

CENSUS DATA RELIABILITY

Examining Accuracy of ACS Data to Map
Environmental Justice Demographics

Rachel Pierstorff

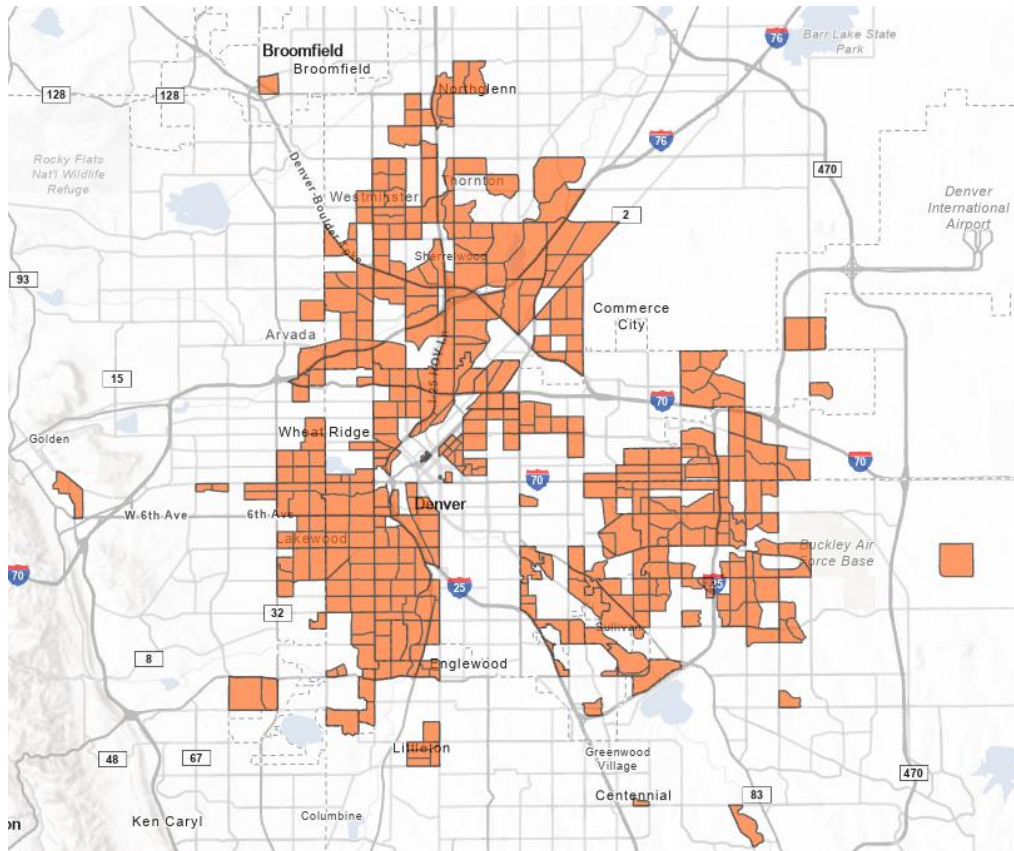
Byron Schuldt

OUTLINE



- Project motivation and DRCOG's Env. Justice dataset
- The source data - census
- Investigating census data reliability (accuracy of estimates)
 - % Minority Population
 - % of Population in Poverty
 - % of Households that are Housing Cost-Burdened
- Key Takeaways

DRCOG'S ENVIRONMENTAL JUSTICE DATASET



Defined As:

- % Households in Poverty > Regional Average
- Or
- % Minority > Regional Average

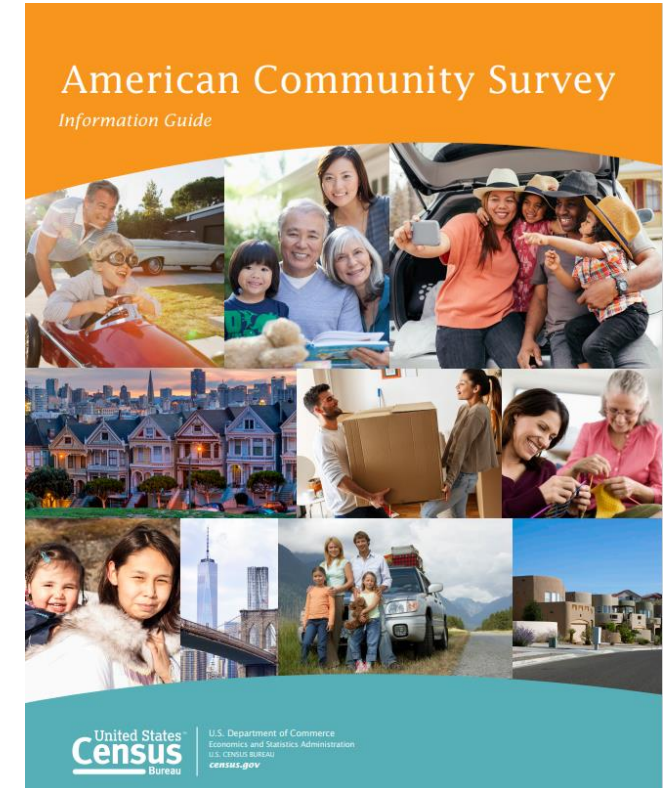
Revisiting Our Methods:

- Different definition?
- Equity Index?

AMERICAN COMMUNITY SURVEY DATA



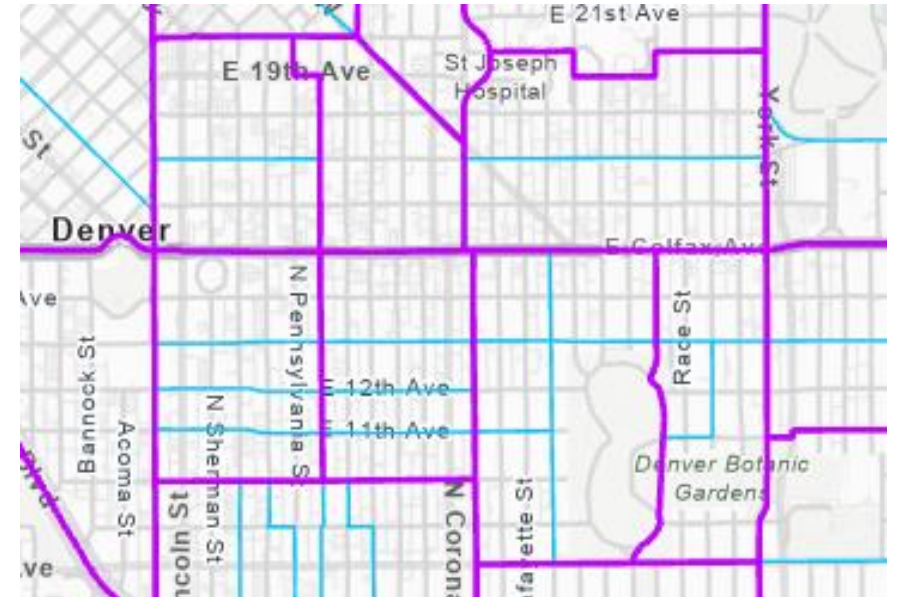
- 5-year “American Community Survey” (ACS) Census Data
- Samples 3.5 million households/year over 5-year period (Jan 2015 – Dec 2019)
- Estimates for:
 - Population
 - People under various poverty thresholds
 - Households that are housing cost-burdened
 - Minority population
- Better reliability than 1-year estimates



