

An online dashboard for tracking, comparing, and forecasting COVID-19 in counties across the US

Danqing Wang^{4, 8}, Maya Shen⁶, James Duncan³, Chandan Singh², Tiffany Tang¹, Yu Wang¹,
Nick Altieri¹, Xiao Li¹, Abhineet Agarwal⁴, Robert Netzorg², Raaz Dwivedi²,
Peter Norvig⁸, Bin Yu^{1, 2, 3, 5, 7}

¹ Department of Statistics, ² Department of EECS, ³ Division of Biostatistics

⁴ Department of Physics, ⁵ Center for Computational Biology
University of California, Berkeley

⁶ Department of Statistics, Carnegie Mellon University

⁷ Chan Zuckerberg Biohub, San Francisco

⁸ Google

Abstract

Objective: This work seeks to provide an online resource to help the public understand current and future county-level COVID-19 outbreaks in the US.

Materials and Methods: This website aggregates data from multiple resources and provides short-term COVID-19 forecasts in an interactive, online dashboard.

Results: The web resource can be accessed online at covidseverity.com. It displays COVID-19 data from every county along with short-term (7-day) forecasts and forecast intervals. Furthermore, we provide many ways to search and explore the data such as allowing users to interact with the data by looking at cases, deaths, new cases, and new deaths. It also allows for interactive, searchable visualizations of different counties and allows users to cluster counties using various metrics. All code and data sources are made available on Github at github.com/Yu-Group/covid19-severity-prediction.

Discussion: With a variety of academic sources providing different information, it has become increasingly important to provide easily accessible and rigorous COVID-19 data and forecasts. We hope the resource introduced here can help fill this need.

Conclusions: Our website covidseverity.com can help provide the public with information and short-term forecasts on COVID-19 outbreaks for individual counties.

Keywords: COVID-19, dashboard, forecasting, interactive visualization

1 Introduction

COVID-19 continues to spread across the United States. As of Oct 29, 2020, more than 9 million confirmed cases and 225 thousand deaths have been reported due to COVID-19 in the US [1]. Outbreaks are not isolated to small geographic regions in the US and continue to appear all across the country.

Though many visualization tools exist to track these outbreaks [2, 3, 4], there is still a need for a rigorous and accessible interface for short-term forecasting and comparative exploration of COVID-19 outbreaks at the county-level. We develop such an interface at covidseverity.com to provide the public with information and short-term forecasts on COVID-19 outbreaks for individual counties. The interface allows users to explore eight different metrics for every US county as well as compare each county with counties similar to it based on several characteristics.

2 Materials and Methods

All source code is available on [GitHub](https://github.com) and is built using Plotly and Dash [5]. The dashboard is hosted on an Amazon Web Services (AWS) EC2 instance, which is configured to automatically run a number of scripts in the GitHub repository on a daily schedule. These scripts download the latest data from a number of sources, run pre-processing and forecasting pipelines, and refresh the dashboard with the latest interactive visualizations. We describe the main components of the dashboard and its data below.

2.1 Dashboard

Eight metrics The dashboard provides eight metrics to visualize the COVID-19 outbreak at the county-level: cumulative cases, cumulative deaths, daily new cases, daily new deaths, cumulative cases per 100k population, cumulative deaths per 100k population, daily new cases per 100k, and daily new deaths per 100k. Each of these metrics provides a different perspective on the dynamics of the COVID-19 outbreak in the US.

Forecasting and data The dashboard also shows time-series curves for cases and deaths for each county along with historical and current seven-day forecasts and forecast intervals from previous work [6]. Data is aggregated from a variety of sources in that work, including COVID-19 case and death data from the New York Times [1] and USAFacts [7]. An overview of the datasets is provided on the [GitHub](#) repository. For more details, we refer interested readers to the previous paper [6].

