Table of Contents

[Data: Inputs, Outputs, and Reference Files 2](#_Toc182581178)

[Inputs 2](#_Toc182581179)

[Database 2](#_Toc182581180)

[Tables in the Database 2](#_Toc182581181)

[Outputs 3](#_Toc182581182)

[ArcGIS 3](#_Toc182581183)

[Codebook 4](#_Toc182581184)

[Preliminary Data Collection: 4](#_Toc182581185)

[Exploratory Data Analysis 4](#_Toc182581186)

[ArcGIS Mapping 4](#_Toc182581187)

[Instruction Manual 6](#_Toc182581188)

# Data: Inputs, Outputs, and Reference Files

## Inputs

### Database

acsse\_2022.db: The SQLite database containing four tables.

##### Links

GitHub Database: <https://github.com/brutus-the-homeschooler/Capstone/blob/main/Database/acsse_2022.db>

### Tables in the Database

#### census\_data

This dataset contains data fetched from the ACSSE API, including variables with associated measurements of error. Each row includes the FIPS code for the state and place, enabling identification of geographic areas. The variables and their error measurements provide insight into various metrics, though specific columns may vary depending on the data pulled from the API.

#### place\_dictionary

This dataset provides information about places included in the ACSSE data, containing place names along with their corresponding FIPS codes. The 'place' column holds the FIPS code for each place, with additional columns providing metadata related to each location. This file was generated via the Census API.

##### Links

Census API Places: [https://api.census.gov/data/2022/acs/acsse?get=NAME,K200101\_001E&for=place:\*](https://api.census.gov/data/2022/acs/acsse?get=NAME,K200101_001E&for=place:*)

#### state\_dictionary

This dataset contains information about states along with their corresponding FIPS codes. The 'state' column includes the FIPS code for each state, with additional columns providing metadata relevant to each state. This file is an unduplicated version of the place\_dictionary, designed to allow for faster connection to state-level data.

#### variable\_dictionary

This dataset includes all current variables from the 2022 survey, along with their associated measurements of error. It provides summary statistics for each variable, including minimum, maximum, mean, median, and the count of missing values. Note that all values are stored as strings, with 'None' representing missing values instead of '0'.

Most of the variable information is based on details from the Census API documentation. This documentation was cleaned and refined during the Data Dictionary assignment, with those results also stored in this location.

##### Links

Census API Documentation: <https://api.census.gov/data/2022/acs/acsse/variables.html>

Data Dictionary: <https://github.com/brutus-the-homeschooler/Capstone/blob/main/Dictionary/Variable%20Dictionary.csv>

## Outputs

### ArcGIS

#### All Places

These files include shapefiles, GeoJSON, and CSV files for all 2,399 places within the ACS data. They enable the visualization and pinpointing of these locations in mapping software such as ArcGIS.

##### Links

GitHub: <https://github.com/brutus-the-homeschooler/Capstone/tree/11616002c7af4fa692a2ba92d60f46348fca0c84/ArcGIS/2022%20ACS%20All%20Places>

#### Subset

These files include shapefiles, GeoJSON, and CSV files for the 543 places that passed the reliability check across 19 variables. This subset enables accurate mapping and analysis in software like ArcGIS.

##### Links

GitHub: https://github.com/brutus-the-homeschooler/Capstone/tree/main/ArcGIS/2022%20ACS%20Subset

#### Story Map

Based on feedback from the capstone sponsor, ArcGIS will serve as the primary visualization tool for the final product. To support this, we created a Story Map as part of the Exploratory Data Analysis (EDA) assignment. The Story Map showcases various maps and visualizations designed to help nontechnical users understand our technical approach and problem-solving methodology.

##### Links

Story Map: <https://github.com/brutus-the-homeschooler/Capstone/tree/11616002c7af4fa692a2ba92d60f46348fca0c84/ArcGIS/2022%20ACS%20All%20Places>

# Codebook

### Preliminary Data Collection:

#### Acs\_data\_to\_sqlite.py

This script is used to scrape the American Community Survey data for the 1-year Supplemental from 2022. This includes towns and places with 20,000 residents or more. The script creates the four tables listed above. To run successfully, one needs a Census API key. To better streamline this need, the Census API key used for the initial creation is housed within a secret in GitHub.

##### Links

Code: <https://github.com/brutus-the-homeschooler/Capstone/blob/main/Scripts/Workflow/acs_data_to_sqlite.py>

### Exploratory Data Analysis

#### City of Dublin EDA.R

This script generates the plots and analyses for the exploratory data analysis assignment. The visualizations created are saved in the GitHub repository and are also featured in the Story Map used to present the analysis.

##### Links

Code: <https://github.com/brutus-the-homeschooler/Capstone/blob/main/Scripts/EDA/City_of_Dublin_EDA.R>

GitHub: <https://github.com/brutus-the-homeschooler/Capstone>

Story Map: <https://storymaps.arcgis.com/stories/329f10f1c2064be987c463b079d29b16>

### ArcGIS Mapping

#### Arcgis\_all\_places.py

This script is used to generate the shapefiles, csv, or geojson that are used to create the map visualization in the Story Map. The files created are also housed within the GitHub repository. The dataset includes all 2,399 places designated within the American Community Survey 2022 Supplemental.

##### Links

Code: <https://github.com/brutus-the-homeschooler/Capstone/blob/main/Scripts/Workflow/arcgis_all_places.py>

GitHub: <https://github.com/brutus-the-homeschooler/Capstone/tree/main/ArcGIS/2022%20ACS%20All%20Places>

#### [Arcgis\_all\_places\_subset.py](https://github.com/brutus-the-homeschooler/Capstone/blob/main/Scripts/Workflow/arcgis_all_places.py)

This script is used to generate the shapefiles, csv, or geojson that are used to create the map visualization in the Story Map. The files created are also housed within the GitHub repository. The dataset includes the 564 places designated within the American Community Survey 2022 Supplemental that have reliable data for variables of interest to the sponsor, the City of Dublin, OH.

##### Links

Code: <https://github.com/brutus-the-homeschooler/Capstone/blob/main/Scripts/Workflow/arcgis_all_places_subset.py>

GitHub: <https://github.com/brutus-the-homeschooler/Capstone/tree/main/ArcGIS/2022%20ACS%20Subset>

# Instruction Manual

To recreate the data collection, processing, validation, and analysis, the codebook should be run in the following order:

1. [Acs\_data\_to\_sqlite.py](#_Acs_data_to_sqlite.py)

This is the base of all analysis as the team was not given a dataset to start from and more directed to homegrown solutions regarding the data needed to find sister cities to the City of Dublin.

1. [City\_of\_Dublin\_EDA.R](#_City_of_Dublin)

This will create the visuals and exploratory data analysis around the different variables within the dataset and remove any unreliable data points.

1. [Arcgis\_all\_places.py](#_Arcgis_all_places.py)

This will create the necessary files to display in ArcGIS to highlight the locations of the 2,399 places within the American Community Survey data.

1. [Arcgis\_all\_places\_subset.py](#_Arcgis_subset.py)

This will create the necessary files to display in ArcGIS to highlight the 564 locations that have reliable data when compared to City of Dublin for the 19 variables identified in the EDA.