Final Report: A County-level Dataset for Informing the United States' Response to COVID-19

Benjamin D. Killeen[1]
killeen@jhu.edu

Mathias Unberath[2]
unberath@jhu.edu

Jie Ying Wu[*]
jieying@jhu.edu

Kinjal Shah[*]
kshah31@jhu.edu

Anna Zapaishchykova[*]
azapais1@jhu.edu

Philipp Nikutta[*]
pnikutt1@jhu.edu

Aniruddha Tamhane [*]
atamhan3@jhu.edu

Shreya Chakraborty [*]
schakr20@jhu.edu

Jinchi Wei[*]
jwei9@jhu.edu

Tiger Gao[*]
tgao11@jhu.edu

Mareike Thies[*]
mthies1@jhu.edu

[1]*Project Member* [2]*Project Mentor* [*]*Dataset Collaborators*
Computer Integrated Surgery, Russell H. Taylor
Johns Hopkins University,
Baltimore, MD, United States

***Abstract*—On March 11, 2020, the World Health Organization declared the outbreak of novel coronavirus 2019 (COVID-19), the disease caused by SARS-CoV2, a global pandemic. At present, the most effective tools for combating the virus are aggressive diagnostic testing and non-pharmaceutical interventions (NPIs), such as social distancing. In the United States, the availability of tests continues to fall short of what's required to identify hot-spots and predict the spread of the disease, a fact that is exacerbated by the prevalence of asymptomatic cases which are nevertheless vectors for transmission. As such, despite many states relaxing or removing NPIs, epidemiologists and public health experts warn that, absent a vaccine, the potential for resurgence is ever-present. Data driven approaches may inform local regions' implementation or safe rollback of NPIs which, although an effective tool for saving lives in a pandemic, have adverse effects on individuals' economic and mental wellbeing. This report relates to two efforts toward this end: a county-level dataset with machine readable features and a clustering method to augment epidemiological modeling, with interactive visualizations. Here, we outline the project member's role in the former, which also relied on exhaustive efforts of the listed collaborators, and present the latter, member-only task.**

This report is in part based on published work by the listed authors, detailing the county-level dataset, available at arxiv.org/pdf/2004.00756.pdf. The dataset has been made public at github.com/JieYingWu/COVID-19_US_County-level_Summaries, and the clustering visualizations at github.com/benjamindkilleen/covid19-county-descriptors-dashboard.

I. Technical Summary

A. Background

The ongoing outbreak of COVID-19 has wreaked havoc on nearly every facet of life in United States, with ongoing outbreaks in every state across the country as shown in Fig. 1. In the absence of a vaccine or sophisticated contact tracing, our best weapon against the virus continues to be non-pharmaceutical interventions, *e.g.* bans on public gatherings or stay-at-home orders. Schools and universities across the country have closed their main campuses, with some programs shuttered altogether as remote learning poses too great a challenge for students and teachers alike [1]. Non-essential commerce and entertainment venues have halted in-person service or closed indefinitely, and the unemployment rate in April skyrocketed to 14.7%, compared to 4.4% in March and 3.5% in February [2]. Yet these measures are necessary to protect vulnerable individuals in the community, such as the elderly or immune-compromised, to whom the virus appears particularly dangerous—although it must be noted that COVID-19 poses significant risk to otherwise healthy individuals as well. Safely and responsibly managing these interventions to limit their harmful effects while still protecting the public is of the utmost importance.

B. Problem

Epidemiological modeling, or predicting the spread of disease, is a powerful tool for anticipating outbreaks. However, many of these modeling approaches, such as the popular "susceptible, infected, and recovered" (SIR) model, rely on accurate measurements for the number of ongoing cases [3]. In the United States (and other countries), the lack of available testing has cast doubt on the accuracy of official models, and even the official death count may be severely underestimated, based on comparison with previous years [4], [5]. Data-driven approaches may augment epidemiological models, using information such as local demographics, health care capacity, and economic productivity to supplement direct observations of the disease. Toward this end, we have collected a county-level dataset detailed in [6] and leveraged that data to better inform NPI implementation decisions at the county level.

