### **Analysis of Zip Code Characteristics: Source Selection and Description**

### **Overall Description of Data Sources**

### The datasets integrated into the VDSS\_Final\_Data.R script collectively focus on the characteristics of ZIP codes across Virginia, with the primary objective of facilitating the analysis of ZIP code types and their relationship to the proximity of Local Departments of Social Services (LDSS) and other related research outcomes. Each dataset contributes unique information that, when combined, offers a comprehensive view of the geographic, demographic, and service-related aspects of ZIP codes.

### **Exhaustive List of ZIP-County Pairings**

A key aspect of the VDSS\_Final\_Data.R script is ensuring a thorough and consistent tracking of ZIP-county pairings across each year-quarter in Virginia. This approach is crucial for maintaining a reliable longitudinal analysis that reflects both current and historical changes in ZIP codes.

#### **Data Sources**

1. **USPS HUD ZIP County Crosswalk**:
   * **Purpose**: Tracks ZIP-county pairings on a quarterly basis from 2010 to 2024, offering a detailed view of how these relationships evolve over time.
   * **Method**: Unique ZIP-county pairings are extracted from each year-quarter and applied uniformly across all quarters. This standardization ensures that every quarter has approximately 1,170 unique ZIP codes and 1,700 unique ZIP-county combinations, providing consistency in the dataset.
2. **Bailey & Helmuth ZIP Code Changes**:
   * **Purpose**: Captures ZIP code removals reported in the USPS bi-weekly bulletin board, focusing on changes from 2010 to 2020.
   * **Method**: ZIP codes identified for removal during this period are excluded from the final dataset based on their effective removal dates. This ensures that the dataset accurately reflects historical ZIP code changes.

#### **Why These Data Sources Were Chosen**

These sources were selected to create the most comprehensive and accurate list of ZIP codes and their associated counties over time. The USPS HUD ZIP County Crosswalk provides a reliable foundation for tracking ZIP-county pairings, while the Bailey & Helmuth data ensures that historical changes are incorporated, making the dataset particularly robust for longitudinal studies.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### **ZIP Code Type Assignments**

Another critical decision in constructing the VDSS\_Final\_Data.R script was assigning a ZIP code type to each entry. ZIP codes can fall into one of three categories: Standard, Unique, or PO Box. These designations were provided by zipdatamaps.com, a resource that is updated on a weekly basis.

#### **ZIP Code Types Explained**

* **Standard ZIP Code**: This is the most common type of ZIP code, used for general residential and business addresses. It typically covers a specific geographic area and includes multiple delivery points.
* **Unique ZIP Code**: This type is assigned to a single high-volume address, such as a large corporation, government agency, or university. Unique ZIP codes are specific to that address and do not serve the general public.
* **PO Box ZIP Code**: These ZIP codes are associated with post office boxes at a particular location. They are used exclusively for PO Box addresses and do not correspond to physical street addresses.

#### **Pros and Cons of Using zipdatamaps.com**

* **Upside**: The primary advantage of using zipdatamaps.com is its exhaustive coverage and the frequency of updates. Weekly updates ensure that the ZIP code types are current, and individual checks of the designations have confirmed their accuracy in most cases.
* **Downside**: However, there are some limitations. Not all ZIP codes in the dataset may have a TYPE designation, which could introduce gaps in the analysis. Additionally, the source does not provide transparency about how these designations are determined, which raises questions about their sourcing and accuracy.

#### **Considerations for Future Improvements**

To enhance the reliability of the ZIP code type assignments, we could cross-reference the designations from zipdatamaps.com with another source or seek out a more authoritative database entirely. Doing so would address the current limitations and ensure that every ZIP code in the dataset is accurately categorized.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### **Residential vs. Non-Residential Distinction**

In determining whether a ZIP code is classified as residential or non-residential, we relied on the RES\_RATIO from the HUD USPS ZIP County dataset. This metric refers to the proportion of a ZIP-county pairing that comprises the total number of residential addresses within that ZIP code.

#### **Methodology**

* **RES\_RATIO**: If any ZIP-county pairing had a RES\_RATIO greater than 0, we classified that entire ZIP code as residential. This classification was applied uniformly across all years, ensuring that once a ZIP code was deemed residential, it remained so throughout the dataset. While ZIP codes may theoretically shift from residential to non-residential over time, such occurrences are rare and were not observed in our analysis.

