

An Introduction to Tidy R Spatial Packages: Incorporating Historic Sociodemographic Data from the US Census and Visualizing Geographic Distributions

Stephen Uong, MPH, PhD(c)

Department of Epidemiology, Mailman School of Public Health, Columbia University

Christina A. Mehranbod, MPH, PhD(c)

Department of Epidemiology, Mailman School of Public Health, Columbia University

Agenda



Introduction (15 min)

Census Data
Spatial Data
R for Spatial Data



R tutorial (30 min)

Importing Census data
Joining tabular data
Exploring data
Mapping data



Questions (5 min)

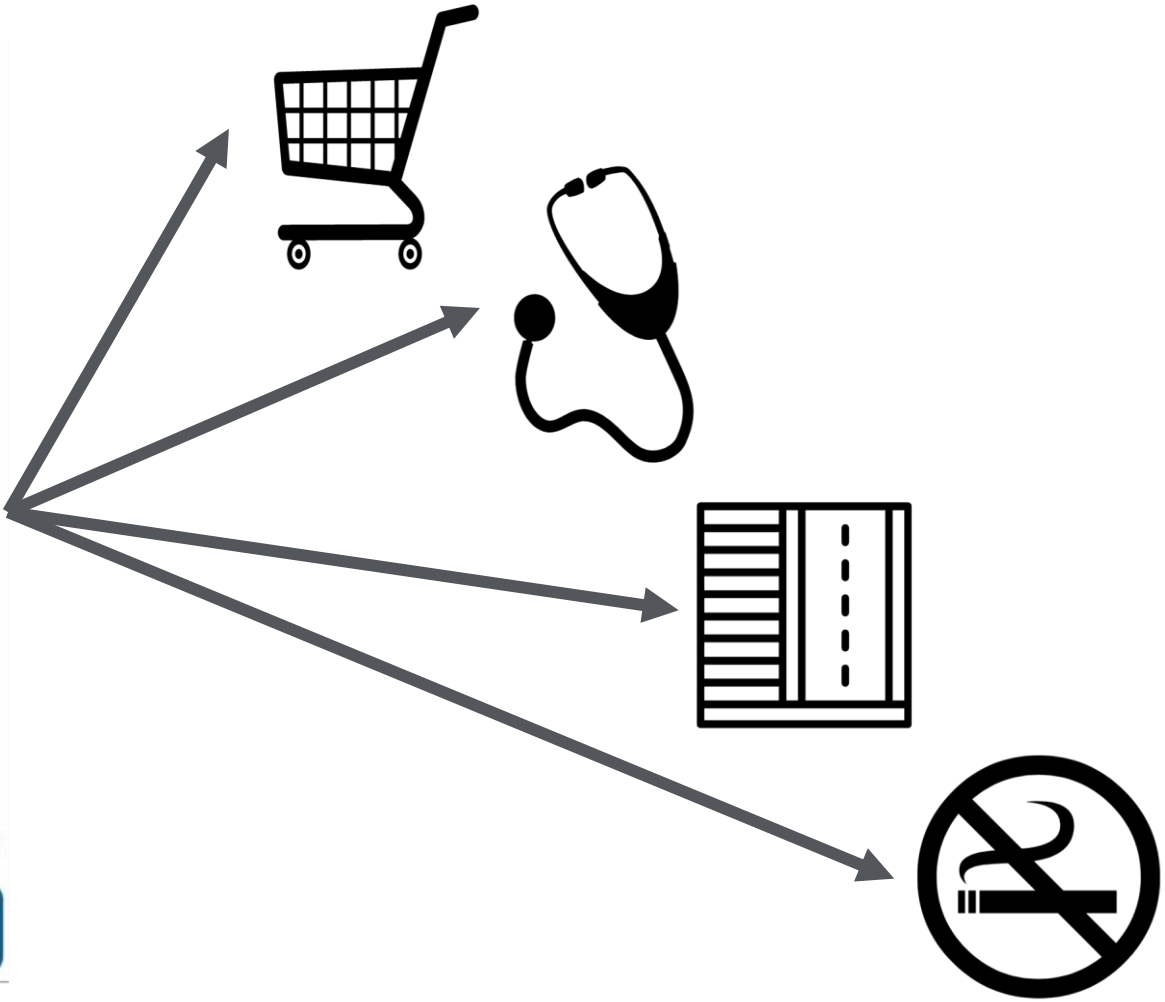
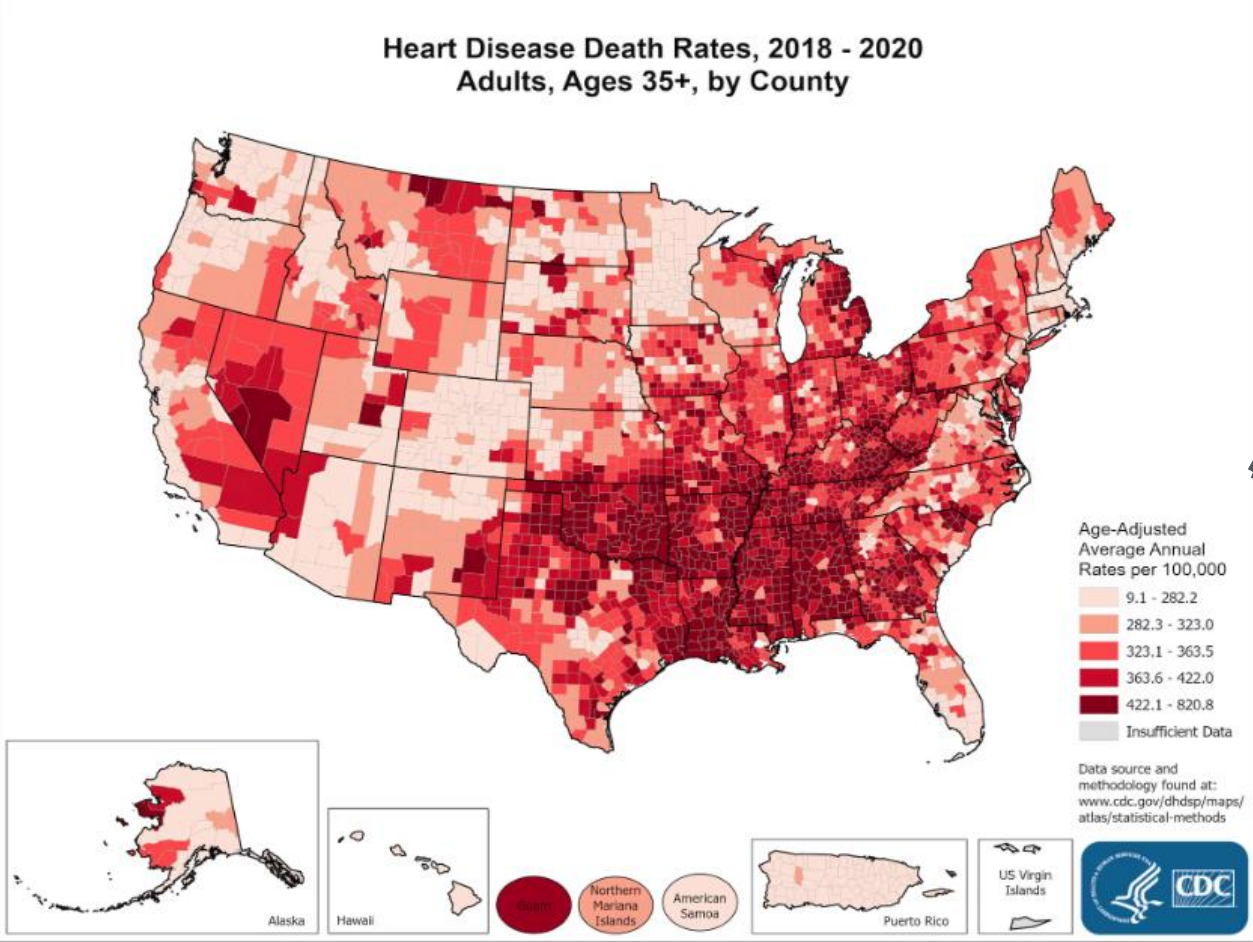
Social determinants of health

