	8,294	272	2,023,256	39,714	50,945,148	121	3,682	25
9	Chile	380,034	+1,866	10,299	+94	353,131	+1,712	16,604	1,259	19,860	538	1,932,592	100,997	19,135,201	50	1,858	10
10	Spain	379,799	+2,935	28,605	+26	N/A	N/A	N/A	617	8,123	612	7,472,031	159,806	46,756,944	123	1,635	6
11	Iran	336,324	+2,625	19,162	+174	292,058	+1,814	25,104	3,971	3,998	228	2,788,027	33,145	84,115,247	250	4,390	30
12	UK	313,798		41,347	+18	N/A	N/A	N/A	70	4,620	609	18,868,566	277,774	67,927,863	216	1,643	4
13	Saudi Arabia	294,519	+1,482	3,303	+34	260,393	+3,124	30,823	1,805	8,445	95	4,071,857	116,756	34,874,813	118	10,559	9
14	Pakistan	286,674	+753	6,139	+10	264,060	+867	16,475	783	1,295	28	2,205,664	9,964	221,366,281	772	36,059	100
	Bangladesh	269,115	+2,617	3,557	+44	154,871	+1,762	110,687		1,632	22	1,315,901	7,981	164,878,278	613	46,353	125

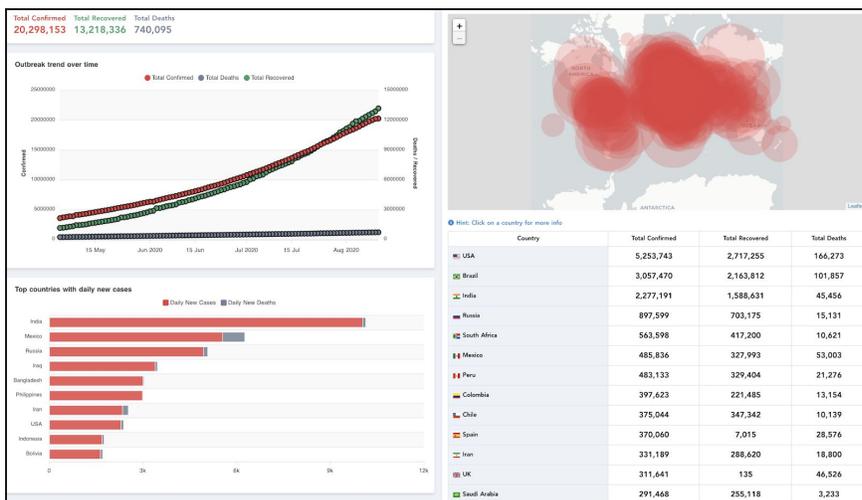
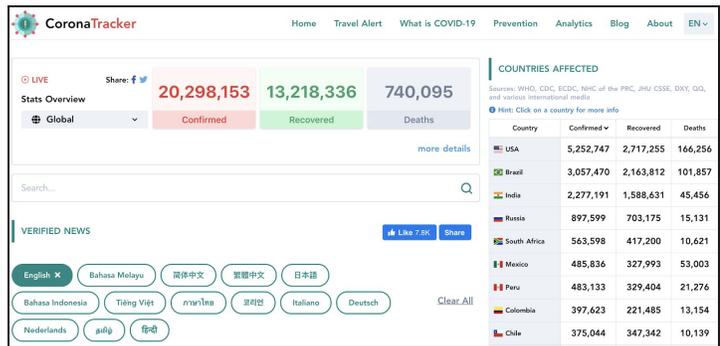
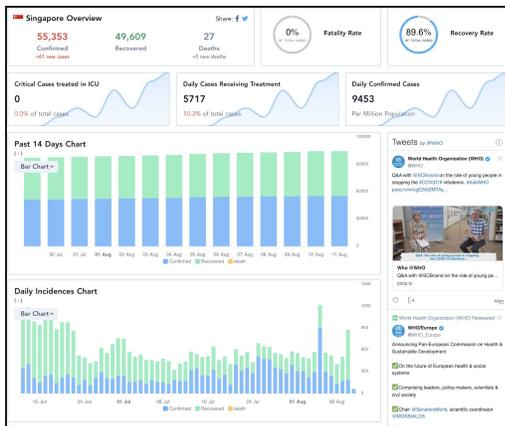
Pros:

- Different statistics like total tests, active and serious cases
- Several comprehensive graphs that compares data of countries
- Extensive information of infected countries

Cons:

- Limited news articles
- No information on preventing infection
- Data sources include non-governmental agencies which may be less reliable.

b. CoronaTracker:



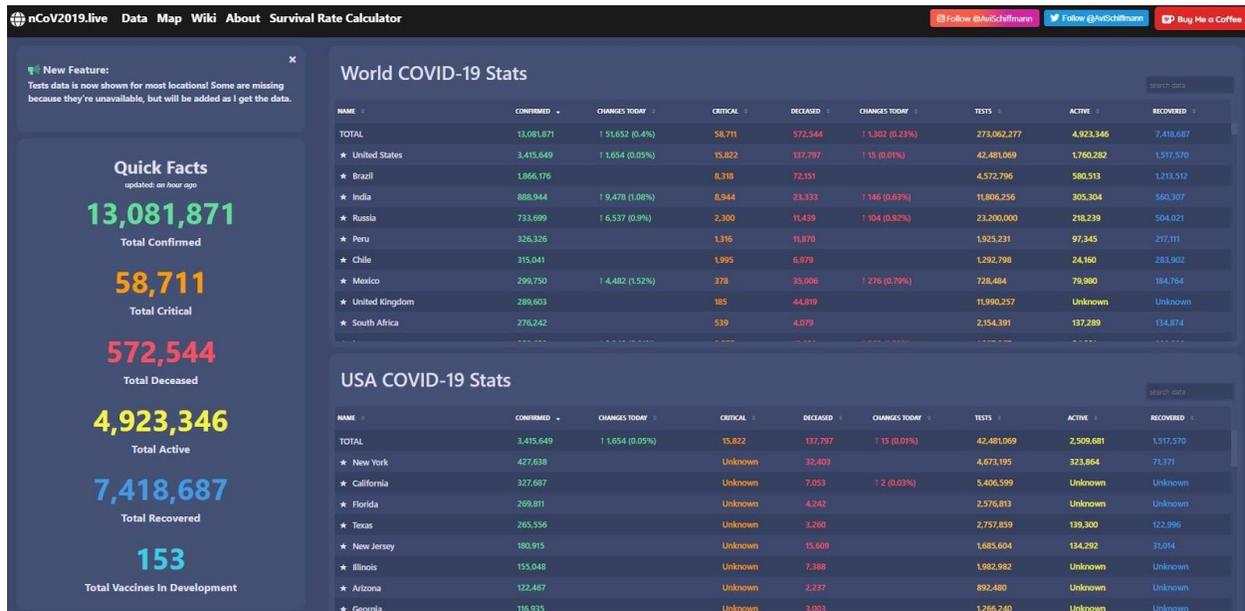
Pros:

- Extensive information for specific countries
- Contains many graphs, heatmaps, news updates, global travel information and methods to avoid infection
- Available in different languages

Cons:

- Table for global data only contains statistics on 3 categories, which are total confirmed, recovered and deaths, so users are unable to compare at one glance statistics on new cases, deaths and recovered between countries

c. ncov2019live:



Pros:

- Regularly updated
- Contains heatmaps and survival rate calculator
- Displays data by region
- Includes tips on preventing infection

Cons:

- No graphs

3 The Study & Methodology

a. Methodology

Our platform uses Worldometers, Corona Tracker, and ncov2019live as our global data sources as they are widely used by people worldwide to monitor COVID-19 statistics. By comparing all 3 of them, reliability of our website would be significantly enhanced. Not only that, but Worldometers was chosen as it has been providing live world statistics for other data, long before COVID-19 started, increasing its reliability. However, as Corona Tracker uses Worldometers as one of its data sources, its reliability would only be as high as Worldometers', which is why ncov2019live was scraped, to increase our platform's reliability. Data compared between the 3 sources would inevitably have some differences due to the difference in the timing at which they were updated. Hence, we decided on solely using Worldometers for countries with differences in their data as Worldometers is the most reliable source.

