Hospital Surge Prediction using Google Trends

Laura, Casey, Tzu-Hsuan, Aparajita, Solvejg

Team 4

Roadmap

1 **Motivation** - Why hospital surges? Why Google Trends?

2 **Data** - What datasets are we using?

3 **Model & Performance** - What is the best predictive model?
- Does performance differ by city?

4 **Caution & Conclusion** - What are the assumptions and conclusions of the analysis?

Why Hospital Surges?

SCIENCE

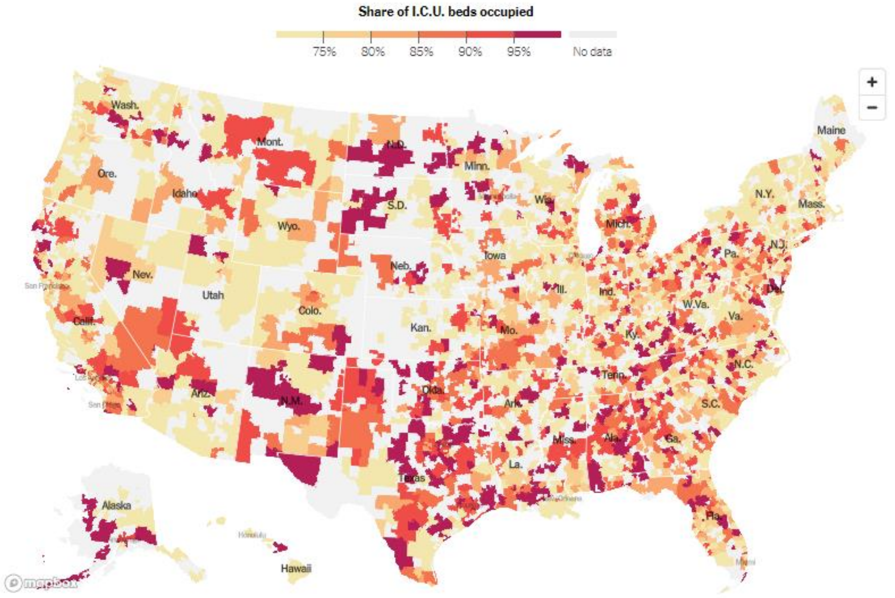
America's Hospitals Have Never Experienced Anything Like This

If they run out of space, where will all the sick people go?

SARAH ZHANG MARCH 25, 2020

'There's No Place for Them to Go': I.C.U. Beds Near Capacity Across U.S.

By Lauren Leatherby, John Keefe, Lucy Tompkins, Charlie Smart and Matthew Conlen Dec. 9, 2020



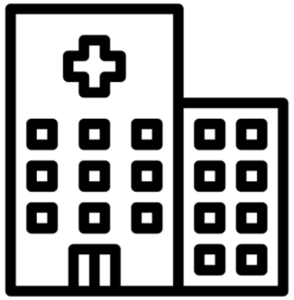
Source: New York Times analysis of U.S. Department of Health and Human Services data. · Note: Shows 7-day average patient count by hospital service area.

13 Deaths in a Day: An 'Apocalyptic' Coronavirus Surge at an N.Y.C. Hospital

Hospitals in the city are facing the kind of harrowing increases in cases that overwhelmed health care systems in China and Italy.

Sources: [1,2,3]

