Documenting Social, Geographic, and Economic Inequities in COVID-19 Mortality at the County Level in the U.S. Using Generalized Additive Models

Christian Testa[1,*]

November 13, 2022

Abstract

We present three types of applications of generalized additive models (GAMs) to COVID-19 mortality rates in the US for the purpose of advancing methods to document inequities with respect to which communities suffered disproportionate COVID-19 mortality rates. First, GAMs can be used to describe the changing relationship between COVID-19 mortality and county-level covariates (sociodemographic, economic, and political metrics) over time. Second, GAMs can be used to perform spatiotemporal smoothing that pools information over time and space so that county level mortality rates that tend to be noisy due to small population counts or stochasticity can be summarized by a smooth, dynamic latent surface describing the mortality risk associated with geographic locations over time. Third, estimation of the associations between county-level covariates and COVID-19 mortality after controlling for spatiotemporal effects allow for the distinguishing of what patterns in COVID-19 mortality were more plausibly due to geography than variation in county-level covariates. Each of these approaches provides a valuable approach and perspective to documenting inequities in COVID-19 mortality by addressing the question of which populations have suffered the worst burden of COVID-19 mortality taking into account the nonlinear spatial, temporal, and social patterning of disease.

Abbreviations used: United States (US), Coronavirus Disease 2019 (COVID-19), Generalized Additive Model (GAM), Centers for Disease Control and Prevention (CDC)

[1] Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health

[*] Correspondence: Christian Testa <ctesta@hsph.harvard.edu>

Introduction

As we enter the third winter with the novel coronavirus disease COVID-19 in the United States, evidence documenting the intense disparities in COVID-19 mortality rates comparing socially advantaged and disadvantaged populations continues to mount. Eliminating inequities in health outcomes has been stated as a major policy goal of the Biden administration (The White House, 2021, 2022), representing a revitalized focus on health justice and underlining the importance of adequate data reporting systems that measure the prevalent health inequities. To this end, we use generalized additive models (GAMs) in a flexible regression framework to illustrate the evolving roles and relationships sociodemographic, geographic, and economic conditions have with respect to trends in COVID-19 mortality. The code, data, and documentation necessary to reproduce the analyses contained in this paper are online and free to access at https://github.com/ctesta01/covid.gradient.estimation.

Background

Having passed over 1 million COVID-19 deaths in the United States earlier this year (Donovan, 2022), and facing uncertain prospects for the third COVID-19 winter looming even as new iterations on the COVID-19 vaccines become available, it remains critical that inequities in COVID-19 outcomes are documented and analyzed to reckon with the unjust and unfair burden of preventable illness. Even though the first vaccines were granted emergency use authorization by the U.S. Food and Drug Administration in 2020 (Mayo Clinic, 2022), with the first shots going in arms in December 2020, COVID-19 is still continuing to cause hundreds of deaths a day in the U.S. in the fall of 2022 ("United States COVID - Coronavirus Statistics - Worldometer," 2022). The new bivalent vaccines released at the end of August 2022 contain mRNA sequences from both the original strain as well as the recently emergent BA.4 and BA.5 lineages in an effort to make the nation's immunity more up-to-date and robust against the myriad of phylogenetic directions the COVID-19 virus is evolving to explore (Office of the Commissioner, 2022). Despite the updated bivalent boosters representing a significant step forward in prevention strategy, less than 4% of eligible Americans had taken the booster in the first month after it became available (Bendix, 2022; Lambert, 2022). As such, and with an enduring history of inequities in health care availability (Bailey et al., 2021; Blendon et al., 2002; Carpenter, 2021; Chrisler et al., 2016; Feldman et al., 2021; Okonkwo et al., 2021; Ortega & Roby, 2021; Rapp et al., 2022; Whitehead et al., 2016), it is clear that without further intervention not all communities will be equally able to benefit from the new vaccines available and inequities in COVID-19 illness and mortality may persist despite the technological innovations in vaccine technology.

