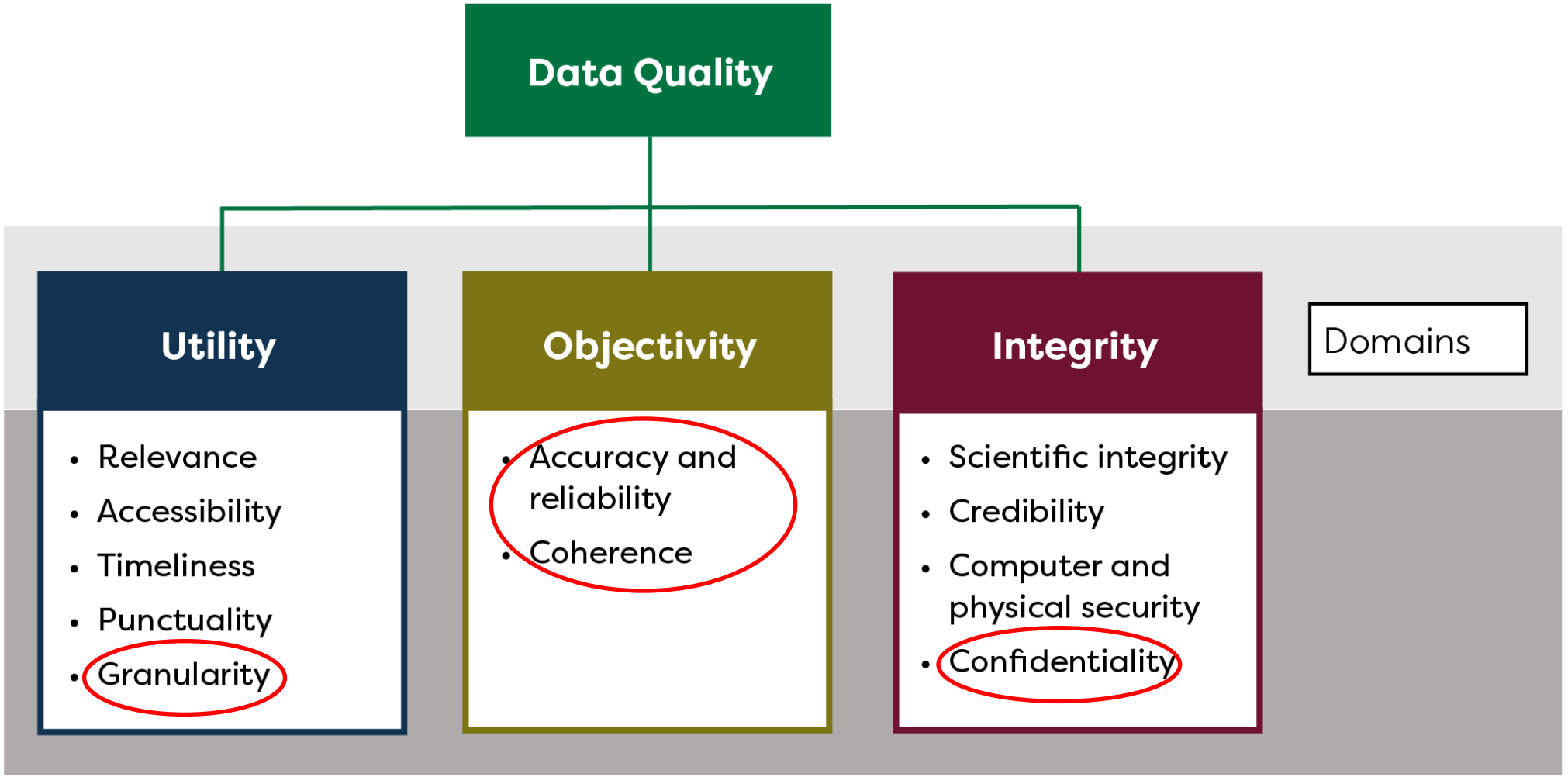
Bureau of Transportation Statistics

The Effects of Geography on Data Quality

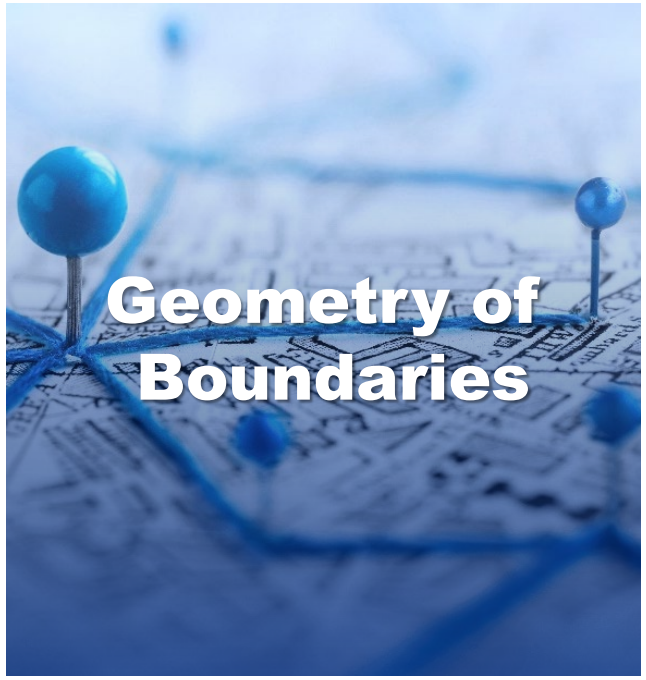
Rolf Schmitt
Ed Strocko

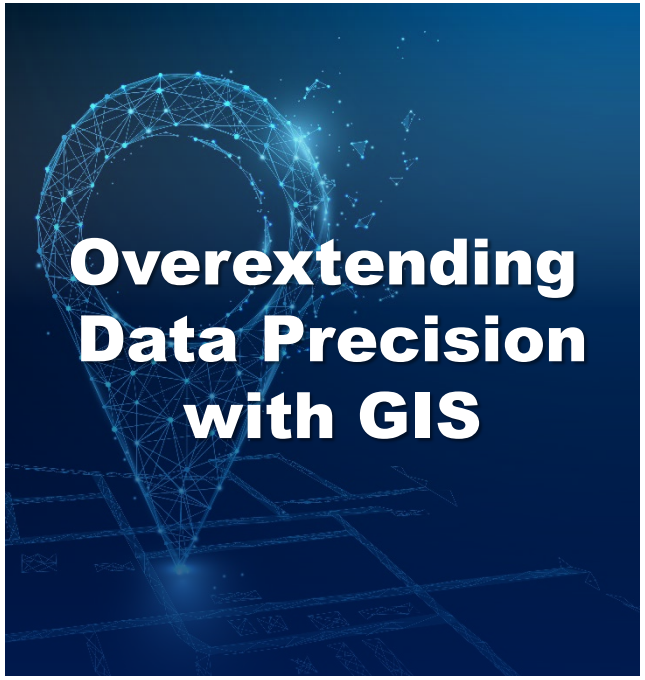
FCSM Data Quality Framework

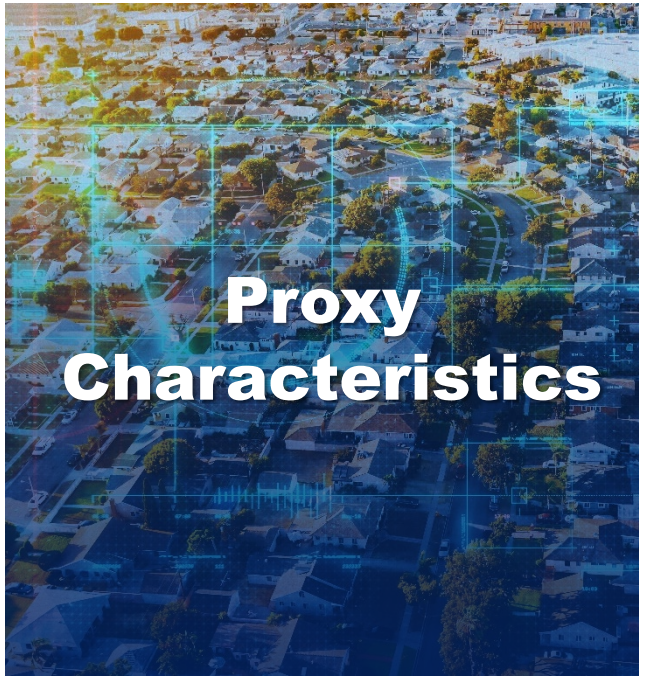