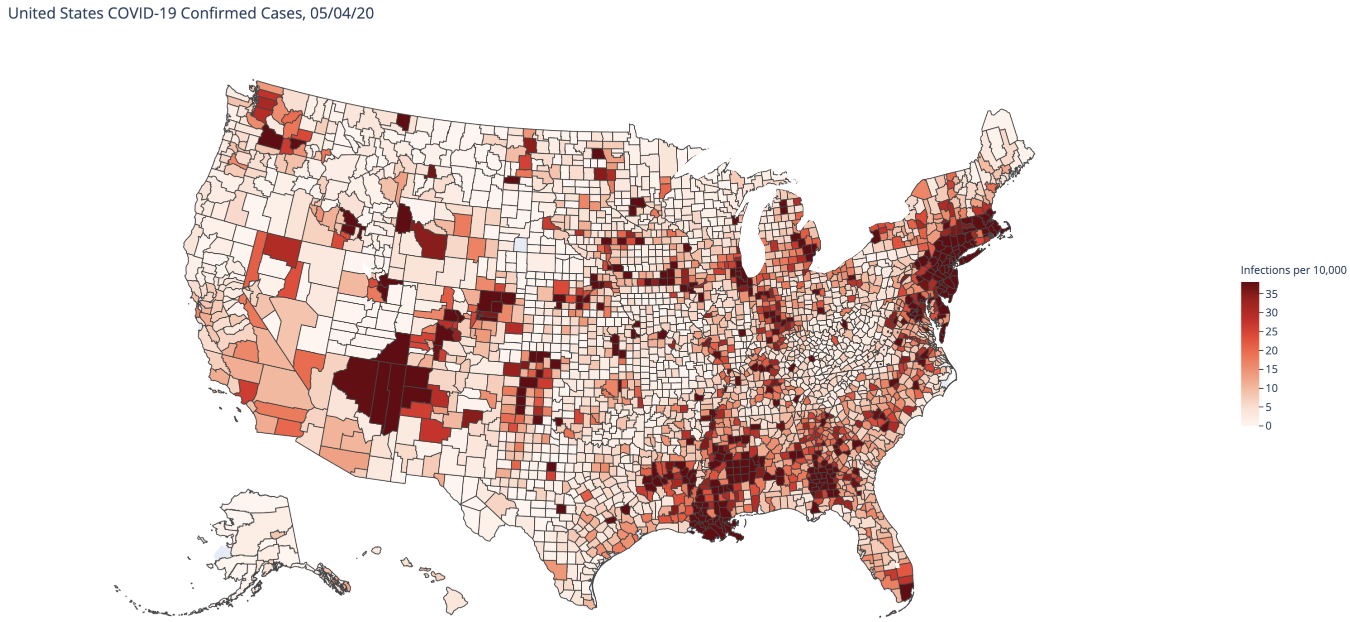

Fig. 1. Confirmed cases of COVID-19 per 10,000 individuals in the United States. For interpretability, the colorscale attains its maximum at the 90th percentile.

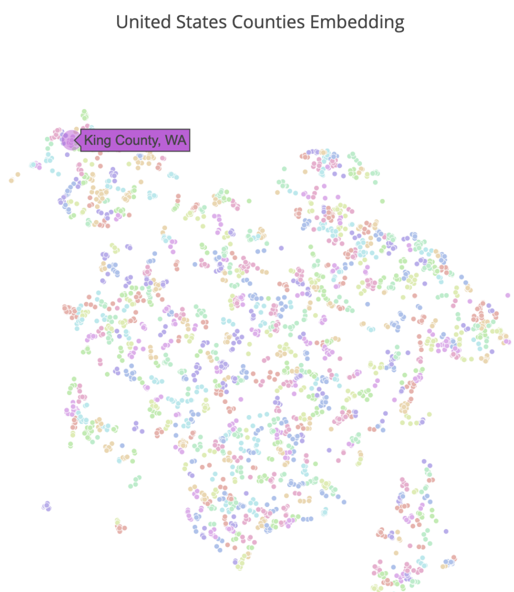

Fig. 2. An unsupervised embedding of United States counties based on statistics including unemployment, median household income, number of intensive care unit (ICU) beds, and age and gender demographics. Each dot corresponds to a county for which the chosen features are readily available, with adjacent representations colored by cluster. Colors are reused. For instance, King County, WA (shown) is in the same cluster as Los Angeles County, CA; San Diego County, CA; and Prince George's County, MD, which borders Washington, D.C..