The Role of Geography in COVID-19

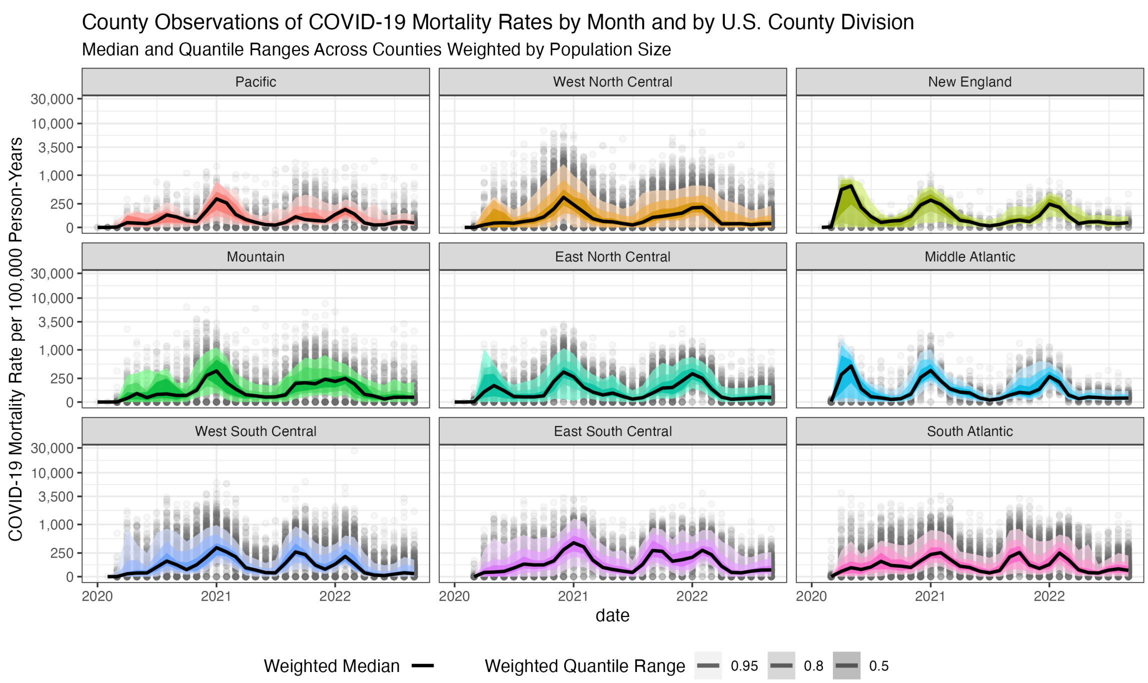


Figure 1: Estimates of monthly COVID-19 mortality rates per 100,000 person-years by county organized by Census Division. For each division, the median trendline and quantile ranges are shown weighted by county population size.

Prior literature has demonstrated that geography has played a significant role in the spread and impact of of COVID-19. Methods employed to highlight the role geography plays have included quantile regression (Sigler et al., 2021), Besag-York-Molli'e mixed models (Whittle & Diaz-Artiles, 2020), spatial cluster analysis (Sugg et al., 2021), geographically weighted regression (Mollalo et al., 2020; Park et al., 2021), and others.

In the U.S. context, one of the key aspects to the geographic story of COVID-19's spread and diffusion was the early surge of cases and epicenter in New York City during March 2020 (Thompson, 2020) followed by subsequent waves of cases in the South and Midwest (Glenza, 2020; Scott, 2020; Shumaker & Wu, 2020). As Park et al. stated in their article *COVID-19 Deaths in the United States: Shifts in Hot Spots over the Three Phases of the Pandemic and the Spatiotemporally Varying Impact of Pandemic Vulnerability* where they summarized US trends from March 2020 to May 2021, “hot spots have shifted from densely populated cities and the states with a high percentage of socially vulnerable individuals to the states with relatively relaxed social distancing requirements, and then to the states with low vaccination rates” (Park et al., 2021).

When considering the drivers of the COVID-19 pandemic, it’s necessary to note that geography and social conditions are inextricably linked. In July 2021, the CDC reported that “the COVID-19 cumulative death rate in non-metropolitan areas has exceeded that of metropolitan areas since December 2020,” noting that of the approximately 1/5th of Americans who live in rural areas, many “are considered highly vulnerable according to CDC’s Social Vulnerability Index (SVI), which includes factors such as housing, transportation, socioeconomic status, race, and ethnicity” (CDC, 2021). Moreover, rural communities often have lower health insurance rates, higher disability rates, older populations, and limited access to health care. One of CDC’s Morbidity and Mortality Weekly Reports found that vaccination against COVID-19 was lower in rural communities than in urban communities between December 2020 and April 2021 (Murthy, 2021).

