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

**Memorandum**

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

To: Derek Wu, Neil Cholli

From: Dylan Craig

Date Created: 12/20/2024

Subject: Data Management Memo on Office Distances (CommonHelp VA)

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

**Purpose:** To demonstrate the data cleaning and management process for deriving distances between zip codes residents live in and the VDSS local offices they may be required to go to based on their county of residence.

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

**Step 1: Addresses of VDSS Local Offices** (2000-2024)

Using historical screenshots (via the WayBack Machine) of [VDSS office locations](https://www.dss.virginia.gov/localagency/index.cgi) for each county, a dataset was created with county/office address/date as the unit of observation. This was split into two separate datasets, one for VDSS Offices across all years and one for VDSS offices in October of 2012.Later work was completed to create start and end dates as well, with the initial dates recorded as “start dates” given that they were the first instance of an address appearing on the wayback machine.

**Step 2: Assigning Zip Codes to VDSS Local Offices**

Because we were interested, in part, in the relationship between residents’ distance to their local VDSS and their propensity to use application assistance such as CommonHelp or the Enterprise Call Center, we determined which zip codes were at least partially served by a given VDSS office. This was completed using the [HUD USPS ZIP CODE CROSSWALK FILES](https://www.huduser.gov/portal/datasets/usps_crosswalk.html) which in part link zip codes to all associated FIPS codes.

**Step 3: Calculating Distance Between VDSS Local Offices and Midpoints of Zip Codes Served**

*VDSS Local Office Coordinates*:

In order to obtain our main measurement of interest – how far away is a given resident from their local VDSS – we first obtained the latitude/longitude measurement of each VDSS office. This could only be completed for VDSS addresses in which there was a physical address (typically occurring by the late 2000s on the VDSS office website) and was done manually using Google Maps. This was completed for both 2012 data and all years.

*VDSS Zip Code Coordinates:*

Because of the vast number of zip codes, Google API was used in R to calculate the midpoint of each zip code. This process was largely accurate (verified by spot-checking) but had some serious outliers. These outliers were corrected manually in Google Maps.

*VDSS Office to Zip Code Distances:*

Using a formula in R, Haversine distances measuring the direct distance (“as the crow flies”) between each office and the midpoints of zip codes it serves were calculated. These distances stood up to manual spot-checking in Google Maps.

Then, using Google API, other distances including driving distance in miles, transit distance in miles, driving time in minutes, and transit time in minutes were carried out to offer alternative measurements of distance that may more accurately reflect zip code residents’ real proximity to their local VDSS office. See the data analysis memo for a more thorough breakdown of this exercise.

A .dta file was created designating “treated” zip-office pairs, or those that have an above median distance to their VDSS office compared to other zip-office pairs.