Second Half Progress Report 2

Date: November 7, 2021

Accomplishments:

- Created a new asthma studies Jupyter notebook file and read in the CDC Chronic Disease Indicators Asthma dataset. Although the dataset website https://chronicdata.cdc.gov/Chronic-Disease-Indicators/U-S-Chronic-Disease-Indicators-Asthma/us8e-ubyj reports it was last updated in February, 2021, the most recent year for which there are entries is 2019. 2012 is the only year in common with the air quality studies I've completed. Therefore, I will limit the asthma studies to 2012.
- Completed exploratory analysis of the asthma dataset.
- Cleaned the CDC Chronic Disease Indicators: Asthma dataset. Removed observations from the District of Columbia, Guam, Puerto Rico, the Virgin Islands, and the United States. Narrowed the dataset to 2012 observations and to one question: Current asthma prevalence among adults aged 18 and older. Further narrowed the dataset to overall incidence (not stratified by gender or race) and used the crude prevalence rather than the age adjusted prevalence.
- Completed analysis on the asthma studies target questions: (1) What does the reported incidence of asthma look like by state? (2) Which states have the highest and lowest reported incidence of asthma.
- Searched for 2012 asthma datasets for states with the highest reported prevalence of asthma. Located asthma datasets for Kentucky (tied with Maine for highest) and Vermont (number 3). The state asthma datasets provide granularity at the county level, although the reporting year is not limited to 2012 in either dataset. Kentucky reports asthma in two or three-year periods. Options are 2011-2012 and 2011-2013. Vermont reports asthma for two-year periods. Options are 2011-2012 and 2012-2013.
- Uploaded the asthma Jupyter notebook file to the GitHub project repository.
- Uploaded the Kentucky and Vermont asthma datasets to the GitHub project repository.
- Uploaded the Second Half Progress Report 2 to the GitHub project repository.

Current Activities: I am currently working on air quality and asthma visualizations with U.S. maps.

Challenges: Searching for appropriate asthma datasets has been the most time consuming. Creating visualizations with U.S. maps using Python is very challenging, and although I am confident that I can do those visualizations with Tableau, I want to do them in Python first.

Work to be Completed: For the next project milestone, I will finish the AQI and asthma visualizations, join the AQ and asthma datasets, and complete the exploratory analysis of the combined datasets.