BACKGROUND

Excess deaths are the deaths occurring above and beyond the expected number of deaths for an area during the same time frame, in this case, in the absence of the COVID-19 pandemic (U.S. Centers for Disease Control and Prevention, 2021). In one estimate of excess deaths, Ackley et al. (2021) discovered that the proportion of excess deaths associated with COVID-19 was higher in counties in the Midwest and Northeast of the United States vs. the South and West. Specifically, the Southern and Western regions of the United States had less excess deaths directly attributed to COVID-19 compared to the excess deaths not directly assigned to COVID-19 (Stokes et al., 2021). What accounts for these regional differences? For one, communities were possibly impacted on different scales by state and/or county policies, hospital capacity, differential access to healthcare, and more (Stokes et al., 2021). Thus it's important to elucidate the role of these contextual factors in impacting excess deaths across the United States.

TASK

I aim to understand which variables describing community characteristics (infrastructure, health care system, the built environment-related) serve as explanatory variables for excess COVID-19 deaths through multilevel modeling. One model will analyze the variables at the county level and metropolitan statistical area (MSA) level. As an alternative, another model will analyze the variables at the county and state level.

DATA SOURCES

My three primary data sources will be the 2021 County Health Rankings, COVID-19 Excess Deaths dataset, and Google Street View images. All three datasets primarily contain numeric data. Some data cleaning that will be required includes dealing with missing values, changing data types, and joining the dataset(s) with a shapefile to provide county geometries.

The 2021 County Health Rankings "measures the health" of counties in the United States and ranks them in comparison to other counties in the same state (University of Wisconsin Population Health Institute, n.d.). The variables I will utilize from the dataset measure broadband access, the number of uninsured adults, the number of primary care physicians, and home ownership. With 3194 observations, the 2021 County Health Rankings draws upon the

2015-2019 American Community Survey, 2018 Small Area Health Insurance Estimates, and 2018 Area Health Resource Files. The Google Street View dataset captures built environment characteristics of a county such as the proportion of greenery and sidewalks. With 2980 observations, these data were constructed by Nguyen et al. (2019) by processing Google Street View images. The Excess Deaths dataset was constructed by Rahman and Panozzo (2020) from the University of Toledo. With 485 observations, the dataset was constructed by creating county-level estimates of excess deaths by drawing from CDC National Center for Health Statistics' Provisional Mortality Data and Provisional COVID-19 Death Counts by Week for 2012-2018. One limitation is that in counties with less than 10 cases, the numbers are suppressed and reported as zero. Thus, there are significantly less observations in the Excess Deaths dataset when compared with the other two datasets.

