The purpose of this document is to keep track of the development of the third incarnation of the Asthma Equity Explorer, broken down by sub task. This document will most likely be developed into an add-on for the README.

#### To-Do:

* Check in with Karen on status of gentrification work
* Down the line: Look into box/box-syncing for moving towards hosting data

Prototype:

* In rmd check out prev vs ed visits for specific areas.
  + Dataset: https://data.chhs.ca.gov/dataset/asthma-emergency-department-visit-rates
* Make sure the colors match the skirt
* Combined risk rank, not risj rank. Longer converstion. Figure out what to put there.
* Make highlighted geo a lighter shade of red. Get rid of yellow highlight.
* Make slide with no selection, then slide for each geographic level.
* Level and place fixed for geography selection; dropdown for factors.
* Match MITRE hex codes: <https://comm.mitre.org/strategiccommunications/self-service/logos-fonts-colors/>
* Use SDOH as an example; redlining score, housing stress

#### Summary:

Blah

#### Useful past repositories:

* Explore-data: A repository for exploring, processing, and analysing health equity data.
* Asthma-dashboard (v1): The first version of the asthma equity explorer.
* Simplified-asthma-dashboard (v2): The second version of the asthma equity explorer.

#### The Data:

There are a total of 8 publicly available datasets used for the asthma equity explorer. The table below tracks the details of the datasets.

Data Sources Overview for Eight Hosted Data Sources

|  |  |  |  |
| --- | --- | --- | --- |
| *NAME* | *CONTENT DESCRIPTION* | *YEAR* | *GEOGRAPHIC LEVELS* |
| American Community Survey (ACS)  COMPLETE | Socioeconomic features | 2019 | Census tract, county, state, national |
| Behavioral Risk Factor Surveillance System (BRFSS)  COMPLETE | Health measures, including asthma and mental health (self-reported); telephone survey | 2019 | State |
| PLACES  COMPLETE | Self-reported health measures categorized by health outcomes, prevention and unhealthy behaviors modeled from BRFSS data | 2018-2019 data and 2010, 2014-2019 population estimates | County, incorporated and census designated places, census tract, and zip code tabulation area (zcta) |
| Redlining  COMPLETE | Home Owners’ Loan Corporation (HOLC) grades in 138 US cities | 1935-40 | Historical neighborhood (mapped to modern census tracts) |
| ACAGPM Particulate Matter  COMPLETE | Annual particulate matter 2.5 concentrations in micrograms per cubic meter from WUSTL’s ACAG data | 2018 | Census tract, county, state |
| Housing Stress (HUD)  COMPLETE | Measures which demonstrate the extent of housing problems and needs, particularly for low income households | 2018 | Census tract, county, state |
| Access to Parks (NEPHTN)  COMPLETE | Percentage of people living within ½ mile of a park for various geographic levels | 2015, 5 year interval | Census tract, county, state, national |
| Access To Healthy Food (USDA)  COMPLETE | Food access indicators which measure the availability of healthy food | 2019 | Census tract, county, state, national |

Here is a breakdown of the datasets, broken down by their link, data collection year, geographic levels, sampling methods notes describing the datasets, updates from the last round of data collection, and files in repositories which make use of the data.