TRACTS VS BLOCK GROUPS

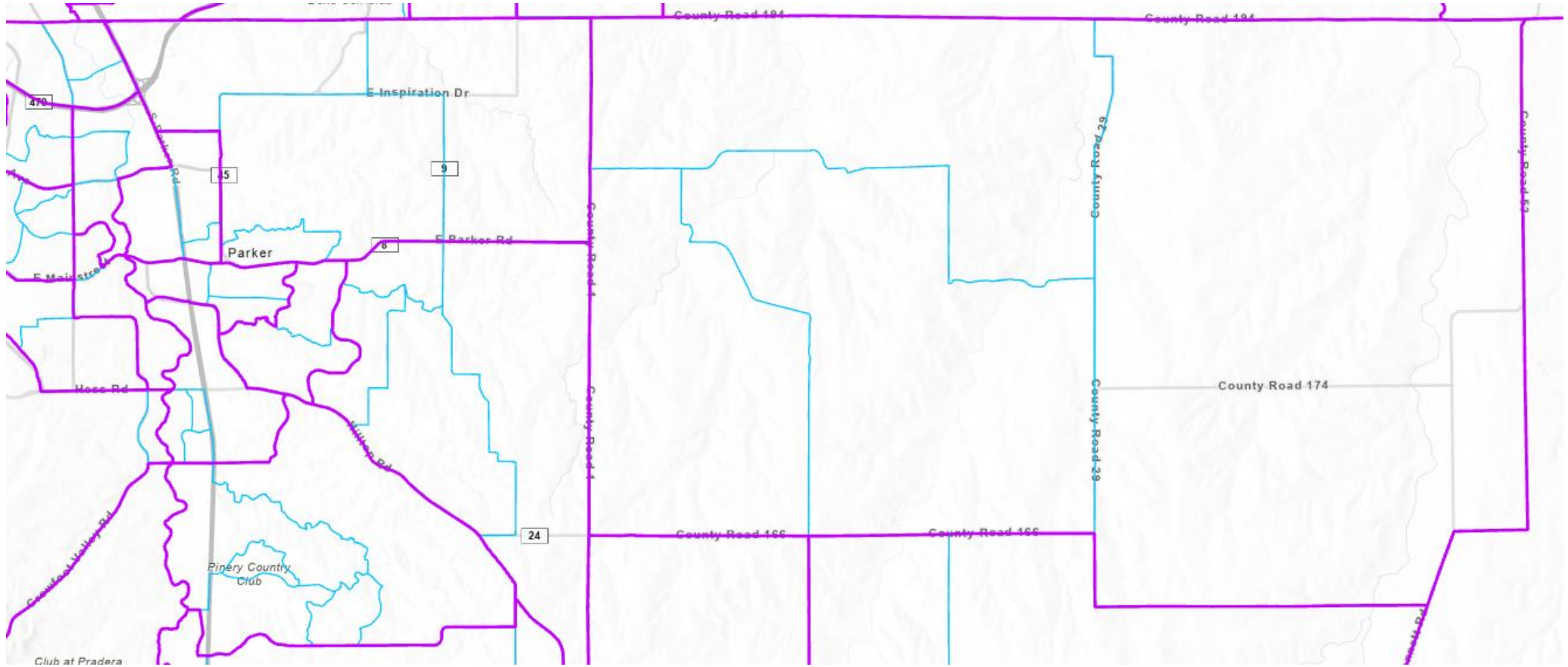


- Tract: subdivision of a county or city, between 1,000 and 8,000 people
- Block Group: subdivision of tracts, contain between 600-3,000 people



