

An An Interactive Introduction to R Spatial Packages: Incorporating Historic Sociodemographic Data from the US Census and Population Life Expectancy in Static and Animated Maps

Stephen Uong, MPH, PhD(c)

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

Slides adapted from a previous presentation co-created with Christina Mehranbod.

Workshop Schedule

Introduction and Setup (~ 1 hour)

Introduction to Census Data (~ 30 minutes)

Break (~ 15 minutes)

Introduction to Spatial Data (~ 1 hour)

Break (~ 15 minutes)

Mapping (~ 1 hour)

Questions (~ 5 minutes)

Learning Objectives

At the end of this workshop, participants will be able to:

- Recognize the basics of Census and spatial data.
- Apply GIS concepts to import, clean, and export spatial data.
- Create maps in R, using best practices for mapping and writing tidy R code.

Introduction

Presenter: Stephen Uong

- **Current role:** PhD candidate in Epidemiology, Columbia University Mailman School of Public Health
- **Work experience:** Kaiser Permanente, Johnson & Johnson, CDC, Council for State and Territorial Epidemiologists
- **Interests:**
 - Spatial epidemiology, urban planning and health
 - Machine learning applications in health equity
 - Health equity: sexual and gender minority, immigrant health
 - Mental health, injury, violence
 - R

I love R and GIS!



Why Geospatial Data for Population Health Sciences?

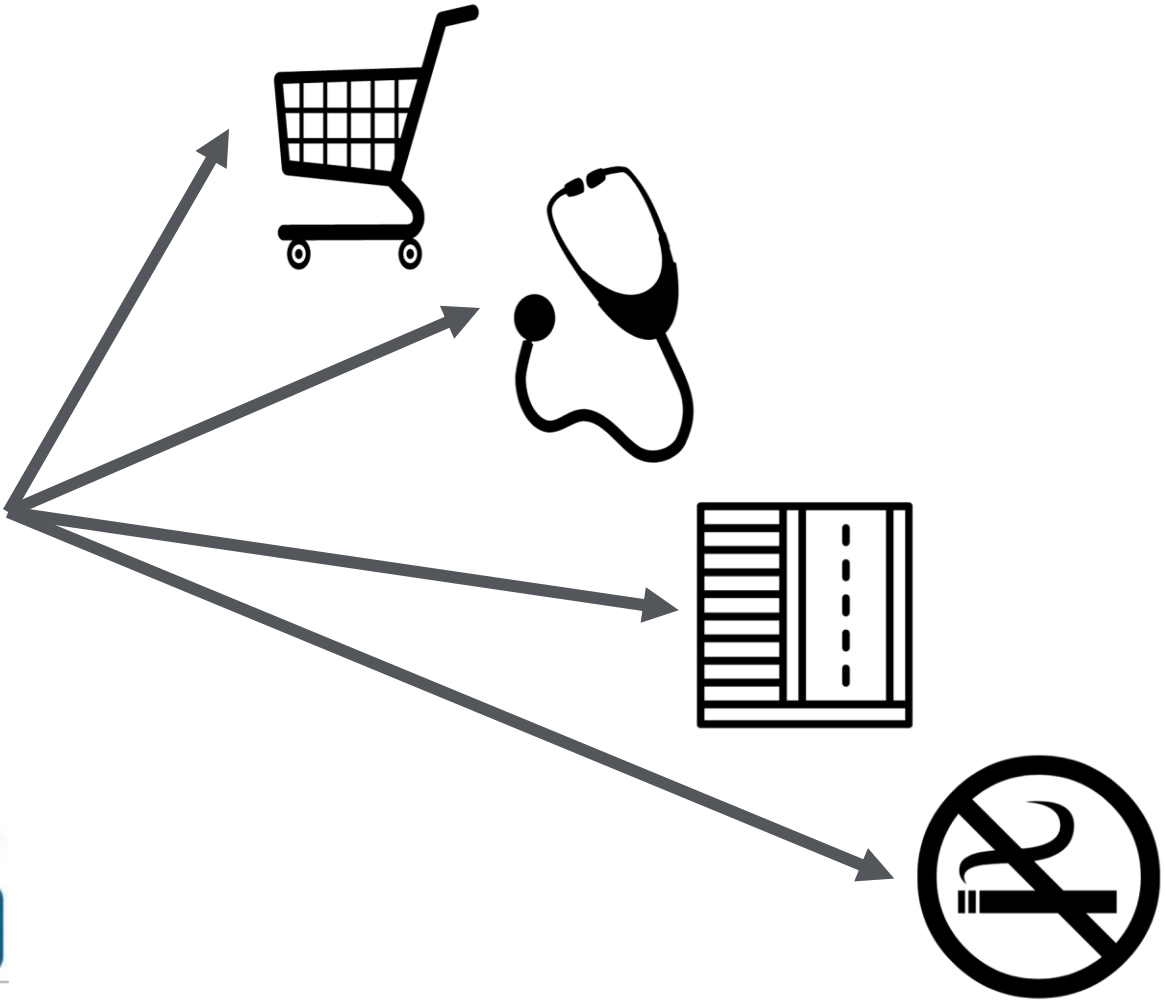
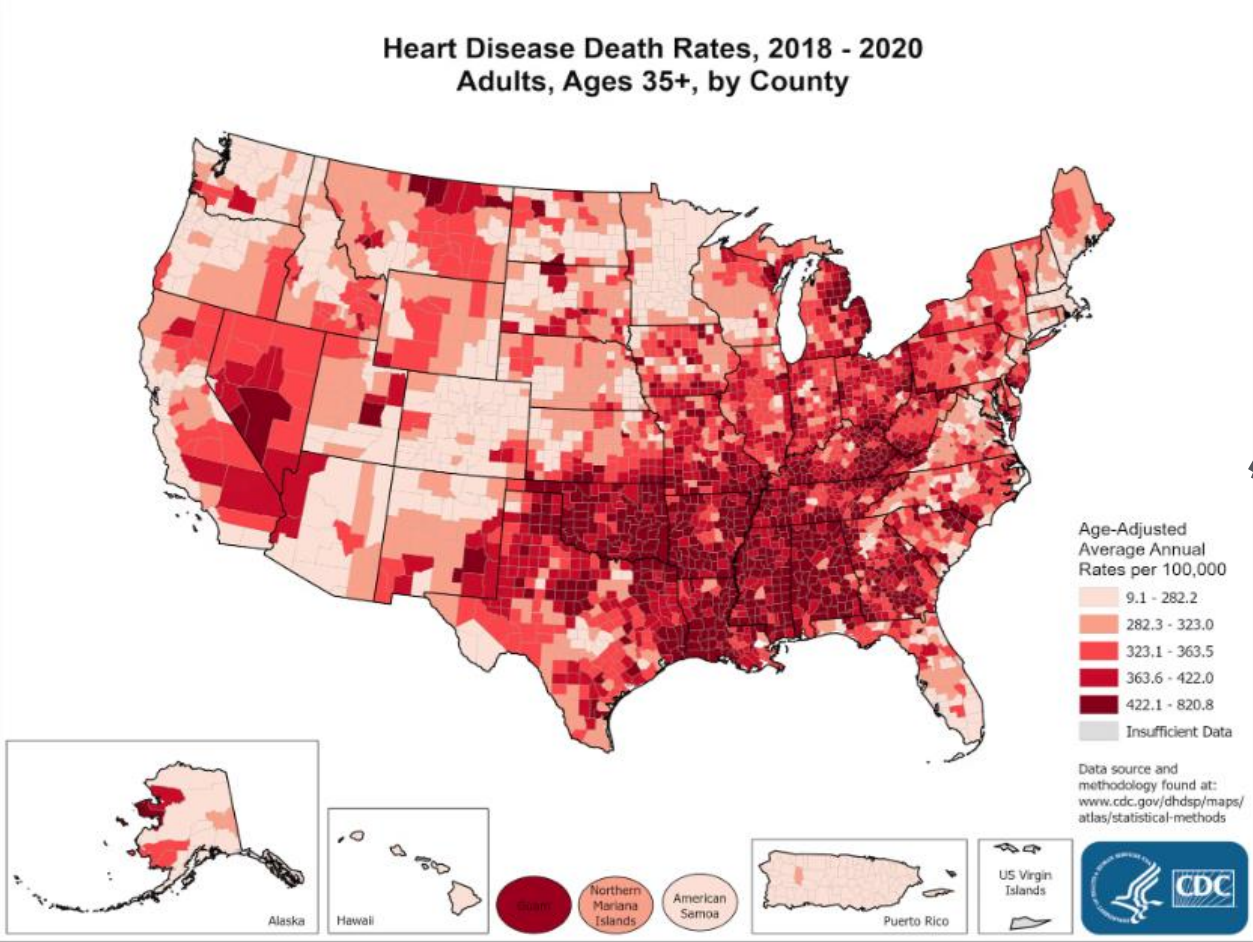
Social determinants of health

“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



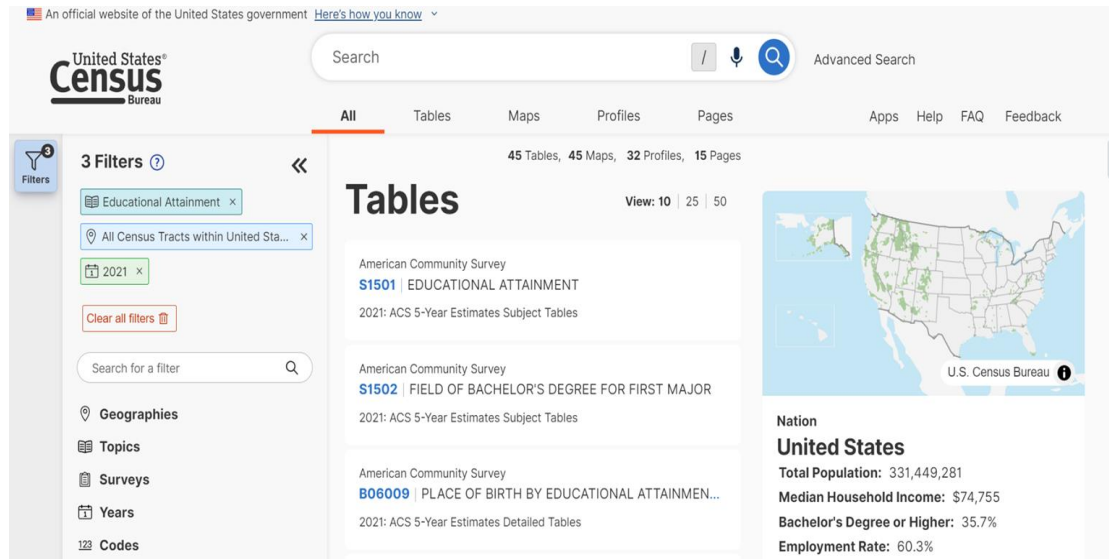
Source: CDC, 2022

Why use R for Spatial Data?

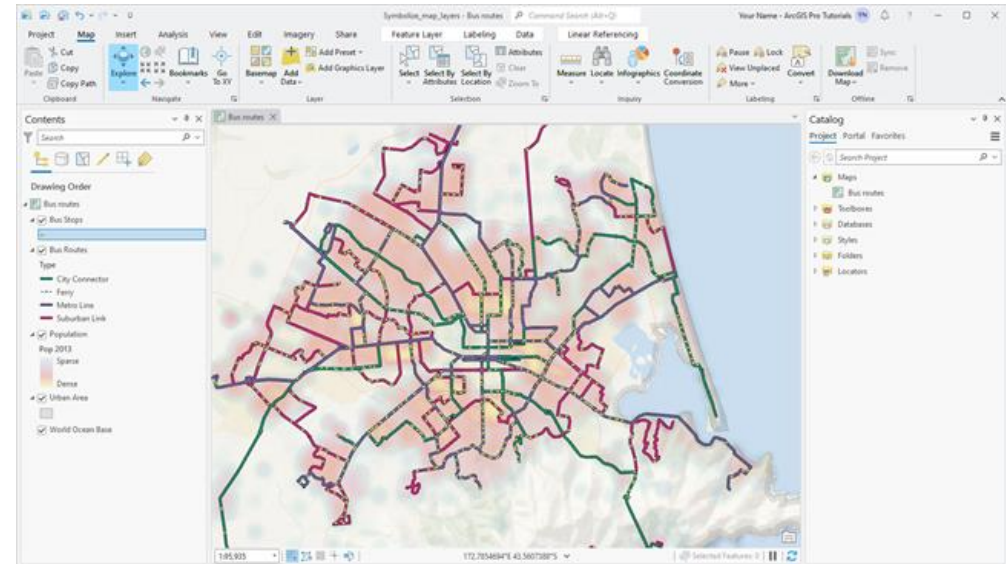
Traditional Approaches

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')
```

Tools Used in this Workshop

R and RStudio

We will learn geoprocessing using R and RStudio.

- **R:** Programming language and software environment used for statistical computing, data analysis, and visualization.
- **RStudio:** Integrated development environment designed to make working with R easier by providing a user-friendly interface and tools



Files are Available through Github



Git: version control system that tracks changes in files and allows multiple people to collaborate on a project without overwriting each other's work.



GitHub: an online platform for hosting Git repositories, enabling collaboration and code sharing over the internet.

R Tools: Tidyverse and Quarto