,

C. Approach

1) Dataset: We describe our dataset in brief, referring to [6] for a complete description. Table I provides a description of the type and availability of the dataset, taken directly from [6]. The chief aim of the dataset is to provide machine-readable features aggregating disparate data sources, including federal agencies, news publications, and private companies. This is to facilitate further research based on county descriptors and response to the virus, such as NPIs. The data is publicly available on GitHub with associated documentations and column descriptions also in a machine-readable format. Since this dataset was an exhaustive effort among multiple collaborators (listed with *)

2) Embedding and Clustering: We leverage our county-level dataset to inform the county-level response to COVID-19 by embedding and clustering similar counties. By identifying counties which, because they share relevant demographic or economic properties, may also share epidemiological properties, one could infer how the virus' spread may be affected by implementing or removing NPIs. This can be done qualitatively using visualizations, as in Fig. 3 and as provided by our COVID-19 county-level embedding dashboard, or it can be done quantitatively, as discussed in II-C, by training epidemiological models based on these clusters.

For the clustering shown in Fig. 2, we use a UMAP embedding [63] to reduce raw-data representations of U.S. counties to two dimensions, allowing for easy visualization. The UMAP (Uniform Manifold Approximation and Projection) embedding encodes the high-dimensional manifold of data points in a low-dimensional vector space by preserving the cross-entropy between the two representations [63]. For a more complete review of the UMAP algorithm, we refer to [63].

For Fig. 2, we used the following features from our dataset:

```
Some college or associate's degree 2014-18
POVALL_2018
Unemployed_2018
Median_Household_Income_2018
Housing units
```

Data Type	Source	Availability
COVID-19 Infections COVID-19 Related Deaths Time-series	[7]	—
2020 Date of COVID-19 Interventions, *e.g.* stay-at-home order	[8]–[46]	—
March, 2020 Out-of-home Activity Time-series	SafeGraph [47]	—
2018 Population Estimates	Census [48]	97-100%
2014-2018 Educational Attainment	Census [49]	100%
2018 Estimated Poverty Level	USDA [50]	97%
2018 Employment and Income	USDA [51]	99%
2019 Precipitation and Temperature	NOAA [52]	86% (37.8% imputed)
2010 Housing and Density	Census [53]	99%
2018 Age Group Demographics	Census [54]	97%
2018 Household Demographics	Census [54]	25%
2018 Ethnic Group Demographics	Census [55]	97%
2019 Healthcare Capacity: Physicians, NPs, PAs	AAMC, KFF [56]–[58]	86-97%
2019 Healthcare Capacity: ICU Beds	KFF [59], [60]	92-97%
2019 Public Transit Scores	CNT [61]	95%
2016 Crime Rates	DOJ [62]	97%

TABLE I
DATA SOURCE DESCRIPTIONS AND PERCENTAGE OF COUNTIES INCLUDED FOR STATIC DATA

```
Male_age0to17
Female_age0to17
Male_age18to64
Female_age18to64
Male_age65plus
Female_age65plus
Density per square mile of land area
  - Housing units
ICU Beds
crime_rate_per_100000
transit_scores - population weighted
  averages aggregated from town/city
  level to county
```

where certain features have been normalized by population as appropriate. Additionally, features have been scaled to zero mean and unit variance to prevent over-fitting on columns with large numerical values.

Finally, we obtain 650 clusters using hierarchical agglomerative clustering [64]. The choice of 650 clusters is somewhat arbitrary and was made to keep cluster sizes small enough to be meaningful while still large enough for effective epidemiological models. Fig. 2 shows these clusters by color (with repetition).

D. Results

We present here a visualization dashboard for comparing the rate of COVID-19 transmission among counties in the same cluster. The dashboard, which can be run from source at present and will be hosted publicly in the future, allows a user to visualize which counties are contained in each cluster on a county map. Fig. 3 shows the dashboard's primary analysis tool, which displays the cumulative number of cases in each county for the selected cluster, with dates for particular interventions shown. The user can select different interventions to show as well as whether to view confirmed cases or deaths, and switch between accumulated total or per-capita rate. Additionally, the user can align each plot by date or by the number of days since 50 confirmed cases, which allows for more direct comparison between counties. In this manner, the user can interpret the effect of interventions on counties which may have behaved similarly otherwise. Finally, a visualization of the smoothed rate of transmission, as a derivative of confirmed cases, provides more direct visualization of the effect of NPIs, although it should be noted that any innacuracies in testing are amplified in the first derivative, resulting in a noisy plot for most counties.

E. Significance

Since its publication on GitHub and Kaggle, our county-level dataset has experienced significant engagement from the community. On Kaggle, the dataset has been downloaded over 600 times at the time of this writing, winning the COVID-19 Dataset Award ($1000). On GitHub, it has been downloaded by over 11,000 users. Our paper describing the dataset, [6], has been cited 3 times as of this writing, indicating its usefulness to the community. We hope that it will continue to inform data-driven approaches for combating the COVID-19 pandemic.

Additionally, we hope that our interactive dashboard for visualizing COVID-19 spread in county clusters will serve to underline the importance of NPIs even as many states move to reopen public spaces and businesses. Although these measures have serious adverse effects, it is essential to move forward in a responsible, fact-based manner with lives as the foremost concern.

II. MANAGEMENT SUMMARY

A. Responsibilities

It is important, since this project relied on many more individuals than the one receiving credit in Computer Integrated Surgery, to differentiate those efforts which are particular to him. First, in relation to the county-level dataset, these are:

- Determined a sensible organization strategy for coalescing data from disparate sources in order to facilitate machine-readability.
- Implemented this strategy in Python code and documenting its usage.

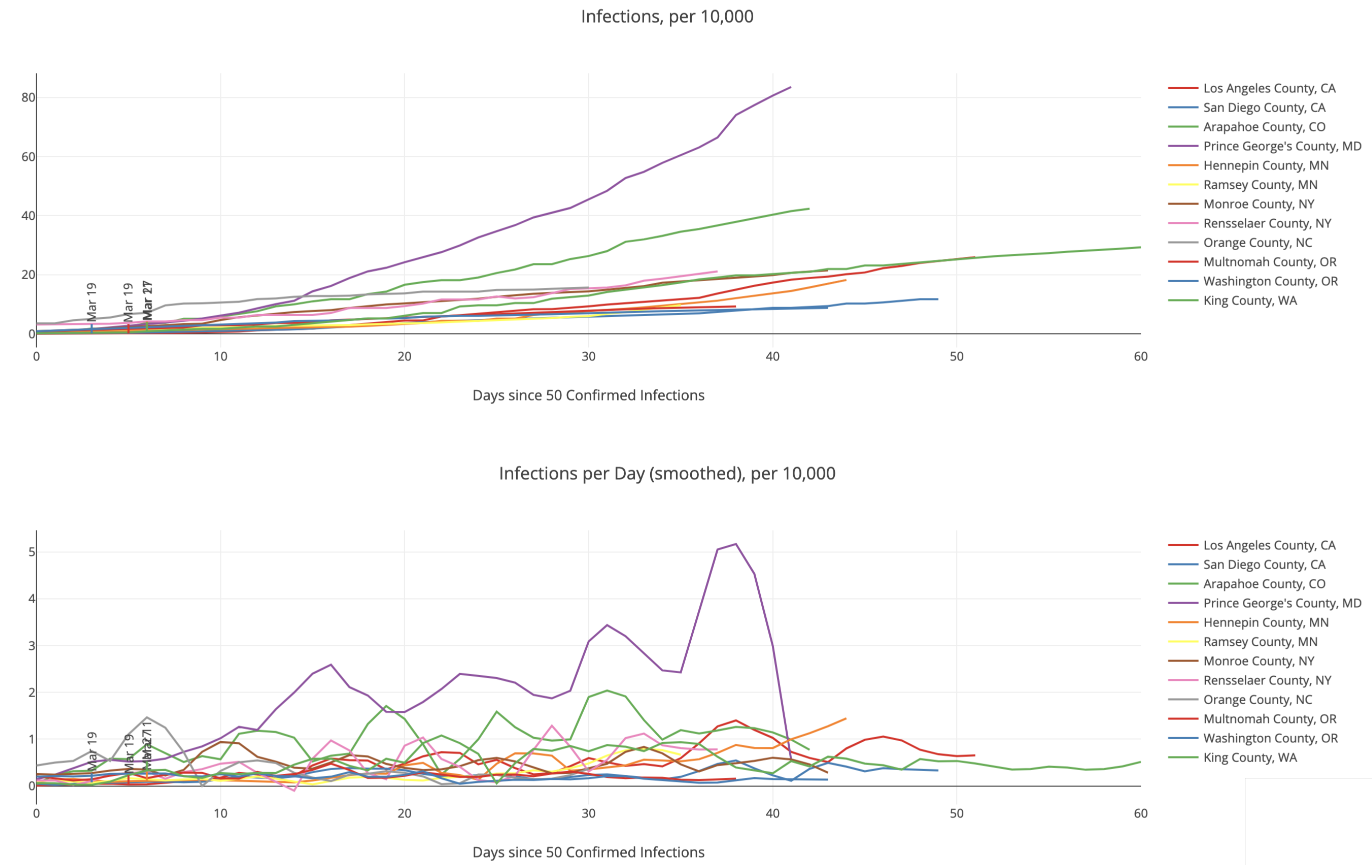

Fig. 3. The number (top) and growth rate (bottom) of confirmed cases for counties in the same cluster as King County, WA, per 10,000. For comparability, each timeseries is aligned by the date of the first 50 cases. Here, the implementation dates for interventions banning large gatherings (more than 500 people) are present.