- Census Tract Bounds
- Census Block Group Bounds

TRACTS VS BLOCK GROUPS

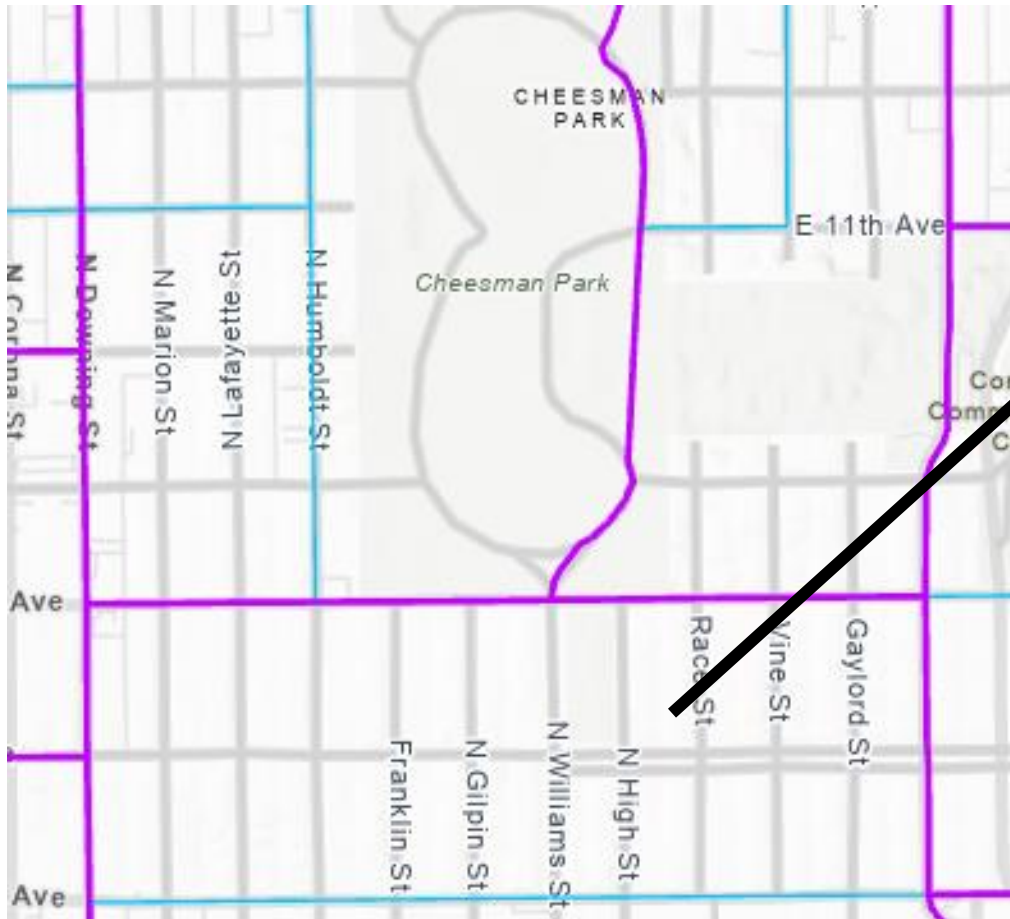


INITIAL QUESTIONS



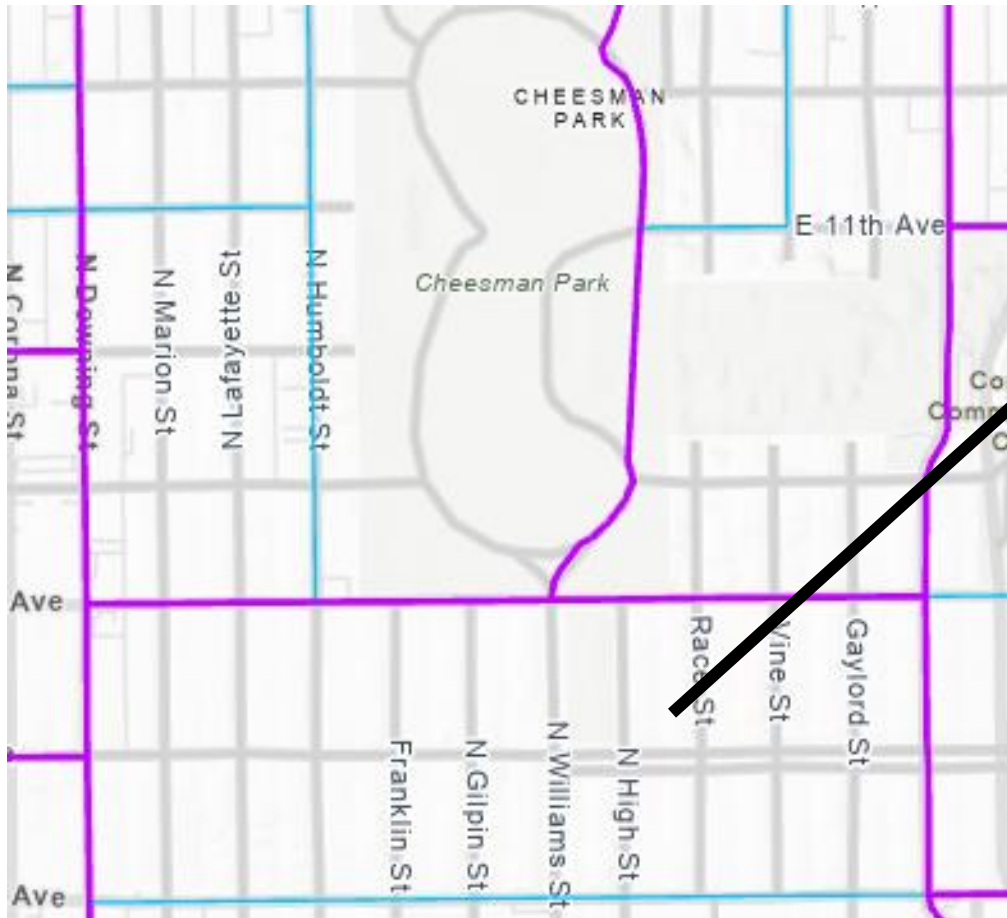
- How reliable are ACS estimates?
- Does this change significantly for:
 - Different variables?
 - Block Groups vs Tracts?
 - 2015-2019 vs 2016-2020 ACS?

INVESTIGATING RELIABILITY: MARGIN OF ERROR



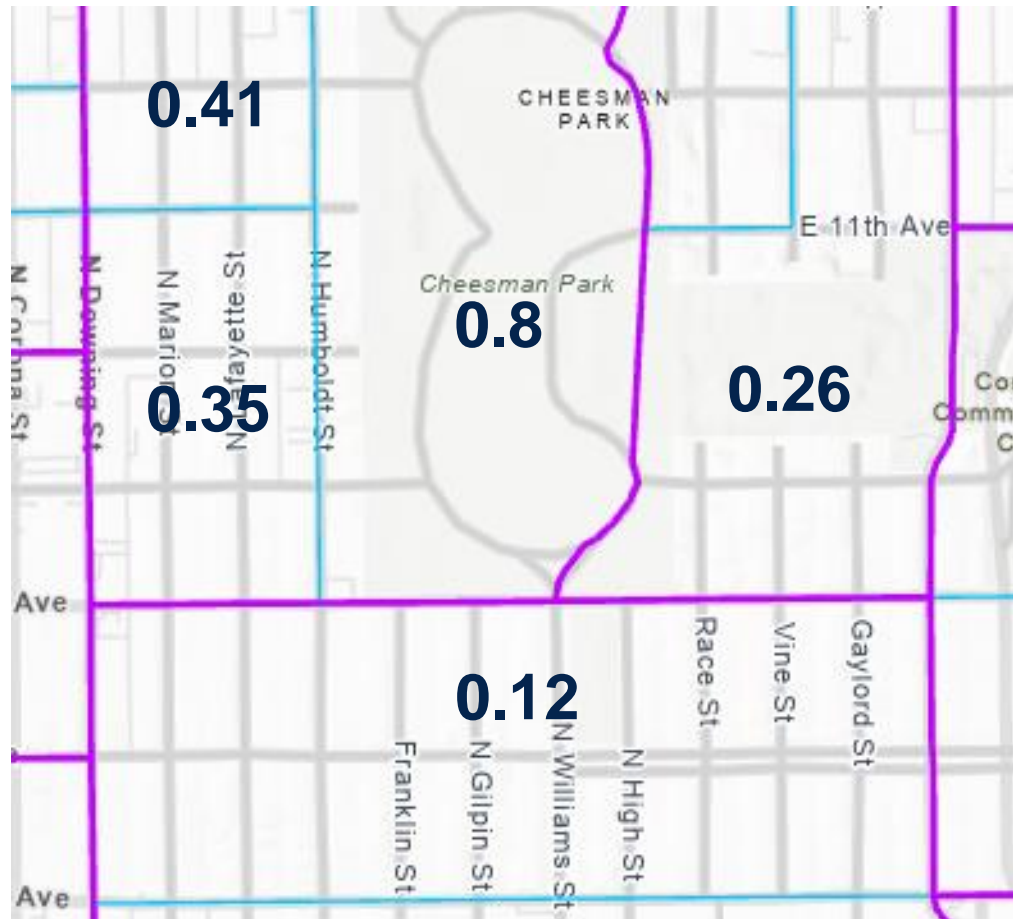
- Estimated Population Below Poverty Level: 100
- Margin of Error: 20
- (90% Likelihood that pop below poverty is between 80 and 120)
- *Is this good or bad?*