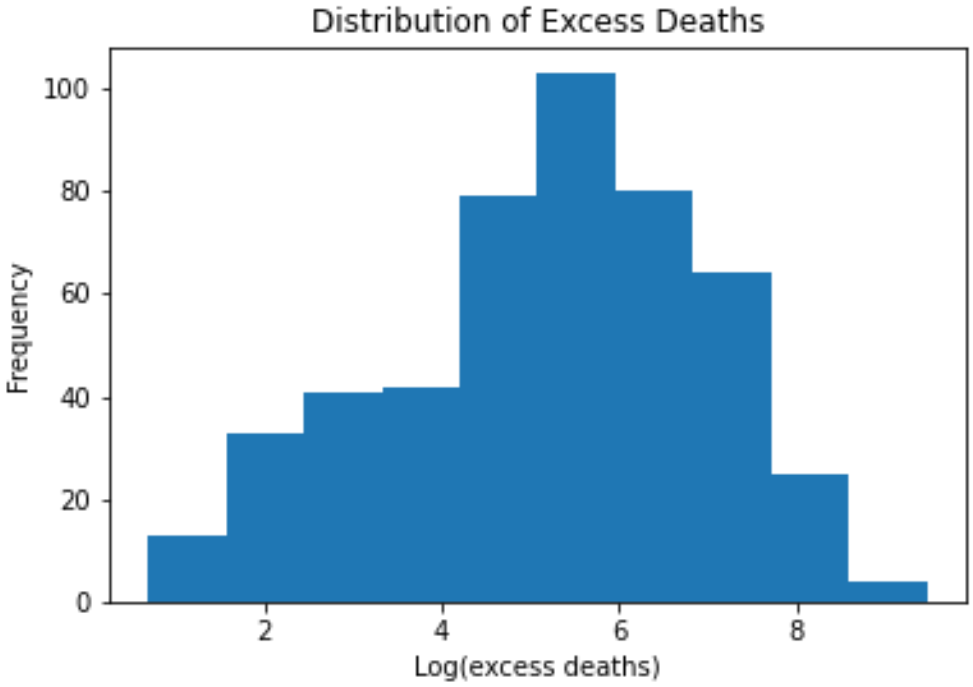
PYTHON TOOLS

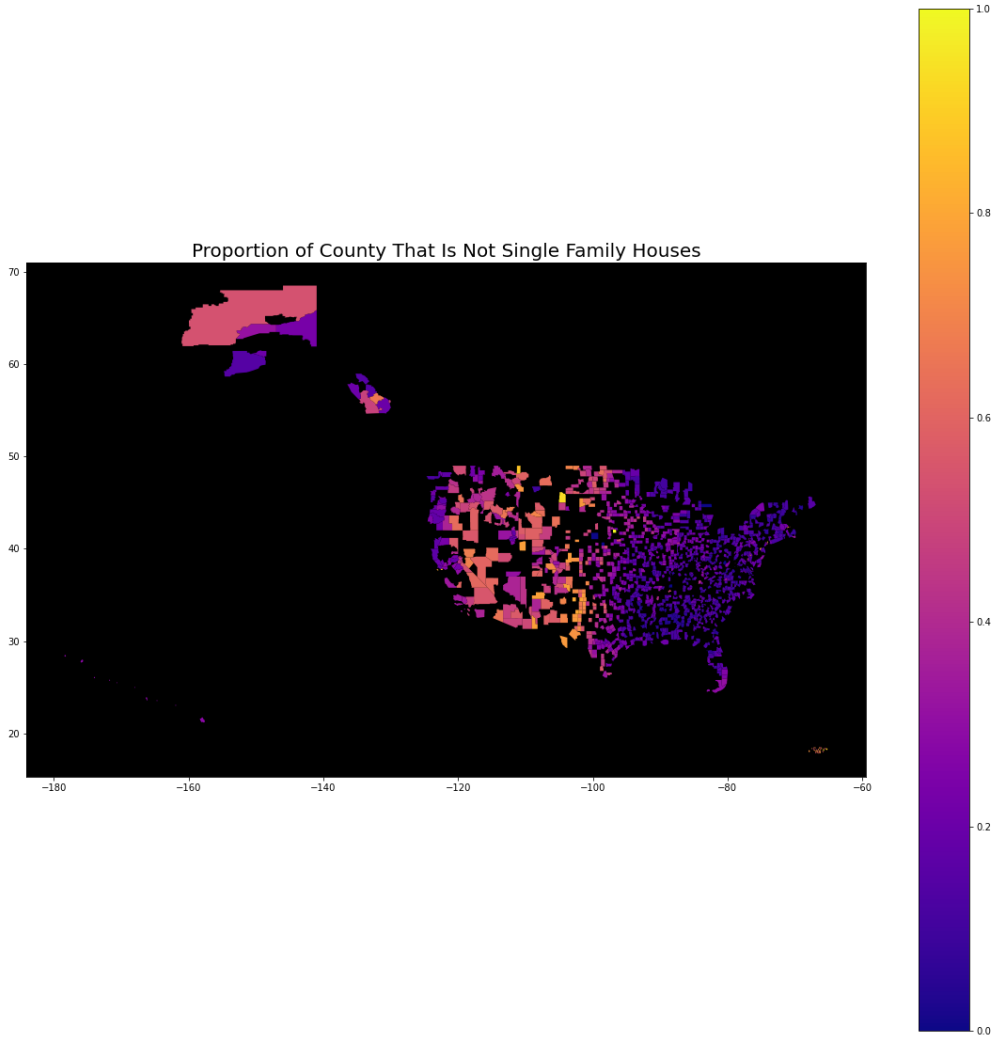
Available modules include pandas, geopandas, numpy, and statsmodels. There are no competing modules. There is no need to write custom code beyond that which is needed to analyze the dataset.

ENVISIONED CHALLENGES

One envisioned challenge is that there may not be enough observations in certain geographic areas or the observations will primarily be available for densely-populated areas. If the data is sparse across the United States, then the analysis would not offer meaningful insights. Based on the data availability, the multi-level modeling may need to be restricted to being applied to only a few geographic regions or states.

EXPLORATORY DATA ANALYSIS


Distribution of Excess Deaths
Frequency
100
80
60
40
20
0
2
4
6
8
Log(excess deaths)



REFERENCES

Ackley, C. A., Lundberg, D. J., Ma, L., Elo, I. T., Preston, S. H., & Stokes, A. C. (2021). *County-Level Estimates of Excess Mortality Associated with COVID-19 in the United States*. https://doi.org/10.1101/2021.04.23.21255564

Nguyen, Q. C., Khanna, S., Dwivedi, P., Huang, D., Huang, Y., Tasdizen, T., Brunisholz, K. D., Li, F., Gorman, W., Nguyen, T. T., & Jiang, C. (2019). Google Street View Features by County [Data file].

Rahman, Md. Ishfaq Ur., Panozzo, Kimberly A. (2020). Excess Deaths by County [Data file].

Stokes, A. C., Lundberg, D. J., Elo, I. T., Hempstead, K., Bor, J., & Preston, S. H. (2021). COVID-19 and excess mortality in the United States: A county-level analysis. *PLOS Medicine*, *18*(5), e1003571. https://doi.org/10.1371/journal.pmed.1003571

University of Wisconsin Population Health Institute. (n.d.). *Explore Health Rankings | Our Methods*. County Health Rankings & Roadmaps; University of Wisconsin Population Health Institute. https://www.countyhealthrankings.org/explore-health-rankings/our-methods

U.S. Centers for Disease Control and Prevention. (2021, October 13). Excess Deaths Associated with COVID-19. Centers for Disease Control and Prevention; U.S. Department of Health & Human Services. https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm