**Memorandum: Script Direcotry**

**To:** Project Contributors  
**From:** Dylan Craig  
**Date Created:** 12/29/2024  
**Subject:** Script Directory

This project is organized as an R Project, ensuring all scripts and related files are structured for reproducibility and ease of use. The folder structure reflects key components of the project, with clearly defined directories for data, scripts, and outputs. All scripts are designed to run within this framework, requiring no modifications if the R Project is used as intended.

Below is the breakdown of the project folder structure:

* **Raw Data**: Contains all unprocessed datasets used as inputs for analyses.
* **Data Outputs**: Includes cleaned and processed datasets as well as analysis results.
* **Plots**: Stores visualizations and graphical outputs.
* **Scripts**: Houses R scripts for data cleaning, analysis, and visualization. Includes write-ups explaining data management/cleaning practices as well as data analysis as necessary.
* **Write-Ups**: Holds reports, memos, and other project-related documents.

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

# Theme: VDSS\_Zip\_Office\_Distances

**Script Name: *VDSS\_Zip\_GGMAP\_Geocoding\_All.R***

**Description**:  
This script geocodes unique ZIP codes from a merged dataset of VDSS offices and USPS ZIP-FIPS mappings. The resulting dataset includes latitude and longitude for each ZIP code and is saved as an Excel file.

**Inputs**:

* *"Raw Data/VDSS\_Zip\_Office\_Distances/VDSS\_Offices\_All.xlsx"*
* *"Raw Data/VDSS\_Zip\_Office\_Distances/USPS\_Zip\_County.xlsx"*

**Outputs**:

* *"Data Outputs/VDSS\_Zip\_Office\_Distances/VDSS\_Office\_Zip\_GeoCode\_All.xlsx"*

**Plots**:  
None.

**Script Name: *VDSS\_Zip\_GGMAP\_Geocoding\_2012.R***

**Description**:  
This script geocodes unique ZIP codes from a merged dataset of VDSS offices (2012 data) and USPS ZIP-FIPS mappings. The resulting dataset includes latitude and longitude for each ZIP code and is saved as an Excel file.

**Inputs**:

* *"Raw Data/VDSS\_Zip\_Office\_Distances/VDSS\_Offices\_2012.xlsx"*
* *"Raw Data/VDSS\_Zip\_Office\_Distances/USPS\_Zip\_County.xlsx"*

**Outputs**:

* *"Data Outputs/VDSS\_Zip\_Office\_Distances/VDSS\_Office\_Zip\_GeoCode\_2012.xlsx"*

**Plots**:  
None.

**Script Name:**  
*VDSS\_Zip\_Code\_Changes\_By\_Year.R*

**Description:**  
This script analyzes and visualizes changes in 5-digit ZIP codes by locality over time. It identifies the first instance of ZIP code changes, aggregates the number of changes by year, and generates a bar plot. The output is saved as a PDF.

**Inputs:**

* *"Raw Data/VDSS\_Zip\_Office\_Distances/VDSS\_Offices\_All.xlsx"*

**Outputs:**  
None

**Plots:**

* *"Plots/VDSS\_Zip\_Office\_Distances/VDSS\_Zip\_Code\_Changes\_By\_Year.pdf"*

**Script Name:**  
*VDSS\_Distances\_All.R*

**Description:**  
This script calculates haversine, driving, and transit distances and times between LDSS offices and ZIP codes using geospatial data, converts them to miles and minutes, and outputs an updated Excel file.

**Inputs:**

* *"Data Outputs/VDSS\_Zip\_Office\_Distances/VDSS\_Office\_Zip\_GeoCode\_All\_Corrected.xlsx"*

**Outputs:**

* *"Data Outputs/VDSS\_Zip\_Office\_Distances/VDSS\_Office\_Zip\_GeoCode\_All\_Distances.xlsx"*

**Plots:**  
None

**Script Name:**  
*VDSS\_Distances\_2012.R*

**Description:**  
This script calculates haversine, driving, and transit distances and times between VDSS office midpoints and ZIP code midpoints for the 2012 dataset, converts them to miles and minutes, and saves the updated Excel file.