Accuracy & Reliability Involving Geographic Data

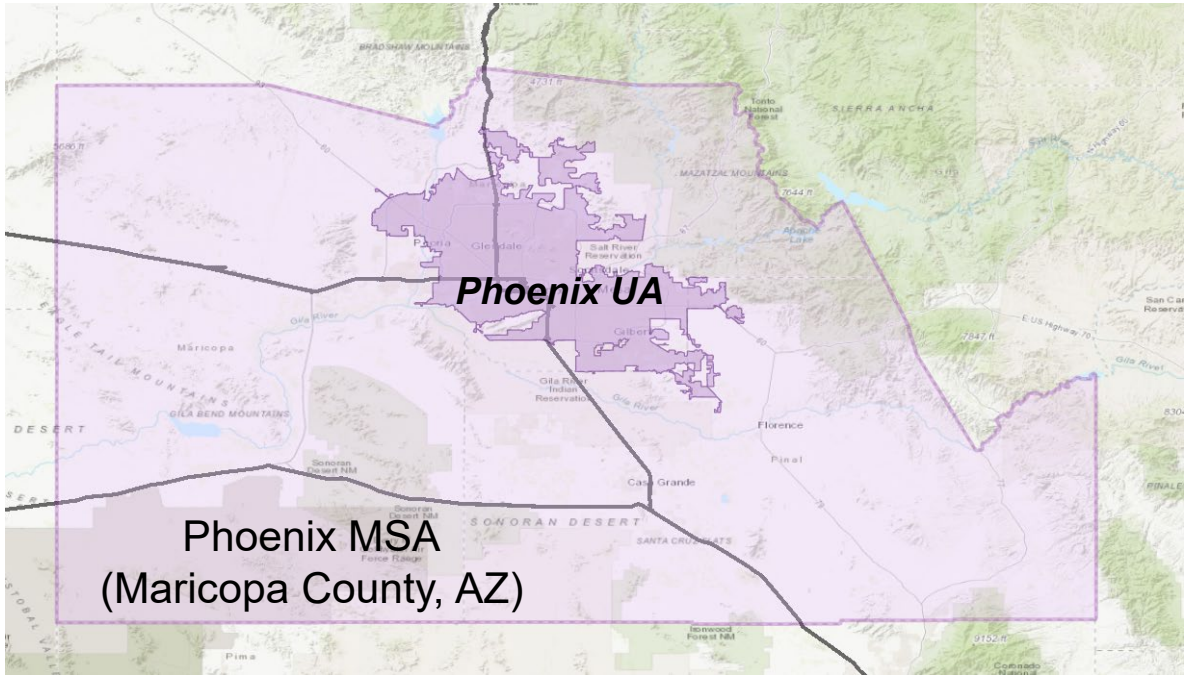
Data that characterize geographic locations are affected by:





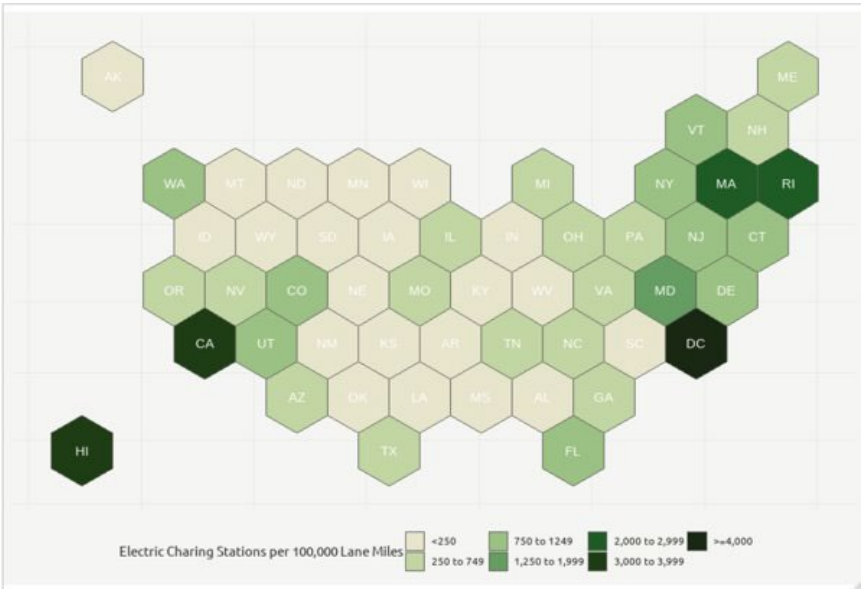


Statistics about geographic areas are affected by how the areas are defined and subdivided.

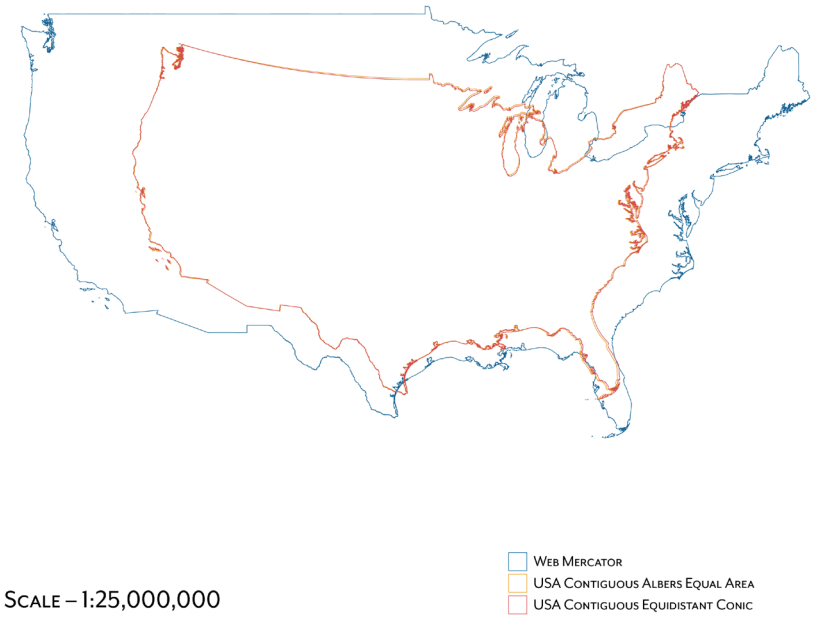


So far in 2022, the average vehicle speed on the interstates of the Phoenix MSA (light purple) is **63.2 mph**. If limited to the Phoenix UA (dark purple) -- eliminating the many rural interstate segments technically within the MSA -- this *decreases* to **56.8 mph**.

When the number of electric charging stations in each state are divided by its lane miles, small states (**DC**, MA, RI, HI) show up disproportionately



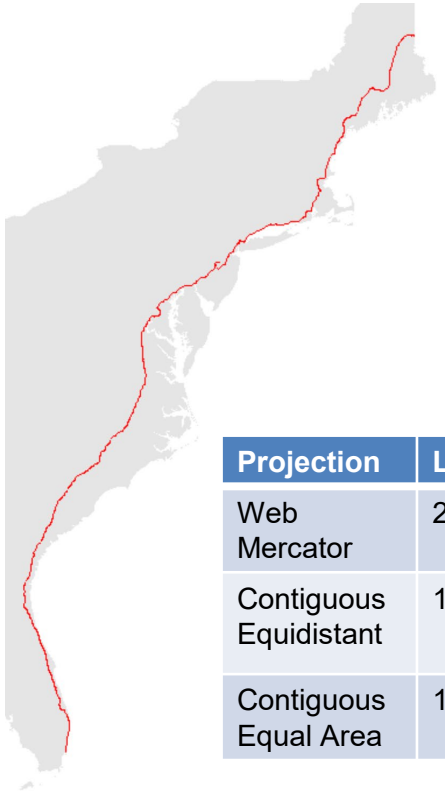
Projections Matter for measuring from a map.



Why is this problem important?

- Geostatistical analyses, maps not replicable
  - The same analysis produces different results when using an alternate or incorrect projection
- Leads to misleading or incorrect statistics or maps
  - Improperly-selected projections artificially inflate or deflate the size, area, and/or length of critical features

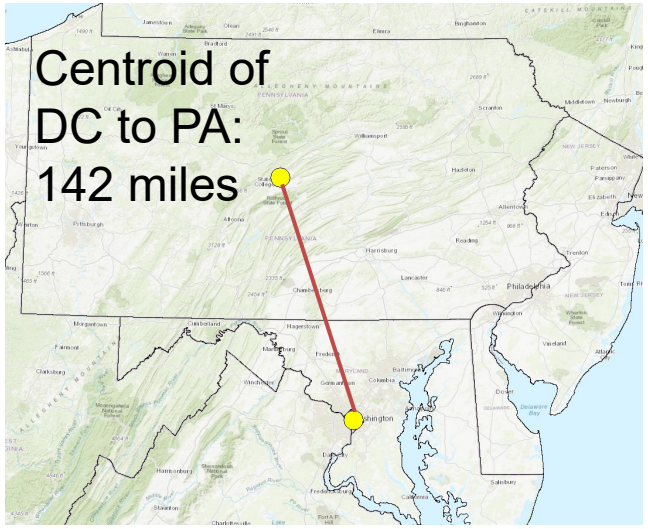
NHS – Interstate 95



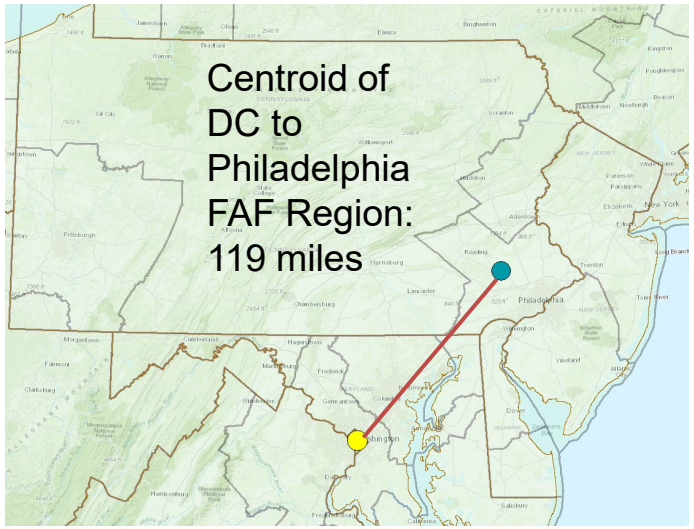
Projection	Length (mi)
Web Mercator	2442.659
Contiguous Equidistant	1931.487
Contiguous Equal Area	1933.901

