Using ACS data in a GIS course: Challenges and opportunities

Yelena Ogneva-Himmelberger, Ph.D.
Clark University

2021 ACS Data Users Conference
May 18, 2021

AMERICAN COMMUNITY SURVEY
DATA USERS GROUP

Outline

- Context: University GIS courses
- Data uncertainty: Guidance from the Census Bureau and research community
- GIS-friendly ACS data sources and instructional resources
- Examples of student work
- Round table discussion

- Private research University located in Worcester, Massachusetts

- Offers Bachelors and Ph.D. degrees in Geography

- Offers Masters of Science degree in GIS

Context

Courses where ACS data is used:

- Advanced vector GIS
- Social applications of GIS
- Spatial analysis for health

Mix of graduate/undergraduate students

- Masters in GIS
- Geography majors

How is ACS data used?

- class demonstrations
- final projects

Which ACS data is used?

- Usually, 5-year estimates for census tracts

Opportunities

Table 1.1. **Population and Housing Data Included in the American Community Survey Data Products**

Social Characteristics
Ancestry
Citizenship Status
Disability Status[1]
Educational Attainment
Fertility
Grandparents as Caregivers
Language Spoken at Home
Marital History[2]
Marital Status
Migration/Residence 1 Year Ago
Period of Military Service
Place of Birth
School Enrollment
Undergraduate Field of Degree[3]
Veteran Status[2]
Year of Entry

Economic Characteristics
Class of Worker
Commuting (Journey to Work)
Employment Status
Food Stamps/Supplemental Nutrition Assistance Program (SNAP)[4]
Health Insurance Coverage[2]
Income and Earnings
Industry and Occupation
Place of Work
Poverty Status
Work Status Last Year

Housing Characteristics
Computer and Internet Use[5]
House Heating Fuel
Kitchen Facilities
Occupancy/Vacancy Status
Occupants Per Room

Plumbing Facilities[6]
Rent
Rooms/Bedrooms
Selected Monthly Owner Costs
Telephone Service Available
Tenure (Owner/Renter)
Units in Structure
Value of Home
Vehicles Available
Year Householder Moved Into Unit
Year Structure Built

Demographics Characteristics
Age and Sex
Group Quarters Population
Hispanic or Latino Origin
Race
Relationship to Householder
Total Population

[1] Questions on Disability Status were significantly revised in the 2008 survey to cause a break in series.
[2] Marital History, Veterans' Service-Connected Disability Status and Ratings, and Health Insurance Coverage were added in the 2008 survey.
[3] Undergraduate Field of Degree was added in the 2009 survey.
[4] Food Stamp Benefit amount was removed in 2008.
[5] Computer and Internet Use was added to the 2013 survey.
[6] One of the components of Plumbing Facilities, flush toilet, and Business or Medical Office on Property questions were removed in 2016.
Source: U.S. Census Bureau.

Challenges

- Which variables to select
- How to aggregate several variables
- How to incorporate uncertainty of the estimate into the analysis

Data uncertainty: Guidance from the Census Bureau

Understanding and Using American Community Survey Data
What Researchers Need to Know

Issued March 2020

United States Census Bureau
U.S. Department of Commerce
U.S. CENSUS BUREAU
census.gov

Understanding and Using American Community Survey Data
What All Data Users Need to Know

Issued September 2020

United States Census Bureau
U.S. Department of Commerce
U.S. CENSUS BUREAU
census.gov

7. Understanding Error and Determining Statistical Significance

8. Calculating Measures of Error for Derived Estimates

https://www.census.gov/programs-surveys/acs/guidance/handbooks.html

Coefficients of variation

$$CV = \frac{\left(\frac{MOE}{1.645}\right)}{ESTIMATE} \times 100$$

- "A coefficient of variation (CV) measures the *relative* amount of sampling error that is associated with a sample estimate (…) and is usually expressed as a percent."

- "There are no hard-and-fast rules for determining an acceptable range of error in ACS estimates. Instead, data users must evaluate each application to determine the level of precision that is needed for an ACS estimate to be useful."

Source: U.S. Census Bureau, Understanding and Using American Community Survey Data: What All Data Users Need to Know, U.S. Government Publishing Office, Washington, DC, 2020, page 55.

Data uncertainty: Guidance from research community

CrossMark

Spatial Variation in the Quality of American Community Survey Estimates

David C. Folch[1] · Daniel Arribas-Bel[2] ·
Julia Koschinsky[3] · Seth E. Spielman[4]

Applied Geography

ELSEVIER

APPLIED GEOGRAPHY

CrossMark
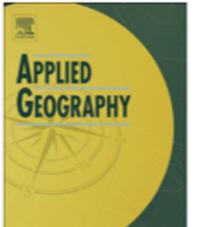
Patterns and causes of uncertainty in the American Community Survey

Seth E. Spielman [a,b,*], David Folch [b], Nicholas Nagle [c]

[a] Geography Department, University of Colorado, 110 Guggenheim Hall, Box 260 UCB, Boulder, CO 80309, USA
[b] Institute of Behavioral Science, University of Colorado, 110 Guggenheim Hall, Box 260 UCB, Boulder, CO 80309, USA
[c] Department of Geography, University of Tennessee, USA

