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

**Memorandum**

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

To: Derek Wu, Neil Cholli

From: Dylan Craig

Date Created: 12/26/2024

Subject: Data Management Memo on Bad ZipCounty Rates by Zip Code

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Purpose:**

**\*\*Read READ\_ME as well as Script Directory for technical breakdown including file names**

**Step 1: Import .dta file on Bad ZipCounty Rates by Zip Code** (Year Unknown)

Neil Cholli and Derek Wu collected and cleaned zip code level data, assumedly provided by the VDSS, which indicated the percent of mismatches between residents of a zip code and the VDSS local office they received services from. VDSS offices serve counties/cities but not every resident ends up receiving services from the VDSS office that serves their county/city.

**Step 2: Import ZCTA shapefile using TIGRIS**

Using R, ZCTA shapefiles were downloaded from the Census Bureau TIGER shapefiles. ZCTAs approximate Zip Codes and most share the same 5-digit numeric indicator. By not using a zip code shapefile, however, this can introduce mismatches as well as missing data into our data cleaning.

**Step 3: Calculate Percentile Quantiles**

Since mismatch rates are clustered between 0.01 and 0.1, we set the categories up as follows:

* 0.00 to 0.01 (i.e., 0% to 1% mismatch rate)
* 0.01 to 0.018
* 0.018 to 0.1
* 0.1 to 0.4
* 0.4 to 1

There should be some email correspondence better highlighting our reasoning for this layout, but it’s based mostly on creating a more normal distribution of the data rather than a heavily right-skewed distribution.

**Step 4: Generate an interactive heat map**

After merging the .dta file containing mismatch rates by zip codes and the ZCTA’s listed in the shapefile dataframe, an interactive heatmap was generated. See data analysis memo for breakdown. The plot was saved in .html, .pdf, and .png to plots folder.