Figure 1 shows the monthly COVID-19 mortality rates for the counties grouped within each of the nine U.S. Census Divisions (U.S. Department of Commerce Economics and Statistics Administration & U.S. Census Bureau, 2000). The figure summarizes each division’s median mortality rates weighted by county population size. Quantile ranges are included to illustrate the range of mortality rates observed, again weighted by county population size. Weighting based on county population size is used as opposed to equal weighting for calculating the median and quantile ranges so that in divisions where low population counties predominate, mortality rate estimates that are noisy due to small population counts are not weighted equally to more reliable mortality rate estimates from larger population counties. Notably, the mortality associated with the early surge of cases starting in NYC and spreading through New York, New Jersey, and Massachusetts is visible in the Middle Atlantic and New England division figures. The figure also illustrates how the first peak in the mortality time-series for states in the Midwest (West North Central, East North Central) occurred later, in late 2020 and going into early 2021.

The Social Determinants of COVID-19 Mortality

- Insert section on prior modeling results
- Insert discussion motivating the need for spatiotemporal modeling

Methods

Data Sources

The following variables were retrieved at the county level:

- Counts of COVID-19 deaths (The New York Times, 2021).
- Population size estimates for 2020 from the U.S. Census (Redistricting File Public-Law 94-171 Dataset) (US Census Bureau, 2021).
- Median age, median household income, racial/ethnic composition, population density, percent below the federal poverty line, and number of households with high ($100k+)/low (<$25k) household income by racial/ethnic group from the 2014-2019 5-year American Community Survey (US Census Bureau, 2020) through the `tidyverse` R package (Walker & Herman, 2022).
- Votes cast in the 2020 presidential election (**mit_county_2022?**)

Generalized Additive Models

Traditional generalized linear models fit a regression model using the functional relation $g(\mu_i) = \mathbf{X}_i\beta$ where Examples of common link functions include the identity function for linear regression, log for Poisson regression, and the logit function for logistic regression.

Generalized additive models (GAMs) improve upon generalized linear models by allowing for the fitting of smooth functions that transform the x variables. This is a convenient means to include nonlinear relationships between the outcome and predictor variables. Whereas a generalized linear model may have looked like

$$g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} ...$$

a generalized additive model could look like

$$g(\mu_i) = \mathbf{A}_i \theta + f_1(x_{i1}) + f_2(x_{i2}) + f_3(x_{i3}, x_{i4}) + \dots$$

where $\mu_i \equiv \mathbb{E}(Y_i)$, $\mathbf{A}_i$ is a row of the model matrix for any strictly non-parametric model components, θ is the corresponding parameter vector, and the f_j are smooth functions of the covariates x_k (Wood, 2017).