“CV SCORES” AS MEASURE OF RELIABILITY



- Estimated Population Below Poverty Level: 100
- Margin of Error: 20
- (90% Confidence that pop below poverty is between 80 and 120)
- $20/1.645 = 12.16$ (Standard Error)
- $12.16/100 = \mathbf{0.12}$ (CV Score)
- CV score allows comparison across block groups (low = more reliable)

“CV SCORES” AS MEASURE OF RELIABILITY

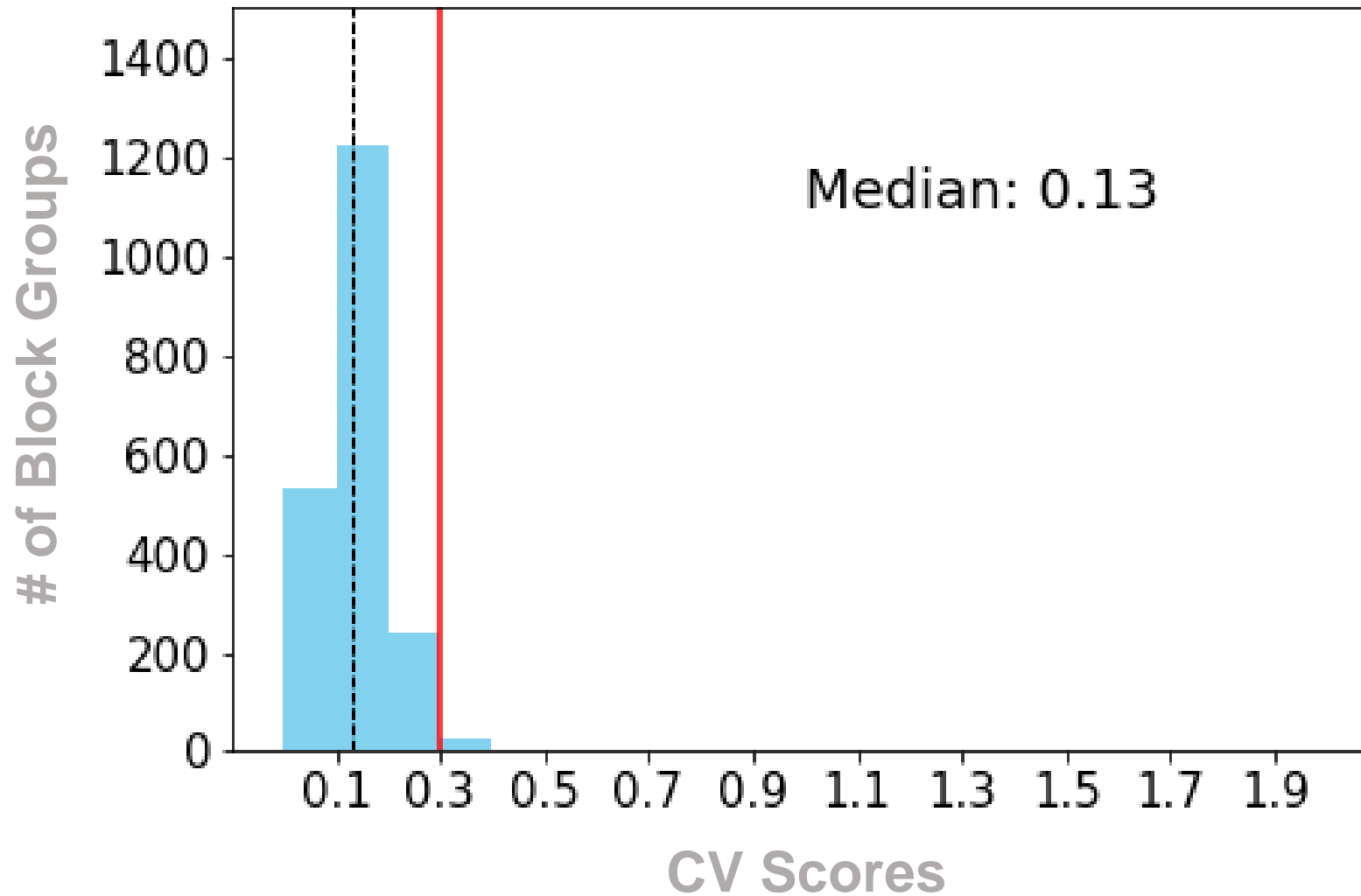


- Estimated Population Below Poverty Level: 100
- Margin of Error: 20
- (90% Confidence that pop below poverty is between 80 and 120)
- $20/1.645 = 12.16$ (Standard Error)
- $12.16/100 = \mathbf{0.12}$ (CV Score)
- CV score allows comparison across block groups (low = more reliable)