- Implemented example notebooks to facilitate easy engagement with the dataset.
- Produced visualizations for the data.
- First-authored an Arxiv preprint detailing the dataset.

Second, the clustering effort, associated visualizations, and analysis tools are the primary effort of the project member, but frequent feedback from listed collaborators and the project mentor has been essential to their development. Additionally, these efforts are informed by ongoing epidemiological modeling, for which they will be used.

B. Deliverables

Table II lists our minimum, expected, and maximum deliverables for this project, as laid out in the beginning. With regard to the county-level dataset, we met all our deliverables. The deliverables not meant pertain to advanced epidemiological modeling, to which this project's contribution is the county-level clustering and visualization. That work is still in progress, as a large collaboration, and will be submitted in the coming weeks for peer review.

C. Future Work

Future work focuses on leveraging county-level data and the associated clusters for advanced epidemiological modeling. This is in the hopes of informing regions which have either not experienced sufficient spread to model the disease or are interested in learning how rolling back NPIs might affect the transmission rate.

D. Lessons Learned (Personal Reflection)

Throughout this project, I have had to set aside what prior experience may have prepared me for, *i.e.* computer vision, for the sake of fulfilling a pressing need in the midst of an emergency. I have had to coordinate with many collaborators remotely, maintaining a sense of urgency while not succumbing to the overwhelming nature of this pandemic. I have had to adapt, as this project took over my initial project for the semester: "Improved Generalization for Pelvis X-ray Landmark Detection," and I have corresponded with other researchers, developing my networking skills in that area. Overall, it has been an invaluable experience, with the foremost lesson being the value of contributing, even in a small way, to so great an endeavor as combatting the pandemic.

III. Technical Appendices

See the repositories listed in the abstract for code, data, and publication.

	Type	Deliverable	Status
Minimum	Dataset	Structured county-level dataset including COVID-19 cases, out-of-home activity, and healthcare capacity.	Done
	Implementation	Inline-documented formatting tools using Python, available on GitHub.	Done
	Analysis	Exponential model illustrating rapid spread.	Done
	Publication	Medium article describing the dataset in a general overview.	Done
Expected	Dataset	Constantly-maintained and up-to-date county-level data available on GitHub and Kaggle.	Done
	Implementation	Well-documented example scripts for using the dataset, including visualizations of county-level time series.	Done
	Analysis	Detailed epidemiological Models	Ongoing
	Publication	Arxiv dataset paper providing a detailed description of the dataset.	Done
Maximum	Dataset	Constantly-maintained and up-to-date county-level data available on GitHub and Kaggle.	Done
	Implementation	Advanced epidemiological modeling, based on county clustering.	Ongoing
	Analysis	Advanced modeling and interactive visualizations for web page.	Done
	Publication	Publicly available web page highlighting results and further modeling analyses publications, next two weeks.	Ongoing

TABLE II
PROJECT DELIVERABLES

REFERENCES

[1] O. Krauth, "Kentucky schools closed for remainder of school year, Beshear says," *Louisville Courier Journal*, Apr. 2020.

[2] "Bureau of Labor Statistics Data," https://data.bls.gov/timeseries/LNS14000000.

[3] D. Smith and L. Moore, "The sir model for spread of disease: The differential equation model," *Loci.(originally Convergence.) https://www. maa. org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model*, 2004.

[4] I. Scher, "The US is severely under-testing for coronavirus as death toll and new cases rise," https://www.businessinsider.com/the-us-is-not-testing-enough-people-for-covid-19-2020-3, library Catalog: www.businessinsider.com.

[5] J. Katz, D. Lu, and M. Sanger-Katz, "U.S. Coronavirus Death Toll Is Far Higher Than Reported, C.D.C. Data Suggests," *The New York Times*, Apr. 2020.