Results

Evaluating Spatial and Temporal Autocorrelation

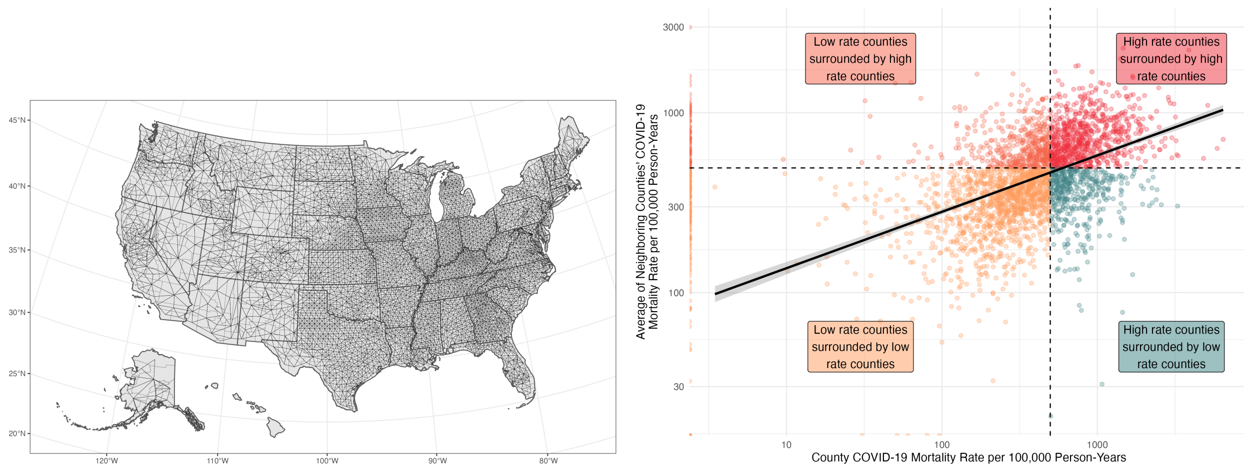


Figure 2: A) The spatial neighbor graph for counties in the US. In this graph, county centroids are connected if the two counties are adjacent to one another (i.e., share a boundary). This and similar graphs that represent neighbor relationships or geographic proximity are often used in modeling contexts to create spatial weights that account for which regions help to explain a given region's observed data.
(#fig:figure county neighboring graph)

Regression Modeling Results

We used GAMs to describe the following:

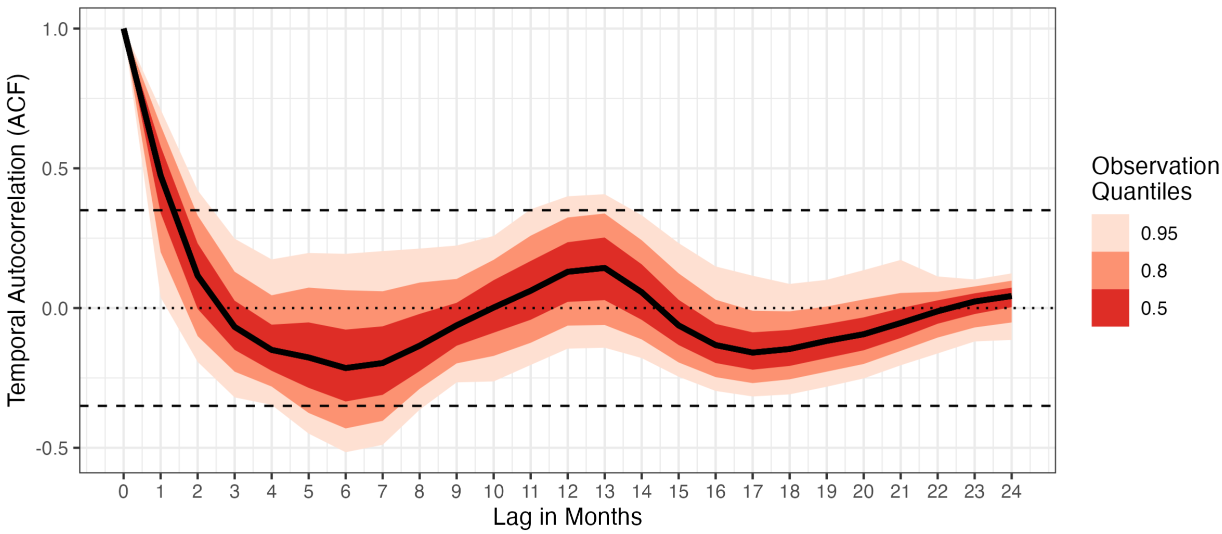
- the changing associations between individual area-based measures and COVID-19 mortality over time;
- the changing associations between bivariate area-based measures and COVID-19 mortality over time;
- associations with area-based measures and COVID-19 mortality after controlling for spatio-temporal autocorrelation.

Additionally, as sensitivity analyses, random forest models were fit to compare with the coefficient estimates from the GAMs.