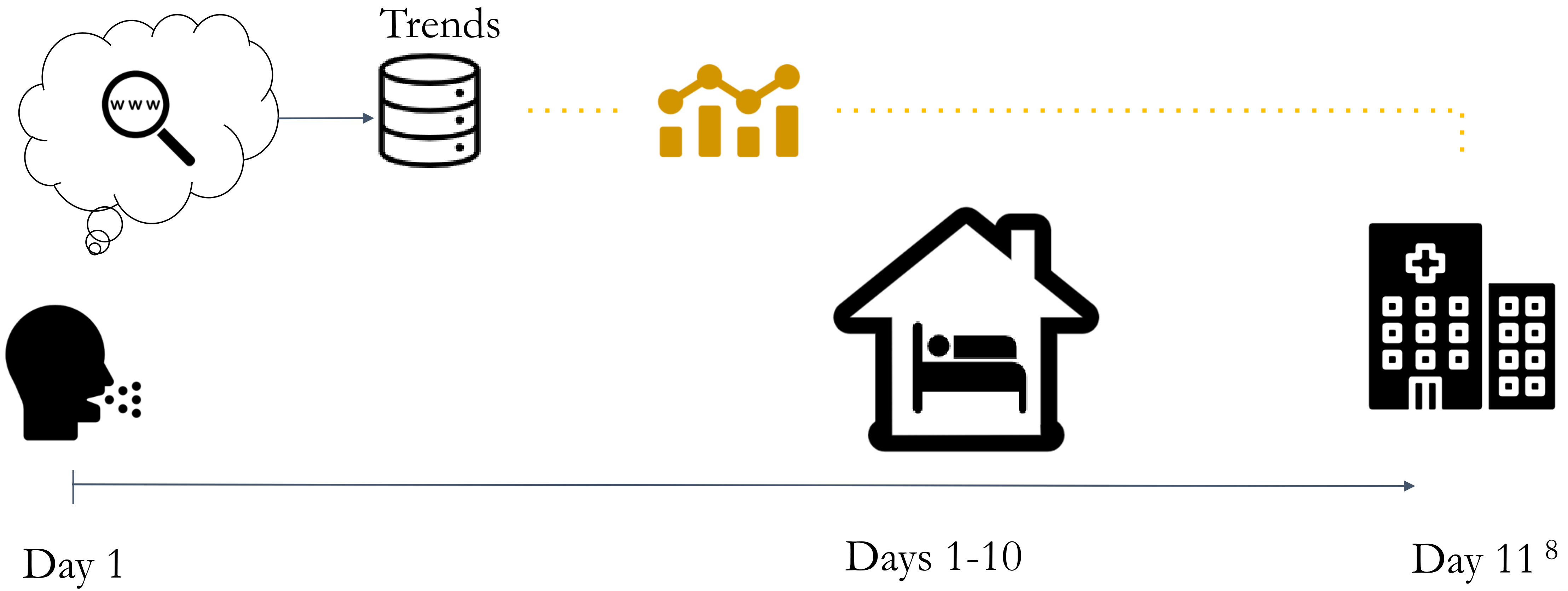
Despite available data, **hospital surge planning remains a challenge**



- Hospitals have been **preparing for the worst** for over a year
  - Increase bed capacity
  - Postpone elective surgeries [4]
  - Move beds to waiting rooms and parking garages [5]

- Reported data has **not** always been **dependable** [6,7]
  - Varies state by state
- Novel data source to help predict needs?

Google searches could offer **novel, timely data** for hospital surge planning



COVID-19 prediction using Google Trends has been done at the **state level**

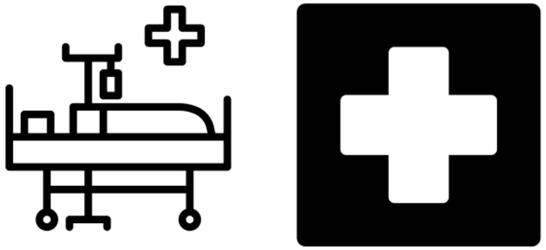
- Researchers in Asia, Europe, and the US have conducted similar work [9,10]

- Previous domestic work mainly focuses on national and state geographies [11]

- Target is COVID-19 case prediction

Inspired by this previous work, we conduct our analysis at the **metropolitan city level** and attempt to **predict hospital surges**

Previous prediction work focuses on **cases and deaths**



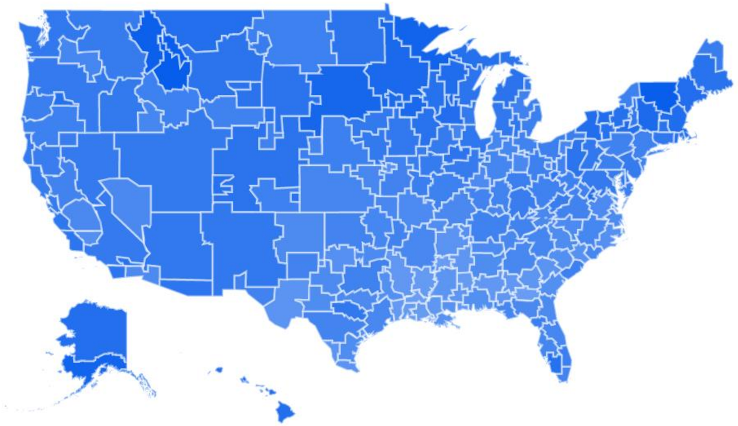
We focus on hospitalization instead [12]