VISUALIZING OVERALL RELIABILITY



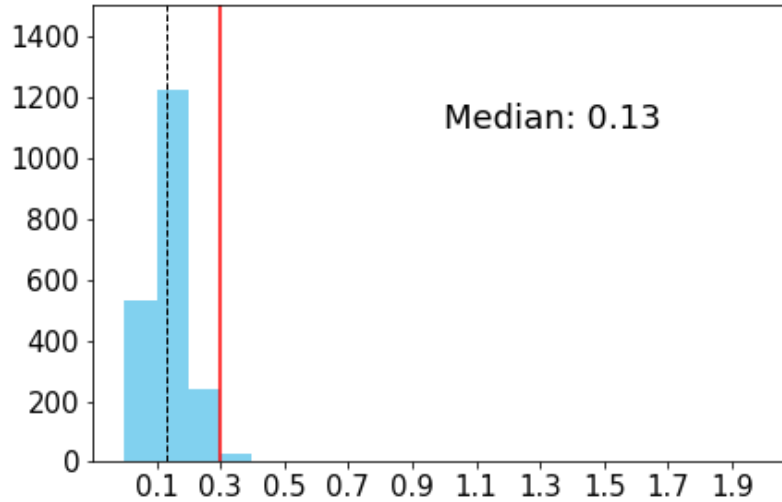
Block Groups: Total Population Estimate



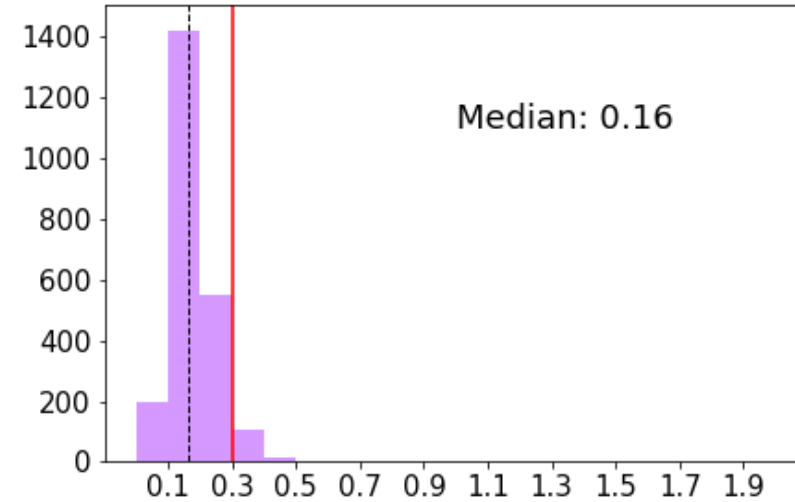
TOTAL POPULATION



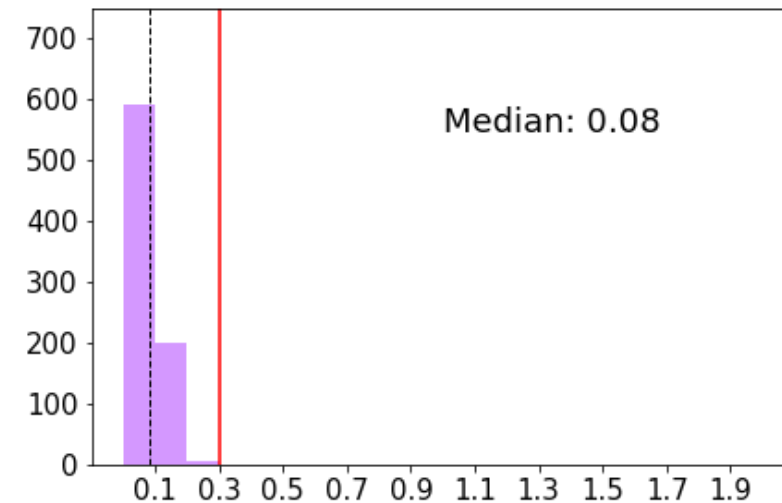
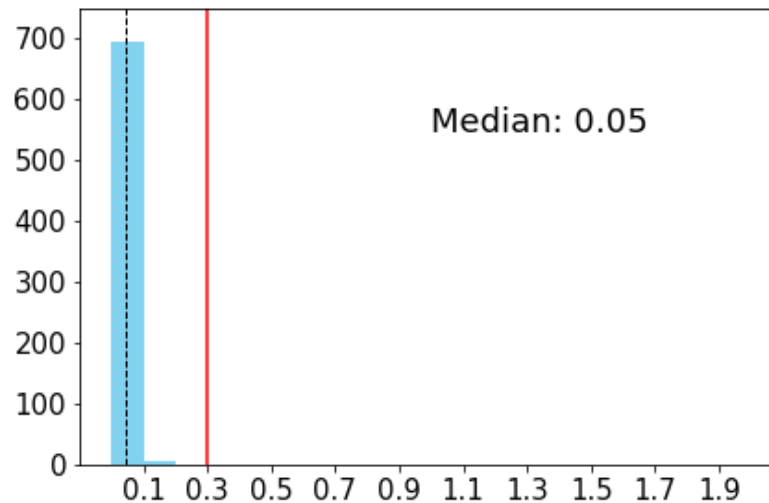
2019 5yr ACS



