Table IV: Description of ACS Social, Economic, and Housing - Kevin Hayes

Name	Definition	Data Type	Possible Values	Required ?
Geography	A unique number associated with each county. This will not be used because I am only looking at the US, so the names of states and counties are sufficient for identification.	Text	0500000US01003 0500000US01015 0500000US01043 0500000US01049 0500000US01051 0500000US01055 0500000US01069	Yes.
Geographic Area Name	The name of the county, followed by a comma and the full name of its state.	Text	Baldwin County, Alabama Calhoun County, Alabama Cullman County, Alabama DeKalb County, Alabama Elmore County, Alabama Etowah County, Alabama	Yes
Estimate	The estimated value of a given label	Number (Either Integer or Float. Depends on row)	327,167,439, 97.0, 38.2, 52,423,114, 79.9	Yes
Margin of Error	Absolute expected deviation from a given estimate.	Number (Either Integer or Float. Depends on row)	27,812, 27,815, 0.1, 0.2, 34,697	No
Percent	Percentage of a given sublabel. (Note: base level labels have either their initial value or (X) for this. This will be fixed in data cleaning)	Float	49.2, 50.8, 6.0, 6.1, 6.5,	No

Percent Margin of Error	relative expected deviation from a given estimate (Note: labels where this does not apply have either their initial value or (X) for this. This will be fixed in data cleaning)	float	0.4 0.3 0.6 0.9 0.5	No

- We can join this to our main dataframe by using Geographic Area Name to generate the correct index.

- All of these are followed by !!<label>!!<sublabel> to denote where the value applies to. These labels are SEX AND AGE, RACE, Race alone or in combination with one or more other races, HISPANIC OR LATINO AND RACE, and CITIZEN, VOTING AGE POPULATION. The sublabels refer to various races or qualifiers to the main label. In practice I plan to break up each label into its own dataframe.

- This enrichment data can help analyze the spread of COVID-19 by allowing us to see if areas with people of given ages or races were more or less likely to have people infected with or killed by COVID-19.

- Potential hypothesis questions:
    - Is there a correlation between mean age in a county and the number of COVID-19 deaths in said county?
    - Is there a correlation between mean age in a county and the number of COVID-19 infections in said county?
    - Is there a correlation between the percentage of people of age 85 or over in a county and the number of COVID-19 deaths in said county?
    - Is there a correlation between the percentage of people in a county that are hispanic or latino and deaths from COVID-19?