SPATIAL
DEMOGRAPHY

RESEARCH

Handling Data Quality Information of Survey Data in GIS: A Case of Using the American Community Survey Data

David W. Wong[a] and Min Sun[a]
George Mason University

Computers, Environment and Urban Systems

ELSEVIER

Spatial aggregation as a means to improve attribute reliability

Min Sun, David W.S. Wong [1]

Figure 4. ArcGIS overlays the coefficients of variation (CVs) on ACS estimates using the ACS Mapping Extension function
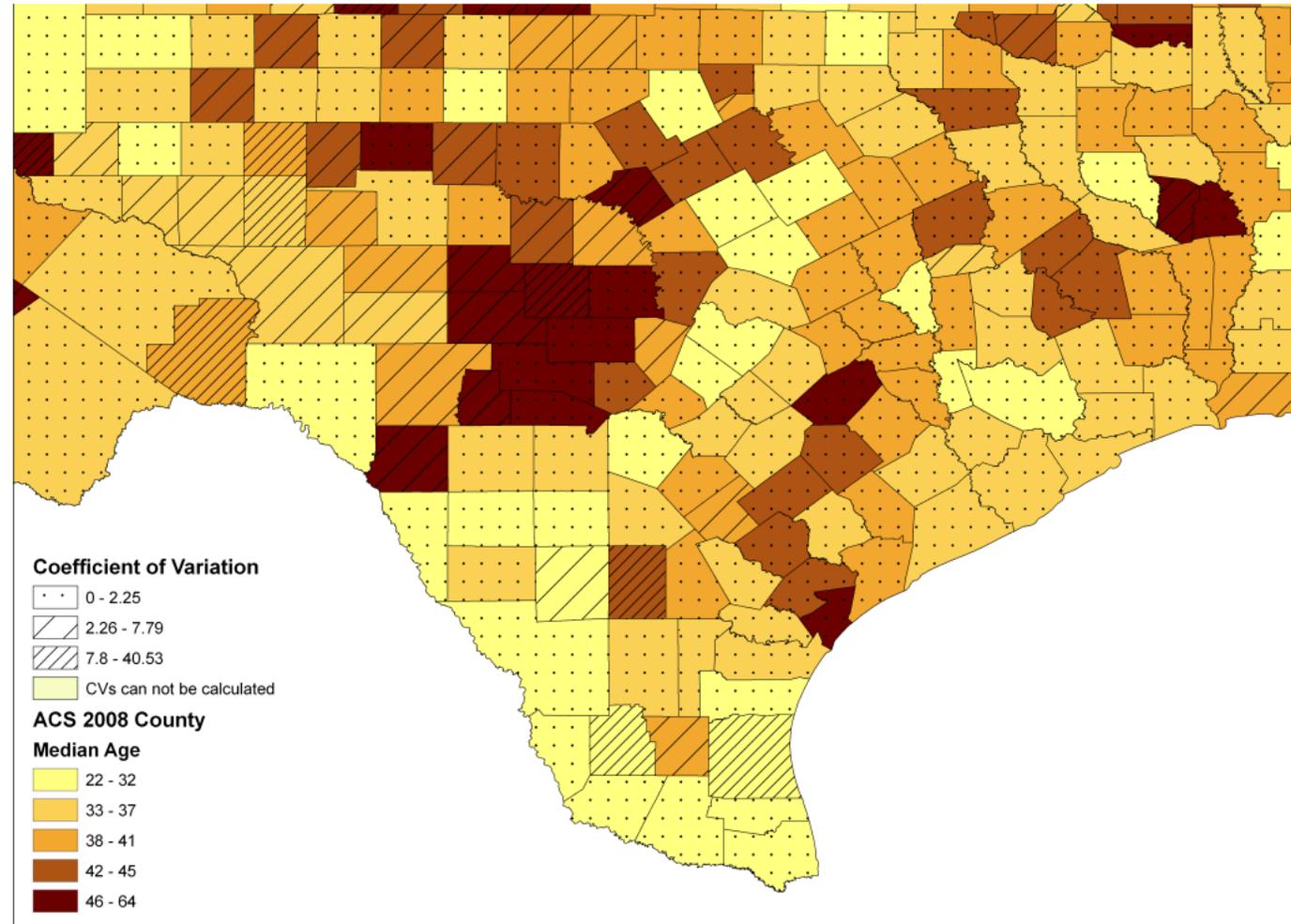
GIS-friendly ACS data sources
and instructional resources

ArcGIS Living Atlas of the World

386 Results

ACS Population Variables - Boundaries

Feature Layer By esri_demographics

This layer contains the most current release of data from the American Community Survey (ACS) about total population count by sex and age group. These are 5-year estimates shown by tract, county, and state boundaries.

Authoritative

ACS Race and Hispanic Origin Variables - Boundaries

Feature Layer By esri_demographics

This layer contains the most current release of data from the American Community Survey (ACS) about population broken down by race and Hispanic origin. These are 5-year estimates shown by tract, county, and state boundaries.

Authoritative
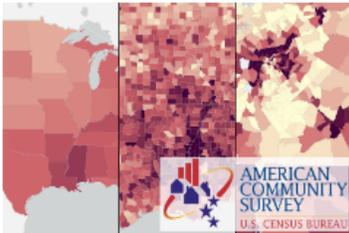
ACS Poverty Status Variables - Boundaries

Feature Layer By esri_demographics

This layer contains the most current release of data from the American Community Survey (ACS) about poverty status by age group. These are 5-year estimates shown by tract, county, and state boundaries.

Authoritative

ACS Median Household Income Variables - Boundaries

Feature Layer By esri_demographics

This layer contains the most current release of data from the American Community Survey (ACS) about median household income by race and by age of householder. These are 5-year estimates shown by tract, county, and state boundaries.

Authoritative

Great, but only the latest 5-year estimates are available
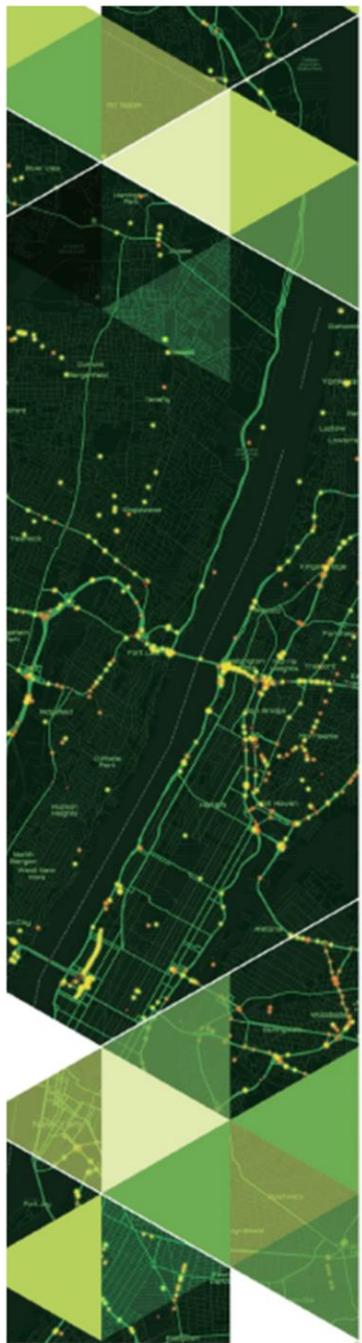
AN ESRI
TECHNICAL PAPER

MARCH 2021

Methodology Statement: 2015–2019 American Community Survey

esri | THE SCIENCE OF WHERE

The reliability is based on thresholds that Esri has established based on the usability of the estimates. Users should be aware that these are generalized thresholds:

- ■ 📊 High Reliability: Small CVs (less than or equal to 12 percent) are flagged green to indicate that the sampling error is small relative to the estimate, and the estimate is reasonably reliable.[5]

- ■ 📊 Medium Reliability: Estimates with CVs between 12 and 40 are flagged yellow—use with caution.

- ■ 📊 Low Reliability: Large CVs (over 40 percent) are flagged red to indicate that the sampling error is large relative to the estimate. The estimate is considered very unreliable.

[5] National Research Council, *Using the American Community Survey: Benefits and Challenges* (Washington, DC: The National Academies Press, 2007).

DOWNLOAD U.S. CENSUS DATA TABLES & MAPPING FILES

IPUMS NHGIS

ABOUT

REGISTER

DONATE TO NHGIS

DATA

BROWSE AND SELECT DATA

DOWNLOAD OR REVISE MY DATA

API

The **National Historical Geographic Information System** (NHGIS) provides easy access to summary tables and time series of population, housing, agriculture, and economic data, along with GIS-compatible boundary files, for years from 1790 through the present and for all levels of U.S. census geography, including states, counties, tracts, and blocks. *Read more*.

https://www.nhgis.org/

https://usa.ipums.org/usa/

Estimating and Mapping Reliability for American Community Survey Data

Written by Barbara M. Parmenter and Janet Lau, revised 2/14/2013

Source: Tufts University GIS Center

Examples of student work

American Community Survey 2016 Estimates for NYC Black Population
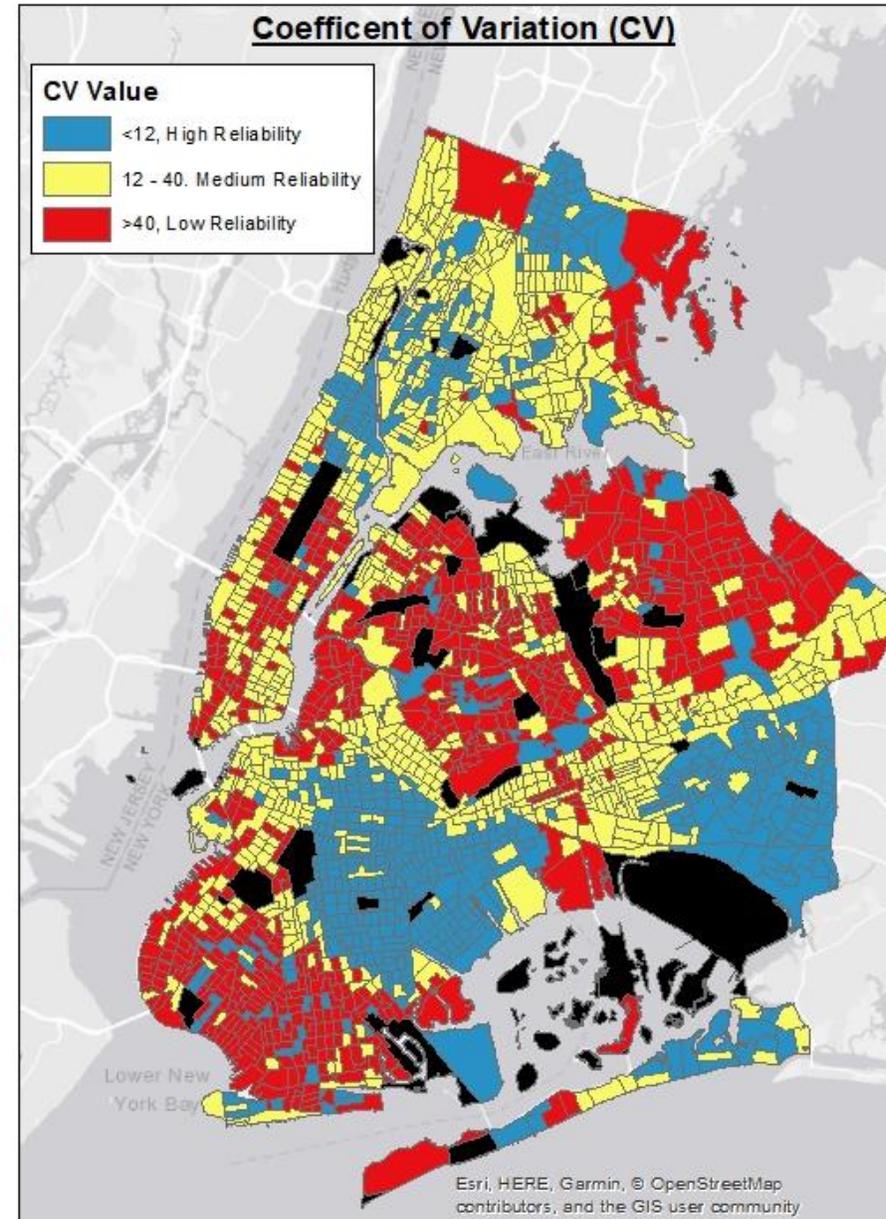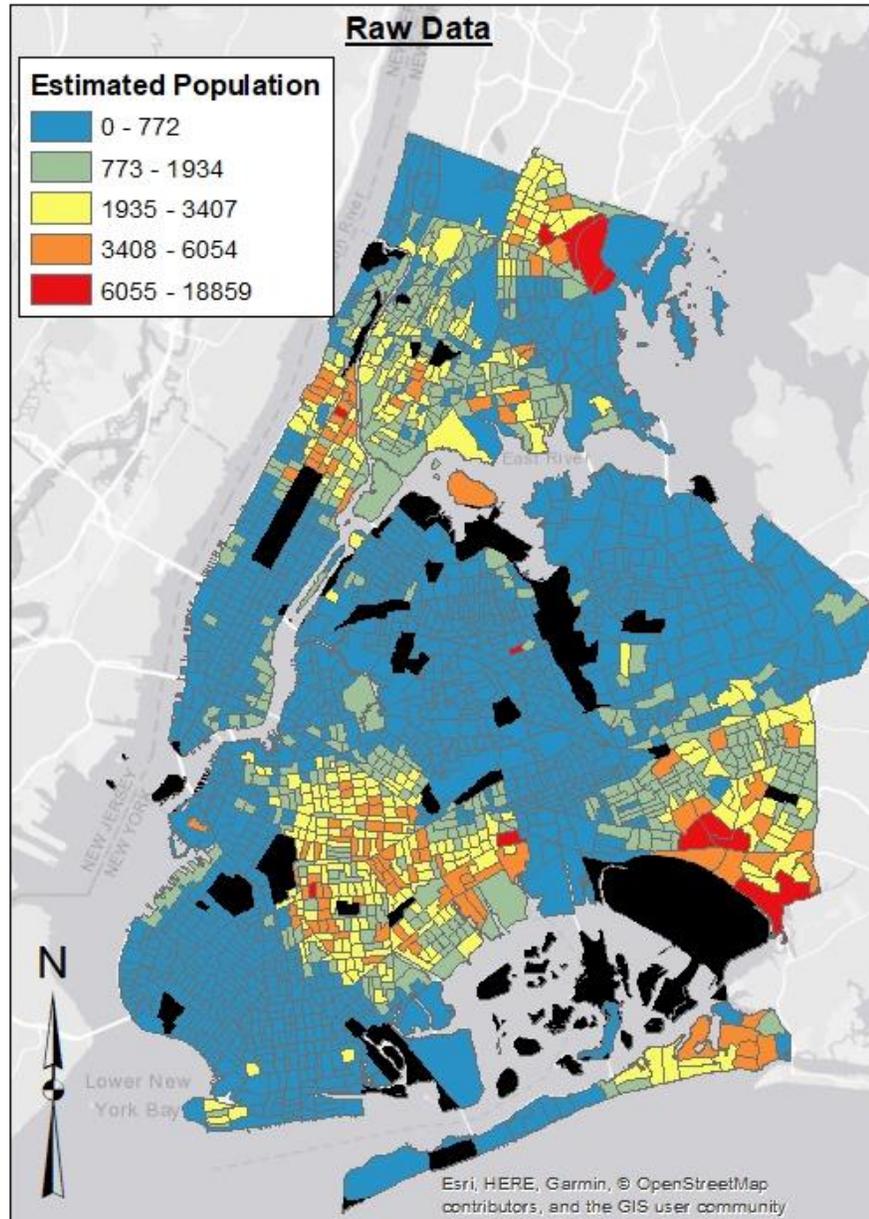
Raw Data

