COVID-19 Data Project: Tracking the Pandemic

Nashae Prout[1] and Tracy Flood, MD Ph.D.[2]

[1]Morgan State University & [2]BroadStreet Health

Abstract

This pandemic is unlike anything encountered in recent history. BroadStreet Health has implemented a volunteer-driven COVID-19 Data Project with the ultimate goal of producing a publicly available database on the COVID-19 pandemic in the US. With 400+ volunteers training in data entry, data validation, data visualization, and epidemiologic methods the BroadStreet team has produced 1000+ graphs, 2400+ policies, and several intern led projects. The resulting publically available and free dataset includes county and state level data for all 50 states, DC and Puerto Rico. The next steps for the project are to continue to assess policies, monitor cases, and publish COVID-19 data at the county and state level; as well as to continue providing free, research-ready datasets for the public good.

Background/Hypothesis

- The first official US cases of COVID-19 were reported February 2020 (AJMC 2020)
- With millions of cases and hundreds of thousands of deaths (CDC, 2020) immensely large data sets are generated rapidly.
- Accuracy in COVID-19 data collection and reporting is needed to inform medical and policy decisions regarding treatment and response efforts.

We hypothesized that without guidance for case definitions or strict protocols for data publication, daily COVID-19 data would highly vary in quality. Therefore the project aims include tracking the pandemic at county levels and above, assessing issues in data publication, providing a publicly available dataset on COVID-19, and showing shortfalls in the US' pandemic response.

Methods

Study design: Observational
Population: US Population
Instruments: Google Drive, Tableau, Maptiler, Adobe Xd, & Slack
Analytic Strategy: Using comparison data sets & volunteer-led special projects
Procedure:

1. Volunteers input primary data for COVID-19 cases and deaths at the county- and state-level based on publicly available data from local health departments into proper Google document.
❖ Volunteers rotate states while conducting primary data entry for 4 weeks before being eligible to join teams or special projects.
2. Quality Assurance Team volunteers analyze data and address anomalies by comparing our data to other sources like JHU, NYT, and the CDC to validate it.
3. Data Visualization Team volunteers take validated data to create charts and maps using: Tableau, Maptiler, Adobe XD, etc
❖ Volunteers can conduct special projects or initiatives related to COVID-19 for publication on the project website after getting their project approved and reviewed by program administrators.
4. Lastly, all COVID-19 data products and volunteer-led projects are reviewed and published on the project's official website.
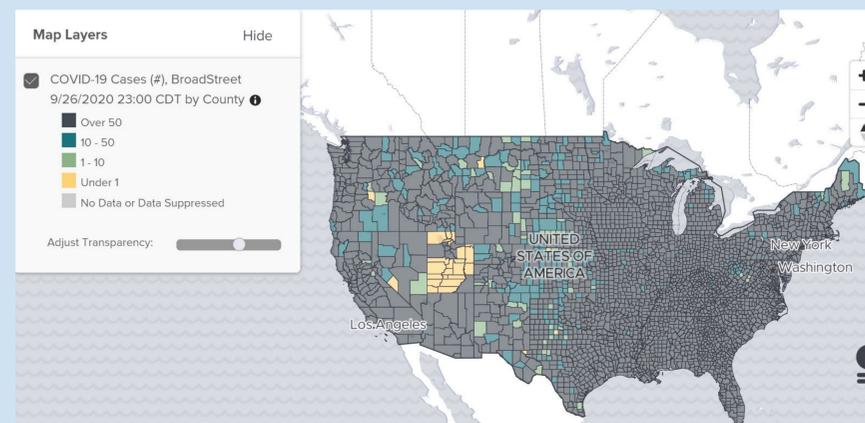
Results

Figure 1. Map of US indicates the widespread presence of COVID-19. All 50 states, the District of Columbia, and Puerto Rico have confirmed cases and deaths.
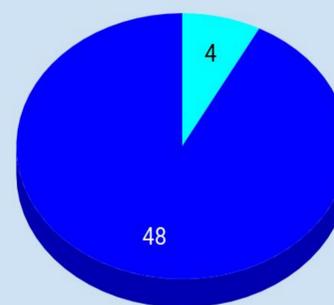

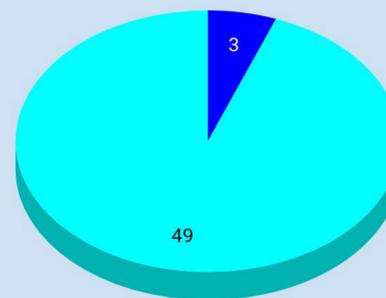Figure 2. Pie chart of states/territories that include any demographics data in testing data 4 out of 52.

Figure 3. Pie chart of states/territories that include any demographics data in confirmed case data 49 out of 52.
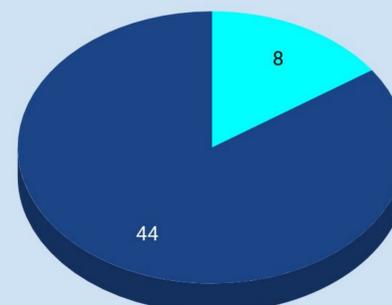

Figure 4. Pie chart of states/territories that include any demographics data in mortality case data 8 out of 52.

Conclusion

- States have responded to the pandemic differently, and therefore have had differing outcomes when it comes to cases and deaths.
- Some have more uncertain COVID-19 rates than others like: examples include Texas, California, and Florida.
- It was found that states use different definitions for reported numbers. This includes different definitions of positive cases. Varying definitions may also contribute to observed differences in regional numbers trends.
- Non-uniform COVID-19 reporting has also been noticed. These findings suggest the need to standardize COVID-19 data collection and reporting practices nationwide.
- Based on the findings, it can be determined that the pandemic is ongoing and has **significant** data and policy shortfalls.

Future Work

The next steps for the project are to continue to assess policies, monitor cases, and publish COVID-19 data at the county and state level; as well as to continue providing free, research-ready datasets for the public good.

References

AJMC Staff. (2020, July 3). A Timeline of COVID-19 Developments in 2020. The American Journal of Managed Care Retrieved September 28, 2020, from https://www.ajmc.com/view/a -timeline-of-covid19-developments -in-2020

Coronavirus Disease 2019 (COVID-19). (2020). Centers for Disease Control and Prevention. Retrieved September 28, 2020, from https://www.cdc.gov/coronavirus/2019-ncov/index.html

BroadStreet COVID-19 Data Project. (2020, August 27). BroadStreet Health. Retrieved September 28, 2020, from https://covid19dataproject.org/

Why It Matters. (2020).The Covid Tracking Project. Retrieved September 28, 2020, from https://covidtracking.com/about/why-it-matters

Acknowledgments