Visualizing the lagged connection between COVID-19 cases and deaths in the United States: An animation using per capita state-level data (January 22, 2020 – July 8, 2020)

The Harvard Center for Population and Development Studies

Working Paper Series

Visualizing the lagged connection between COVID-19 cases and deaths in the United States: An animation using per capita state-level data (January 22, 2020 – July 8, 2020)

Christian C. Testa, BS[1], Nancy Krieger, PhD[1], Jarvis T. Chen, ScD[1], and William P. Hanage, PhD[2]

July 10, 2020

HCPDS Working Paper Volume 19, Number 4

Title: Visualizing the lagged connection between COVID-19 cases and deaths in the United States: An animation using per capita state-level data (January 22, 2020 – July 8, 2020)

Authors: Christian C. Testa, BS[1]
Nancy Krieger, PhD[1]
Jarvis T. Chen, ScD[1]
William P. Hanage, PhD[2]

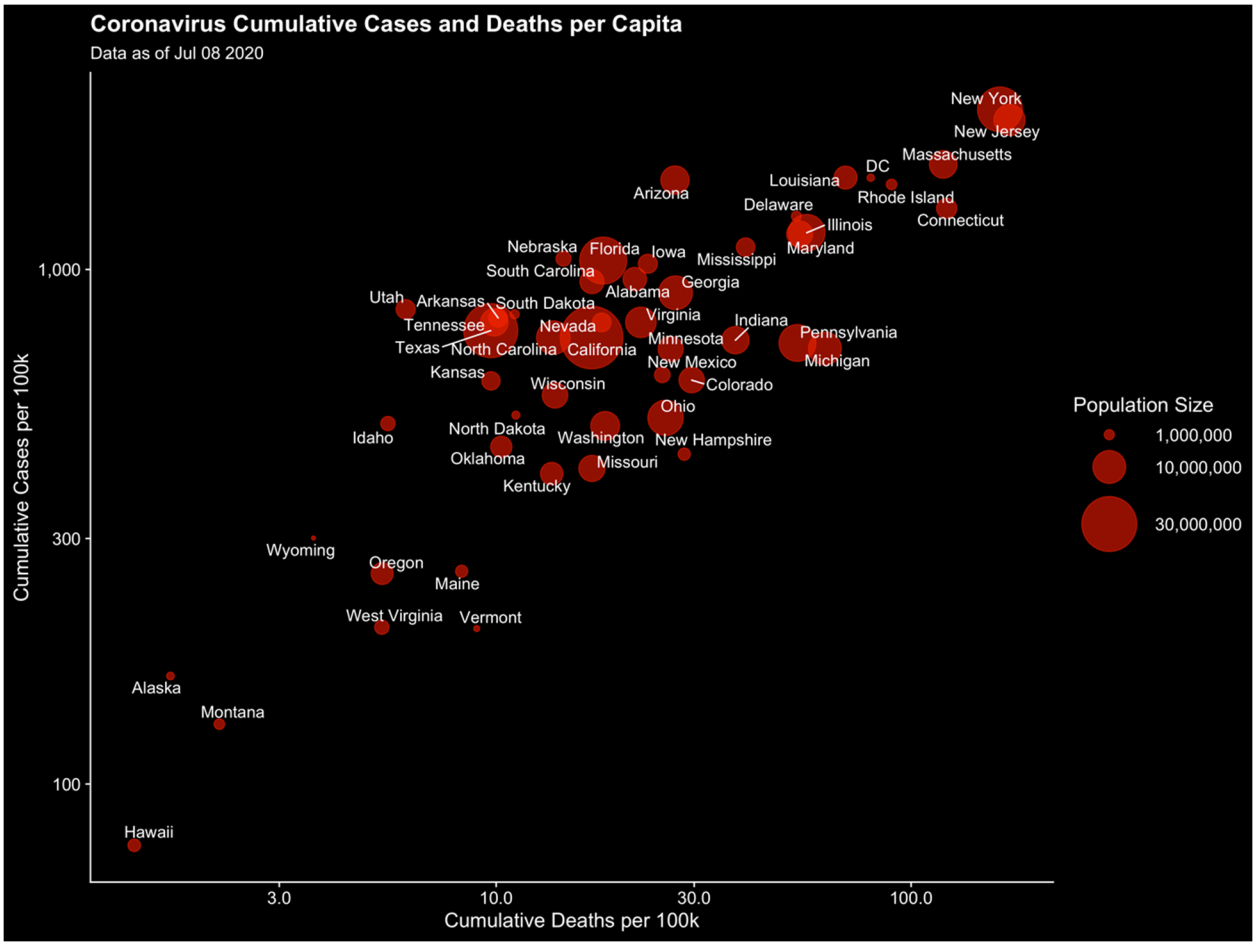
Corresponding Author: Christian C. Testa, BS
Statistical Data Analyst,
Department of Social and Behavioral Sciences,
Harvard T.H. Chan School of Public Health
Email: ctesta@hsph.harvard.edu

Affiliations:

1. Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health
2. Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard TH Chan School of Public Health

ABSTRACT

Data visualizations of the COVID-19 pandemic in the United States often have presented case and death rates by state in separate visualizations making it difficult to discern the temporal relationship between these two epidemiological metrics. By combining the COVID-19 case and death rates into a single visualization we have provided an intuitive format for depicting the relationship between cases and deaths. Moreover, by using animation we have made the temporal lag between cases and subsequent deaths more obvious and apparent. This work helps to inform expectations for the trajectory of death rates in the United States given the recent surge in case rates.