2.2 County comparisons

We provide an interactive visualization which allows users to easily compare counties based upon the various metrics listed above. For example, we can visualize cases and deaths for different counties for side-by-side comparisons. We can also view, for any county, the five counties most similar to it (i.e., nearest neighbors) with respect to different county-level features including the number of ICU beds, recent daily deaths, median age in a county, and total deaths. The similarity between counties given the user-selected features is measured using the Euclidean norm of the difference between features, after they are normalized to have mean 0 and standard deviation 1.

3 Results

Our dashboard is available online at covidseverity.com. It shows data for all US counties and is updated daily. Below, we provide three snapshots from our interactive dashboard to highlight its core visualizations and features.

3.1 Dashboard visualizations

Fig 1 shows a screenshot of the web interface. By selecting tabs at the top, one can interactively visualize different metrics for COVID-19 cases and deaths on the map. The left table shows the counties with the highest values for the chosen metric. The user can zoom into the map to focus on a particular region, state, or county. When a cursor hovers over a particular county, a text box appears with the county’s name, state, and the values of each metric listed in the tabs at the top.

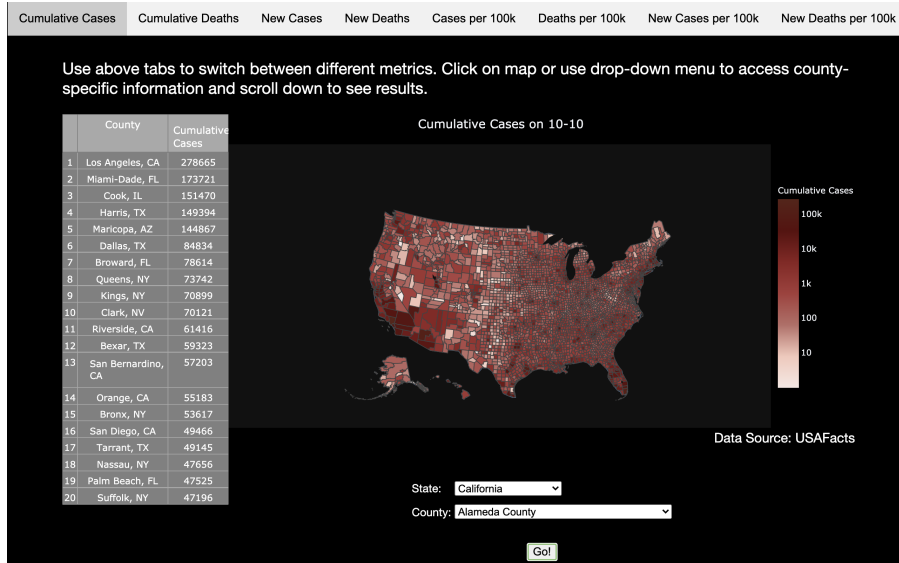


Figure 1: Screenshot of the interactive dashboard map showing cumulative cases for each county. The left panel shows the 20 counties with the highest values of the metric selected in the tabs at the top.

Below the map, a drop-down menu allows the user to choose a specific county to further explore (see Fig 2) in terms of cases, deaths, and past or future forecasts. This visualization includes the observed time-series and seven-day forecast (left plot in Fig 2), as well as the method’s historical forecast and intervals since its inception (right plot in Fig 2). This allows users to judge our model’s historical performance for any county and is an important feature often missing from dashboards that provide forecasts.