- Projections are rarely cited in cartographic and geo-statistical/tabular products
- Projections are often selected by individual discretion, regardless of project purpose or scale
- Projections are easily and unknowingly used inappropriately, especially with modern point-and-click GIS software

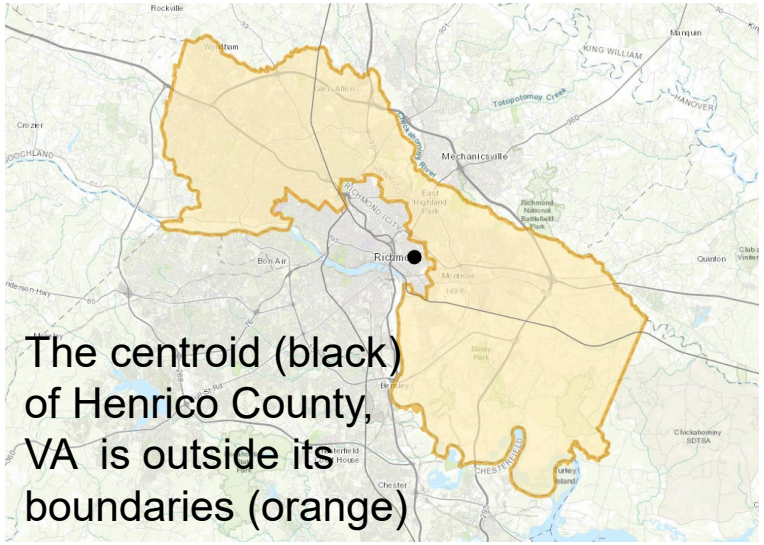
Distances between locations are often measured between centroids of the geographic units.



Many smaller FAF Regions are within a state. Larger and fewer geographic units result in longer distances between units than smaller and more numerous geographic units for the same actual spatial distribution of activity.



Not all activities are well represented by centroids:



Accessibility of a Location

The Geographic center of states such as State College, PA but can be farthest away from every city in the state. Does this make such places most or least accessible?