Animated version available at
https://github.com/ctesta01/animated_covid_rates/blob/master/covid19_cumulative_cases_and_deaths_animated.gif

Keywords: COVID-19, cases, deaths, data visualization, animation

Introduction

To date, leading visualizations for COVID-19 in the US have presented COVID-19 reported case and death rates in side-by-side or separate visualizations [1-3]. However, using animation could potentially provide a compelling format and narrative depicting how COVID-19 cases and deaths have progressed in the United States. We explore this by depicting the cumulative cases and deaths rates by state with rates per capita over time. Our animation thus shows not only the relationship between case and death rates, but also their lagged relationship over time, reflecting that deaths often occur 2-8 weeks after onset of symptoms [4].

Methods

We analyzed COVID-19 case and death rates as reported by USA Facts for the dates January 22, 2020 through July 8, 2020 [1]. Our analysis is exempt from IRB review, since only publicly available de-identified case and death data aggregated to the state level are employed.

Data are presented as proportions of cumulative cases and deaths per 100,000 residents. Data manipulation and visualization was carried out in the R programming language [5] using packages from the tidyverse (including dplyr, readr, ggplot2, magrittr, and lubridate) [6], as well as the janitor [7], ggrepel [8] and gganimate packages [9].

Our animated visualization, static visualizations, and the code to reproduce them are available at: https://github.com/ctesta01/animated_covid_rates

Data visualized are available for download from https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/

Discussion

Our visualization makes clear the lag between cases and deaths in COVID-19 outcomes in the United States. In the initial stages of the epidemic, for many states case rates increase while death rates remained low. After this initial surge in case rates across the United States, death rates subsequently increased in all states. As a result of this lagged relationship between COVID-19 cases and deaths and given the recent surge in COVID-19 cases, we expect that deaths attributable to COVID-19 will increase 4-6 weeks after observed surges in case rates. However, the magnitude of the subsequent surge in deaths attributable to COVID-19 in a given geography will be affected by the age distribution of the geography and its age-specific COVID-19 case rates.

A major limitation of our analysis is that it relies on reported COVID-19 cases and deaths. Underestimation of cases, due to lack of access to testing, and related underestimation of confirmed COVID-19 deaths, could thus potentially bias results [10]. If, however, the relative underestimation is similar for both outcomes during the same time period, the correlation of the lag will not be biased.

In common with all similar analyses, we are limited by the fact that we are only able to consider the pandemic up to the time of writing. We are anxious that lessons from places with high rates of transmission, disease and death, which in the US have been overwhelmingly concentrated among communities already without adequate healthcare and not provided with needed

protection from the pandemic (among many other existing inequities) should be learned and communicated: whatever the policy and response failures in the past, we can work toward a better future [11-14]. Our work suggests the consequences of what may be expected in the near future, given the increased opportunities for transmission occurring in numerous US states, and the resulting disease and death.

References

1. Coronavirus Locations: COVID-19 Map by County and State. USA Facts. Available at: https://usafacts.org/visualizations/coronavirus-covid-19-spread-map Accessed July 9 2020.

2. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. https://www.covidtracker.com/ Accessed July 9 2020.

3. NYTimes. Coronavirus in the U.S.: Latest Map and Case Count. Available at: https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html

4. WHO. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf Accessed July 9 2020.

5. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

6. Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

7. Sam Firke (2020). janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 2.0.1. https://CRAN.R-project.org/package=janitor

8. Kamil Slowikowski (2020). ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.8.2. https://CRAN.R-project.org/package=ggrepel

9. Thomas Lin Pedersen and David Robinson (2020). gganimate: A Grammar of Animated Graphics. R package version 1.0.5. https://CRAN.R-project.org/package=gganimate

10. Pearce N, Vandenbroucke JP, VanderWeele TJ, Greenland S. Accurate Statistics on COVID-19 Are Essential for Policy Guidance and Decisions. *Am J Public Health*. 2020; 110(7):949-951. doi:10.2105/AJPH.2020.305708

11. Pilkington E. As 100,000 die, the virus lays bare America's brutal fault lines – race, gender, poverty, and broken politics. *The Guardian*, May 28, 2020. https://www.theguardian.com/us-news/2020/may/28/us-coronavirus-death-toll-racial-disparity-inequality ; accessed: July 10, 2020.

12. Bailey ZD, Moon JR. Racism and the political economy of COVID-19: will we continue to resurrect the past? *J Health Politics, Policy, Law* 8641481, May 28, 2020. https://read.dukeupress.edu/jhppl/article/doi/10.1215/03616878-8641481/165296/Racism-and-the-Political-Economy-of-COVID-19-Will ; accessed: July 10, 2020.

13. Bambra C, Riordan R, Ford J, Matthews F. The COVID-19 pandemic and health inequalities. *J Epidemiol Community Health* 2020 [epub ahead of print]; doi:10.1136/jech-2020-214401 ; https://jech.bmj.com/content/jech/early/2020/06/13/jech-2020-214401.full.pdf ; accessed: July 10, 2002

14. Chotiner I. The interwoven threads of inequality and health. The coronavirus crisis is revealing the inequities inherent in public health due to societal factors, Nancy Krieger, a professor of social epidemiology, says (Interview with Nancy Krieger). *The New Yorker*, April 14, 2020. https://www.newyorker.com/news/q-and-a/the-coronavirus-and-the-interwoven-threads-of-inequality-and-health ; accessed: July 10, 2020.