COVID-19 Research Group
Eddy, Chenyang, Jackson

Stage 2: Variable Codebook

For our project, we are using a self-compiled dataset, which draws upon datasets from multiple sources. In our compiled data set, each observation is a county in one of the ten states with the most coronavirus cases (as of 4/26). These states are California, New Jersey, New York, Pennsylvania, Illinois, Michigan, Louisiana, Texas, Florida, Massachusetts. In total, there are 792 counties in our data set, although not every county has data for each variable.

Name	Label	Range/Code	Missing Data
State	Name of State		
County	Name of County		
totalCases	Total confirmed COVID-19 cases in county.[1]	0-160000	NA
totalDeaths	Total confirmed COVID-19 deaths in county.[2]	0-18000	NA
Pop.2019	Estimated county population in 2019.[3]	150-10040000	NA
death.perCapita	Number of confirmed COVID-19 deaths per capita (calculated)	0-0.0012	NA
Beds	Estimated number of Hospital Beds in county[4]	0-30500	NA
Income	Per capita income in 2018, USD.[5]	19000-200000	NA
StayHomeDate	Date of statewide stay at home order.[6] Categorical - will be split into some number of dummy 1/0 variables for regression.	2020-03-19 through 2020-04-04	

Sources and Notes

1) Confirmed Case Count by State, time series. Johns Hopkins via github. Link
2) Confirmed Death Count by State, time series. Johns Hopkins via github. Link
3) US Census Bureau, estimated population for 2019. Link
4) Homeland Infrastructure Foundation-Level Data. Last updated October 2019. Does not reflect any additional or temporary bed capacity constructed during the pandemic. link
5) Personal Income by County, bea.gov. link. Per capita income in 2018.
6) "See Which States and Cities Have Told Residents to Stay at Home." *New York Times*, updated April 20, 2020. <https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html>