**VDSS\_ZIP\_BORDER\_POPULATION\_FILTERED.R - README**

**Author: Dylan Craig**

**Date Last Updated: 12/29/2024**

**File Name: VDSS\_ZIP\_BORDER\_POPULATION\_FILTERED.R**

**Purpose**

**This script processes ZIP code, FIPS, and bordering county data for Virginia, refining associations based on population share data from Quarter 4, 2012. It outputs a filtered dataset that retains only FIPS codes sharing more than 1% of a ZIP code's population and includes their bordering FIPS codes.**

**Overview**

**The script filters, transforms, and structures ZIP code-FIPS-bordering county data to:**

1. **Include only Virginia FIPS codes and ZIP-FIPS pairs with at least 1% population share.**
2. **Expand the dataset to incorporate bordering FIPS codes while removing redundancies.**
3. **Save the final structured data for analysis.**

**Datasets and Tools**

**Input Files:**

1. **USPS Zip County Data: Includes ZIP codes and their associated FIPS codes.**
2. **Virginia Bordering Counties Data: Lists Virginia counties and their bordering FIPS codes.**
3. **Quarter 4, 2012 Population Share Data: Provides population share ratios for ZIP-FIPS associations.**

**Output File:**

1. **VDSS\_Zip\_FIPS\_Border\_Population\_Filtered.xlsx: Contains the final structured and filtered data.**

**Dependencies:**

* **R Libraries:** 
  + **readxl: For reading Excel files.**
  + **dplyr: For data manipulation.**
  + **tidyr: For reshaping data.**
  + **openxlsx: For writing data to Excel files.**
  + **purrr: For functional programming.**

**Script Functionality**

**1. Loading and Filtering Data:**

* **Reads ZIP-to-county data, bordering county FIPS data, and population share data.**
* **Filters the USPS Zip County data to include only Virginia FIPS codes (prefix 51).**
* **Retains only ZIP-FIPS pairs with a total ratio (TOT\_RATIO) of at least 1%.**

**2. Reshaping Bordering County Data:**

* **Converts the Virginia bordering counties data from a wide to a long format.**
* **Focuses on relevant counties and their associated bordering FIPS codes.**

**3. Processing ZIP Codes:**

* **Iterates through each unique ZIP code to:** 
  + **Identify and retain associated FIPS codes meeting the population share threshold.**
  + **Retrieve bordering FIPS codes while excluding redundant associations.**
  + **Organize data into structured lists for FIPS and bordering FIPS codes.**

**4. Expanding and Finalizing Data:**

* **Transforms structured lists into separate columns.**
* **Ensures data consistency by removing duplicates and cleaning redundant columns.**

**5. Saving the Output:**

* **Exports the final structured dataset to an Excel file for further use.**

**File Output**

**VDSS\_Zip\_FIPS\_Border\_Population\_Filtered.xlsx:**

* **Contains ZIP codes, associated FIPS codes, and bordering FIPS codes filtered by population share.**

**Usage**

1. **Place the input files in the specified directories.**
2. **Install the required R libraries.**
3. **Execute the script in an R project environment.**
4. **Access the output file in the Data Outputs/VDSS\_Zip\_Border\_Status/ directory.**

**Notes**

* **Ensure the input datasets are complete and correctly formatted.**
* **The accuracy of the filtering depends on the correctness of the population share and ZIP-FIPS data.**
* **Review the output file for consistency and accuracy before further analysis.**