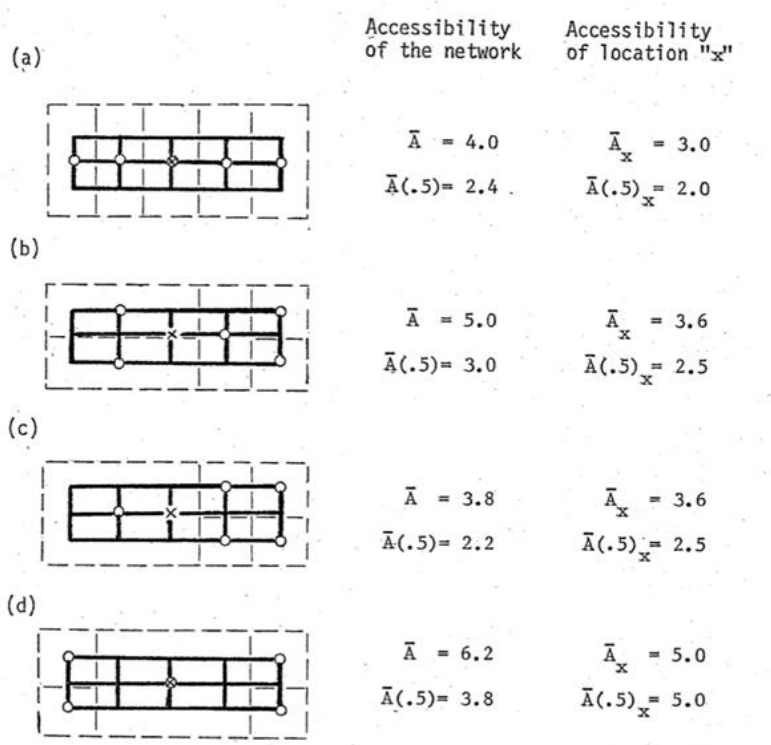
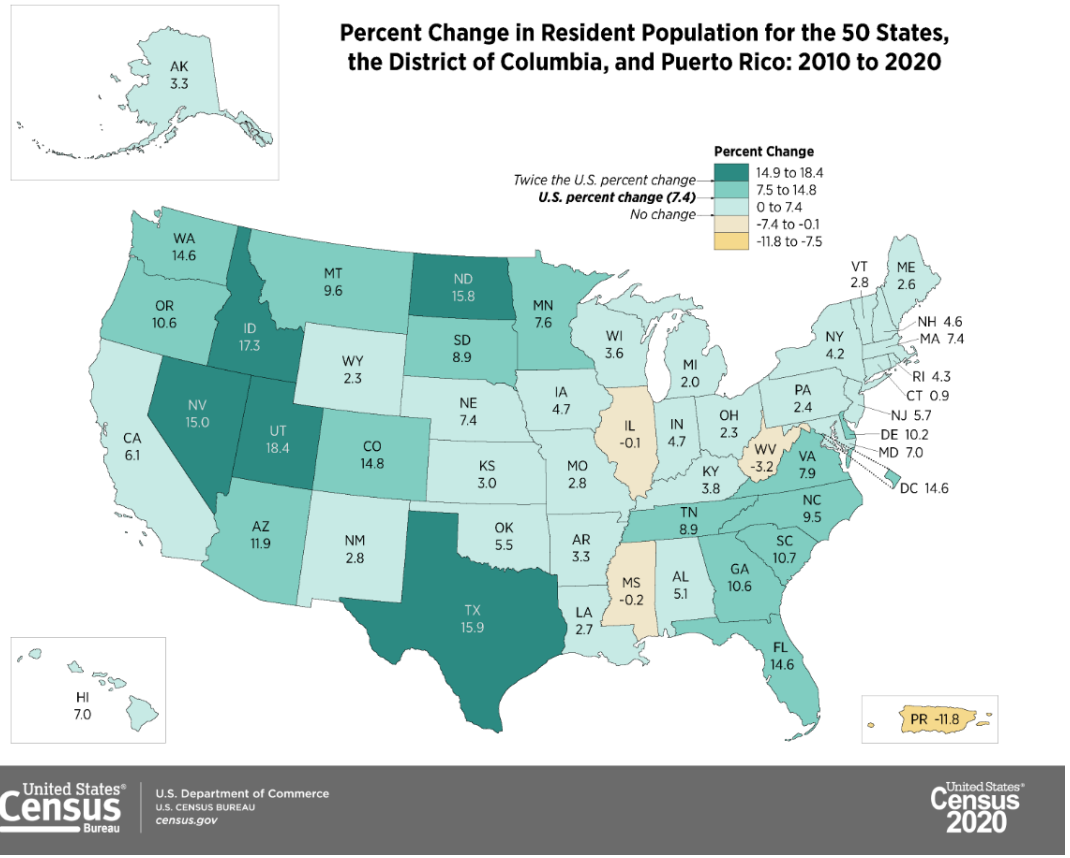