Tidyverse: A collection of R packages designed for data science that share a common philosophy, syntax, and data structures.

Quarto: Open-source publishing system that allows you to create reports, presentations, blogs, books, and more using R and other programming languages.

- Quarto Markdown is an extension of Rmarkdown (.qmd instead of .rmd files)



Census API

API = Application Programming Interface

- A way to access the Census data in a program.
- API key: Code that identifies and authenticates a user or application when it tries to access. Basically a password

Data Used in this Workshop

***US Census
American
Community Survey***



Source: CDC, 2021

Health Outcome

Life expectancy



Source: CDC, 2021

Setup

Request 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

Clone the git repository

Go to our git repository:

<https://github.com/phispu/workshop-2024-iaphs-rspatial>

Clone the git repository:

1. On top of this page, click on `Code` and copy the link to this git repository.
2. Open RStudio.
3. In RStudio, click on `File` → `New Project...` → `Version Control` → `Git`
4. Under "Repository URL", paste the link of the git repository.
5. Under "Project directory name", name your project directory.
6. Under "Create project as subdirectory of:", select the folder to save the files.
7. Click on `Create Project` when you are done to clone your repository!

Introduction to U.S. Census Data

The United States Census

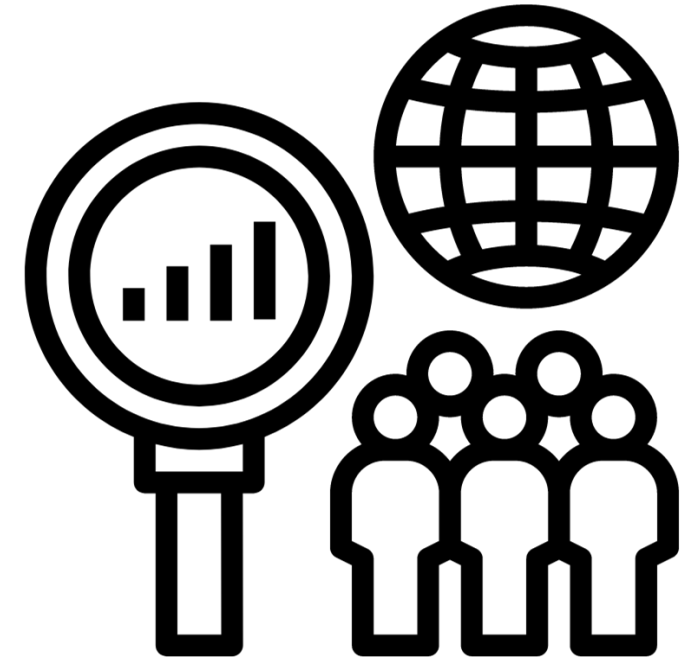
Population count

Mandated by the
Constitution

Critical for
representation and
funding

Confidentiality and
privacy

Evolution and