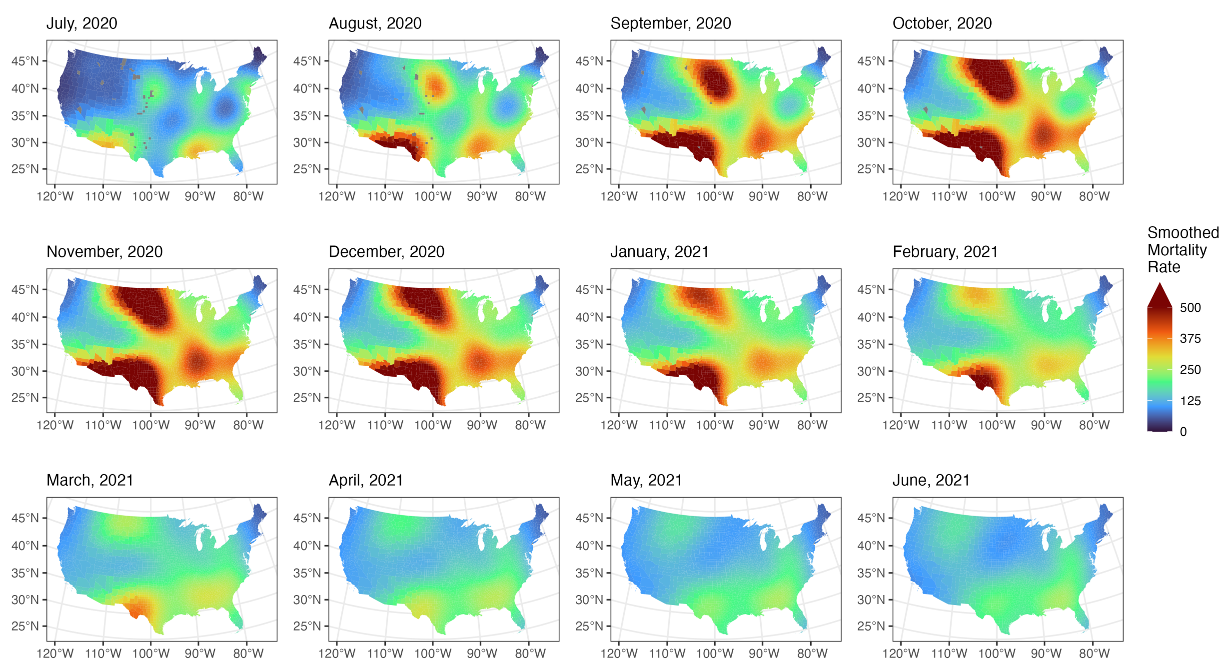
Discussion

Limitations

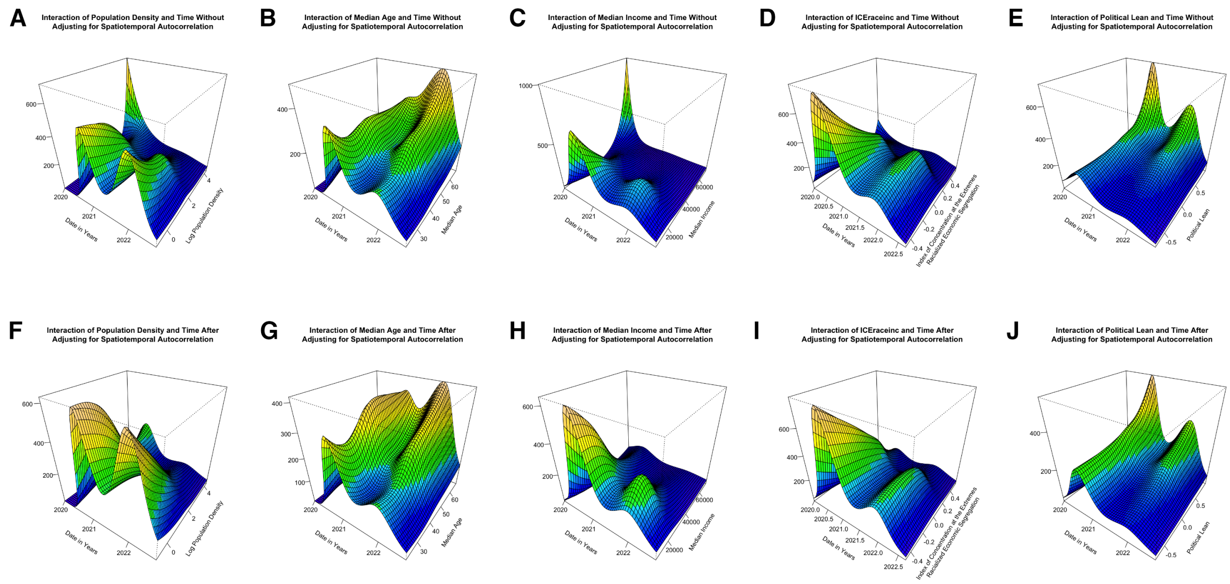
- Missing data
- Ecological fallacy



Temporal Autocorrelation (ACF)
Lag in Months
Observation Quantiles
0.95
0.8
0.5



July, 2020
August, 2020
September, 2020
October, 2020
November, 2020
December, 2020
January, 2021
February, 2021
March, 2021
April, 2021
May, 2021
June, 2021
Smoothed Mortality Rate



