Exploratory Analysis of the use of ArcGIS Online to Expand Data Analysis and Accessibility for Multidisciplinary Cancer Research

Jennifer Ciminelli[1], Katherine Y. Tossas[1,2]

[1]Massey Comprehensive Cancer Center, Virginia Commonwealth University, Richmond, VA;

[2] Department of Health Behavior and Policy, Virginia Commonwealth University School of Public Health, Richmond, VA

BACKGROUND

- Community profile data provide important information in the understanding of catchment area populations, but often data is collected at varying areal units.

- Properly allocating data across different geographies or areal units can be a barrier for ensuring proper community characteristic data inclusion in cancer-related research, resulting in generalized or under-reporting of community characteristics.

- To address this, we tested three methodologies to apportion, allocate and aggregate U.S. Census Tract population data to U.S. Zip Code Tabulation Boundary Areas (ZCTA) using a GIS approach.

METHODS

- Esri GeoEnrichment was used to calculate the population for each ZCTA using data apportionment, a geographic retrieval method that estimates population using a weighted block proportion. (**Fig. 1**).

- HUD USPS Zip Code Crosswalk files were used to allocate the residential proportion of Census tracts to the geographically associated ZCTA, then summed to calculate total population for the ZCTA. (**Fig. 2**).

- U.S. Census tracts were assigned to the nearest zip code, each tract population associated with the ZCTA was aggregated, and the total population associated with the ZCTA calculated. (**Fig. 3**).

Figure 1. ZCTA symbolized by total population (2020 U.S. Census P1). Esri GeoEnrichment used to calculate values.

KEY FINDING

Esri GeoEnrichment provides an accessible, accurate data apportion tool for calculating population across varying geographies.

U.S. Census 2020 Total Population
- 0 to 11,000
- >11,000 to 22,000
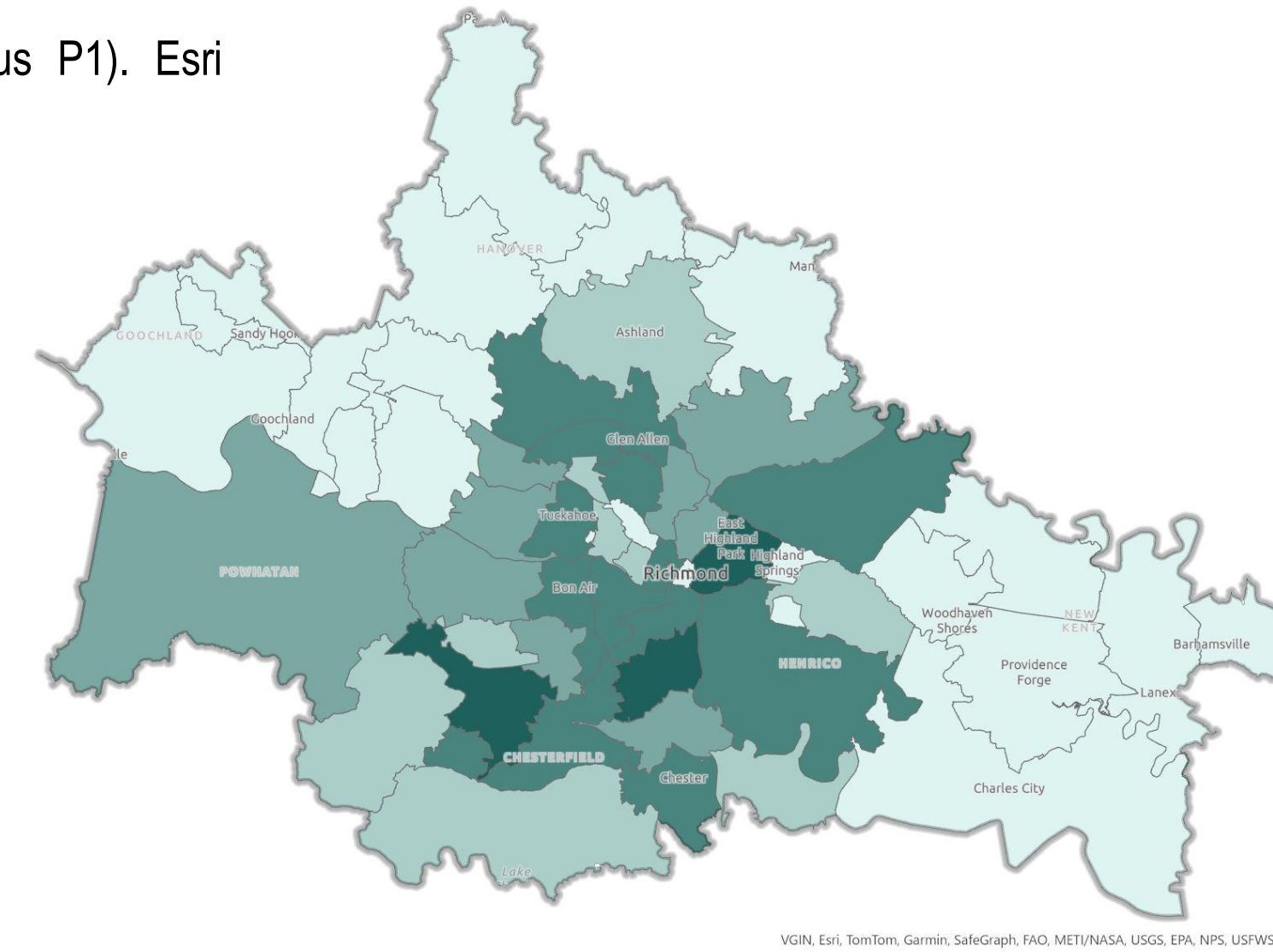