[6] B. Killeen, J. Y. Wu, K. Shah, A. Zapaishchykova, P. Nikutta, A. Tamhane, S. Chakraborty, J. Wei, T. Gao, M. Thies, and M. Unberath, "A County-level Dataset for Informing the United States' Response to COVID-19," *arXiv:2004.00756 [physics, q-bio]*, Apr. 2020.

[7] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *The Lancet Infectious Diseases*, vol. 0, no. 0, Feb. 2020.

[8] Carter, "'It's over.' NC bars, restaurants close to the public, leaving employees with uncertainty," https://www.newsobserver.com/news/coronavirus/article241284781.html, library Catalog: www.newsobserver.com.

[9] S. Carroll, "62 coronavirus cases in Arkansas; governor extends school closure, bans restaurant dine-in," https://katv.com/news/local/coronavirus-cases-rise-to-46-in-arkansas, Mar. 2020, library Catalog: katv.com.

[10] C. Hansen, "Alabama Governor Closes Nonessential Businesses as Coronavirus Spreads," https://www.usnews.com/news/national-news/articles/2020-03-27/alabama-gov-kay-ivey-closes-nonessential-businesses-as-coronavirus-spreads, library Catalog: www.usnews.com.

[11] R. Rettner, "Arkansas: Latest updates on coronavirus," https://www.livescience.com/coronavirus-arkansas.html, library Catalog: www.livescience.com.

[12] S. Mook, "Burgum closes bars, restaurants amid coronavirus concerns; schools to stay closed indefinitely," /news/education/5007393-Burgum-closes-bars-restaurants-amid-coronavirus-concerns-schools-to-stay-closed-indefinitely, library Catalog: www.grandforksherald.com.

[13] Z. Anderson, "Coronavirus Florida: Governor closes all restaurants and gyms," https://www.heraldtribune.com/news/20200320/coronavirus-florida-governor-closes-all-restaurants-and-gyms, library Catalog: www.heraldtribune.com.

[14] K. Staff, "Coronavirus in Nebraska, Iowa: The latest headlines and resources to keep you informed," https://www.ketv.com/article/coronavirus-covid19-nebraska-omaha-latest/31213658, Mar. 2020, library Catalog: www.ketv.com.

[15] A. Soga, "Coronavirus in Vermont: Governor orders bars, restaurants closed," https://www.burlingtonfreepress.com/story/news/2020/03/16/coronavirus-vermont-burlington-mayor-orders-24-hour-restaurant-bar-closure/5062491002/, library Catalog: www.burlingtonfreepress.com.

[16] A. Kite, K. Hardy, J. Smith, and S. Vockrodt, "Coronavirus shuts Kansas City restaurants, leaving staff unemployed, yearning for work," https://www.kansascity.com/news/coronavirus/article241494536.html, library Catalog: www.kansascity.com.

[17] W. B. L. J. Feuer, Noah Higgins-Dunn, "Coronavirus: NY, NJ, CT coordinate restrictions on restaurants, limit events to fewer than 50 people," https://www.cnbc.com/2020/03/16/new-york-new-jersey-and-connecticut-agree-to-close-restaurants-limit-events-to-less-than-50-people.html, Mar. 2020, library Catalog: www.cnbc.com.

[18] G. Hiatt, "D.C. Adds 'Stay-at-Home' Order on Same Day as Maryland and Virginia," https://dc.eater.com/2020/3/15/21180673/dc-mayor-muriel-bowser-coronavirus-response-elminate-bar-seats-limit-table-size, Mar. 2020, library Catalog: dc.eater.com.

[19] P. Svitek, "Gov. Greg Abbott closes bars, restaurants and schools as he anticipates tens of thousands could test positive for coronavirus," https://www.texastribune.org/2020/03/19/texas-restaurants-bars-closed-greg-abbott/, Mar. 2020, library Catalog: www.texastribune.org.

[20] KCCI, "Gov. Reynolds issues state of public health disaster emergency, closing Iowa businesses," https://www.kcci.com/article/gov-reynolds-issues-state-of-public-health-disaster/31700874, Mar. 2020, library Catalog: www.kcci.com.

[21] A. J. Capuano, "In Missouri, no dining-in at restaurants, groups of 10 or more banned amid coronavirus," https://ktvo.com/news/local/in-missouri-no-dining-in-at-restaurants-groups-of-10-or-more-banned-amid-coronavirus, Mar. 2020, library Catalog: ktvo.com.