- Age effects
- Fixed population denominators -> real-time denominator estimates
- Vaccination and mobility effects

Conclusions

References

Bailey, Z. D., Feldman, J. M., & Bassett, M. T. (2021). How Structural Racism Works — Racist Policies as a Root Cause of U.S. Racial Health Inequities. *New England Journal of Medicine*, *384*(8), 768–773. https://doi.org/10.1056/NEJMms2025396

Bendix, A. (2022). Less than 4% of eligible people have gotten updated Covid booster shots, one month into the rollout. In *NBC News*. https://www.nbcnews.com/health/health-news/updated-covid-booster-shots-doses-administered-cdc-rcna48960

Blendon, R. J., Schoen, C., DesRoches, C. M., Osborn, R., Scoles, K. L., & Zapert, K. (2002). Inequities In Health Care: A Five-Country Survey. *Health Affairs*, *21*(3), 182–191. https://doi.org/10.1377/hlthaff.21.3.182

Carpenter, E. (2021). "The Health System Just Wasn't Built for Us": Queer Cisgender Women and Gender Expansive Individuals' Strategies for Navigating Reproductive Health Care. *Women's Health Issues*, *31*(5), 478–484. https://doi.org/10.1016/j.whi.2021.06.004

CDC. (2021). Location, location, location – COVID Data Tracker Weekly Review. In *Centers for Disease Control and Prevention*. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/past-reports/07162021.html

Chrisler, J. C., Barney, A., & Palatino, B. (2016). Ageism can be Hazardous to Women's Health: Ageism, Sexism, and Stereotypes of Older Women in the Healthcare System. *Journal of Social Issues*, *72*(1), 86–104. https://doi.org/10.1111/josi.12157

Donovan, D. (2022). U.S. Officially Surpasses 1 Million COVID-19 Deaths. In *Johns Hopkins Coronavirus Resource Center*. https://coronavirus.jhu.edu/from-our-experts/u-s-officially-surpasses-1-million-covid-19-deaths

Feldman, J. L., Luhur, W. E., Herman, J. L., Poteat, T., & Meyer, I. H. (2021). Health and health care access in the US transgender population health (TransPop) survey. *Andrology*, *9*(6), 1707–1718. https://doi.org/10.1111/andr.13052

Glenza, J. (2020). Covid cases increase across US as upper midwest sees rapid rise. *The Guardian*. https://www.theguardian.com/world/2020/oct/22/covid-cases-coronavirus-us-midwest

Lambert, J. (2022). Most Americans Don't Know About the Omicron Covid Boosters. In *Grid News*. https://www.grid.news/story/science/2022/10/04/most-americans-dont-know-about-the-omicron-covid-boosters-that-spells-trouble-for-the-coming-winter/

Mayo Clinic. (2022). History of COVID-19: Outbreaks and Vaccine Timeline. In *MayoClinic.org*. https://www.mayoclinic.org/coronavirus-covid-19/history-disease-outbreaks-vaccine-timeline/covid-19

Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of The Total Environment*, *728*, 138884. https://doi.org/10.1016/j.scitotenv.2020.138884

Murthy, B. P. (2021). Disparities in COVID-19 Vaccination Coverage Between Urban and Rural Counties — United States, December 14, 2020–April 10, 2021. *MMWR. Morbidity and Mortality Weekly Report*, *70*. https://doi.org/10.15585/mmwr.mm7020e3

Office of the Commissioner. (2022). Coronavirus (COVID-19) Update: FDA Authorizes Moderna, Pfizer-BioNTech Bivalent COVID-19 Vaccines for Use as a Booster Dose. In *FDA*. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-moderna-pfizer-biontech-bivalent-covid-19-vaccines-use

Okonkwo, N. E., Aguwa, U. T., Jang, M., Barré, I. A., Page, K. R., Sullivan, P. S., Beyrer, C., & Baral, S. (2021). COVID-19 and the US response: Accelerating health inequities. *BMJ Evidence-Based Medicine*, *26*(4), 176–179. https://doi.org/10.1136/bmjebm-2020-111426