**Inputs:**

* *"Data Outputs/VDSS\_Zip\_Office\_Distances/VDSS\_Office\_Zip\_GeoCode\_2012.xlsx"*

**Outputs:**

* *"Data Outputs/VDSS\_Zip\_Office\_Distances/VDSS\_Office\_Zip\_GeoCode\_2012.xlsx"*

**Plots:**  
None

**Script Name:**  
*VDSS\_Distances\_Analysis\_All.R*

**Description:**  
This script analyzes geospatial distance and time data from LDSS offices to ZIP codes. It includes data transformations, summary statistics, histogram visualizations, missing data analysis, and correlation coefficient calculations. Outputs are saved as Word documents and PDF plots.

**Inputs:**

* *"Data Outputs/VDSS\_Zip\_Office\_Distances/VDSS\_Office\_Zip\_GeoCode\_All\_Distances.xlsx"*

**Outputs:**  
None

**Plots:**

* *"Plots/VDSS\_Zip\_Office\_Distances/Distance\_Summary\_Statistics\_Miles.docx"*
* *"Plots/VDSS\_Zip\_Office\_Distances/Missing\_Data\_Analysis.docx"*
* *"Plots/VDSS\_Zip\_Office\_Distances/Correlation\_Matrix.docx"*
* *"Plots/VDSS\_Zip\_Office\_Distances/Haversine Distance Distribution.pdf"*
* *"Plots/VDSS\_Zip\_Office\_Distances/Driving Distance Distribution.pdf"*
* *"Plots/VDSS\_Zip\_Office\_Distances/Transit Distance Distribution.pdf"*
* *"Plots/VDSS\_Zip\_Office\_Distances/Driving Time Distribution.pdf"*
* *"Plots/VDSS\_Zip\_Office\_Distances/Transit Time Distribution.pdf"*

# Theme: VDSS\_Zip\_Bad\_ZipCounty\_Rate\_Heat\_Map

**Script Name:** *VDSS\_Zip\_Bad\_ZipCounty\_Rate\_Heat\_Map.R*

**Description:**  
This script creates an interactive heat map showing the "Bad ZipCounty Rate" for zip codes in Virginia. The map overlays county boundaries and is exported as HTML, PDF, and PNG files.

**Inputs:**

* *"Raw Data/VDSS\_Bad\_ZipCounty\_Rate\_Heat\_Map/bad\_zipcounty\_rate.dta"*

**Outputs:**  
None

**Plots:**

* *"Plots/VDSS\_Bad\_ZipCounty\_Rate\_Heat\_Map/VDSS\_Bad\_ZipCounty\_Rate\_Heat\_Map.html"*
* *"Plots/VDSS\_Bad\_ZipCounty\_Rate\_Heat\_Map/VDSS\_Bad\_ZipCounty\_Rate\_Heat\_Map.pdf"*
* *"Plots/VDSS\_Bad\_ZipCounty\_Rate\_Heat\_Map/VDSS\_Bad\_ZipCounty\_Rate\_Heat\_Map.png"*

# Theme: VDSS\_Zip\_Treat\_Map

**Script Name:**  
*VDSS\_Zip\_Treat\_Map.R*

**Description:**  
This script creates a thematic map displaying ZIP codes in Virginia categorized by treatment status based on the treat\_g50 variable. The map is exported in PNG, PDF, and HTML formats.

**Inputs:**

* *"Raw Data/VDSS\_Zip\_Treat\_Map/zip\_treat.dta"*

**Outputs:**  
None

**Plots:**

* *"Plots/VDSS\_Zip\_Treat\_Map/Zip\_Treat\_Map.png"*
* *"Plots/VDSS\_Zip\_Treat\_Map/Zip\_Treat\_Map.pdf"*
* *"Plots/VDSS\_Zip\_Treat\_Map/Zip\_Treat\_Map.html"*

# Theme: VDSS\_Zip\_Type

**Script Name:**  
*VDSS\_Zip\_Type.R*

**Description:**  
This script merges USPS ZIP-County data with VDSS ZIP data types. It identifies ZIP codes that failed to merge and exports both the merged dataset and the failed merge data.