##### Dataset Descriptions

|  |  |
| --- | --- |
| *American Community Survey (ACS)*  COMPLETE | |
| Link | Homepage: [https://www.census.gov/programs-surveys/acs](https://www.census.gov/programs-surveys/acs" \t "_blank)  Design and methodology: [https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html](https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html" \t "_blank) |
| Year | 2019 |
| Geographic Levels | Census tract, county, state |
| Sample | Random sample of housing unit (HU) addresses and residents of group quarters (GQ) facilities from the Census Bureau’s Master Address File (MAF). Selected independent HU address samples for each of the 3,143 counties and county equivalents in the U.S., including the District of Columbia. There were approximately 2.9 million HU addresses in the U.S. chosen for 2010. |
| Notes | Accessed through the “tidycensus” package in R  Useful links:  [https://www.census.gov/programs-surveys/acs/guidance/handbooks/general.html](https://www.census.gov/programs-surveys/acs/guidance/handbooks/general.html" \t "_blank)  [https://api.census.gov/data/2019/acs/acs5/variables.html](https://api.census.gov/data/2019/acs/acs5/variables.html" \t "_blank)  [https://www.census.gov/data/developers/data-sets.html](https://www.census.gov/data/developers/data-sets.html" \t "_blank)  [https://www.census.gov/glossary/](https://www.census.gov/glossary/" \t "_blank)  [https://www.census.gov/programs-surveys/acs/guidance/estimates.html](https://www.census.gov/programs-surveys/acs/guidance/estimates.html" \t "_blank)  [https://walker-data.com/tidycensus/articles/basic-usage.html](https://walker-data.com/tidycensus/articles/basic-usage.html" \t "_blank) |
| Updates | Has data through 2020.  Try to include B01003\_001 Total Pop, B02001\_001 Total Pop, B02001\_002 White alone, B02001\_003 Black or African American alone, B02001\_004 American Indian and Alaska Native alone, B02001\_005 Asian alone, B02001\_006 Native Hawaiian and Other Pacific Islander alone, B02001\_007 Some other race alone, B02001\_008 Two or more races, B03001\_002 Not Hispanic or Latino, B03001\_003 Hispanic or Latino  OR  B03002\_001 Total Pop, B03002\_002 Not Hispanic or Latino, B03002\_003 White alone, B03002\_004 Black or African American alone, B03002\_005 American Indian and Alaska Native alone, B03002\_006 Asian alone, B03002\_007 Native Hawaiian and Other Pacific Islander alone, B03002\_008 Some other race alone, B03002\_009 Two or more races, B03002\_012 Hispanic or Latino  ACS has all the data from decennial census but for 2020. However, it is mapped to 2020 census tracts. Will stick to 2019.  Income info also sourced from here. |
| Files | Asthma\_dashboard\_data\_processing\_functions.R   * Get\_acs\_info: Pull desired acs variables * Get\_ice\_metric: Pull income info * Get\_merged\_shp: above 2 functions are called |
| *Behavioral Risk Factor Surveillance System (BRFSS)*  COMPLETE | |
| Link | BRFSS website <http://www.cdc.gov/brfss>  Methodology: <http://www.cdc.gov/brfss/factsheets/pdf/DBS_BRFSS_survey.pdf>  Glossary: <https://chronicdata.cdc.gov/Behavioral-Risk-Factors/Behavioral-Risk-Factor-Surveillance-System-BRFSS-H/iuq5-y9ct/data>   2018 documentation: <https://www.cdc.gov/brfss/annual_data/2018/pdf/overview-2018-508.pdf> |
| Year | 2011-present (2018 used: <https://www.cdc.gov/brfss/annual_data/2018/pdf/codebook18_llcp-v2-508.pdf> ) |
| Geographic Levels | State |
| Sample | Random sampling of telephones from states |
| Notes |  |
| Updates | … |
| Files | … |
| *PLACES*  COMPLETE | |
| Link | Homepage: [https://www.cdc.gov/places/about/index.html](https://www.cdc.gov/places/about/index.html" \t "_blank)  Measure Definitions: [https://www.cdc.gov/places/measure-definitions/index.html](https://www.cdc.gov/places/measure-definitions/index.html" \t "_blank) |
| Year | Data sources used to generate these model-based estimates include Behavioral Risk Factor Surveillance System (BRFSS) 2019 or 2018 data, Census Bureau 2010 population estimates, and American Community Survey (ACS) 2015-2019 or 2014-2018 estimates. The 2021 release uses 2019 BRFSS data for 22 measures and 2018 BRFSS data for 7 measures |
| Geographic Levels | county, places (incorporated and census designated places), census tract, and ZIP Code Tabulation Area (ZCTA) |
| Sample | Model-based population estimates. 3,142 counties within 50 states and District of Columbia (DC), including 9 county and city consolidated cities. 28,484 of 29,321 Places or Census Designated Places (CDP) with a population of 50 or more from Census Tiger 2019 Places and CDPs based on the Census 2010 population count. 72,337 of 73,057 census tracts in the 50 states and Washington, DC that have a Census 2010 population of 50 or more people. 32,409 of 32,989 ZCTAs in the 50 states and Washington, DC with Census 2010 population of 50 or more people. |