[22] B. Tobin, D. Ghabour, D. Costello, and M. Glowicki, "Kentucky Derby postponed, restaurants restricted as state tries to control spread of virus," https://www.courier-journal.com/story/news/2020/03/16/coronavirus-kentucky-beshear-orders-restaurants-bars-close/5057062002/, library Catalog: www.courier-journal.com.

[23] HNN, "LIST: Here's how the state and each island is responding to coronavirus," https://www.hawaiinewsnow.com/2020/03/18/list-bar-closures-cruise-ship-screening-here-are-all-iges-covid-directives/, library Catalog: www.hawaiinewsnow.com.

[24] "LIVE UPDATES: Here's the latest on the coronavirus in Forsyth County and Georgia," https://www.forsythnews.com/news/health-care/heres-latest-coronavirus-georgia/, library Catalog: www.forsythnews.com.

[25] J. Helminiak, "Local bars, resaurants, gyms, theaters react to coronavirus," https://www.peninsulaclarion.com/news/local-bars-resaurants-gyms-theaters-react-to-coronavirus/, Mar. 2020, library Catalog: www.peninsulaclarion.com.

[26] WGME, "Maine bars, restaurants ordered to close to dine-in customers, coronavirus cases increase," https://wgme.com/news/coronavirus/gov-mills-mandates-maine-bars-restaurants-close-to-dine-in-customers, Mar. 2020, library Catalog: wgme.com.

[27] "Map: Coronavirus and School Closures - Education Week," *Education Week*, Mar. 2020.

[28] A. Ganucheau, "Mayors scramble to know: Does Gov. Reeves' coronavirus declaration clash with local orders?" https://mississippitoday.org/2020/03/25/mayors-scramble-to-know-

does-gov-reeves-coronavirus-declaration-clash-with-local-orders/, Mar. 2020, library Catalog: mississippitoday.org.

[29] A. Press, "Montana Extends School, Restaurant Closures 2 More Weeks," https://www.usnews.com/news/best-states/montana/articles/2020-03-22/evidence-of-community-spread-in-montanas-gallatin-county, library Catalog: www.usnews.com.

[30] M. Etehad and L. K. Peterson, "Nevada orders all casinos, bars, restaurants closed as U.S. coronavirus cases surge," https://www.latimes.com/world-nation/story/2020-03-17/las-vegas-to-close-all-casinos-at-midnight, Mar. 2020, library Catalog: www.latimes.com.

[31] WHDH, "New Hampshire bans dine-in restaurant meals until April 7," library Catalog: whdh.com.

[32] K. Media, "New restrictions for New Mexico restaurants and bars to begin Monday," Mar. 2020, library Catalog: www.krqe.com.

[33] B. Webb, "Phoenix, Tucson order closures of bars, restaurants," https://www.fox10phoenix.com/news/phoenix-tucson-order-closures-of-bars-restaurants1, Mar. 2020, library Catalog: www.fox10phoenix.com.

[34] A. Kludt and B. Houck, "Restaurants and Bars Shuttered Across the U.S. in Light of Coronavirus Pandemic," https://www.eater.com/2020/3/15/21180761/coronavirus-restaurants-bars-closed-new-york-la-chicago, Mar. 2020, library Catalog: www.eater.com.

[35] R. Nunes, "RI Restaurants Closed Amid Community Spread Of Coronavirus," https://patch.com/rhode-island/newport/coronavirus-ri-dine-restaurants-closed-2-weeks, Mar. 2020, library Catalog: patch.com.

[36] S. Mervosh, D. Lu, and V. Swales, "See Which States and Cities Have Told Residents to Stay at Home," *The New York Times*, Mar. 2020.

[37] L. Ruskin, "State bans restaurant dining as Alaska's confirmed coronavirus cases grow to 6," Mar. 2020, library Catalog: www.alaskapublic.org.

[38] C. Gross, "State to Restrict Bars and Restaurants Statewide Starting at 8PM," https://www.ny1.com/nyc/all-boroughs/coronavirus/2020/03/16/bars-restaurants-gyms-movie-theaters-casinos-new-york-state, library Catalog: www.ny1.com.

[39] Axios, "States order bars and restaurants to close due to coronavirus," https://www.axios.com/ohio-governor-bars-restaurants-coronavirus-26e4b6e3-7f65-4f6a-abf9-f3940220cc6f.html, library Catalog: www.axios.com.