**Inputs:**

* *"Raw Data/VDSS\_Zip\_Type/USPS\_Zip\_County.xlsx"*
* *"Raw Data/VDSS\_Zip\_Type/VDSS\_ZIPDATAMAPS\_TYPE.xlsx"*

**Outputs:**

* *"Data Outputs/VDSS\_Zip\_Type/VDSS\_Zip\_Type.xlsx"*
* *"Data Outputs/VDSS\_Zip\_Type/VDSS\_Zip\_Type\_No\_Type.xlsx"*

**Plots:**  
None

# Theme: VDSS\_Zip\_Residential\_Status

**Script Name:**  
*VDSS\_Zip\_Residential\_Status.R*

**Description:**  
This script merges USPS ZIP-County data with IRS ZIP code data and determines the residential status of ZIP codes based on population values. The output is saved as an Excel file.

**Inputs:**

* *"Raw Data/VDSS\_Zip\_Residential\_Status/USPS\_Zip\_County.xlsx"*
* *"Raw Data/VDSS\_Zip\_Residential\_Status/zip\_code\_IRS\_2020.xls"*

**Outputs:**

* *"Data Outputs/VDSS\_Zip\_Residential\_Status/Zip\_Residential\_Status.xlsx"*

**Plots:**  
None

# Theme: VDSS\_Zip\_Border\_Status

**Script Name:**  
*VDSS\_Zip\_Border\_Status.R*

**Description:**  
This script identifies neighboring counties for Virginia based on geographic boundaries. The outputs include neighboring counties and their FIPS codes, saved as Excel files.

**Inputs:**  
None

**Outputs:**

* *"Data Outputs/VDSS\_Zip\_Border\_Status/virginia\_bordering\_counties.xlsx"*
* *"Data Outputs/VDSS\_Zip\_Border\_Status/virginia\_bordering\_counties\_fips.xlsx"*

**Plots:**  
None

**Script Name:**  
*VDSS\_Zip\_Border\_Population\_Filtered.R*

**Description:**  
This script processes ZIP code, FIPS, and bordering county data for Virginia, filtering and transforming the data into a structured format. The output is a filtered dataset with relevant FIPS and bordering FIPS codes saved to an Excel file.

**Inputs:**

* *"Raw Data/VDSS\_Zip\_Border\_Status/usps\_zip\_county.xlsx"*
* *"Data Outputs/VDSS\_Zip\_Border\_Status/virginia\_bordering\_counties\_fips.xlsx"*
* *"Raw Data/VDSS\_Zip\_Border\_Status/ZIP\_COUNTY\_122012.xlsx"*

**Outputs:**

* *"Data Outputs/VDSS\_Zip\_Border\_Status/VDSS\_Zip\_FIPS\_Border\_Population\_Filtered.xlsx"*

**Plots:**  
None

**Script Name:**  
*VDSS\_Zip\_FIPS\_Border\_Status\_Final\_Designation.R*

**Description:**  
This script classifies ZIP codes in Virginia into categories (Interior, Bordering, Overlapping, or No Shapefile) based on their intersection with county boundaries. The final classification is saved to an Excel file.

**Inputs:**

* *"Data Outputs/VDSS\_Zip\_Border\_Status/VDSS\_Zip\_FIPS\_Border\_Population\_Filtered.xlsx"*

**Outputs:**

* *"Data Outputs/VDSS\_Zip\_Border\_Status/VDSS\_Zip\_FIPS\_Border\_Status\_Final\_Designation.xlsx"*

**Plots:**  
None

# Theme: VDSS\_Zip\_Border\_Status\_Type\_Bad\_ZipCounty\_Rate

**Script Name:**  
*VDSS\_Zip\_Border\_Status\_Type\_Bad\_Zipcounty\_Rate.R*

**Description:**  
This script merges datasets related to ZIP types, border statuses, and bad ZIP-County rates. It calculates summary statistics and performs t-tests by BORDER\_STATUS and TYPE, saving the results to an Excel file.