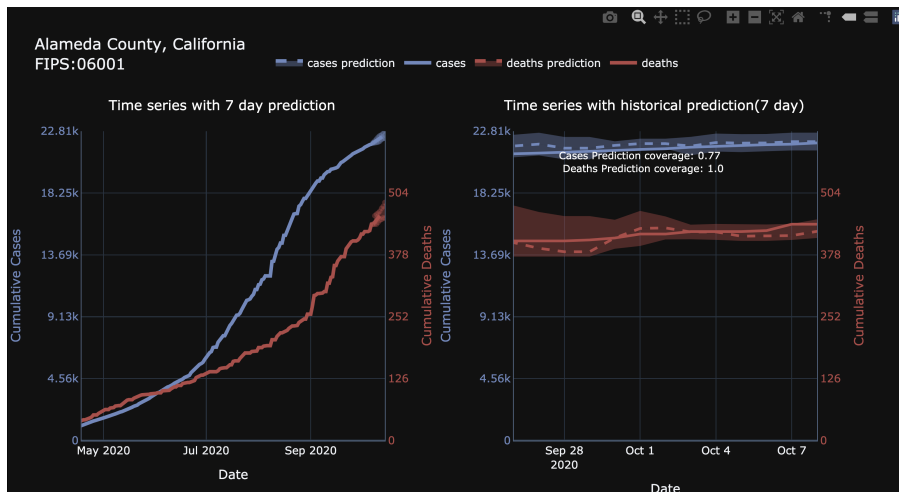


Figure 2: Screenshot of the county-specific dashboard displaying cases, deaths, and forecasts for a selected county. The left time-series shows observed case and death counts as well as the seven-day forecast for the selected county. The right time-series provides the method’s *historical* forecast and intervals.

3.2 County comparison visualizations

We develop a tool for users to conveniently compare COVID-19 severity for counties across the country, shown in Fig 3. We provide time-series curves of COVID-19 cases and deaths in selected counties with and without alignment to their first case or death (bottom and top rows of the time-series plots in Fig 3, respectively). This feature allows for an easy comparison of counties in different phases of their COVID-19 situation.

In the left panel, users have the option to manually select a number of counties for comparison or have the dashboard automatically provide the five most similar counties. The “Manual Selection” mode allows the user to pick one or more counties of interest and corresponding plots are shown in the right panel. In “Nearest Neighbors” mode, after selecting a county and features of interest, the dashboard returns five counties that are most similar to the selected county based upon the selected features (see Fig 3). The similarity metric between counties is measured by first normalizing the selected features and then calculating the L_2 distance between them.

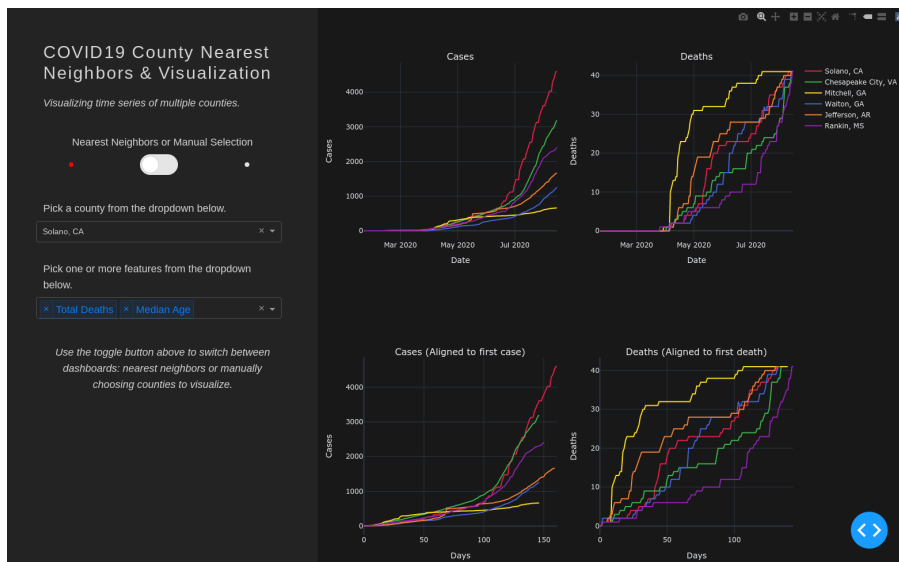


Figure 3: Screenshot of “Nearest Neighbors” mode, which automatically provides similar counties based on chosen metrics; in this case, total deaths and median age.

4 Discussion

We hope the dashboard described herein will be a useful resource to a variety of users, especially to allow the general public to get a more holistic view of COVID-19 outbreaks in the US. The ability to

easily compare different counties and metrics for COVID-19 cases and deaths should enable a better understanding of the outbreak and its trajectory in counties across the country.

