**VDSS\_ZCTA\_ACS\_Characteristics.R - README**

**Author: Dylan Craig**

**Date Last Updated: 12/28/2024**

**File Name: VDSS\_ZCTA\_ACS\_Characteristics.R**

**Purpose**

This script processes and aggregates American Community Survey (ACS) data from multiple sources into a unified dataset, enabling analysis of population characteristics at the ZCTA (ZIP Code Tabulation Area) level.

**Overview**

The script, **VDSS\_ZCTA\_ACS\_Characteristics.R**, combines data from various ACS tables related to population, age, sex, race, ethnicity, income, education, and employment. It standardizes and aggregates these datasets by ZCTA and YEAR for comprehensive analysis.

**Datasets and Tools**

**Input Files:**

Each input dataset corresponds to an ACS table and is stored in separate directories:

1. **Total Population**: B01003\_Total\_Population
2. **Sex by Age**: B01001\_Sex\_By\_Age
3. **Race by Sex**: B02001\_Race\_By\_Sex
4. **Hispanic/Latino by Race**: B03002\_Hispanic\_Latino\_By\_Race
5. **Poverty Last 12 Months by Sex and Age**: B17001\_Poverty\_Last\_12\_Months\_By\_Sex\_By\_Age
6. **Median Income**: B19013\_Median\_Income
7. **Educational Attainment Over 18**: B15001\_Sex\_By\_Age\_By\_Educational\_Attainment\_Over\_18
8. **Marital Status Over 15**: B12002\_Marital\_Status\_By\_Age\_Over\_15
9. **Selected Economic Characteristics**: Selected\_Economic\_Characteristics

**Output Files:**

Aggregated data is saved in both .csv and .xlsx formats for each category, along with a final combined dataset:

1. **Population\_Data**: Population statistics.
2. **Sex\_By\_Age\_Data**: Age and sex distributions.
3. **Race\_By\_Sex\_Data**: Racial and gender composition.
4. **Hispanic\_Latino\_By\_Race\_Data**: Ethnicity distributions.
5. **Poverty\_Status\_Data**: Poverty indicators.
6. **Median\_Income\_Data**: Income data.
7. **Education\_Attainment\_Data**: Educational attainment.
8. **Marital\_Status\_Data**: Marital status by age.
9. **Employment\_Status\_Data**: Employment data.
10. **VDSS\_ZCTA\_ACS\_Characteristics\_Aggregated**: Final aggregated dataset with all categories.

**Dependencies:**

* **R Libraries:**
  + dplyr: Data manipulation.
  + readr: Reading CSV files.
  + stringr: String operations.
  + writexl: Exporting Excel files.
  + tidyverse: Comprehensive data analysis and manipulation.

**Script Functionality**

**1. Reading and Combining Datasets:**

* The read\_combine\_and\_rename function reads CSV files from specified directories, combines them into a single dataset, and renames columns with a prefix for identification.

**2. Processing Individual Datasets:**

* Reads and processes data for each category (e.g., population, income) using the specified directories.
* Converts appropriate columns to numeric for aggregation.
* Saves processed datasets to the output folder in both CSV and Excel formats.

**3. Merging All Data:**

* Merges individual datasets by ZCTA and YEAR to create a unified dataset.
* Saves the final combined dataset in CSV and Excel formats.

**File Output**

1. **Category-Specific Files:**
   * Processed datasets for each ACS category, saved as both CSV and Excel files.
2. **Aggregated Dataset:**
   * A comprehensive dataset combining all ACS categories, saved as VDSS\_ZCTA\_ACS\_Characteristics\_Aggregated.csv and VDSS\_ZCTA\_ACS\_Characteristics\_Aggregated.xlsx.

**Usage**

1. Ensure all required ACS datasets are placed in the corresponding directories.
2. Install the necessary R libraries.
3. Execute the script in an R project environment.
4. Retrieve processed and aggregated data from the output directory.

**Notes**

* Input datasets must be consistently formatted with ZCTA and YEAR columns.
* The script assumes column names in input files align with ACS table definitions.
* Review the final combined dataset for any missing or inconsistent data.