For our local data sources, Mothership's telegram channel was chosen for its efficiency in updating Singapore's statistics, while MOH's website was chosen for its detailed breakdown of local data into several groups like dormitory and imported cases. This is essential as our platform includes multiple graphs that thoroughly analyses such local data, and it is of high importance that we maintain efficiency in our website.

b. Development Tools

a. Python

- i. Tabula, for scraping MOH's Situation reports
- ii. Telethon, for scraping Mothership Telegram Channel
- iii. Requests, CSV, Json, Beautifulsoup, for scraping global data sources
- iv. Pandas, Numpy, Matplotlib, for graphing

b. HTML, Javascript, CSS, Github, Netlify, for product development

- i. Bootstrap, for navigation bar positioning

c. Job Distributions

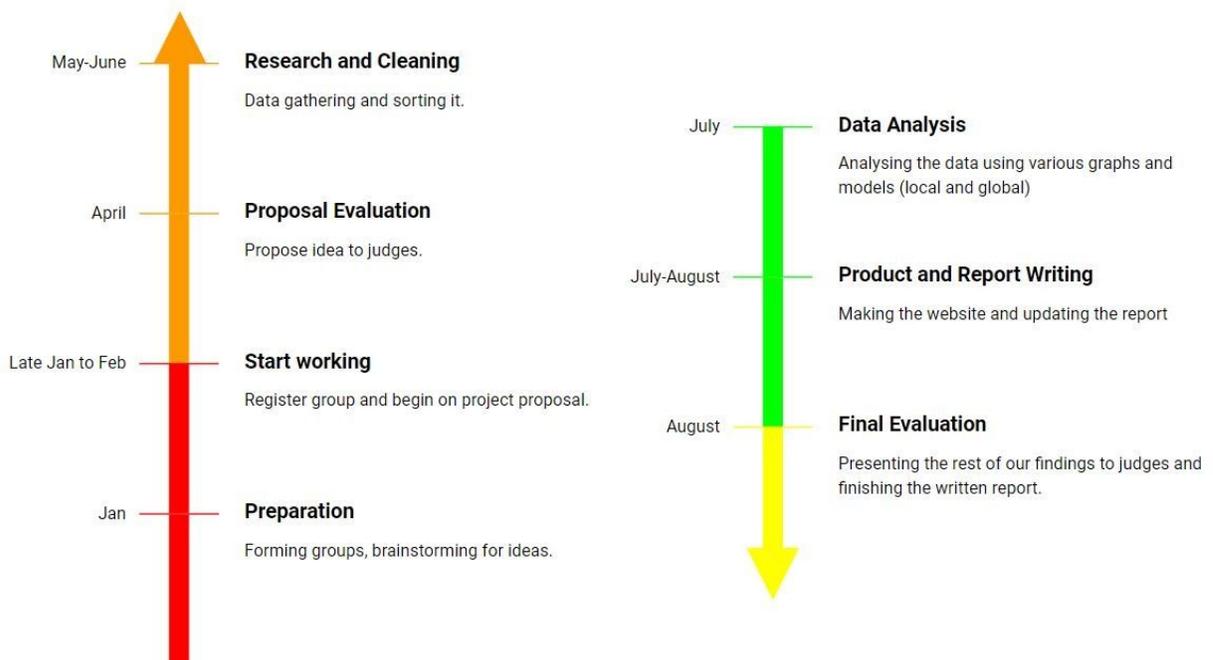
Jun Yi

1. Data Collection (Mothership, Corona Tracker, MOH, MOH's PDFs)
2. Data Cleaning
3. Data Analysis
4. Data Visualisation
5. Product Development
6. Report Writing
7. Slides

Gareth & Cedric:

1. Data Collection (nCoV2019.live, Worldometers)
2. Report Writing
3. Slides

d. Timeline

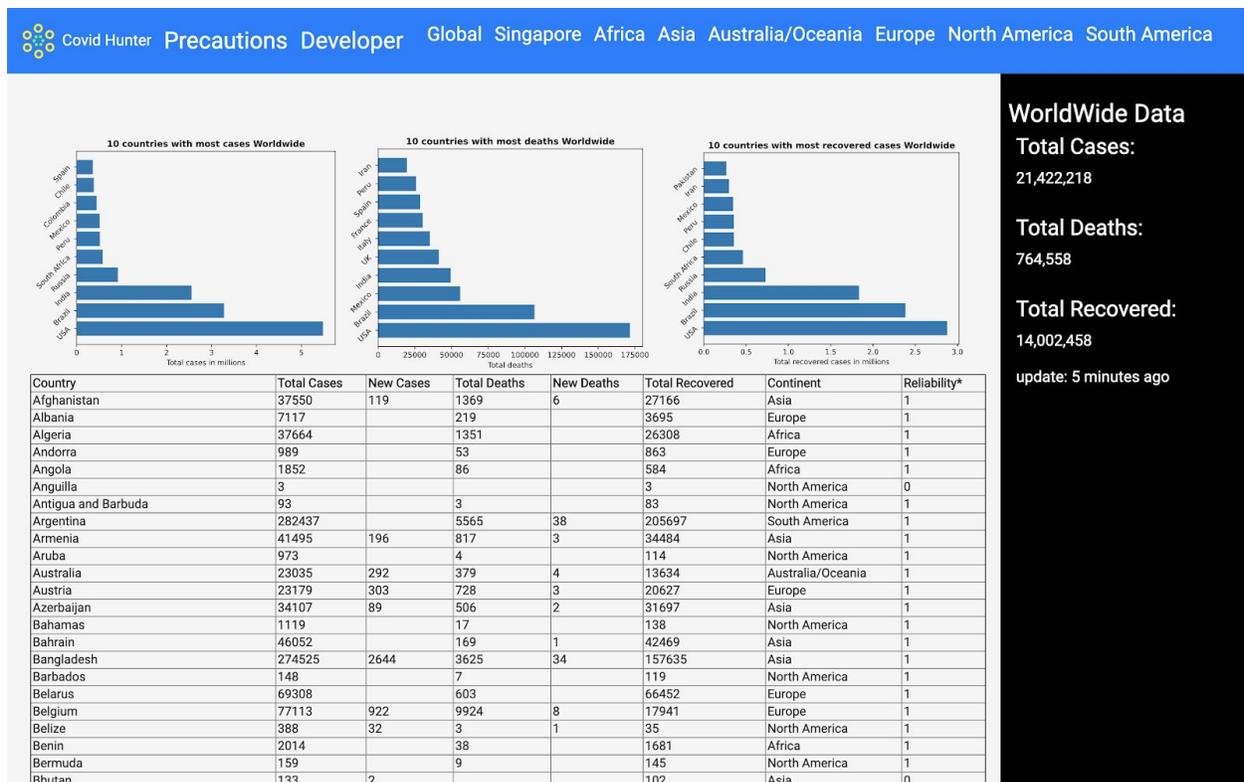


4 Outcomes, Analysis & Discussions

Our final product is a website that comprises coronavirus data worldwide and graphs for clearer visualisation of statistics. Firstly, we included a navigation bar with a link labelled “Precautions” that lead users to a webpage by Centers for Disease Control and Prevention(CDC) containing information on what to do if one falls sick. This navigation bar includes multiple categories, which are global and local data, as well as a data breakdown for all infected continents.

Secondly, a navigation bar is also incorporated at the right. This navigation bar contains data for total cases, deaths and recoveries globally, as well as a self-updating label informing users of the latest time the webpage’s statistics is updated.