2020 5yr ACS



Block Groups

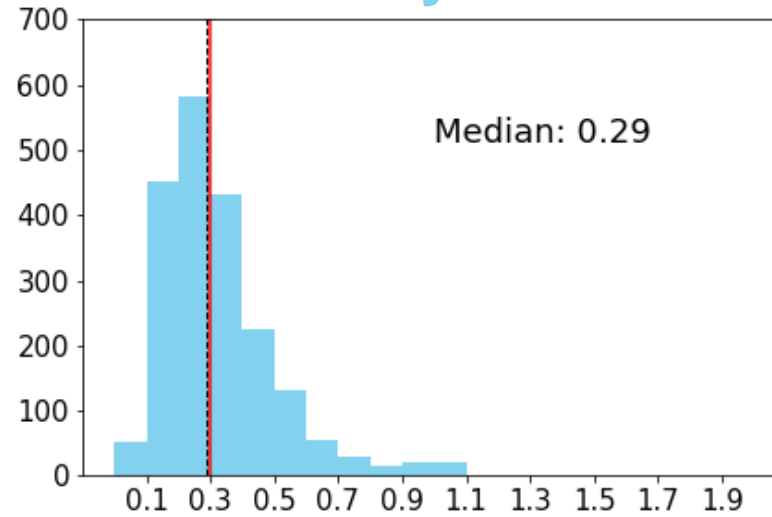


Tracts

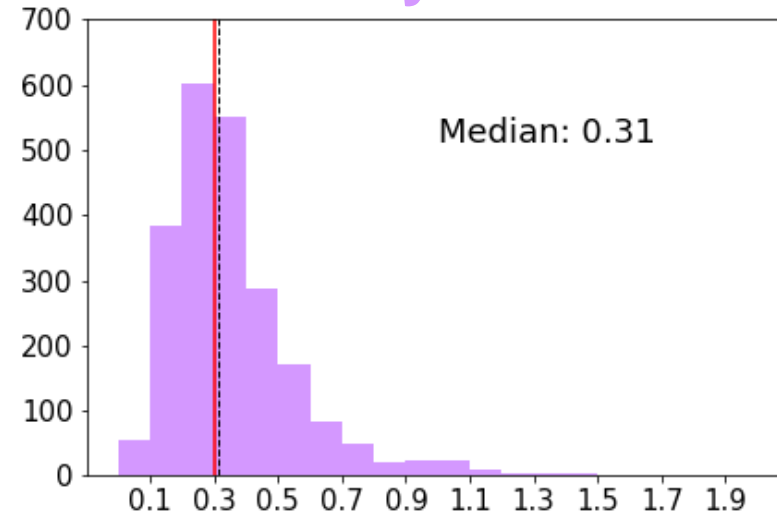
% MINORITY



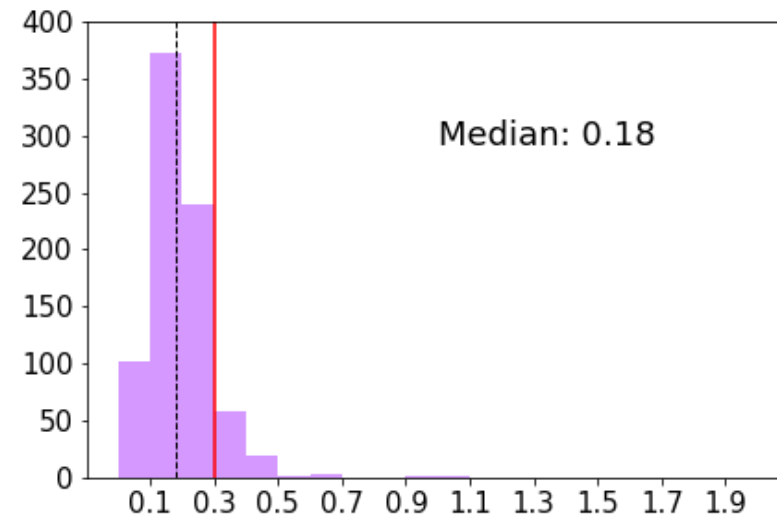
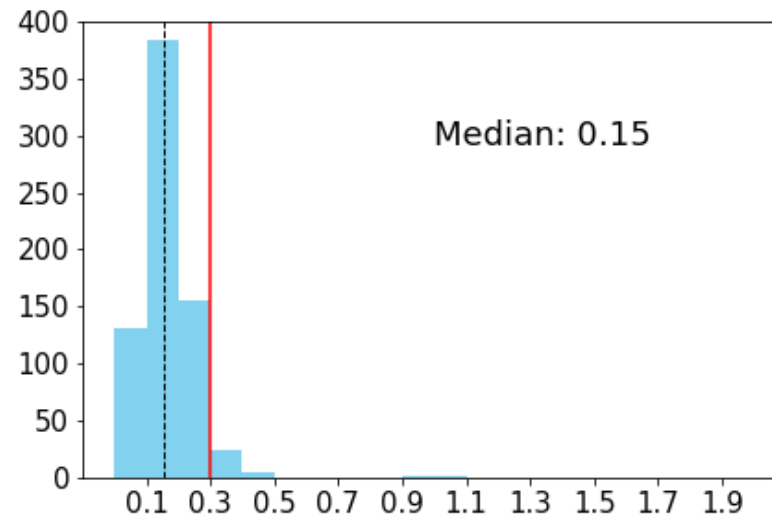
2019 5yr ACS