**Inputs:**

* *"Data Outputs/VDSS\_Zip\_Type/VDSS\_Zip\_Type.xlsx"*
* *"Data Outputs/VDSS\_Zip\_Border\_Status/VDSS\_Zip\_FIPS\_Border\_Status\_Final\_Designation.xlsx"*
* *"Raw Data/VDSS\_Zip\_Border\_Status\_Type\_Bad\_ZipCounty\_Rate/bad\_zipcounty\_rate.dta"*

**Outputs:**

* *"Data Outputs/VDSS\_Zip\_Border\_Status\_Type\_Bad\_ZipCounty\_Rate/VDSS\_Summary\_Statistics\_and\_T\_Tests.xlsx"*

**Plots:**  
None

# Theme: VDSS\_Zip\_ZCTA\_Crosswalk

**Script Name:**  
*VDSS\_Zip\_ZCTA\_Crosswalk\_HRSA.R*

**Description:**  
This script processes HRSA ZCTA data, filtering for Virginia ZIP codes, selecting necessary columns, and renaming them for clarity. The cleaned data is saved to an output file.

**Inputs:**

* *"Raw Data/VDSS\_Zip\_ZCTA\_Crosswalk/HRSA\_ZCTA.xlsx"*

**Outputs:**

* *"Data Outputs/VDSS\_Zip\_ZCTA\_Crosswalk/HRSA\_ZCTA.xlsx"*

**Plots:**  
None

# Theme: VDSS\_Historical\_Zip\_Type\_Changes

**Script Name:**  
*VDSS\_HUD\_USPS\_2011\_2012\_Type\_Comparison.R*

**Description:**  
This script analyzes ZIP data for Q4 2011 and Q1 2012, focusing on TYPE and NUMBER\_TYPES distributions. The results are exported to an Excel file.

**Inputs:**

* *"Data Outputs/VDSS\_Historical\_Zip\_Type\_Changes/VA\_HUD\_USPS\_Zip\_Data\_Collapsed.xlsx"*

**Outputs:**

* *"Data Outputs/VDSS\_Historical\_Zip\_Type\_Changes/VA\_HUD\_USPS\_2011\_2012\_TYPE\_COMPARISON.xlsx"*

**Plots:**  
None

**Script Name:**  
*VDSS\_HUD\_USPS\_Historical\_Zip\_Type\_Missing\_Analysis.R*

**Description:**  
This script analyzes ZIP presence across year-quarters and evaluates missing statuses, including breakdowns by NUMBER\_TYPES and TYPE. The results are saved as Excel files.

**Inputs:**

* *"Data Outputs/VDSS\_Historical\_Zip\_Type\_Changes/VA\_HUD\_USPS\_Zip\_Data\_Collapsed.xlsx"*

**Outputs:**

* *"Data Outputs/VDSS\_Historical\_Zip\_Type\_Changes/VA\_HUD\_USPS\_Zip\_Status.xlsx"*
* *"Data Outputs/VDSS\_Historical\_Zip\_Type\_Changes/VA\_HUD\_USPS\_Zip\_Percent\_Missing\_Present.xlsx"*
* *"Data Outputs/VDSS\_Historical\_Zip\_Type\_Changes/VA\_HUD\_USPS\_Missing\_Analysis\_By\_Number\_Types.xlsx"*
* *"Data Outputs/VDSS\_Historical\_Zip\_Type\_Changes/VA\_HUD\_USPS\_Missing\_Type\_Analysis\_Single.xlsx"*

**Plots:**  
None

**Script Name:**  
*VDSS\_HUD\_USPS\_Historical\_Zip\_Type\_ZipDataMaps\_Comparison.R*

**Description:**  
This script collapses VA HUD USPS ZIP data by ZIP and YEAR\_QUARTER, analyzes type classifications, and compares the results with ZIP data from VDSS\_Zip\_Type.xlsx. It generates summary tables for PO Box and Unique type conversions.