Estimated Population
- 0 - 772
- 773 - 1934
- 1935 - 3407
- 3408 - 6054
- 6055 - 18859

Coefficent of Variation (CV)

CV Value
- <12, High Reliability
- 12 - 40. Medium Reliability
- >40, Low Reliability

Esri, HERE, Garmin, © OpenStreetMap contributors, and the GIS user community

N

0 1 2 4 6 8 Miles

Uninhabitated   Census Tract

Source: American Community Survey 2016 Population Estimates
Projection: Lambert Conformal Conic

American Community Survey 2016 Estimates for NYC Hispanic Population
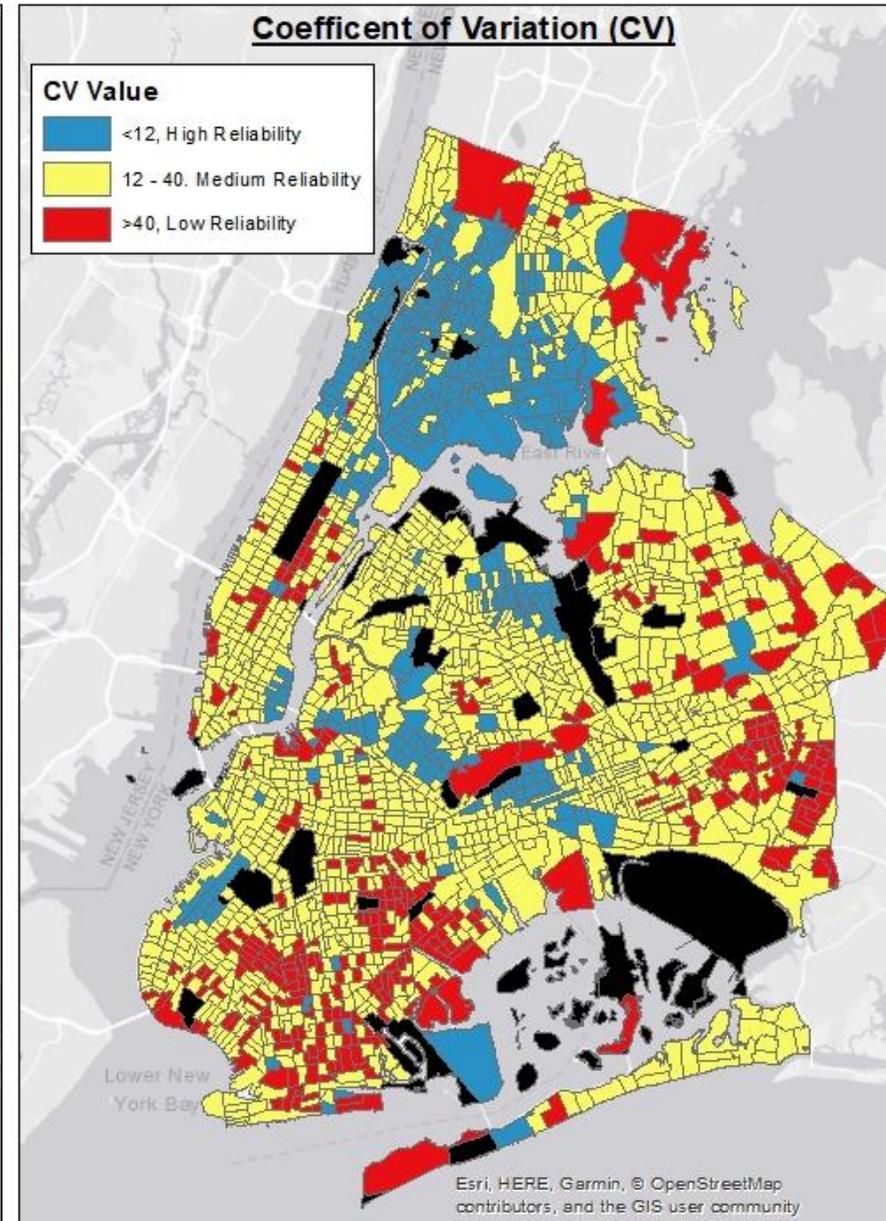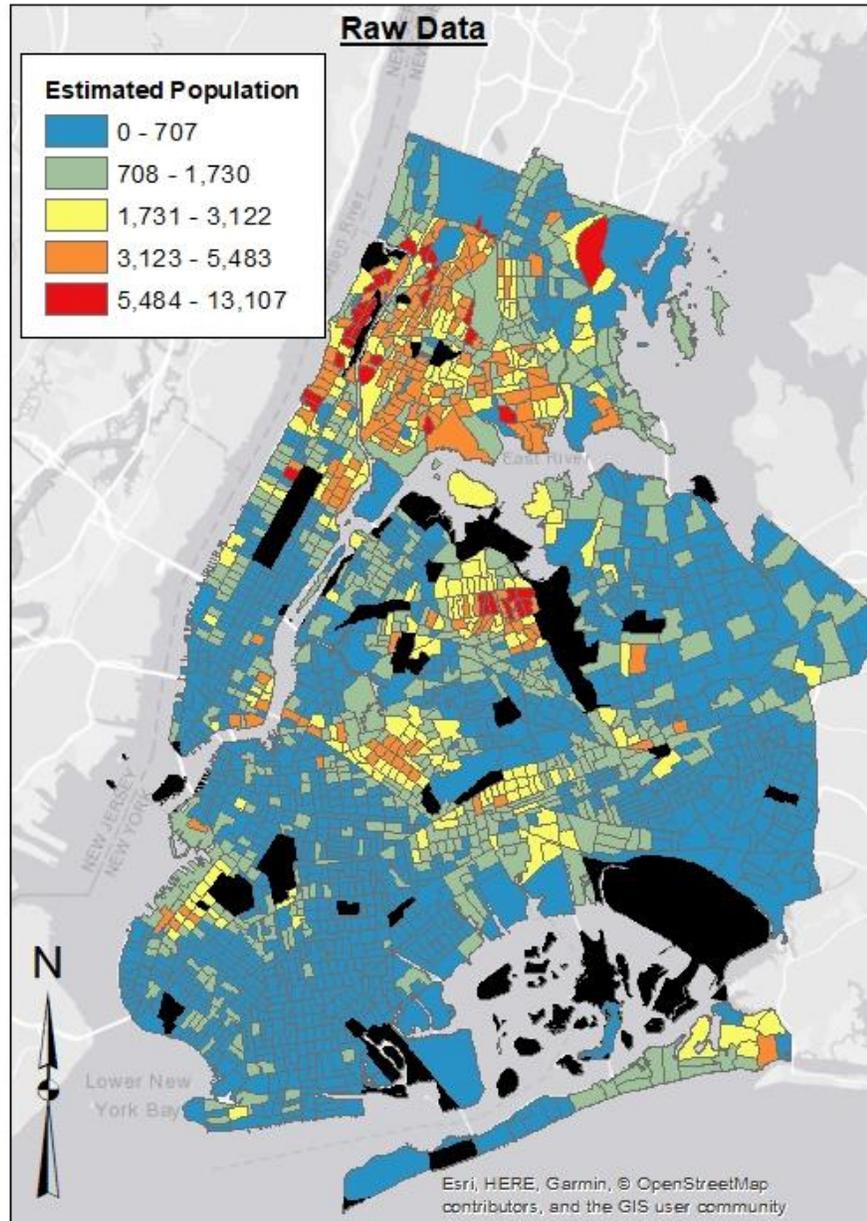
Raw Data

Estimated Population

- 0 – 707
- 708 – 1,730
- 1,731 – 3,122
- 3,123 – 5,483
- 5,484 – 13,107

N

Lower New
York Bay

Esri, HERE, Garmin, © OpenStreetMap
contributors, and the GIS user community

Coefficent of Variation (CV)

CV Value

- <12, High Reliability
- 12 - 40. Medium Reliability
- >40, Low Reliability

Lower New
York Bay

Esri, HERE, Garmin, © OpenStreetMap
contributors, and the GIS user community

0 1 2   4   6   8
Miles

Uninhabitated   Census Tract

Source: American Community Survey 2016 Population Estimates
Projection: Lambert Conformal Conic

Disability Status

Population Identified as Living with One Disability

Languages Spoken

Residents 5+ Who Speak an Asian and Pacific Islander Language at Home

Percent of Population with 1 Disability

- ≤3.39
- ≤5.52
- ≤8.54
- ≤13.77
- ≤25.53

Unreliable Data

Worcester Towns

Worcester Tracts

Worcester County Boundary

Source: Data from ACS 2018, Boundaries from MassGIS
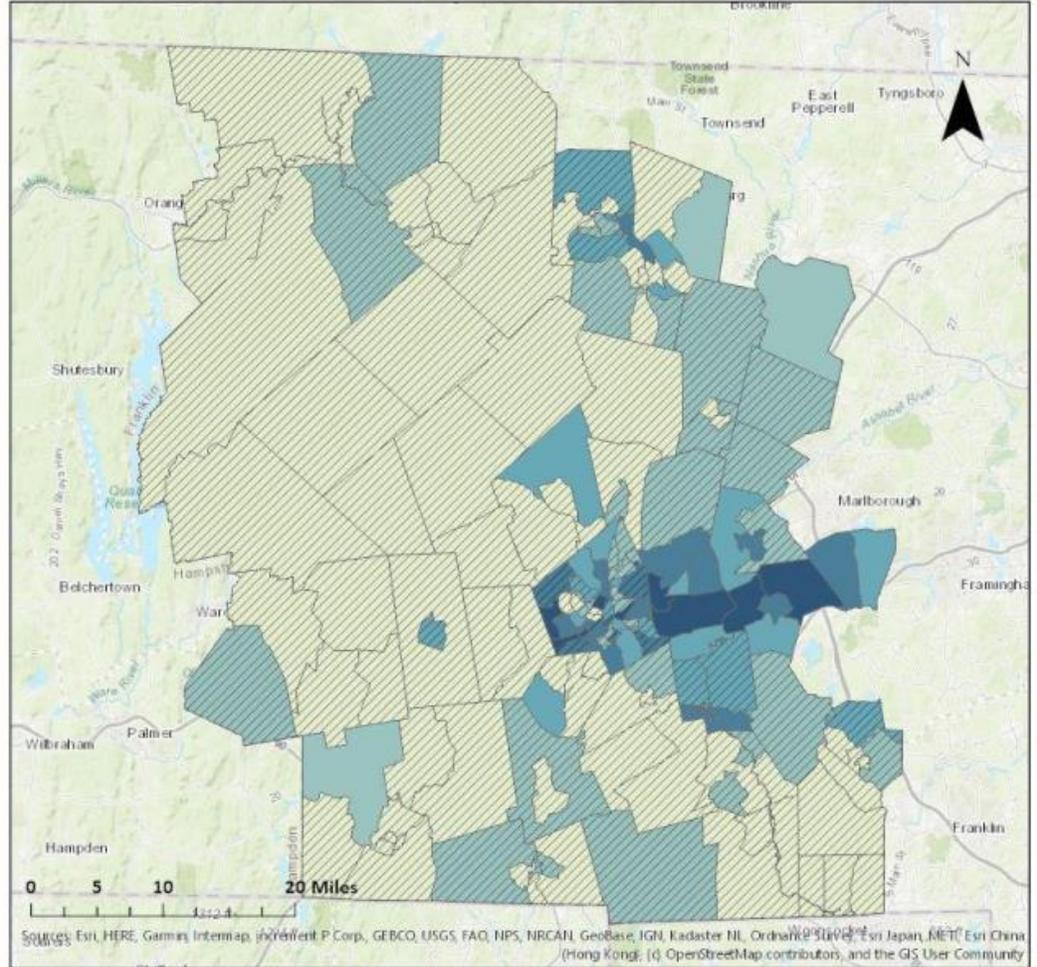
Percent of Population Age 5+
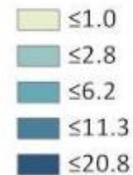
- ≤1.0
- ≤2.8
- ≤6.2
- ≤11.3
- ≤20.8

Unreliable Data

Worcester Towns

Worcester Tracts

Worcester County Boundary

Source: Data from ACS 2018, Boundaries from MassGIS

Education Attainment

Residents 25+ whose Highest Education Completed is Less Than High School

Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), (c) OpenStreetMap contributors, and the GIS User Community
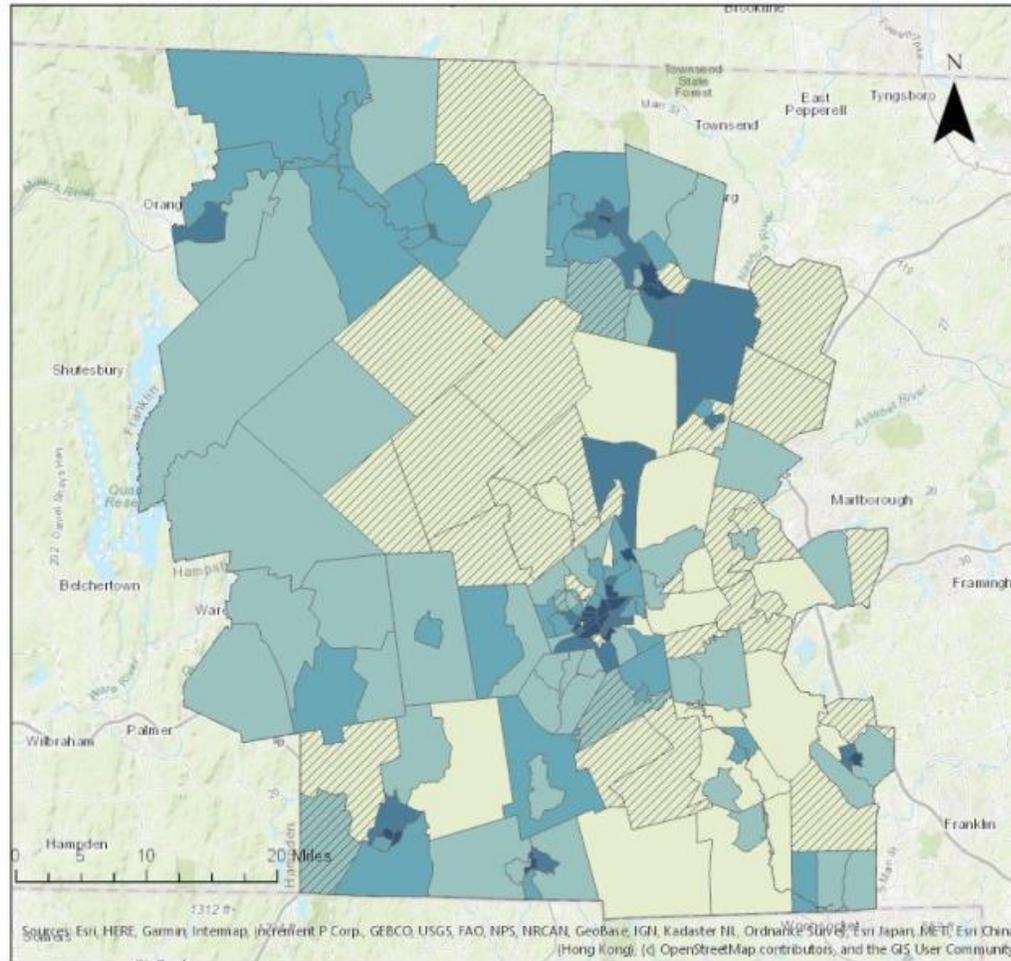
Percent of Population 25+
- ≤4.1
- ≤7.8
- ≤12.8
- ≤19.3
- ≤33.0