“Social determinants of health (SDOH) are the conditions in the environments where people are born, live, learn, work, play, worship, and age that affect a wide range of health, functioning, and quality-of-life outcomes and risks.”



Source: Healthy People 2030, U.S. DHHS

Geographic location impacts health



Source: CDC, 2022

Introduction to the Census

The United States Census

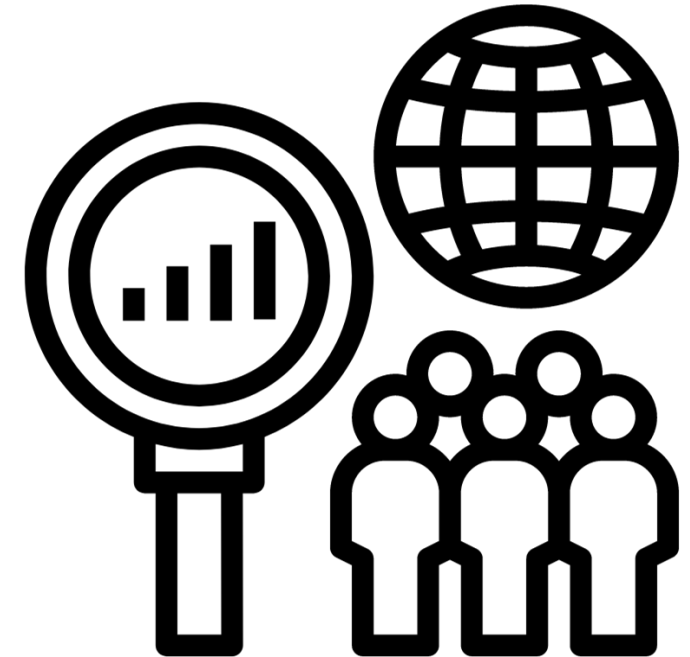
Population count

Mandated by the
Constitution

Critical for
representation and
funding

Confidentiality and
privacy

Evolution and
innovation



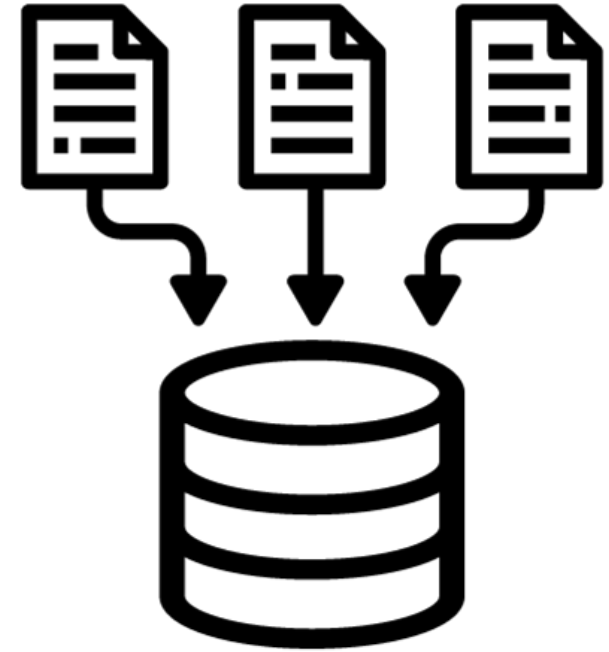
Census Data Collection Techniques



Mail canvass



Internet collection



Central collection
from state sources

Decennial Census

- Conducted every 10 years
- Counts every person
- Less comprehensive, shorter set of questions

OMB No. 0607-1006; Approval Expires 11/30/2021

U.S. DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. CENSUS BUREAU

United States[®]
**Census
2020**

This is the official questionnaire for this address.
It is quick and easy to respond, and your answers are protected by law.

Para completar el cuestionario en español, dele la vuelta y complete el lado verde.

FOR
OFFICIAL
USE ONLY

Start here OR go online at my2020census.gov to complete your 2020 Census questionnaire.

Use a blue or black pen.

Before you answer Question 1, count the people living in this house, apartment, or mobile home using our guidelines.

- Count all people, including babies, who live and sleep here most of the time.
- If no one lives and sleeps at this address most of the time, go online at my2020census.gov or call the number on page 8.

The census must also include people without a permanent place to live, so:

- If someone who does not have a permanent place to live is staying here on April 1, 2020, count that person.

The Census Bureau also conducts counts in institutions and other places, so:

- Do not count anyone living away from here, either at college or in the Armed Forces.
- Do not count anyone in a nursing home, jail, prison, detention facility, etc., on April 1, 2020.
- Leave these people off your questionnaire, even if they will return to live here after they leave college, the nursing home, the military, jail, etc. Otherwise, they may be counted twice.

1. How many people were living or staying in this house, apartment, or mobile home on April 1, 2020?

Number of people =

2. Were there any additional people staying here on April 1, 2020 that you did not include in Question 1?

Mark ☒ all that apply.

☐ Children, related or unrelated, such as newborn babies, grandchildren, or foster children

☐ Relatives, such as adult children, cousins, or in-laws

☐ Nonrelatives, such as roommates or live-in babysitters

☐ People staying here temporarily

☐ No additional people

3. Is this house, apartment, or mobile home — Mark ☒ ONE box.

☐ Owned by you or someone in this household with a mortgage or loan? Include home equity loans.

☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?

☐ Rented?

☐ Occupied without payment of rent?

4. What is your telephone number?

We will only contact you if needed for official Census Bureau business.

Telephone Number

- -

FORM **DI-Q1(E/S)** (05-31-2018)

11100013

Source: US Census Bureau, 2024

American Community Survey

- Conducted every year
- Sent to a *sample* of addresses across the US
- Comprehensive, more questions than the Census
- Topics such as education, housing, employment, etc.

United States[®]
Census
Bureau

American Community Survey

This booklet shows the content of the American Community Survey questionnaire.

Start Here

You have two ways to respond:

Respond online today at: **respond.census.gov/acs**

OR

Complete this form and mail it back as soon as possible.

Your response is required by law.

The American Community Survey is conducted by the U.S. Census Bureau. This survey is one of only a few surveys for which all recipients are required by law to respond. The U.S. Census Bureau is required by law to protect your information.

If you need help or have questions about completing this form, please call 1-800-354-7271.

¿NECESITA AYUDA? Llame al 1-877-833-5625.

For more information about the American Community Survey, visit our website at: census.gov/acs

Please print the name and telephone number of the person who is filling out this form. We will only contact you if needed for official Census Bureau business.

Last Name

First Name MI

Area Code + Number -

How many people are living or staying at this address?

- **INCLUDE** everyone who is living or staying here for more than 2 months.
- **INCLUDE** yourself if you are living here for more than 2 months.
- **INCLUDE** anyone else staying here who does not have another place to stay, even if they are here for 2 months or less.
- **DO NOT INCLUDE** anyone who is living somewhere else for more than 2 months, such as a college student living away or someone in the Armed Forces on deployment.

Number of people

Fill out pages 2-7 for everyone, including yourself, who is living or staying at this address for more than 2 months. Then complete the rest of the form.

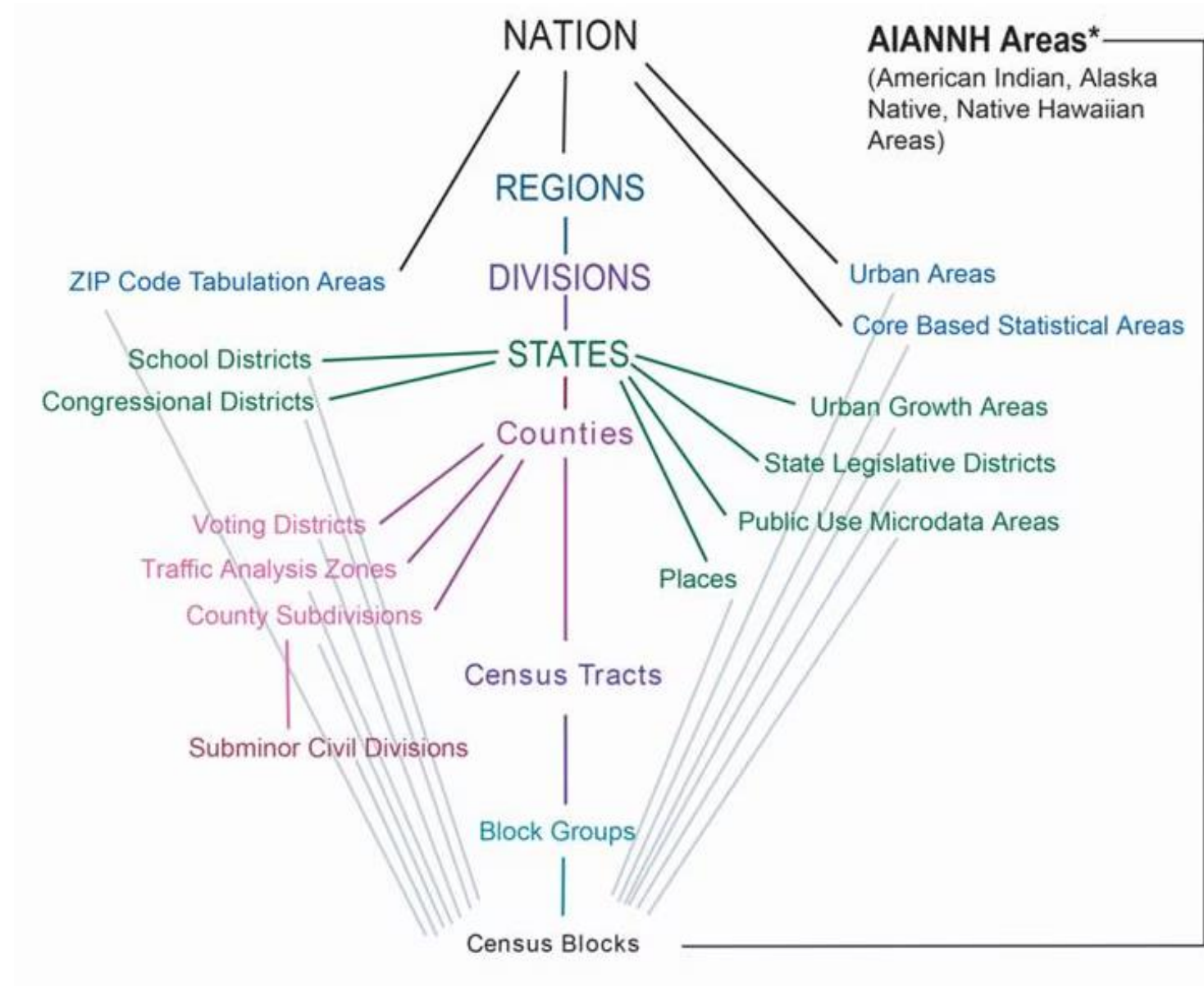
FORM **ACS-1(INFO)(2024)**
(05-15-2023)

OMB No. 0607-0810
OMB No. 0607-0936



Source: US Census Bureau, 2024

Census Geographies



Source: US Census Bureau, 2020

Introduction to Geospatial Data

Types of Geospatial Data

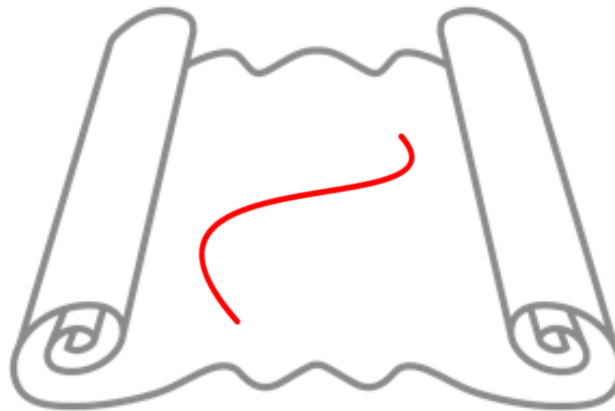
POINTS

Example: GPS coordinates



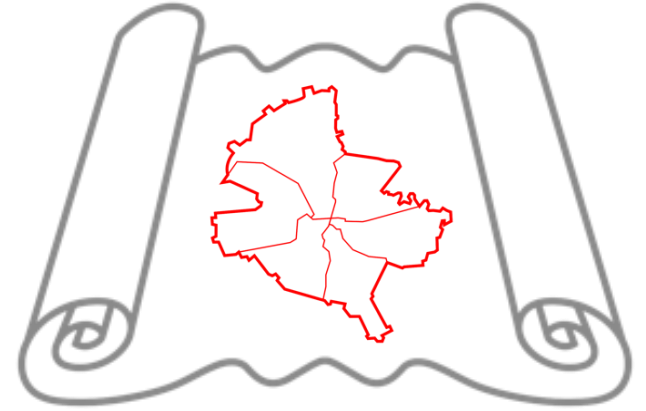
LINES

Example: roads, rivers



POLYGONS

Example: neighborhoods, Census areas



Common Spatial File Types

Shapefiles (Esri):

- .shp: geometry (e.g., polygons)
- .shx: index
- .dbf: attributes (i.e., tabular data)
- Others (not required): .proj (projections)

Other file types:

- .geojson/.json: geographic json
- .gml: geography markup language



Types of Joins

SPATIAL JOIN

Spatial data + spatial data

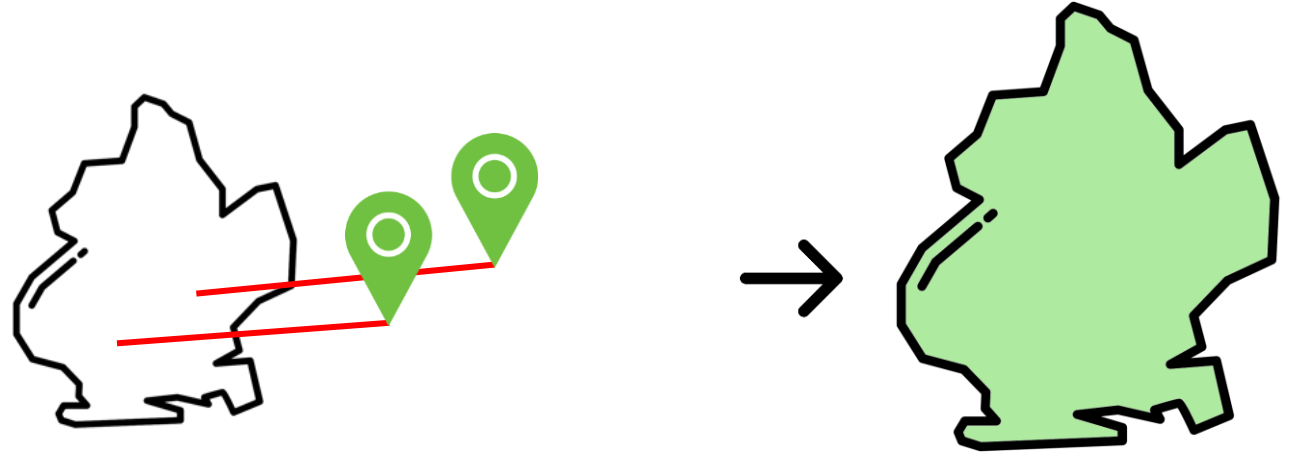
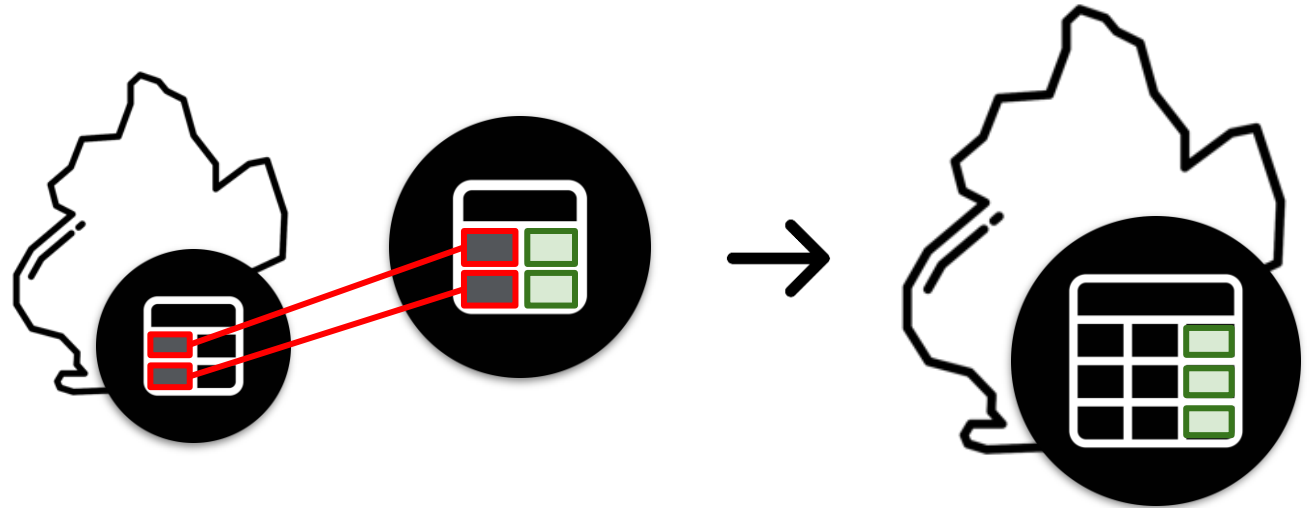