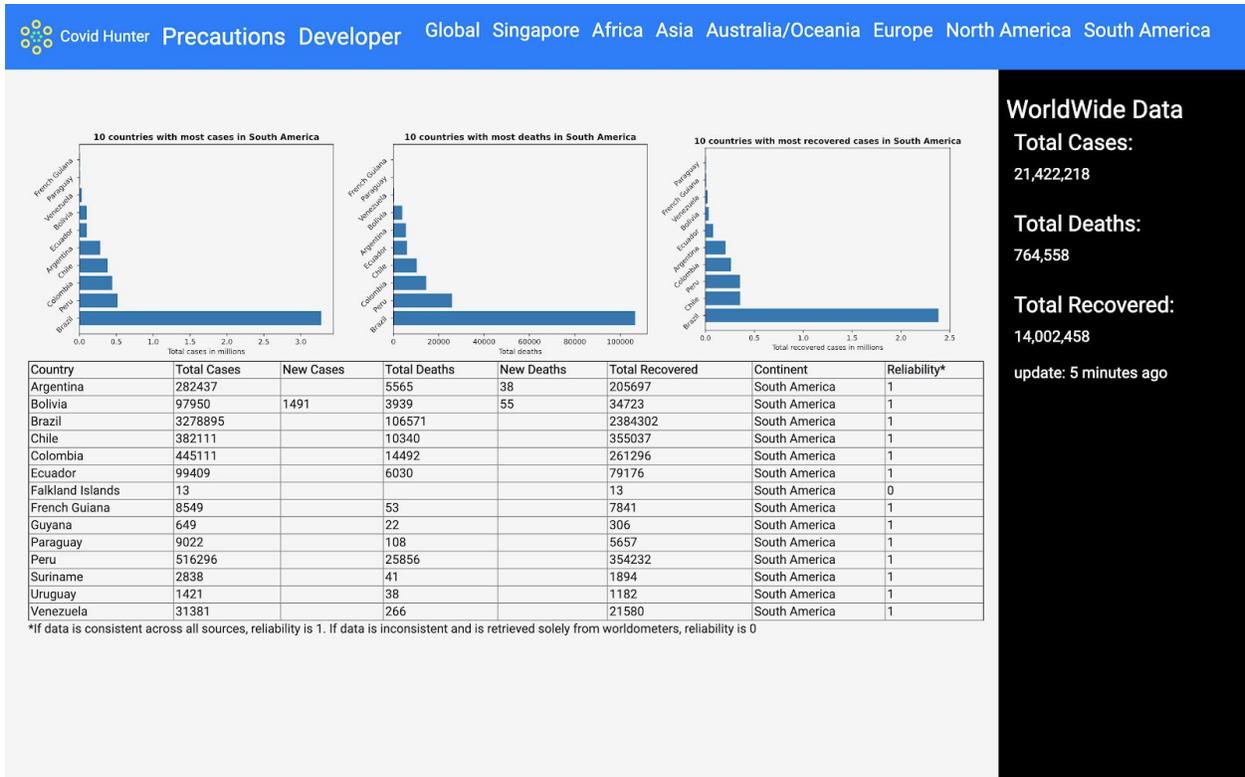
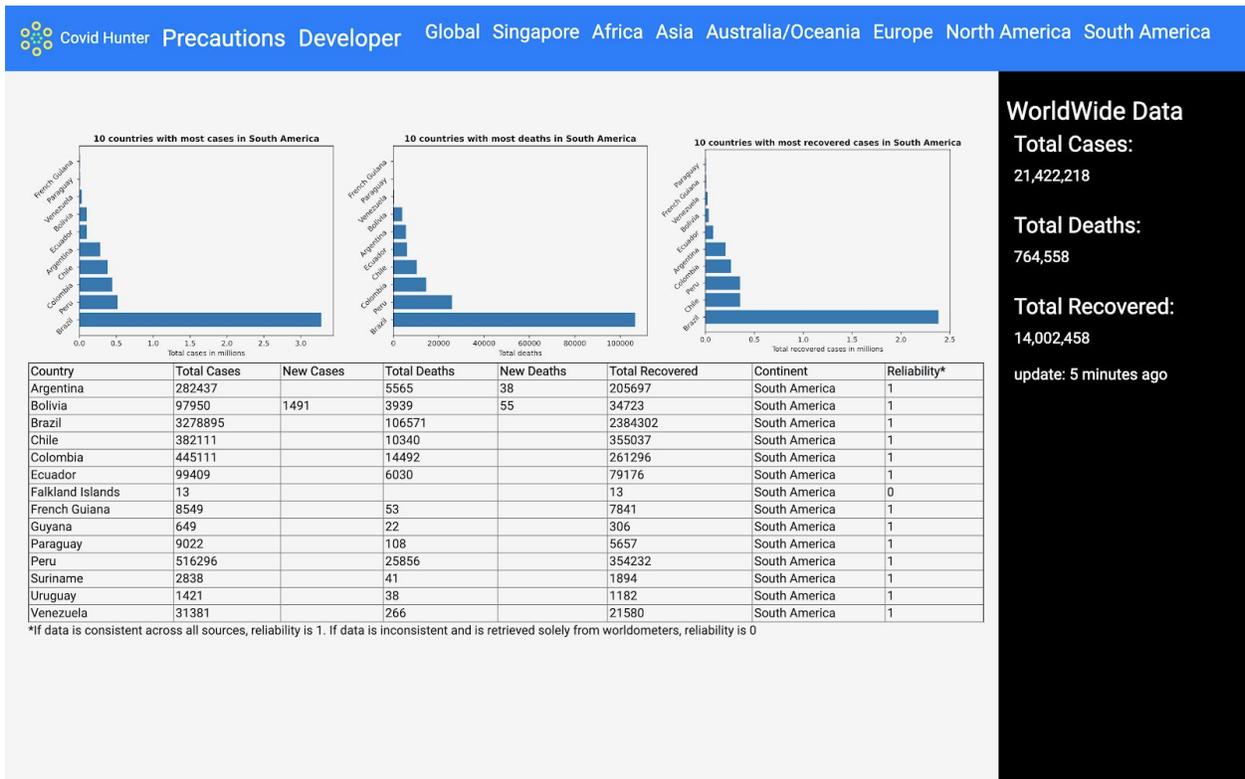
Below is an image of our website’s homepage:



Aside from that, there are also 3 main bar graphs and a table in each of the continents' webpage and the international webpage. These graphs display the top 10 countries for number of cases, deaths and recoveries, allowing users to better understand how badly a country is impacted by the virus. Each table contains data of several types, its continent, and a reliability value. If a country has a value of 1 for its reliability, it shows that the country's data is consistent throughout all 3 sources. If the data for a country is inconsistent across all sources, its data would be retrieved solely from worldometers, displaying a reliability of 0, as Worldometers is the most reliable data source. However, a reliability of 0 would not imply complete lack of reliability, and instead only means that it would be less reliable compared to other countries as its data only comes from 1 source. This definition of reliability is also included at the bottom of every table.

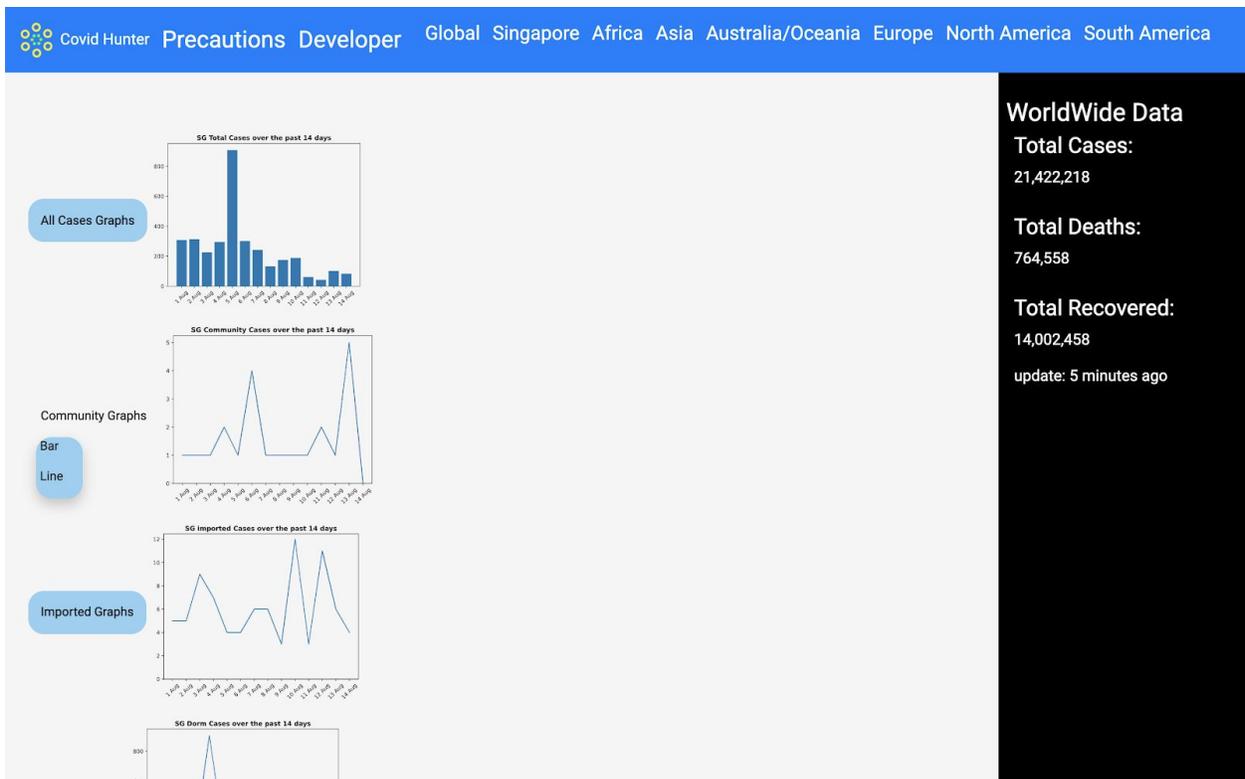
However, one problem faced during collection of global data is that some sources included data for additional states that were not found in the remaining sources. As data for these states were only found in 1 or 2 sources, they were unable to be compared against and were thus removed during data cleaning. An example is Kosovo which is not present in Worldometers and would be extremely unreliable if we were to include its data in our platform.

