Task 2:

* Description of enrichment data and data types - variable dictionary

| Name | Definition | Data type | Possible values | Required? |
| --- | --- | --- | --- | --- |
| Label (Grouping) | Label of subgroups | Object | SEX AND AGE, Total Population, Male, Female, ... | Yes |
| County!!  Estimates | Estimated population for specified category | Object | 1, 2, 3, … | No |
| County!! Margin of Error | Population value for margin of error for given population estimate | Object | ±1, ±2, ±3, … | No |
| County!! Percent | Percent of population that fit into the specified category in the United States | Object | 1, 2, 10%, 20%, (X), … | No |
| County!! Percent Margin of Error | Percent margin of error for specified percentage. | Object | ±0.1, ±0.2, ±0.3, (X), … | No |

* How can you merge the data with the primary COVID-19 dataset? Identify the individual variable which map between the datasets.
  + The variables(County!!Estimates, County!!Margin of Error, …) can be combined with each kind of label(SEX AND AGE, Total Population, …) to create a new column for each county.
  + Ex:
    - Wake County!!Estimates, Age
    - Wake County!!Margin of Error, Age
    - Wake County!!Percent, Age
    - Wake County!!Percent Margin of Error, Age
    - Wake County!!Estimates, Total Population
    - …
* Describe how your enrichment data can help in the analysis of COVID-19 spread. Pose initial hypothesis questions.
  + This data shows the ethnic background of people and some other details such as age, sex, and population density. This can be insightful to seeing what kinds of populations are impacted by covid more than others.
  + Initial hypotheses:
    - If the population of older ages decreases at a more drastic rate, covid is more harmful to people of older ages.
    - If a county with a high population of a certain demographic has high covid numbers, then other counties with a high population of that demographic are more likely to have high covid rates in the future.