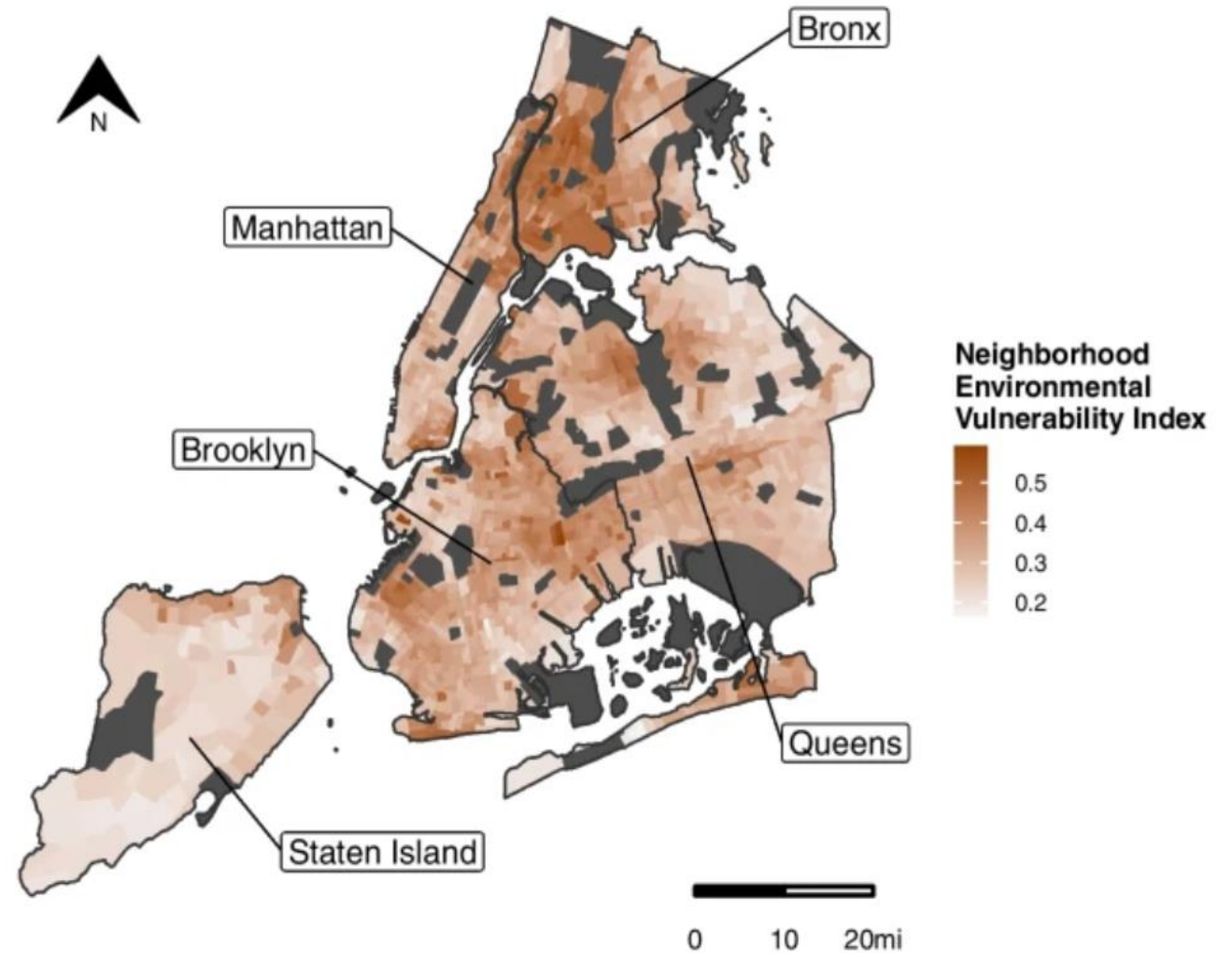
TABLE (TABULAR) JOIN

Spatial data + tabular data



Choropleth Maps

Uses **colors** or **shading** to show variation over geographic areas



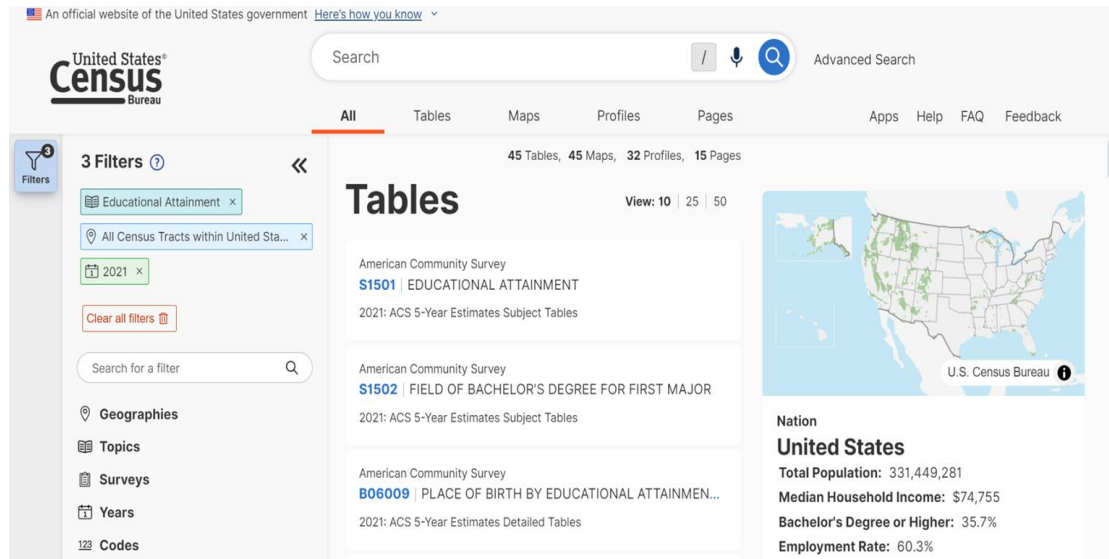
Source: Uong et al., 2023

R for Spatial Data

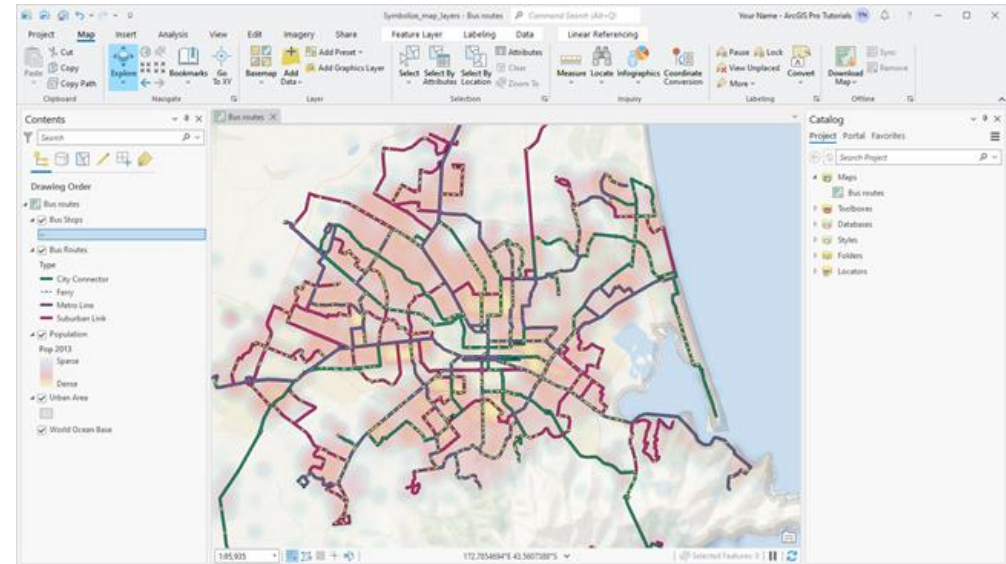
Traditional Approaches

Point-and-click interface: limited replicability.

DOWNLOADING CENSUS DATA



CREATING MAPS



Sources: US Census Bureau, 2024 (left) and ESRI, 2024 (right)

Using R for Spatial Data

Coding in R allows these processes to be **streamlined**, **replicable**, and **transparent**

DOWNLOADING CENSUS DATA

```
acs_data <- get_acs(geography = "tract",
  variables = c(hhincome = "B06011_001",
    total_education = "B06009_001",
    education_less_than_hs = "B06009_002",
    education_hs = "B06009_003",
    education_some_college = "B06009_004",
    education_bachelors = "B06009_005"),
  state = "NY",
  year = 2021,
  output = "wide")
```

CREATING MAPS

```
smap <- merged %>%
  ggplot() +
  ggplot2::geom_sf() +
  geom_sf(aes(fill = perc_hs)) +
  scale_fill_continuous(low = 'white', high = 'blue') +
  theme_void() +
  labs(title = 'High School Educational Attainment in New York') +
  ggspatial::annotation_north_arrow(location = 'tl') +
  ggspatial::annotation_scale(unit_category = 'imperial')
```

R Tutorial

First, you need a Census API Key

https://api.census.gov/data/key_signup.html



Request a U.S. Census Data API Key

Organization Name

Email Address

☐ I agree to the [terms of service](#)

REQUEST KEY

Health Outcome

Coronary Heart Disease



Source: CDC, 2021