- Unreliable Data
- Worcester Towns
- Worcester Census Tracts
- Worcester County Boundary

Source: ACS Living Atlas, Census Data and MassGIS

Financial Status

Residents whose Income in the last 12 Months is Below the Poverty Level

Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), (c) OpenStreetMap contributors, and the GIS User Community
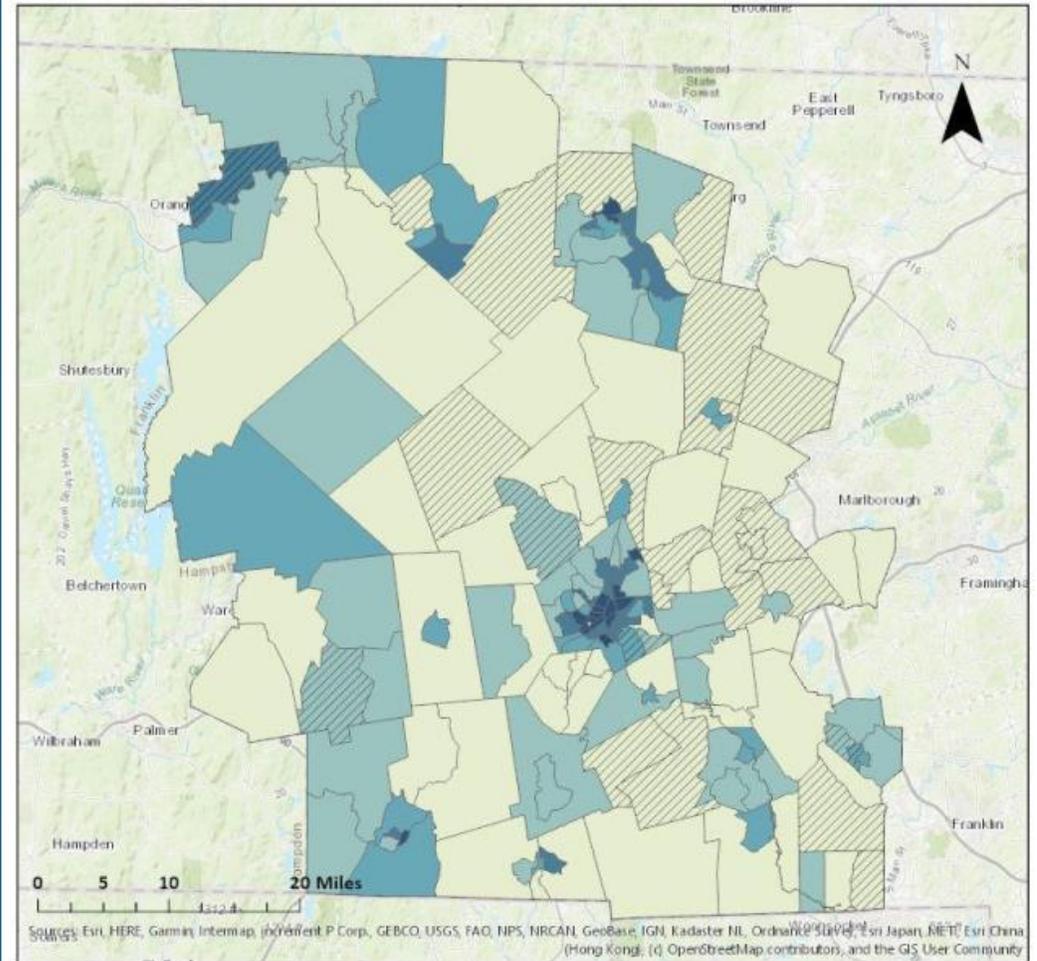
Percent of Population
- ≤5.5
- ≤10.3
- ≤16.9
- ≤29.5
- ≤63.6

- Unreliable Data
- Worcester Towns
- Worcester Tracts
- Worcester County Boundary

Source: Data from ACS 2018, Boundaries from MassGIS

Discussion

- Using ACS data in a GIS course – what are you experiences?

- What instructional resources do you find useful?

- How can ACS data users become more aware of the reliability issues?