The data provided in these dashboards has its limitations (discussed in more detail in [6]). Limited testing suggests that case and death counts are under-reported. Moreover, variations in testing across the US can lead to differences in reported counts for cases and deaths. Nevertheless, we hope the provided visualizations are useful for the public when making comparisons between counties in terms of their COVID-19 severity.

5 Acknowledgements

The authors declare that they have no conflict of interest.

The authors would like to thank many people for help with this effort. Our acknowledgement section is unusually long because it reflects the nature of an ER-like collaborative project. We were greatly energized by the tremendous support and the outpouring of help that we received not only from other research groups, but also high school students, medical staff, ER doctors, and several volunteers who signed up at Response4Life [8].

Bin Yu acknowledges the support of CITRIS Grant 48801, a research award by Amazon, and AWS. We also would like to acknowledge support from Google through a summer intern and her Google supervisors, particularly Cat Allman, Peter Norvig, and Matt Gaunt.

We would like to thank the Response4Life team (Don Landwirth and Rick Brennan in particular), and volunteers for building the base for this project. We thank Aaron Kornblith and David Jaffe for advice from a medical perspective, especially Aaron for a great deal of useful feedback. We want to mention special thanks to Sam Scarpino for sharing data with us. We would also like to thank Max Shen's IEOR group at Berkeley: Junyu Cao, Shunan Jiang, Pelagie Elimbi Moudio for helpful inputs in the early stages of the project. In addition, we would like to thank the high school students Matthew Shen, Anthony Rio, Miles Bishop, Josh Davis, and Dylan Goetting for helping us to collect valuable hospital related information.

Finally, we acknowledge helpful input on various data and modeling aspects from many others including Ying Lu, Tina Eliassi-Rad, Jas Sekhon, Philip Stark, Jacob Steinhardt, Nick Jewell, Valerie Isham, Sri Satish Ambati, Rob Crockett, Marty Elisco, Valerie Karplus, Marynia Kolak, Andreas Lange, Qinyun Lin, Suzanne Tamang, Brian Yandell, and Tarek Zohdi.

References

- [1] The New York Times. COVID-19 Data in the United States. <https://github.com/nytimes/covid-19-data>, 2020. Accessed on 04-01-2020 at <https://github.com/nytimes/covid-19-data>.
- [2] Ensheng Dong, Hongru Du, and Lauren Gardner. An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases*, 20(5):533–534, 2020.
- [3] Benjamin D Wissel, PJ Van Camp, Michal Kouril, Chad Weis, Tracy A Glauser, Peter S White, Isaac S Kohane, and Judith W Dexheimer. An interactive online dashboard for tracking covid-19 in us counties, cities, and states in real time. *Journal of the American Medical Informatics Association*, 2020.
- [4] Douglas Arneson, Matthew Elliott, Arman Mosenia, Boris Oskotsky, Rohit Vashisht, Travis Zack, Paul Bleicher, Atul J Butte, and Vivek A Rudrapatna. Covidcounties-an interactive, real-time tracker of the covid-19 pandemic at the level of us counties. *medRxiv*, 2020.
- [5] Plotly Technologies Inc. Collaborative data science. 2015.
- [6] Nick Altieri, Rebecca L Barter, James Duncan, Raaz Dwivedi, Karl Kumbier, Xiao Li, Robert Netzorg, Briton Park, Chandan Singh, Yan Shuo Tan, Tiffany Tang, Yu Wang, Chao Zhang, and Bin Yu. Curating a covid-19 data repository and forecasting county-level death counts in the united states. *Harvard Data Science Review*, 11 2020. <https://hdr.mitpress.mit.edu/pub/p6isyf0g>.
- [7] USAFacts. COVID-19 Deaths Data. Accessed on 03-31-2020 at <https://www.reuters.com/article/us-health-coronavirus-who/covid-19-spread-map>.
- [8] Response4life. Response4life. <https://response4life.org>, 2020. Accessed on 10-01-2020 at <https://github.com/nytimes/covid-19-data>.