[40] B. Kelman and J. McGee, "Tennessee governor orders restaurants, bars closed except for takeout and delivery; gyms closed over coronavirus," https://www.tennessean.com/story/news/health/2020/03/22/tennessee-governor-restaurants-bars-closed-takeout-and-delivery/2892481001/, library Catalog: www.tennessean.com.

[41] A. Lee, "These states have implemented stay-at-home orders. Here's what that means for you," https://www.cnn.com/2020/03/23/us/coronavirus-which-states-stay-at-home-order-trnd/index.html, library Catalog: www.cnn.com.

[42] T. Semerad, "Utah Orders Restaurants,Bars to Close All Dining to Curb Coronavirus," https://www.sltrib.com/news/2020/03/18/utah-orders-restaurants/, library Catalog: www.sltrib.com.

[43] A. Spiegel, "Virginia Restaurants and Bars Close for Dine-In Service to Help Curb Coronavirus," Mar. 2020, library Catalog: www.washingtonian.com.

[44] E. C. Wida, "Which states have closed restaurants and bars due to coronavirus?" https://www.today.com/food/which-states-have-closed-restaurants-bars-due-coronavirus-t176039, library Catalog: www.today.com.

[45] Star-Tribune, "Wyoming cancellations and closures caused by coronavirus," https://trib.com/news/state-and-regional/health/wyoming-cancellations-and-closures-caused-by-coronavirus/article_228b1e3a-56b6-5e09-a351-3bcb8825e7f3.html, library Catalog: trib.com.

[46] M. Specia, "What You Need to Know About Trump's European Travel Ban," *The New York Times*, Mar. 2020.

[47] SafeGraph, "Footprint data," safeGraph, a data company that aggregates anonymized location data from numerous applications in order to provide insights about physical places. To enhance privacy, SafeGraph excludes census block group information if fewer than five devices visited an establishment in a month from a given census block group.

[48] United States Census Bureau, "2018 Population Estimates." [Online]. Available: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk

[49] ——, "Selected social characteristics in the united states: 2018 acs 1 year data estimate profiles." [Online]. Available: https://data.census.gov/cedsci/table?q=dp02&hidePreview=true&tid=ACSDP1Y2018.DP02&vintage=2018&g=0400000US36.050000&tp=true&y=2018

[50] United States Department of Agriculture Economic Research Service, "Poverty estimates for the U.S., States and counties, 2018." [Online]. Available: https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/

[51] ——, "Unemployment and median household income for the U.S., states and counties, 2007-18." [Online]. Available: https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/

[52] National Oceanic and Atmospheric Administration, "NOAA's Climate Divisional Database." [Online]. Available: ftp://ftp.ncdc.noaa.gov/pub/data/cirs/climdiv/

[53] United States Census Bureau, "Population and housing unit counts: 2010." [Online]. Available: https://www.census.gov/library/publications/2012/dec/cph-2.html

[54] ——, "Selected Social Characteristics in the United States: Households By Type." [Online]. Available: https://data.census.gov/cedsci/table?q=dp02&hidePreview=true&tid=ACSDP1Y2018.DP02&vintage=2018&g=0400000US36.050000&tp=true&y=2018

[55] ——, "Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2010 to July 1, 2018 (CC-EST2018-ALLDATA)." [Online]. Available: https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html#par_textimage_1383669527

[56] American Association of Medical Colleges, "State physician workforce data report," 2019.

[57] "Professionally active primary care physicians by field," *KFF's State Health Facts*, Mar. 2019.

[58] "Professionally active specialist physicians by field," *KFF's State Health Facts*, Mar. 2019.

[59] Henry J. Kaiser Family Foundation, "Sepcial data request," *KFF's State Health Facts*, Mar. 2019.

[60] "Millions of older americans live in counties with no ICU beds as pandemic intensifies," *Kaiser Health News*, Mar. 2020.

[61] Center for Neighborhood Technology, "Alltransit performance score." [Online]. Available: https://alltransit.cnt.org/data-download/

[62] Bureau of Justice Statistics, Department of Justice, "United states crime rates by county." [Online]. Available: https://www.icpsr.umich.edu/icpsrweb/

[63] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," *arXiv:1802.03426 [cs, stat]*, Dec. 2018.

[64] M. L. Zepeda-Mendoza and O. Resendis-Antonio, "Hierarchical Agglomerative Clustering," in *Encyclopedia of Systems Biology*, W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota, Eds. New York, NY: Springer, 2013, pp. 886–887.