2020 5yr ACS



Block Groups

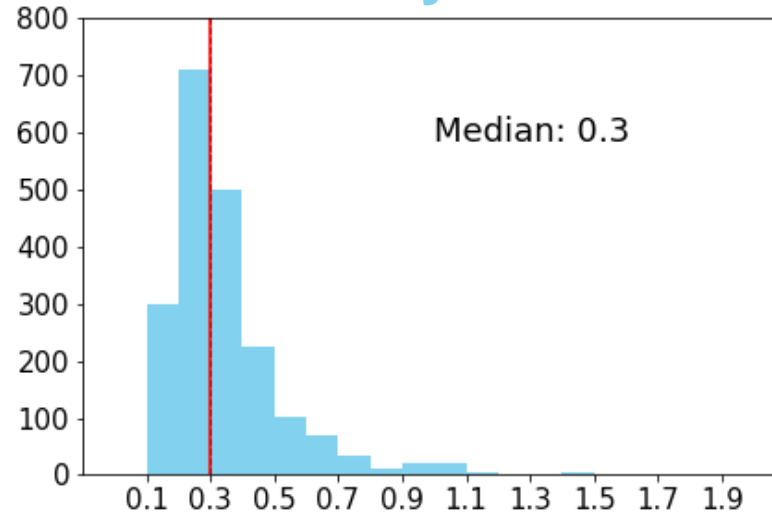


Tracts

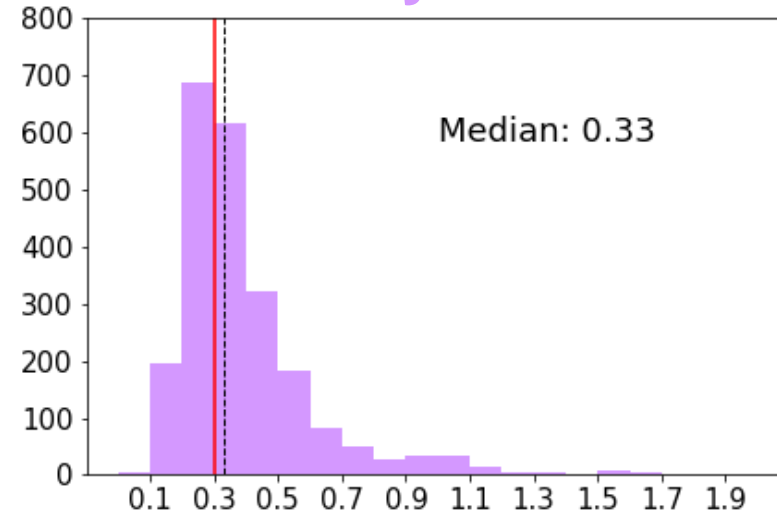
% HOUSEHOLDS HOUSING COST-BURDENED



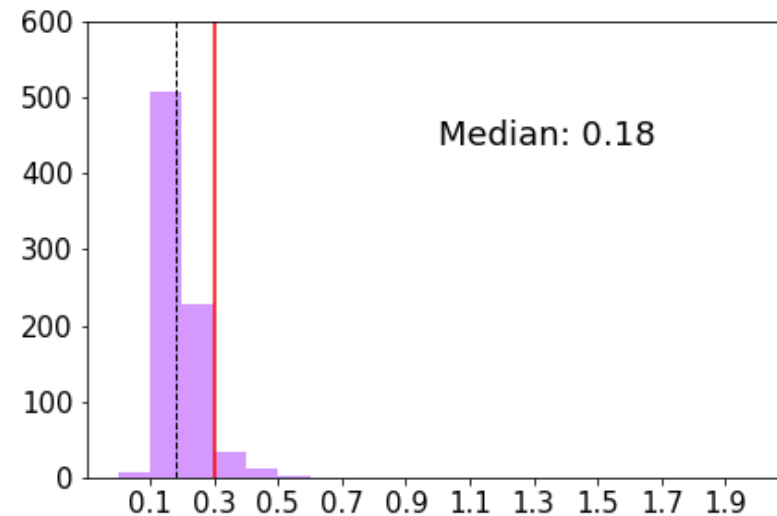
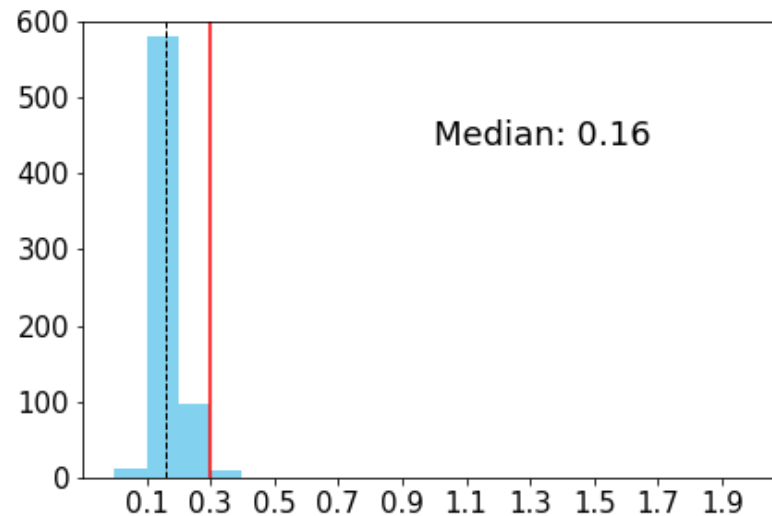
2019 5yr ACS



2020 5yr ACS



Block Groups

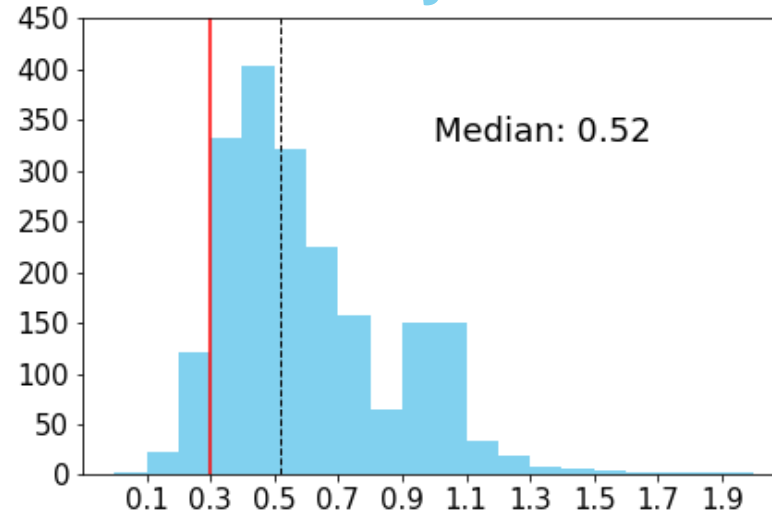


Tracts

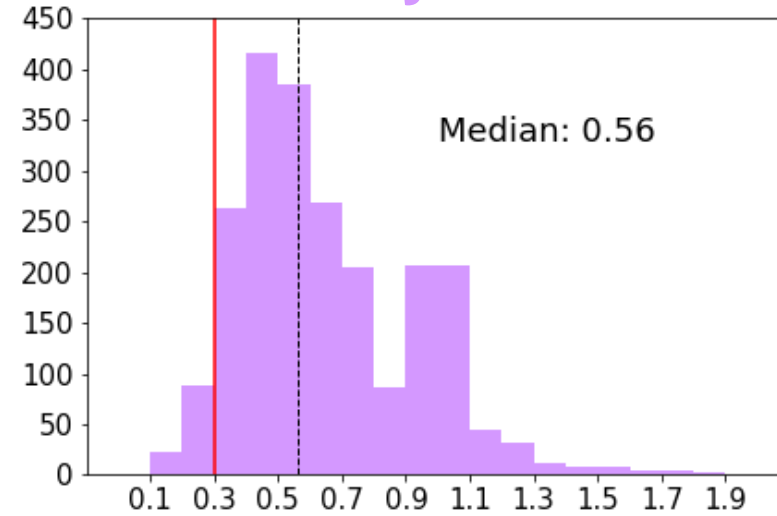
% BELOW POVERTY LEVEL



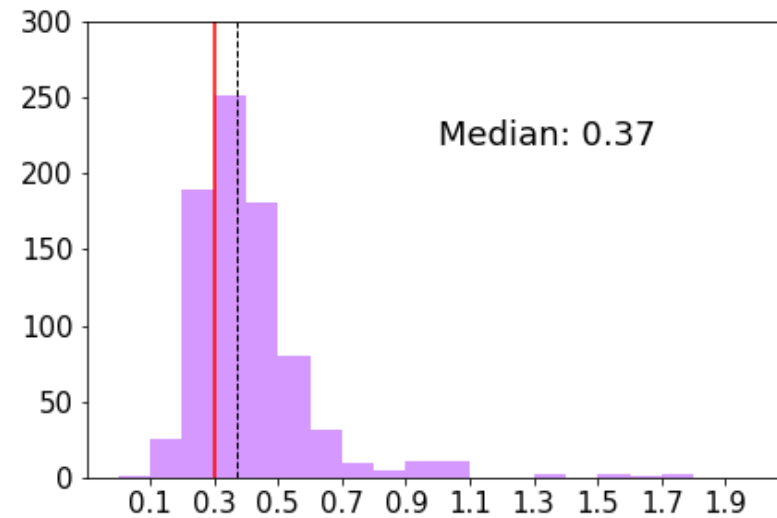
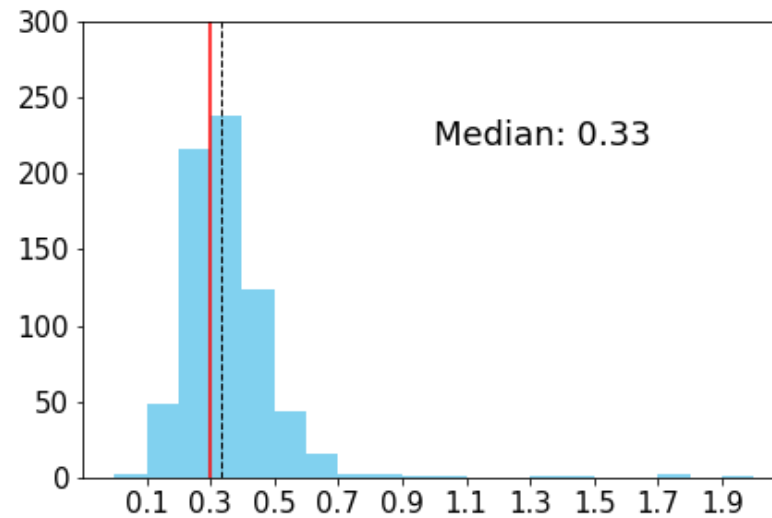
2019 5yr ACS