**Inputs:**

* *"Data Outputs/VDSS\_Historical\_Zip\_Type\_Changes/VA\_HUD\_USPS\_Zip\_Data\_All\_Years.xlsx"*
* *"Data Outputs/VDSS\_Zip\_Type/VDSS\_Zip\_Type.xlsx"*

**Outputs:**

* *"Data Outputs/VDSS\_Historical\_Zip\_Type\_Changes/VA\_HUD\_USPS\_Zip\_Data\_Collapsed.xlsx"*
* *"Data Outputs/VDSS\_Historical\_Zip\_Type\_Changes/Comparison\_Table\_PO\_Box.xlsx"*
* *"Data Outputs/VDSS\_Historical\_Zip\_Type\_Changes/VA\_HUD\_USPS\_ZipDataMaps\_Comparison.xlsx"*

**Plots:**  
None

# Theme: VDSS\_Historical\_Zip\_Code\_Changes

**Script Name:**  
*VDSS\_Historical\_Zip\_Code\_Changes\_Bailey\_Helmuth\_Analysis.R*

**Description:**  
This script analyzes historical ZIP type changes for a specific dataset related to Bailey Helmuth ZIP codes. It examines attributes such as NUMBER\_TYPES, TYPE, presence across YEAR\_QUARTER, and availability of post-2010 data. The results are summarized for both old and new ZIP codes.

**Inputs:**

* *"Data Outputs/VDSS\_Historical\_Zip\_Type\_Changes/VA\_HUD\_USPS\_Zip\_Data\_Collapsed.xlsx"*

**Outputs:**

* *"Data Outputs/VDSS\_Historical\_Zip\_Code\_Changes/Bailey\_Helmuth\_Zips\_Analysis.xlsx"*

**Plots:**  
None

# Theme: VDSS\_ZCTA\_Office\_Distances

**Script Name:**  
*VDSS\_Distances\_All\_ZCTA.R*

**Description:**  
This script calculates distances between ZCTA midpoints and LDSS midpoints using datasets from ZCTA, ZIP-FIPS, and VDSS offices for all years. It merges spatial and tabular data, computes distances, and identifies rows with missing values for review.

**Inputs:**

* *"Data Outputs/VDSS\_Zip\_ZCTA\_Crosswalk/HRSA\_ZCTA.xlsx"*
* *"Raw Data/VDSS\_Zip\_Office\_Distances/VDSS\_Offices\_All.xlsx"*
* *"Raw Data/VDSS\_Zip\_Office\_Distances/USPS\_Zip\_County.xlsx"*

**Outputs:**

* *"Data Outputs/VDSS\_ZCTA\_Office\_Distances/VDSS\_Office\_ZCTA\_GeoCode\_All\_Distances.xlsx"*

**Plots:**  
None

**Script Name:**  
*VDSS\_Distances\_2012\_ZCTA.R*

**Description:**  
This script calculates distances between ZCTA midpoints and LDSS midpoints for the year 2012. It uses spatial data from ZCTA, ZIP-FIPS, and VDSS office datasets to compute the distances and merges them into a structured format for analysis.

**Inputs:**

* *"Data Outputs/VDSS\_Zip\_ZCTA\_Crosswalk/HRSA\_ZCTA.xlsx"*
* *"Raw Data/VDSS\_Zip\_Office\_Distances/VDSS\_Offices\_2012.xlsx"*
* *"Raw Data/VDSS\_Zip\_Office\_Distances/USPS\_Zip\_County.xlsx"*

**Outputs:**

* *"Data Outputs/VDSS\_ZCTA\_Office\_Distances/VDSS\_Office\_ZCTA\_GeoCode\_2012\_Distances.xlsx"*

**Plots:**  
None

# Theme: VDSS\_ZCTA\_ACS\_Characteristics

**Script Name:**

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

**Description:**  
This script processes and aggregates ACS data from multiple sources into a unified dataset. It reads individual datasets, renames columns with specific prefixes for clarity, combines them into a single dataset by ZCTA and YEAR, and exports the results as both CSV and Excel files.