innovation



Census Data Collection Techniques



Mail canvas



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.

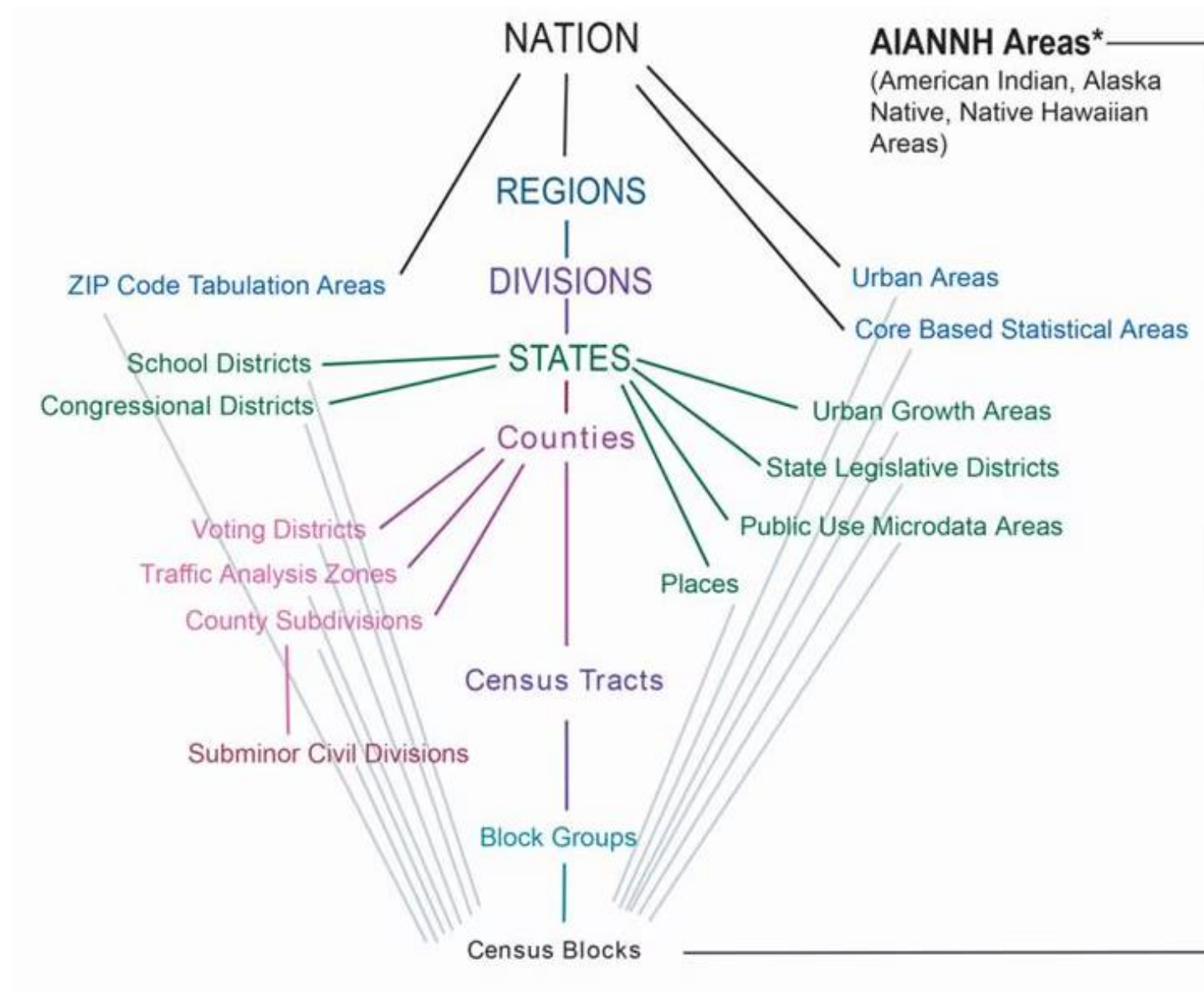
FORM **ACS-1(INFO)(2024)**
(05-15-2023)

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



Source: US Census Bureau, 2024

Census Geographies



Source: US Census Bureau, 2020

The Anatomy of a FIPS Code

Unique identifiers for Census geographies



Source: https://customer.precisely.com/s/article/US-Census-definitions-of-a-Geography-and-FIPS-CODES?language=en_US

Exploring Census Variables: Website

<https://data.census.gov/>

Exploring Census Variables: Website – API Interface

<https://api.census.gov/data/2015/acs/acs5/variables.html>

Code Demo & Exercise

Introduction to Geospatial Data

Types of Geospatial Data