| Notes | Relies on BRFSS and only includes adults >= 18 years of age |
| Updates | Moved from 2020 release to 2021 release.  2 new fields, same geographies.   * *DEPRESSION\_CrudePrev* - Model-based estimate for crude prev of depression among adults >=18, 2019 * *GHLTH\_CrudePrev* - Model-based estimate for crude prev of fair or poor health among adults >=18, 2019 |
| Files | Asthma\_dashboard\_data\_processing\_functions.R   * Get\_merged\_shp: pull disease data |
| *Particulate Matter 2.5* COMPLETE | |
| Link | Data: [https://ephtracking.cdc.gov/showIndicatorPages](https://ephtracking.cdc.gov/showIndicatorPages" \t "_blank) ; [https://data.cdc.gov/browse?category=Environmental+Health+%26+Toxicology&page=2](https://data.cdc.gov/browse?category=Environmental+Health+%26+Toxicology&page=2" \t "_blank)  Air quality monitoring data: EPA Air Explorer: [http://www.epa.gov/airdata/](http://www.epa.gov/airdata/" \t "_blank)  Air quality modeled data: [https://www.epa.gov/aboutepa/about-national-exposure-research-laboratory-nerl](https://www.epa.gov/aboutepa/about-national-exposure-research-laboratory-nerl" \t "_blank)  Census population data: [https://www.census.gov/data/developers/data-sets/popest-popproj/popest.html](https://www.census.gov/data/developers/data-sets/popest-popproj/popest.html" \t "_blank) |
| Year | 2018 |
| Geographic Levels | Census tract, county, state |
| Sample | The monitoring data comes from the U.S. Environmental Protection Agency (EPA) Air quality System (AQS). When AQS data are available from multiple monitors for a given county and day, the highest 24-h average (daily) PM2.5 concentration among all the monitors is selected for purposes of creating daily county level data. EPA provides modeled estimates of PM2.5 using Downscaler (DS) model, which uses a statistical approach to fuse monitoring data in areas where monitors exist, and relies on Community Multiscale Air Quality (CMAQ) modeled output in areas without monitors. DS modeled estimates are available by census tract centroid-the geographic center of the census tract.    Census tract-level datasets contain estimates of the mean predicted concentration and associated standard error. |
| Notes | There are 2 main datasets for PM2.5: monitored and modeled EPA indicator data  and the tract level modeled estimates from EPA. Tract level estimates are given for every tract for every day of the year. We calculate the state level estimates by averaging the county level estimates. Since we are interested in annual estimates, we calculate the tract level estimates by averaging the modeled estimate for every day. Only goes up to 2016    Annual average ambient concentrations of PM 2.5 based on seasonal averages and daily measurement; Beginning 2013, the EPA's revised annual PM2.5 standard of 12​. Daily county level modeled estimates are obtained by selecting the maximum value observed among all the census tracts within each county. County level PM2.5 measures are created using monitor data when available and using modeled estimates for days and locations without such data |
| Updates | Moved from using EPA data from 2016 to using ACAG data mapped to 2019 census tracts in 2018. |
| Files | Asthma\_dashboard\_data\_processing\_functions.R   * Get\_merged\_shp: Pull pm2.5 levels |
| *Redlining* COMPLETE | |
| Link | Homepage: [https://dsl.richmond.edu/panorama/redlining/#loc=5/39.1/-94.58&text=downloads](https://dsl.richmond.edu/panorama/redlining/" \l "loc=5/39.1/-94.58&text=downloads" \t "_blank)  Citation: Robert K. Nelson, LaDale Winling, Richard Marciano, Nathan Connolly, et al., “Mapping Inequality,” American Panorama, ed. Robert K. Nelson and Edward L. Ayers, accessed January 8, 2021, https://dsl.richmond.edu/panorama/redlining/ |
| Year | 1935-40 |
| Geographic Levels | HOLC block (historical neighborhood) |
| Sample | 196 cities in 38 US states |
| Notes | HOLC grades are mapped to census tracts in the given cities for all analyses |
| Updates | … |
| Files | … |
| *Housing Stress: Department of Housing and Urban Development (HUD)* COMPLETE | |
| Link | <https://www.huduser.gov/portal/datasets/cp.html> |
| Year | 2018, 5 year interval |
| Geographic Levels | Census tract, county, state |
| Sample | Each year, the U.S. Department of Housing and Urban Development (HUD) receives custom tabulations of American Community Survey (ACS) data from the U.S. Census Bureau. These data, known as the "CHAS" data (Comprehensive Housing Affordability Strategy), demonstrate the extent of housing problems and housing needs, particularly for low income households. |
