Based on the data given by the United States Census Bureau, we have access to data regarding demographics and housing of the population of the United States. Using this data we can form a data dictionary:

Name	Definition	Data type	Possible Values	Required?
demographic_label	An identifier for the data set that relates to the demographics of the population	Text	Total, Sex (Male/Female), Age, Race, Housing	Yes
demographic_estimate	The estimated number of people in the population that match the demographic label	Integer	161118151, 41498453, 315887408	Yes
margin_of_error	A plus/minus value to adjust for differences between the actual population and estimated population	Integer	±27,812 ±34,712 ±103,258	Yes
percent_estimate	The percentage of number of people in the US accounted for by the demographic label	Tinyint	49.2 50.8 12.7 (are decimals allowed with ints? Ask in class…)	Yes
percent_margin_of_error	A plus/minus value to adjust for differences between the actual percent of the population and the estimated percent of population	Tinyint	0.1	Yes

This data can be merged with the primary Covid-19 dataset by recognizing the presence of the total population data field between the COVID-19 cases and deaths datasets for county-wise populations and the ACS Social, Economic, and Housing enrichment data. Using this we can start to make assumptions about the data based on the percentage of population living within counties outlined in the population dataset for COVID-19 and their demographics.

This enrichment data can help the analysis of COVID-19 spread based on specifying the demographics of the populations listed in the cases and deaths data. By cross referencing the population demographics of the population, and where we see an increase in the number of

COVID-19 cases, we can get a better understanding of which demographics in the US are seeing the most cases. This can further be extrapolated to what age and income are seeing the greatest number of cases and deaths. By utilizing these datasets, we can make educated decisions on which demographics are seeing the highest rates of infection, and can therefore be the first line of defense in controlling the spread.
Initially, there are some questions that can come along with this idea of using demographic data to predict COVID-19 spread:

1. Which demographics are most likely to be affected by COVID-19
2. What is the strongest indicator of effect from COVID-19? Could it be race, age, socioeconomic status or some other confounding variable unaccounted for in the data?
3. How can we use this data to control the spread and death rate of COVID-19?