2020 5yr ACS



Block Groups

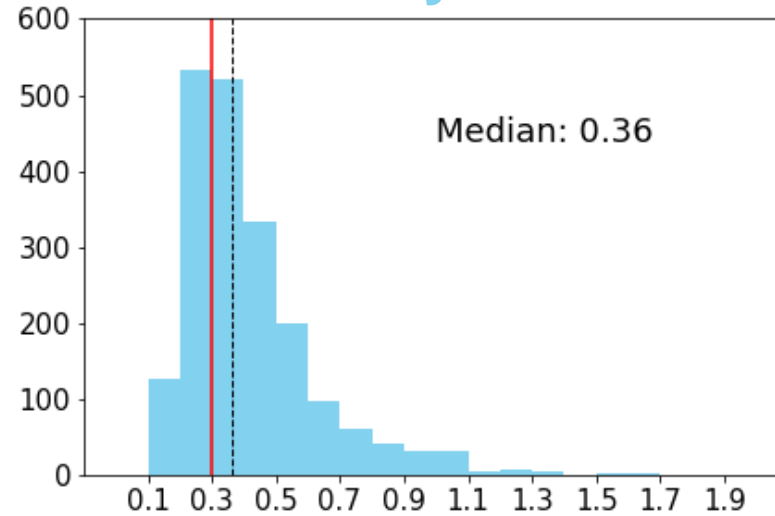


Tracts

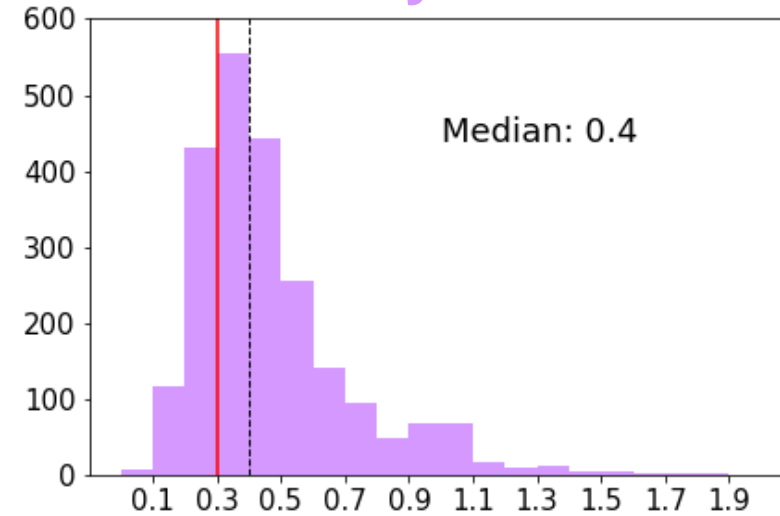
% BELOW 2X THE POVERTY LEVEL



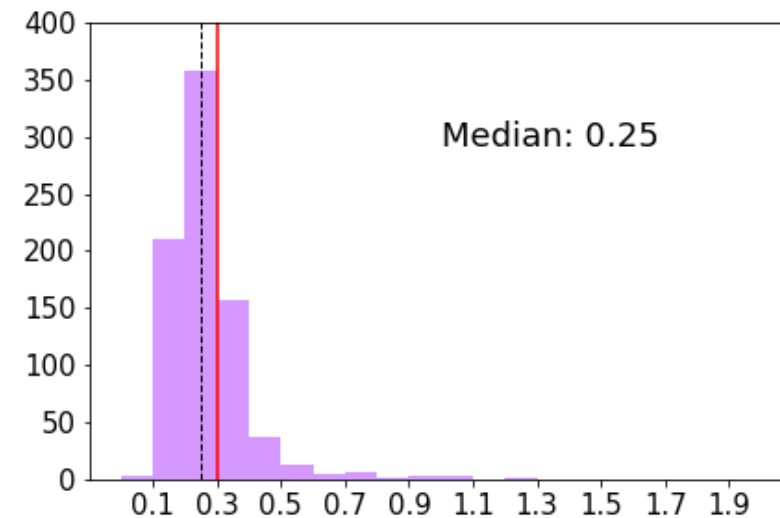
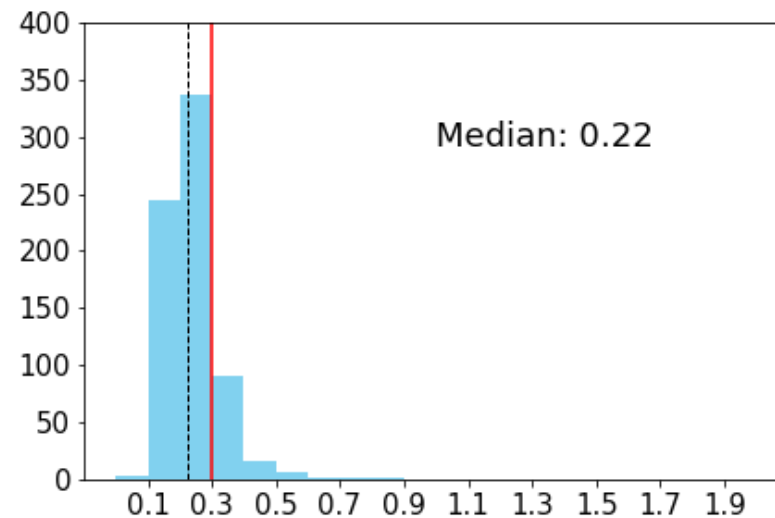
2019 5yr ACS



2020 5yr ACS



Block Groups



Tracts

TAKEAWAYS



- Tracts more reliable than block groups
- 2020 ACS data looks *slightly* less reliable
- Some variables more reliable than others
 - Aggregating variables increases accuracy

DASHBOARD VISUALIZATION



- How does reliability vary across the DRCOG region?
- Is it possible to exclude highly unreliable geographies without excluding too many geographies?
- How do geographies with high concentrations of vulnerable populations compare with areas of high unreliability?