Use COVID-19 Reported Patient Impact and Hospital Capacity by Facility datasource [13]

Calculate surge by 7 day average proportion of ICU capacity

$$= \frac{\text{Sum of avg beds taken}}{\text{Sum of avg beds available}}$$

We use **‘Pytrends’** to extract search data



- ➔ Top google search queries stored in Google Trends
- ➔ Available by city, state, region
- ➔ Normalized on a scale of 1:100

Pytrends is an open-source Application Programming Interface (API)

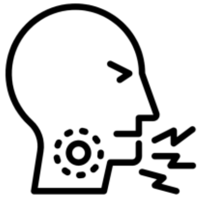
Use **COVID-19 symptoms** as keywords

Vomit
Diarrhea
Fatigue

Smell
Taste

Fever
Chills

Throat
Cough
Breath



10 common COVID-19 symptoms [14]

Create a ~**6 month** panel dataset for the **city-level** geography

Extract data on 210 cities in the US

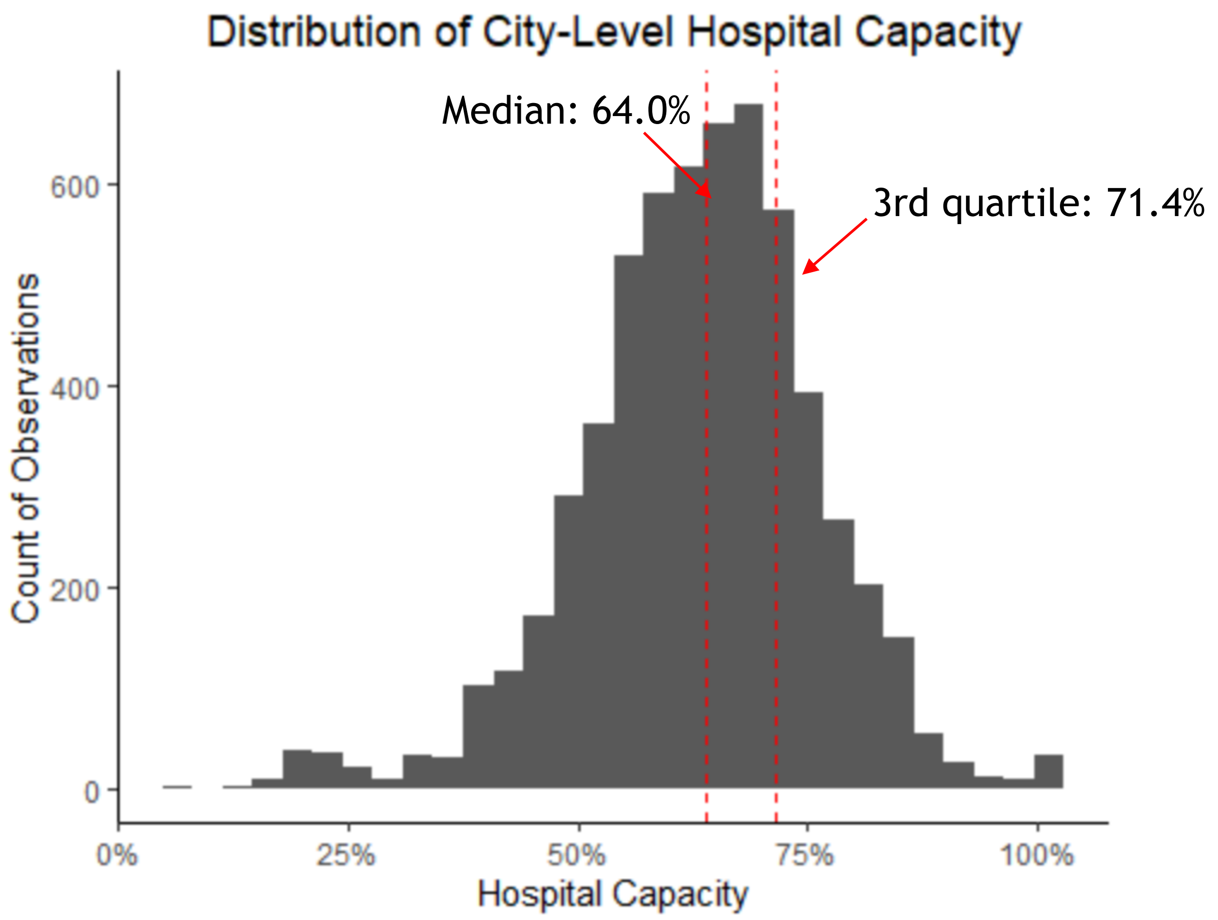
Data ranges 25 weeks, from 07/31/2020 to 1/21/2021

Used fuzzy-wuzzy search to merge metro areas

Metropolitan	**Cities**	**Zip Codes**
Minneapolis-St. Paul/MN	Minneapolis, MN, St. Paul, MN	98424, 98431…
Seattle-Tacoma/WA	Seattle, WA, Tacoma, WA	98424, 98431…

We defined surge using the distribution of city-level hospital capacity

- ➔ Used longitudinal data on **108 cities**
- ➔ Observation is a city for specific week
- ➔ "**Surge**" for a city as at or above 3rd quartile in given week
- ➔ **657** of 2431 **(27%)** observations were at surge



Most frequent search words found at higher rates in states with dangerous levels of capacity

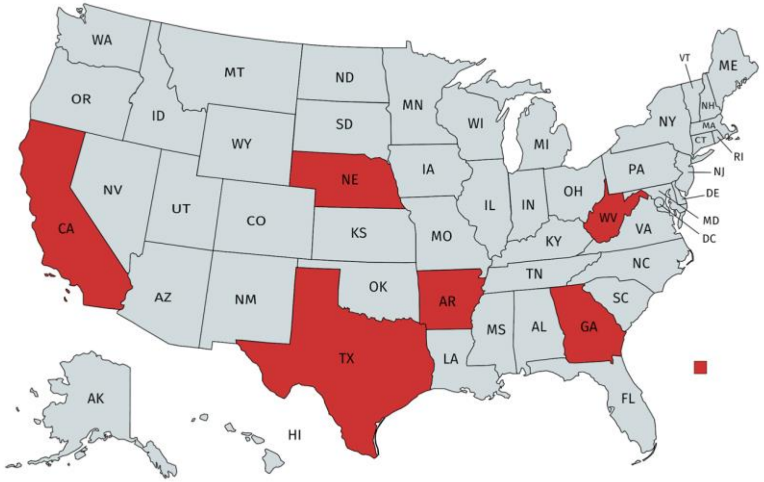
Throat

Fever

Taste

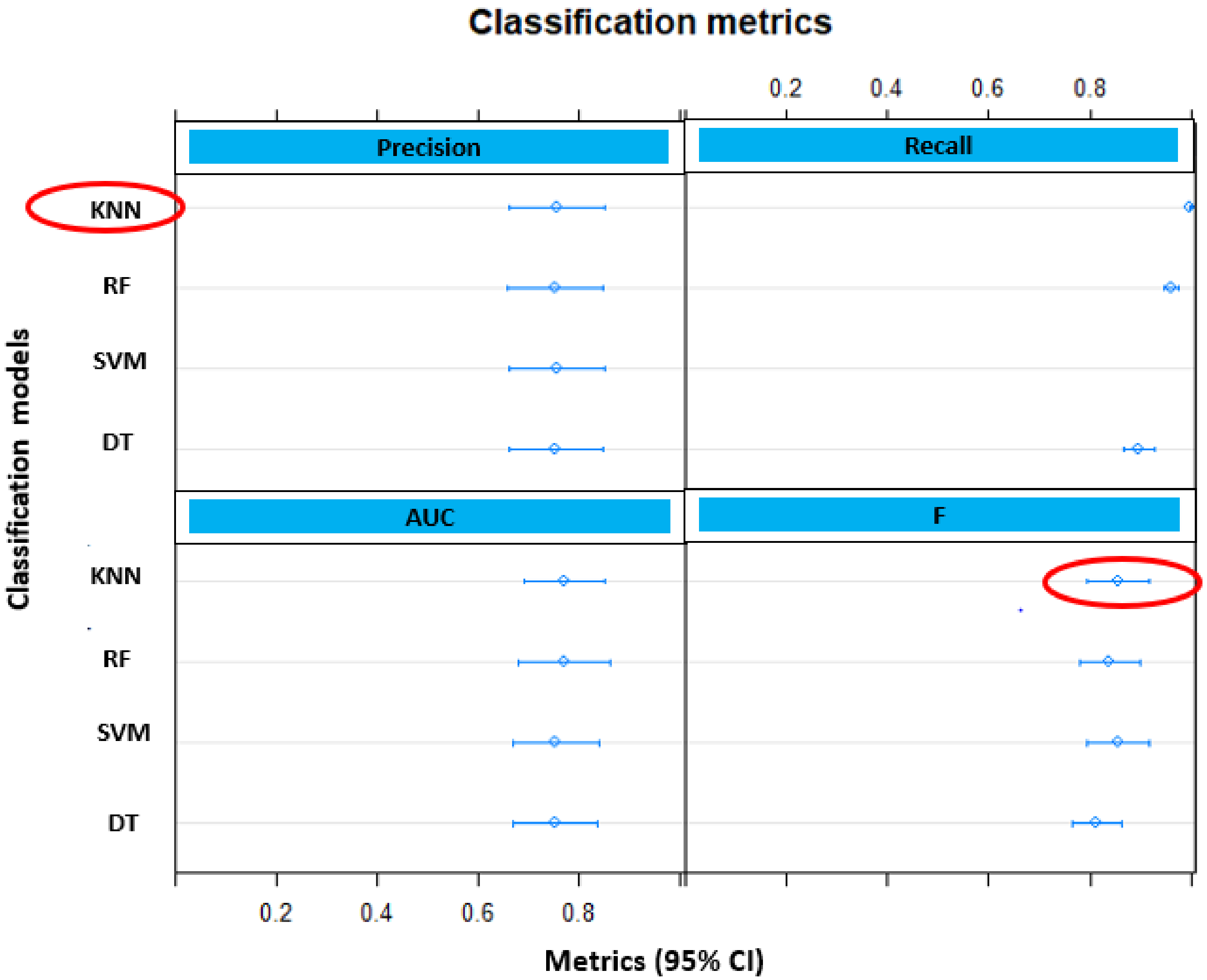
Smell

Most searched COVID-19 terms



Maximum search rates found in states that reached over **95% hospital capacity for at least one week** in the dataset

We built classifiers to predict surge with varying success



<u>Best classification models</u>

- ➔ K-Nearest Neighbors (KNN)
- ➔ Random Forest (RF)
- ➔ Support Vector Machine (SVM)
- ➔ Decision Tree

Explored the impact of city demographics on model performance

Assess variation in predictive model accuracy by county characteristics [15, 16, 17]

- Race
- Age
- Sex
- Median income

Demographic comparison of correctly-predicted metro areas

Found no significant difference between metros with/without 1+ correct surge prediction when comparing characteristics

210 to 108 → Limited by the size of dataset which diminished significantly in merging

Assumptions/Limitations

Assume patients go to hospital in metro of residence

Open-Source API PyTrends is not official

HHS hospital data includes most but not all U.S. facilities

Lose data when merging

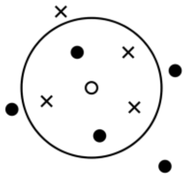
Limited information on rural areas

Conclusion

Hospital surge prediction can save lives and **communities**

Explore the power of an **unconventional datasource**

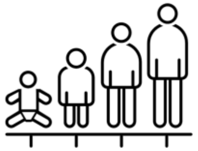
➔ Using city level geography, which has not been done before



K-Nearest Neighbors model performed best on our data

➔ 83 percent accuracy

Conclusion - continued



Hoped to understand differences in **demographic characteristics** for successfully predicted cities but limited by data size

Despite the limitations, had success predicting hospital surges with 83% accuracy using 10 google keywords

References

1. Zhang, Sarah. “America's Hospitals Have Never Experienced Anything Like This.” The Atlantic, Atlantic Media Company, 25 Mar. 2020, www.theatlantic.com/science/archive/2020/03/coronavirus-hospitals-need-more-beds/608677/.
2. Leatherby, Lauren, et al. “'There's No Place for Them to Go': I.C.U. Beds Near Capacity Across U.S.” The New York Times, The New York Times, 9 Dec. 2020, www.nytimes.com/interactive/2020/12/09/us/covid-hospitals-icu-capacity.html.
3. Abelson, Reed. “Covid Overload: U.S. Hospitals Are Running Out of Beds for Patients.” The New York Times, The New York Times, 27 Nov. 2020, www.nytimes.com/2020/11/27/health/covid-hospitals-overload.html.
4. “State Guidance on Elective Surgeries.” Ambulatory Surgery Center Association (ASCA), www.ascassociation.org/asca/resourcecenter/latestnewsresourcecenter/covid-19/covid-19-state.
5. Kremer, Rich. “Mayo Clinic Puts Hospital Beds In Ambulance Garage, Lobbies As COVID-19 Surges.” Wisconsin Public Radio, 30 Nov. 2020, www.wpr.org/mayo-clinic-puts-hospital-beds-ambulance-garage-lobbies-covid-19-surges.
6. White, Easton R, and Laurent Hébert-Dufresne. “State-level variation of initial COVID-19 dynamics in the United States.” PloS one vol. 15,10 e0240648. 13 Oct. 2020, doi:10.1371/journal.pone.0240648
7. “About CDC COVID-19 Data.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 25 Nov. 2020, www.cdc.gov/coronavirus/2019-ncov/cases-updates/about-us-cases-deaths.html#:~:text=CDC%20tracks%20COVID%2D19,territories%20report%20their%20cases.
8. “Remote Monitoring of COVID-19 Symptoms - Early Prediction of Aggressive COVID-19 Progression and Hospitalization.” Mayo Clinic, 4 May 2020, www.mayo.edu/research/remote-monitoring-covid19-symptoms/people-with-covid19.
9. Venkatesh, U, and Periyasamy Aravind Gandhi. “Prediction of COVID-19 Outbreaks Using Google Trends in India: A Retrospective Analysis.” Healthcare informatics research vol. 26,3 (2020): 175-184. doi:10.4258/hir.2020.26.3.175
10. Ciaffi, J., Meliconi, R., Landini, M.P. et al. Google trends and COVID-19 in Italy: could we brace for impact?. Intern Emerg Med 15, 1555–1559 (2020). https://doi.org/10.1007/s11739-020-02371-7
11. Mavragani, A., Gkillas, K. COVID-19 predictability in the United States using Google Trends time series. Sci Rep 10, 20693 (2020). https://doi.org/10.1038/s41598-020-77275-9

References - continued

12. UNDERPAYMENT BY MEDICARE AND MEDICAID FACT SHEET. AMERICAN HOSPITAL ASSOCIATION, Dec. 2017, www.aha.org/system/files/2018-01/medicaremedicaidunderpmt%202017.pdf.

13. U.S. Department of Health & Human Services. "COVID-19 Reported Patient Impact and Hospital Capacity by Facility." HealthData.gov, 15 Mar. 2021, healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/anag-cw7u.

14. "Symptoms of Coronavirus." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 22 Feb. 2021, www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html.

15. Schneider, Jacob. "Google Trends Metro Area GIS Shape File." Google Sites, sites.google.com/view/jacob-schneider/resources.

16. USDA Economic Research Service, https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/

17. U.S. Census Bureau. American Community Survey, 2019 American Community Survey 1-Year Estimates. Accessed using *censusapi* R package.

Thank you!