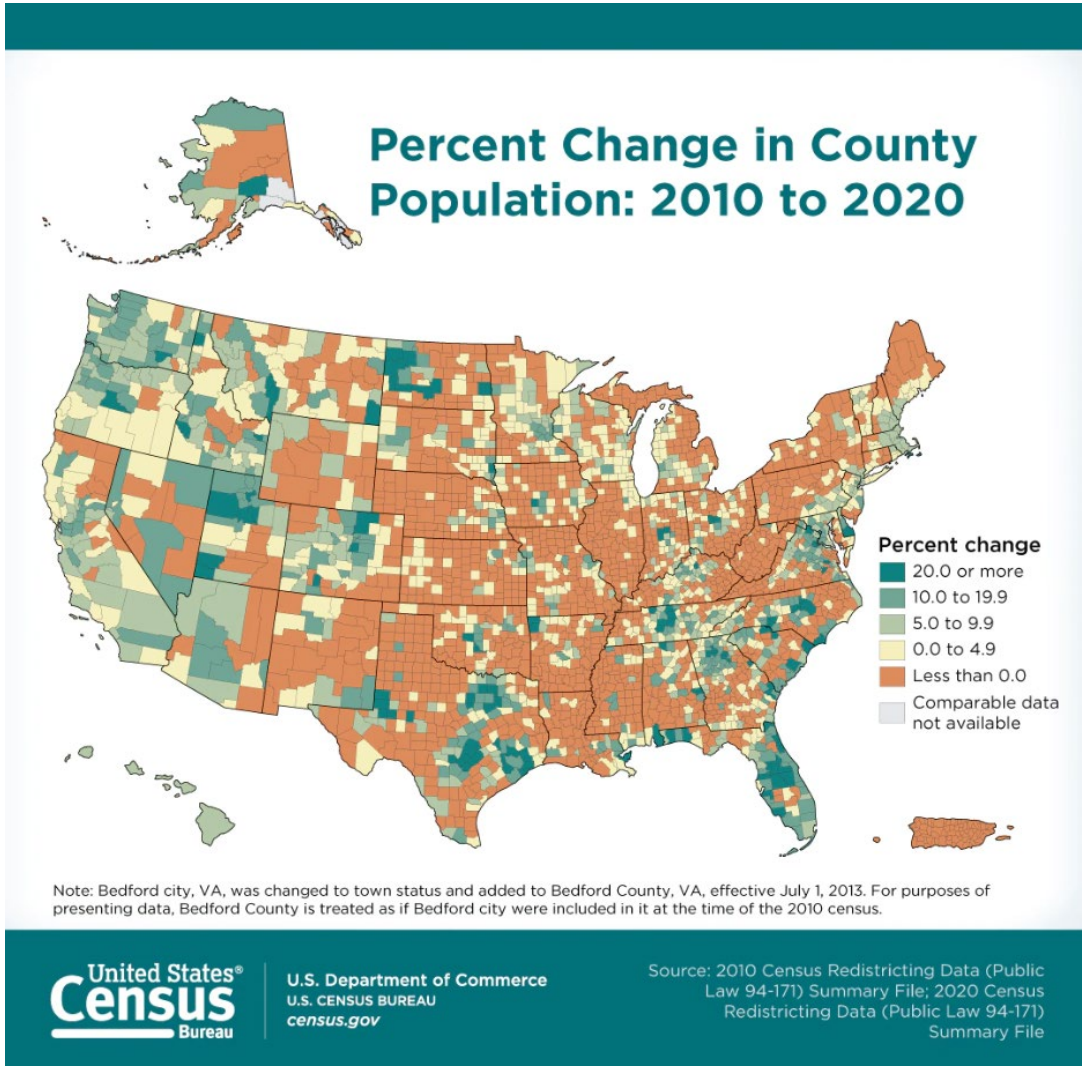


Figure 3-4: DISTORTIONS IN AVERAGE DISTANCE MEASURES CAUSED BY ZONE CONFIGURATIONS

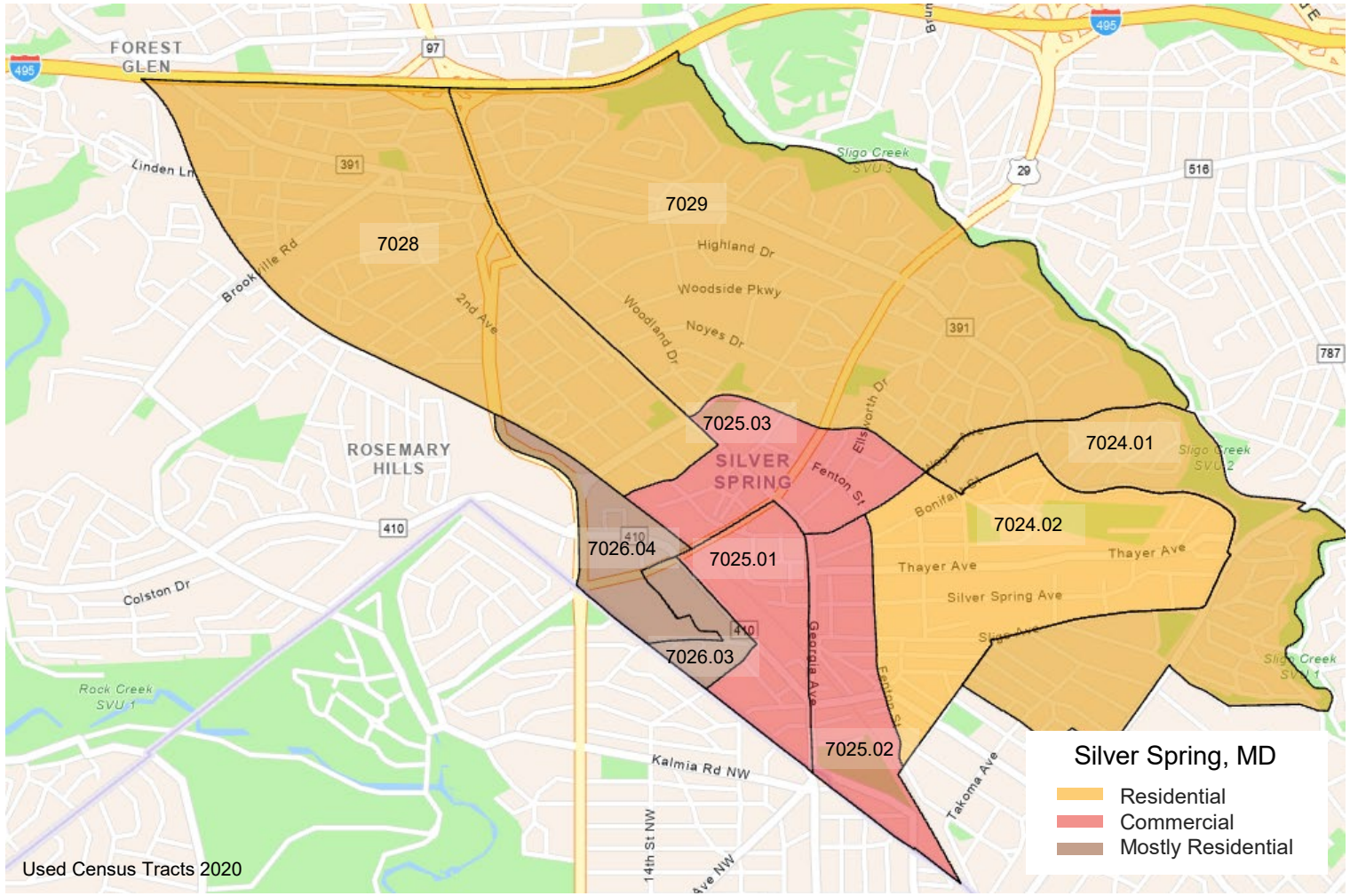
Changing the granularity for a data product does not always solve the statistical problems.