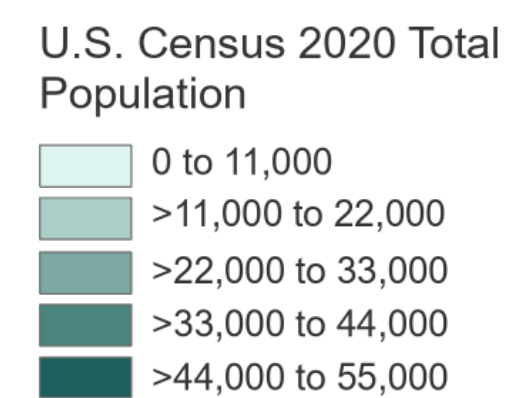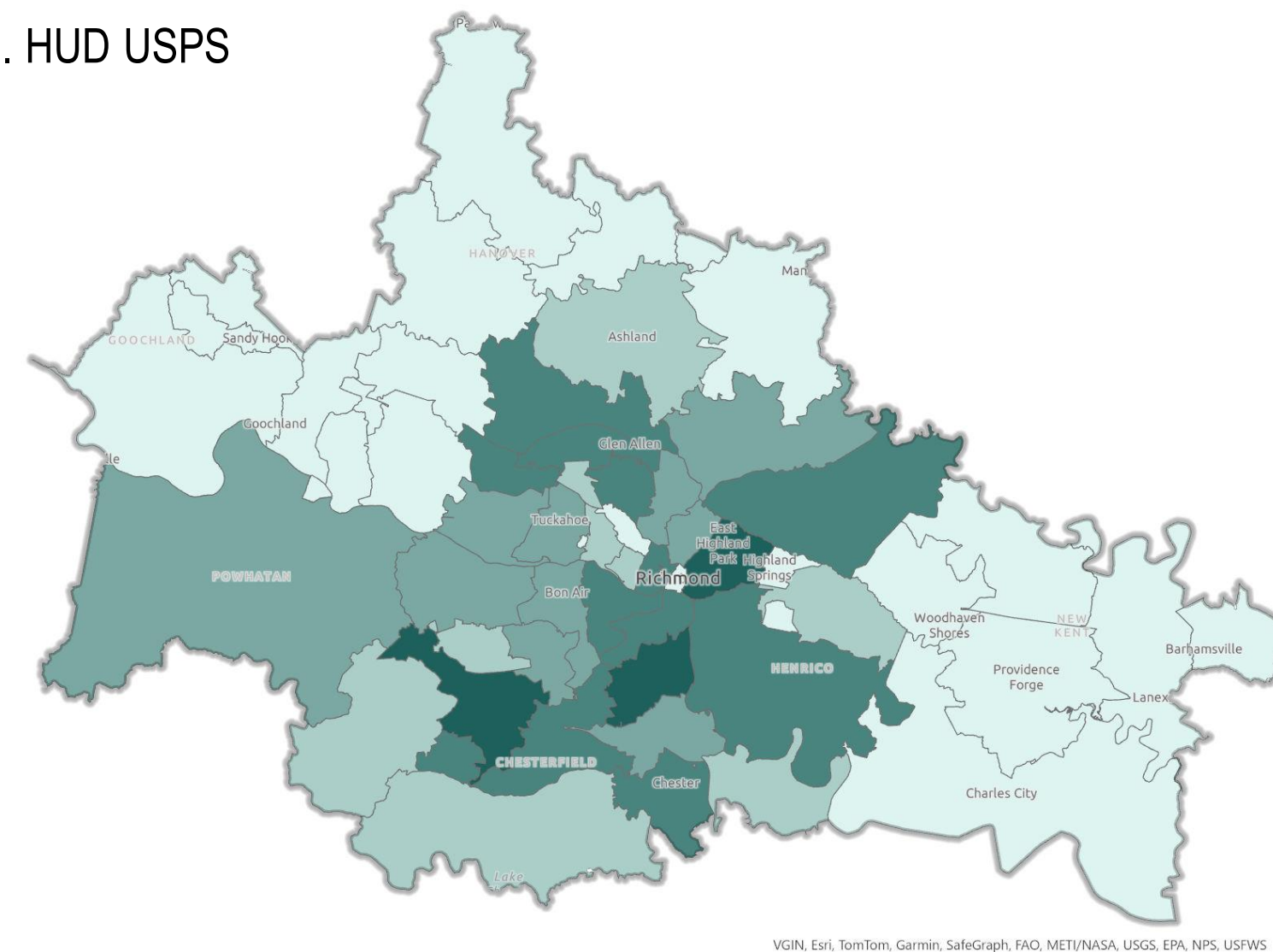- >22,000 to 33,000
- >33,000 to 44,000
- >44,000 to 55,000

Figure 2. ZCTA symbolized by total population (2020 U.S. Census P1). HUD USPS Zip Code Crosswalk Files used to calculate values.

U.S. Census 2020 Total Population
- 0 to 11,000
- >11,000 to 22,000
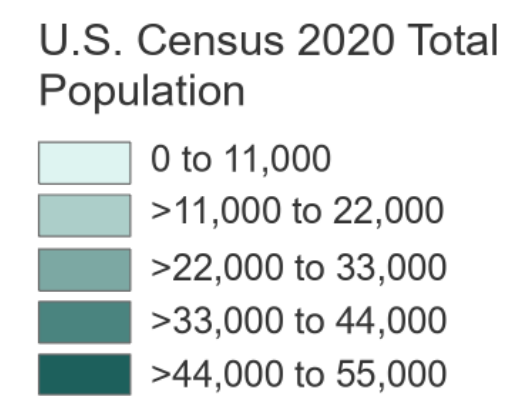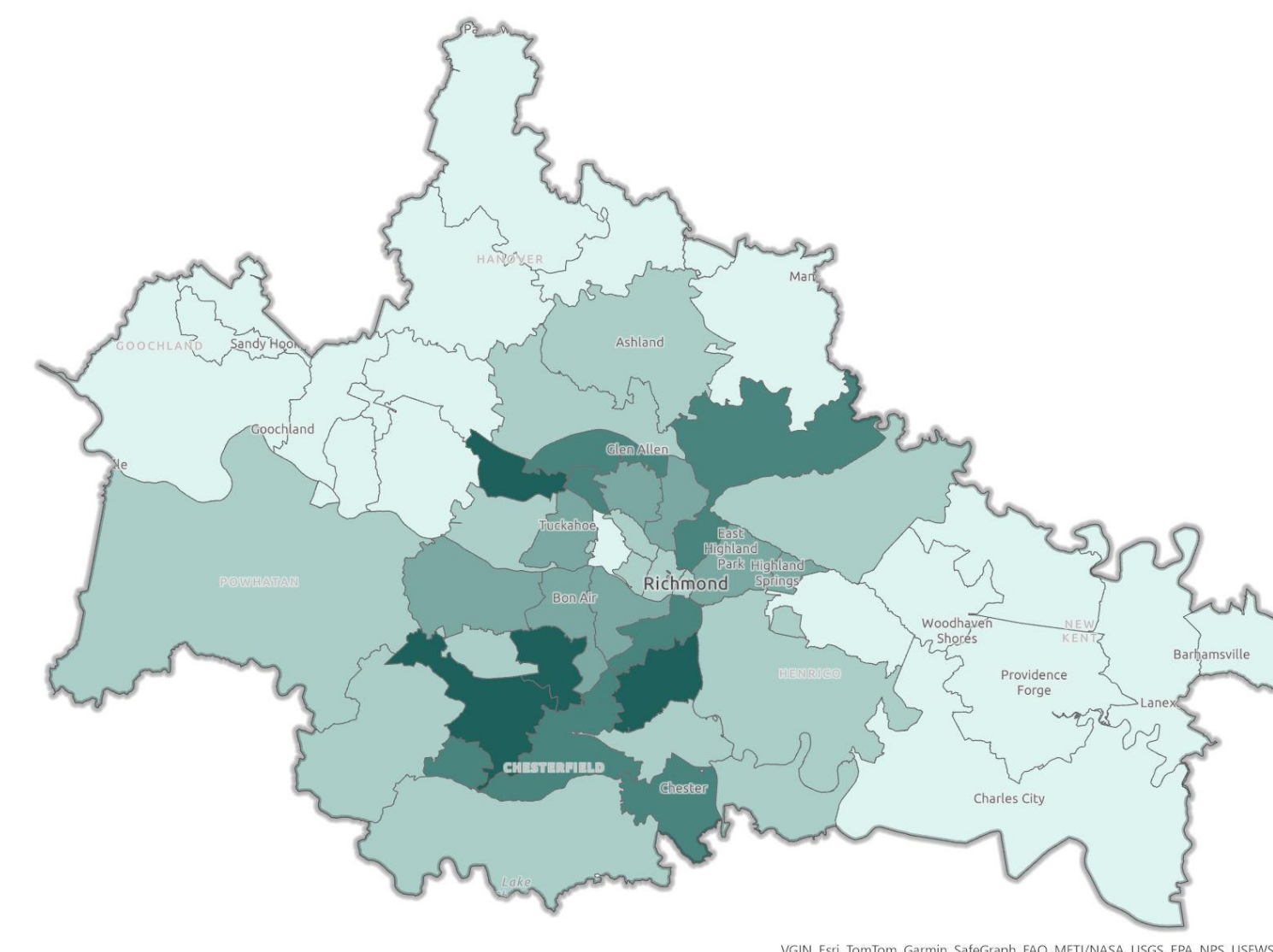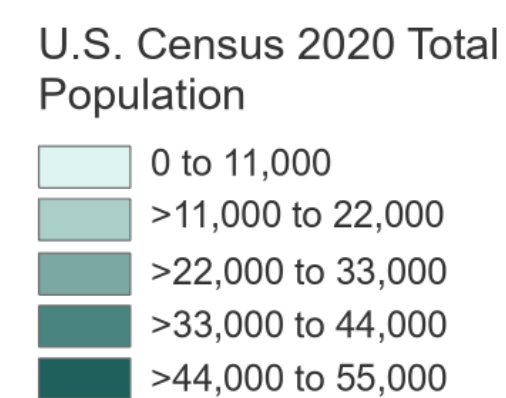- >22,000 to 33,000
- >33,000 to 44,000
- >44,000 to 55,000

Figure 3. ZCTA symbolized by total population (2020 U.S. Census P1). Nearest Census Tract to Zip Code aggregated to calculate values.

U.S. Census 2020 Total Population
- 0 to 11,000
- >11,000 to 22,000
- >22,000 to 33,000
- >33,000 to 44,000
- >44,000 to 55,000

RESULTS

- Total population values were similar among all three methods:

Method	Min	Max	Mean	Median	Range
Closest Tract to Zip Code	0	54,481	18,055	15,773	54,481
HUD USPS Crosswalk Files	0	53,872	18,123	14,487	53,872
Esri GeoEnrichment	0	54,592	18,341	14,742	54,592

- Esri GeoEnrichment 2020 Census population estimation for the ZCTA had an apportionment confidence of 2.576, confidence ranges 1.0 (best) to 5.0 (worst).

- Visualization of results using manual breaks highlight the differences in data among the ZCTA boundaries.

CONCLUSIONS

- The HUD USPS and Esri GeoEnrichment provide an accurate measure of population with the incorporation of proportional estimations to crosswalk varying geography boundaries.

- Esri GeoEnrichment methodology is the only method that can be applied to data collected at different areal units, in addition to the ZCTA (i.e. health district, travel time, food deserts, etc.).

- Understanding place in the context of health outcomes is critical to support cancer research impact.