| Notes | Reference for measure definition: <https://www.hrsa.gov/> ; Official Definition for this housing stress measure: Fraction of households where one or more of the following conditions are met: (1) housing expense/income threshold—monthly housing costs, including utilities, exceed 30% of income, (2) crowding— more household members than rooms, (3) incomplete plumbing—home lacks necessary bathroom facilities, and (4) incomplete kitchen— home lacks essential; Numerator  T1\_est3 : Owner Occupied - has 1 or more of the 4 housing unit problems (lacks kitchen or plumbing, more than 1 person per room, or cost burden greater than 30%)  T1\_est127 : Renter Occupied - has 1 or more of the 4 housing unit problems (lacks kitchen or plumbing, more than 1 person per room, or cost burden greater than 30%)  Denominator  T1\_est2: All Owner Occupied  T1\_est126 : All Renter Occupied |
| Updates | Updated from 2017 to 2018 data. |
| Files | … |
| *Access to Parks (NEPHTN)* COMPLETE | |
| Link | <https://ephtracking.cdc.gov/showIndicatorPages.action?selectedContentAreaAbbreviation=14&selectedIndicatorId=80&selectedMeasureId=> (under community design: population living within half mile of park) |
| Year | 2015, 5 year interval |
| Geographic Levels | Census tract, county, state, national |
| Sample | Parks: Number of people within a buffer of ½ mile radius of a park was determined at the census tract level. These estimates are aggregated to county, state, and national levels. Percentages of people living within ½ mile of the park boundary are calculated for the census tract, county, state, and national levels. The percentage uses the estimated numbers of people as determined via the buffer analysis and then divides this numerator by the total number of people in each geographic unit. |
| Notes | The area-proportion technique used to determine the number of people residing within a park or school's buffer assumes that populations are distributed evenly within census tracts. However, in reality, there are variations in the size, population, and shape of each census tract and in the distribution of populations within them.  Estimates are dependent upon the software employed. Potential variations in the estimates can occur due to rounding of the estimates. We attempted to limit rounding error by reserving rounding until the last step of the aggregation process.  The estimates for the population residing within the half-mile of a park are derived estimates. The area-proportion estimates for race and ethnic subgroups are limited to 2010 data. Hispanic, White Non-Hispanic, Black Non-Hispanic, Other Non-Hispanic, and Multirace Non-Hispanic were summed to determine the total proportion for each block. Each of these estimates were rounded at each of the geographic aggregation units and then summed for the total half-mile population estimate. We employed this approach to maintain internal consistency within the county, state, and national estimates. It is possible a statistically insignificant difference in the estimate could occur due to rounding.  Parks: Using percent of population living within a half mile of a park as a proxy for proximity may underestimate the number of people who have access to parks in populations that are willing to travel farther to reach a park. Similarly, people may use parks located close to their schools or places of work rather than parks close to their homes.  Conversely, this may overestimate access since walk routes to park entrances may be unsafe or much farther than a direct line from a residence to a park boundary. |
| Updates | No changes. |
| Files | … |
| *Limited Access To Healthy Food (USDA)* COMPLETE | |
| Link | <https://www.ers.usda.gov/data-products/food-access-research-atlas/download-the-data/> |
| Year | 2019 |
| Geographic Levels | Census tract, county, state, national |
| Sample | Estimates in the Atlas for 2019 are based on a 2019 list of supermarkets, the 2010 Decennial Census, and the 2014-18 American Community Survey (ACS). |
| Notes | The Food Access Research Atlas maps food access indicators for census tracts using ½-mile and 1-mile demarcations to the nearest supermarket for urban areas, 10-mile and 20-mile demarcations to the nearest supermarket for rural areas, and vehicle availability for all tracts. Atlas users can view census tracts by food access indicators using these different measures to see how the map changes as the distance demarcation or inclusion of vehicle access changes. The map includes updated estimates using 2019 data and allows users to compare these new estimates with those from 2015. |
| Updates | No changes. |
| Files | … |