- Maps show different analytical results at different levels of aggregation.
- Despite using the same discrete data each successive smaller unit consequently changes the pattern





No One System of Geographical Units Serves All Purposes

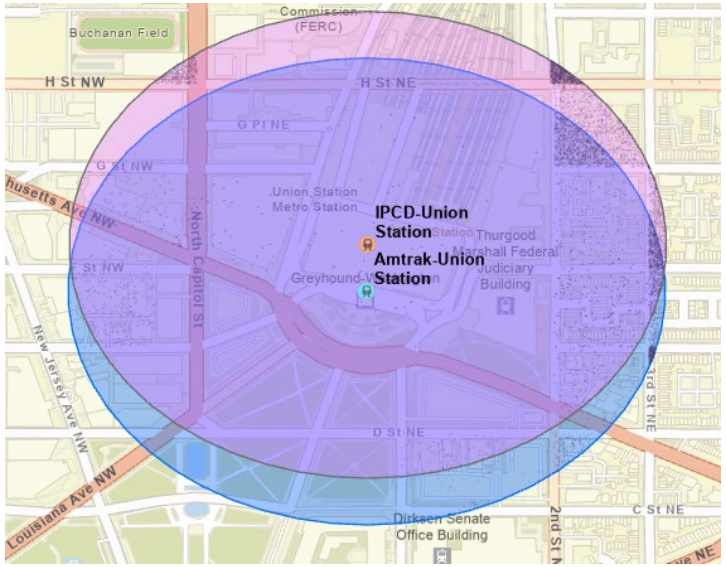


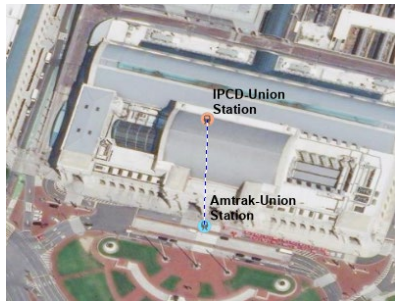
Census tracts can be divided by transportation infrastructure, distinguishing commercial and residential areas.

Many statistics are tied to individual locations and are easily displayed and analyzed with GIS.

vs

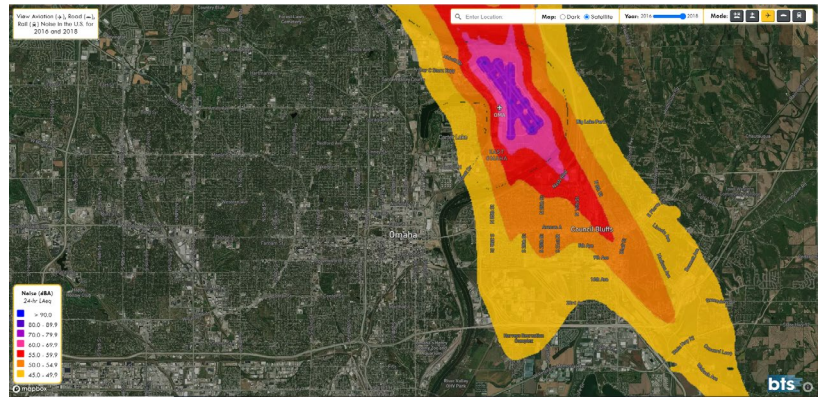
Offset distance from the actual location is more when the ***precision*** of the coordinate systems are lower and is dependent on the scale of the map. *Bridge shown is in Prince George's County, Maryland.*





Total population captured within quarter mile of Stations that are 260 ft apart resulted in a difference of over 1000 people.

On the BTS National Transportation Noise Map, users are not permitted to zoom further into a map beyond the city level (right) because the noise model was not built to be ***precise*** at such a localized scale



Neighborhood characteristics are often used as a proxy for individual characteristics.

- Using neighborhood averages as proxies for individuals is problematic for neighborhoods with low homogeneity and especially for neighborhoods with internal divisions
- Neighborhoods may have affects on the behavior of individual residents
- From ACS Travel time to work, for Maryland Census Tract 1401 in Baltimore Maryland, the average travel time to work is around 30 minutes. However approximately 3.7% (±2.2%) of an estimated 1,928 (±312) commuters had a commute of 90 minutes or more.

Ryan Stavely - MARC 438 #2 © CC BY-SA 2.0


Thank You