**Inputs:**

* *"Raw Data/VDSS\_ZCTA\_ACS\_Characteristics/B01003\_Total\_Population"*
* *"Raw Data/VDSS\_ZCTA\_ACS\_Characteristics/B01001\_Sex\_By\_Age"*
* *"Raw Data/VDSS\_ZCTA\_ACS\_Characteristics/B02001\_Race\_By\_Sex"*
* *"Raw Data/VDSS\_ZCTA\_ACS\_Characteristics/B03002\_Hispanic\_Latino\_By\_Race"*
* *"Raw Data/VDSS\_ZCTA\_ACS\_Characteristics/B17001\_Poverty\_Last\_12\_Months\_By\_Sex\_By\_Age"*
* *"Raw Data/VDSS\_ZCTA\_ACS\_Characteristics/B19013\_Median\_Income"*
* *"Raw Data/VDSS\_ZCTA\_ACS\_Characteristics/B15001\_Sex\_By\_Age\_By\_Educational\_Attainment\_Over\_18"*
* *"Raw Data/VDSS\_ZCTA\_ACS\_Characteristics/B12002\_Marital\_Status\_By\_Age\_Over\_15"*
* *"Raw Data/VDSS\_ZCTA\_ACS\_Characteristics/Selected\_Economic\_Characteristics"*

**Outputs:**

* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Population\_Data.csv"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Population\_Data.xlsx"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Sex\_By\_Age\_Data.csv"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Sex\_By\_Age\_Data.xlsx"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Race\_By\_Sex\_Data.csv"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Race\_By\_Sex\_Data.xlsx"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Hispanic\_Latino\_By\_Race\_Data.csv"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Hispanic\_Latino\_By\_Race\_Data.xlsx"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Poverty\_Status\_Data.csv"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Poverty\_Status\_Data.xlsx"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Median\_Income\_Data.csv"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Median\_Income\_Data.xlsx"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Education\_Attainment\_Data.csv"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Education\_Attainment\_Data.xlsx"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Marital\_Status\_Data.csv"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Marital\_Status\_Data.xlsx"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Employment\_Status\_Data.csv"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/Employment\_Status\_Data.xlsx"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/VDSS\_ZCTA\_ACS\_Characteristics\_Aggregated.csv"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/VDSS\_ZCTA\_ACS\_Characteristics\_Aggregated.xlsx"*

**Plots:**

* None

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

**Description:**  
This script processes and aggregates ACS data from multiple sources into a unified dataset. It reads from an existing aggregated file, performs additional transformations (e.g., employment industry stacking, population checks, calculated rates, weighted averages), and exports the final cleaned data in both CSV and Excel formats.

**Inputs:**

* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/VDSS\_ZCTA\_ACS\_Characteristics\_Aggregated.csv"*

**Outputs:**

* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/VDSS\_ZCTA\_ACS\_Characteristics\_Cleaned.csv"*
* *"Data Outputs/VDSS\_ZCTA\_ACS\_Characteristics/VDSS\_ZCTA\_ACS\_Characteristics\_Cleaned.xlsx"*

**Plots:**  
None

# Theme: VDSS\_Final\_Data

**Script Name:**  
*VDSS\_Final\_Data.R*

**Description:**  
This script processes and integrates multiple datasets to create a final dataset that includes ZIP classification, ZCTA crosswalk, treatment, and quality data. The script cleanses, merges, and validates the data, removes deleted ZIPs, classifies ZIPs as residential or non-residential, adds ZCTA crosswalks, and integrates treatment and quality data. Summary statistics and unique counts by time are also generated.

**Inputs:**

* *"Raw Data/VDSS\_Final\_Data/ZIP\_COUNTY/ZIP\_COUNTY\_*.xlsx"\*
* *"Raw Data/VDSS\_Final\_Data/Bailey\_Helmuth\_Zip\_Changes.xlsx"*
* *"Data Outputs/VDSS\_Zip\_Type/VDSS\_Zip\_Type.xlsx"*
* *"Data Outputs/VDSS\_Zip\_ZCTA\_Crosswalk/HRSA\_ZCTA.xlsx"*
* *"Raw Data/VDSS\_Final\_Data/bad\_zipcounty\_rate.dta"*
* *"Raw Data/VDSS\_Final\_Data/zip\_treat.dta"*

**Outputs:**

* *"Data Outputs/VDSS\_Final\_Data/Final\_Data.xlsx"*

**Plots:**  
None