#### **Alternative Methods**

There are other ways to assess whether a ZIP code is residential, such as using IRS data, which reflects tax returns filed at the ZIP code level. The IRS data could potentially offer a different perspective on ZIP code residency, particularly by highlighting economic activity within the area.

#### **Rationale for HUD USPS Data**

We opted to use the HUD USPS data for several reasons:

* **Consistency**: By sticking to a single source for multiple attributes (e.g., ZIP-county pairings, residential status), we kept the data sources minimal, reducing the complexity of data integration.
* **Indicator Reliability**: The presence of residential addresses is a straightforward and reliable indicator of a ZIP code's residential nature. Using the RES\_RATIO from the HUD USPS dataset allowed us to make clear and consistent distinctions without the need to introduce additional data sources.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### **ZCTA Crosswalk Integration**

To associate ZIP codes with their corresponding ZIP Code Tabulation Areas (ZCTAs), we used a crosswalk provided by the Health Resources and Services Administration (HRSA). The HRSA crosswalk offers a reliable method for mapping ZIP codes to ZCTAs, which is crucial for geographic and demographic analysis.

#### **Data Source**

* **HRSA Crosswalk**: The ZCTA crosswalk was obtained from HRSA, an agency under the U.S. Department of Health and Human Services. You can access the crosswalk here: [HRSA ZIP Code to ZCTA Crosswalk](https://data.hrsa.gov/DataDownload/GeoCareNavigator/ZIP%20Code%20to%20ZCTA%20Crosswalk.xlsx).

#### **Methodology**

* **Exact Matches and Spatial Joins**: HRSA used a combination of exact ZIP to ZCTA matches and spatial joins for ZIP codes that did not have a direct match. The spatial joins were performed using ZIP code points and ZCTA boundaries from the U.S. Census Bureau's TIGER/Line shapefiles. The process involved:
  + **Exact Matches**: Where available, direct ZIP to ZCTA matches were made.
  + **Spatial Joins**: For the remaining ZIP codes, HRSA utilized spatial joins, mapping ZIP code points using [ArcGIS](https://www.arcgis.com/home/item.html?id=dc123f738bf846779c49db6472f82a4b) and ZCTA boundaries from the U.S. Census Bureau's [TIGER/Line shapefiles](https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html).

#### **Validation and Alternatives**

* **Manual Review**: The HRSA crosswalk was the only method to pass a manual review of 14 ZIP codes, where a similar spatial join was conducted using ZIP coordinates from zipdatamaps.com and the TIGER shapefile for ZCTAs. This review confirmed the accuracy of the HRSA method over other available options.
* **Alternative Methods**: One alternative considered was a ZIP to ZCTA crosswalk available on GitHub, created by Joe Germuska. However, this method was found to be less accurate compared to HRSA’s approach, which consistently yielded more reliable results in our testing.

#### **Conclusion**

The HRSA crosswalk was selected due to its robust methodology and validated accuracy, making it the best choice for associating ZIP codes with ZCTAs in our dataset. The crosswalk’s ability to handle both exact matches and spatial joins ensures that our analysis is grounded in accurate geographic data, which is essential for reliable policy and demographic studies.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### **Bad ZIP-County Data**

The bad\_zipcounty data, provided by Neil Cholli, marks two critical figures for each ZIP code in Virginia:

* **Number of SNAP Recipients**: This figure represents the total number of SNAP recipients within each ZIP code.
* **Error Rate**: This percentage indicates the proportion of SNAP recipients in each ZIP code who were incorrectly assigned to the wrong Local Department of Social Services (LDSS).

This data was integrated into the dataset to track potential mismatches in SNAP recipient assignments across Virginia.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### **Treatment Data**

The Treatment Data, also provided by Neil Cholli, assigns a treatment status to each ZIP-county pair based on the associated LDSS office. This treatment status is determined by comparing the Haversine distance between the midpoint of each ZIP-county pair and the coordinates of the LDSS office.

* **Treatment Status**: The treatment status is categorized based on whether the Haversine distance is greater than or less than the global median distance (G50) for all ZIP-county pairs. This classification helps in analyzing the spatial relationship between ZIP codes and their corresponding LDSS offices, offering insights into the geographic distribution of services.

This data was used to add a treatment attribute to each ZIP-county pair, allowing for more detailed spatial and policy analysis within the dataset.