Ortega, A. N., & Roby, D. H. (2021). Ending Structural Racism in the US Health Care System to Eliminate Health Care Inequities. *JAMA*, *326*(7), 613–615. https://doi.org/10.1001/jama.2021.11160

Park, Y. M., Kearney, G. D., Wall, B., Jones, K., Howard, R. J., & Hylock, R. H. (2021). COVID-19 Deaths in the United States: Shifts in Hot Spots over the Three Phases of the Pandemic and the Spatiotemporally Varying Impact of Pandemic Vulnerability. *International Journal of Environmental Research and Public Health*, *18*(17), 8987. https://doi.org/10.3390/ijerph18178987

Rapp, K. S., Volpe, V. V., Hale, T. L., & Quartararo, D. F. (2022). State–Level Sexism and Gender Disparities in Health Care Access and Quality in the United States. *Journal of Health and Social Behavior*, *63*(1), 2–18. https://doi.org/10.1177/00221465211058153

Scott, D. (2020). These 4 Midwestern states are seeing worrying Covid-19 spikes. In *Vox*. https://www.vox.com/2020/9/2/21418812/covid-19-coronavirus-us-cases-midwest-surge

Shumaker, L., & Wu, T. (2020). COVID-19 cases surge in U.S. Midwest and Northeast. *Reuters*. https://www.reuters.com/article/us-health-coronavirus-usa-trends-idUSKBN26Q30B

Sigler, T., Mahmuda, S., Kimpton, A., Loginova, J., Wohland, P., Charles-Edwards, E., & Corcoran, J. (2021). The socio-spatial determinants of COVID-19 diffusion: The impact of globalisation, settlement characteristics and population. *Globalization and Health*, *17*(1), 56. https://doi.org/10.1186/s12992-021-00707-2

Sugg, M. M., Spaulding, T. J., Lane, S. J., Runkle, J. D., Harden, S. R., Hege, A., & Iyer, L. S. (2021). Mapping community-level determinants of COVID-19 transmission in nursing homes: A multi-scale approach. *Science of The Total Environment*, *752*, 141946. https://doi.org/10.1016/j.scitotenv.2020.141946

The New York Times. (2021). *Coronavirus (Covid-19) Data in the United States*. The New York Times. https://github.com/nytimes/covid-19-data

The White House. (2021). Executive Order On Advancing Racial Equity and Support for Underserved Communities Through the Federal Government. In *The White House*. https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/

The White House. (2022). Advancing Equity and Racial Justice Through the Federal Government. In *The White House*. https://www.whitehouse.gov/equity/

Thompson, C. N. (2020). COVID-19 Outbreak — New York City, February 29–June 1, 2020. *MMWR. Morbidity and Mortality Weekly Report*, *69*. https://doi.org/10.15585/mmwr.mm6946a2

United States COVID - Coronavirus Statistics - Worldometer. (2022). In *Worldometers*. https://www.worldometers.info/coronavirus/country/us/

U.S. Department of Commerce Economics and Statistics Administration, & U.S. Census Bureau. (2000). *Census Regions and Divisions of the United States*. https://www2.census.gov/geo/pdfs/maps-data/maps/

reference/us_regdiv.pdf
US Census Bureau. (2021). 2020 Census: Redistricting File (Public Law 94-171) Dataset. In *Census.gov*. https://www.census.gov/data/datasets/2020/dec/2020-census-redistricting-summary-file-dataset.html
US Census Bureau. (2020). American Community Survey 2015-2019 5-Year Data Release. In *Census.gov*. https://www.census.gov/newsroom/press-kits/2020/acs-5-year.html
Walker, K., & Herman, M. (2022). *Tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames.*
Whitehead, J., Shaver, J., & Stephenson, R. (2016). Outness, Stigma, and Primary Health Care Utilization among Rural LGBT Populations. *PLOS ONE*, *11*(1), e0146139. https://doi.org/10.1371/journal.pone.0146139
Whittle, R. S., & Diaz-Artiles, A. (2020). An ecological study of socioeconomic predictors in detection of COVID-19 cases across neighborhoods in New York City. *BMC Medicine*, *18*(1), 271. https://doi.org/10.1186/s12916-020-01731-6
Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Chapman & Hall/CRC Texts in Statistical Science. https://www.routledge.com/Generalized-Additive-Models-An-Introduction-with-R-Second-Edition/Wood/p/book/9781498728331