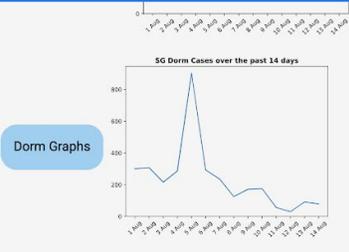
Here are images of webpages of other continents:



For our local data, we included graphs of several types, namely bar and line graphs for our daily statistics over the past 2 weeks on total, dormitory, imported and community cases, as well as 14 pie charts containing a detailed breakdown of such cases. The pie charts also come with a legend that shows the total number of dormitory, imported and community cases along with a percentage of each type of cases that day. All types of graphs are accompanied by a dropdown selection for users to choose their preferred graph, so that users have more options on whichever suits them best.

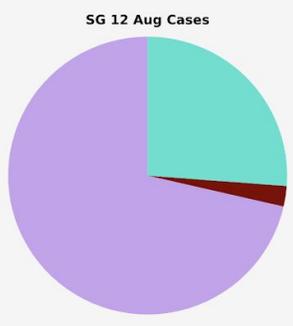
Here are some images of our webpage for local data:





- Date
- 14 Aug
- 13 Aug
- 12 Aug
- 11 Aug
- 10 Aug
- 9 Aug
- 8 Aug
- 7 Aug
- 6 Aug
- 5 Aug
- 4 Aug
- 3 Aug
- 2 Aug
- 1 Aug

13 Aug 2 Aug Total -- 42 Cases
 Dorm -- 71.43% -- 30 Cases
 Community -- 2.38% -- 1 Cases
 Imported -- 26.19% -- 11 Cases



WorldWide Data
Total Cases:
 21,422,218
Total Deaths:
 764,558
Total Recovered:
 14,002,458
 update: 5 minutes ago

Date

14 Aug Total -- 83 Cases
 Dorm -- 95.18% -- 79 Cases
 Community -- 0.00% -- 0 Cases
 Imported -- 4.82% -- 4 Cases

SG 14 Aug Cases



A mobile version of our website was also created for users' convenience as this would allow them to check the statistics more frequently and have all COVID-19 statistics at their fingertips.



5 Implications and Recommendations

a. Limitation

One limitation is the lack of time to include statistics for the number of cases, deaths and recoveries per million population as we would have to scrape multiple sources for the population data to ensure reliability.

b. Further Extension

Data for the total number of tests and critical cases could have been collected, and given more time, instead of collecting data from 3 non-governmental sources, we could have collected data directly from each country's government to allow for higher reliability.

6 Conclusion

In conclusion, the scraping of multiple sources and inclusion of various types of graphs has made our statistics more reliable and our platform more insightful. We have achieved our objective of providing people worldwide with more trustworthy data and have allowed them to better comprehend statistics through usage of charts and diagrams. Lastly, by sorting the data into continents, users are now able to compare at one glance the statistics of different countries in their region.

7 Bibliography

1. Schiffman, A. (2019). Coronavirus Dashboard. Retrieved July 13, 2020, from <https://ncov2019.live/>
2. Weiseng. (n.d.). COVID-19 Corona Tracker. Retrieved July 13, 2020, from <https://www.coronatracker.com/>
3. Coronavirus Cases:. (n.d.). Retrieved July 13, 2020, from <https://www.worldometers.info/coronavirus/>
4. Updates on COVID-19 (Coronavirus Disease 2019) Local Situation. (2020). Retrieved July 13, 2020, from <https://www.moh.gov.sg/covid-19>

5. Centers for Disease Control and Prevention. (n.d.). Retrieved August 11, 2020, from <https://www.cdc.gov/>