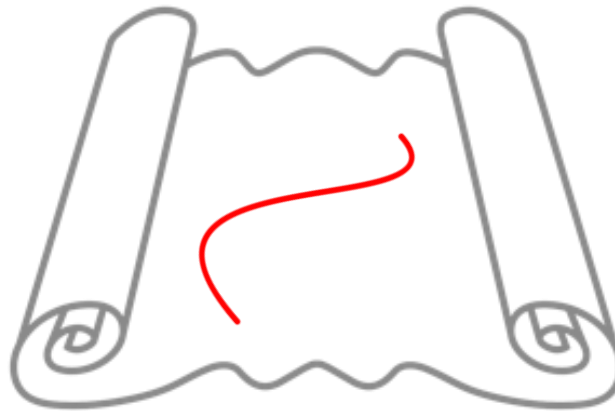
POINTS

Example: GPS coordinates



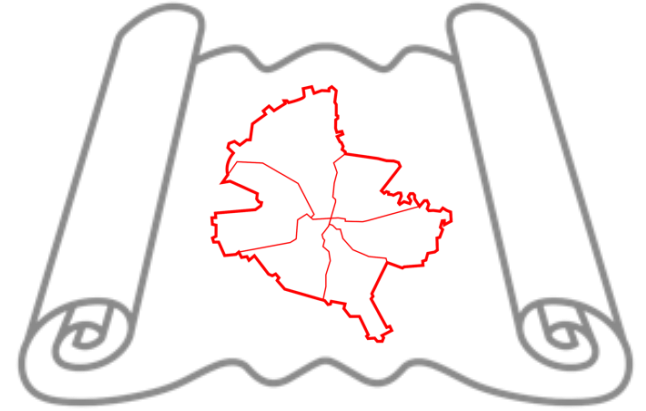
LINES

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)
- *Not required:* .proj (projections)

Simple Features: R

Parquet:

- Open-source file format



Coordinate Reference Systems

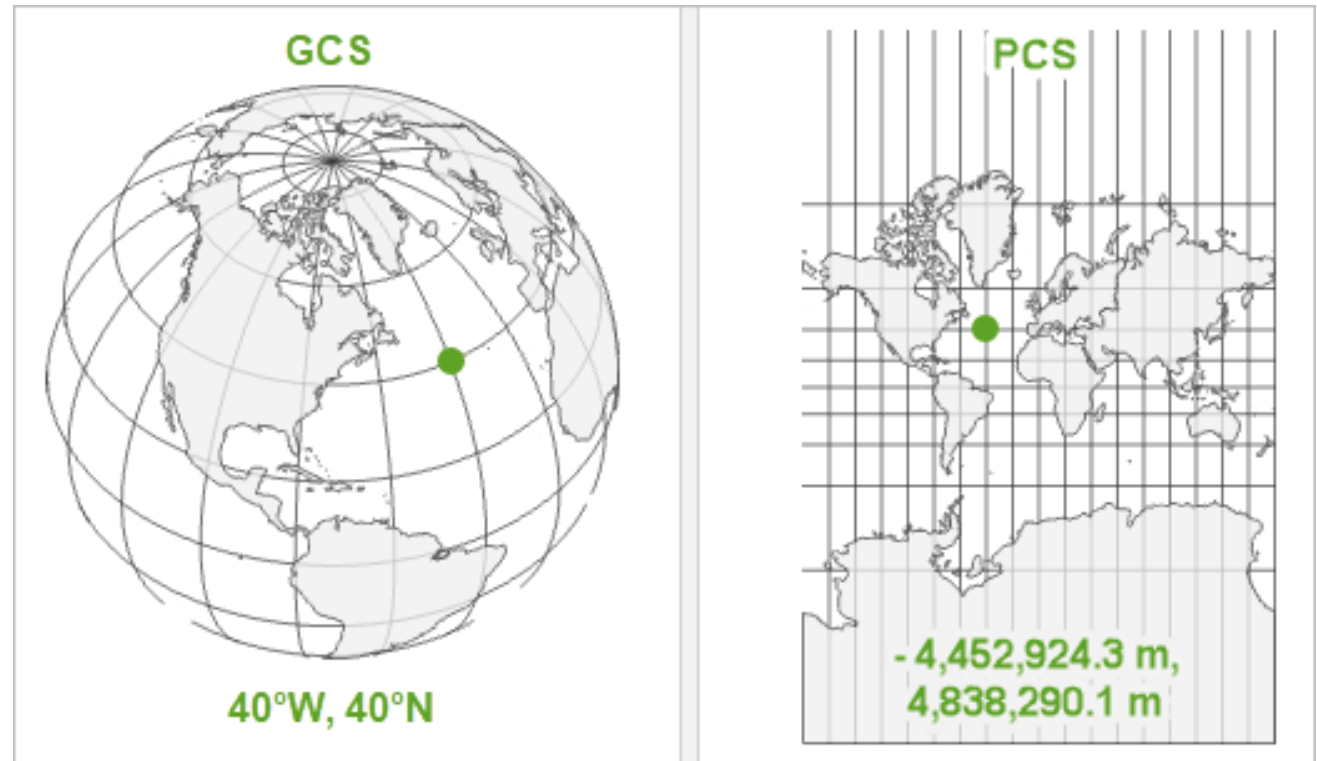
Geographic Coordinate System (GCS):

- 3-D
- How to find the point on the map

Projected Coordinate System (PCS):

- Flat
- How to convert the globe onto a sheet of paper

Different types are optimized for different locations



Source: <https://www.esri.com/arcgis-blog/products/arcgis-pro/mapping/coordinate-systems-difference/#GCS>

Types of Joins

SPATIAL JOIN

Spatial data + spatial data

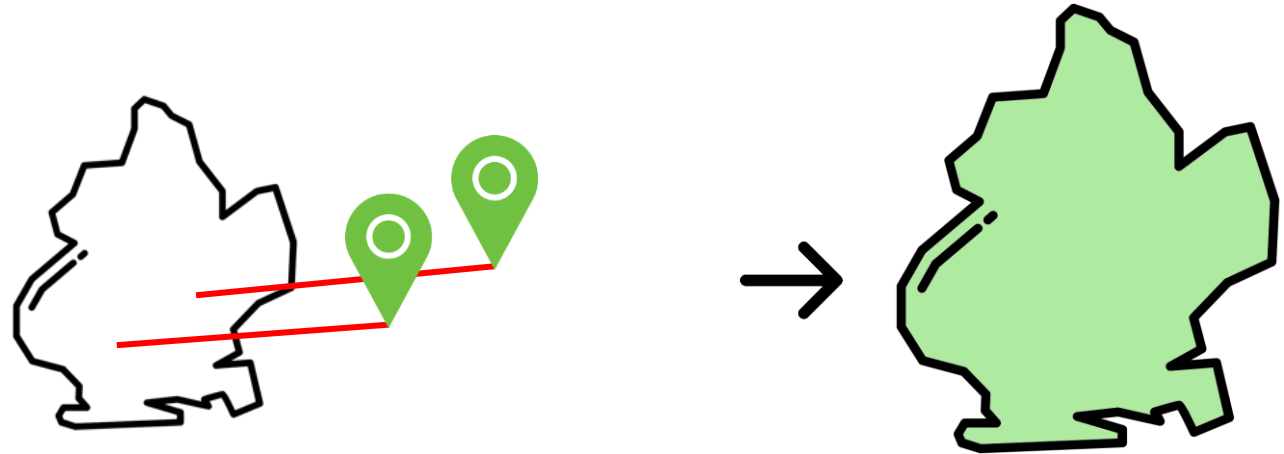
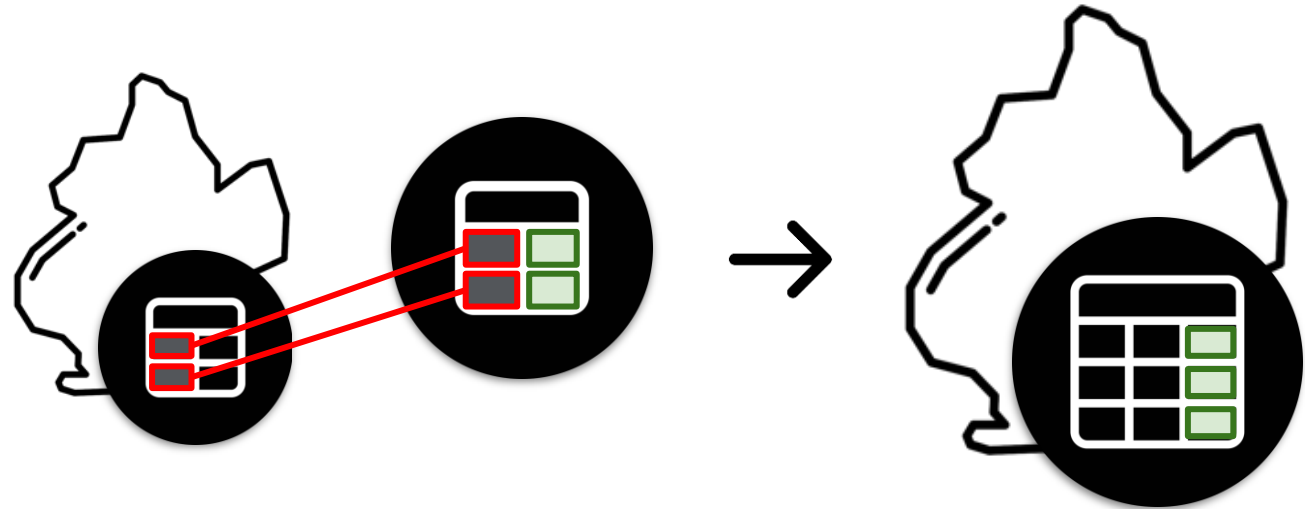


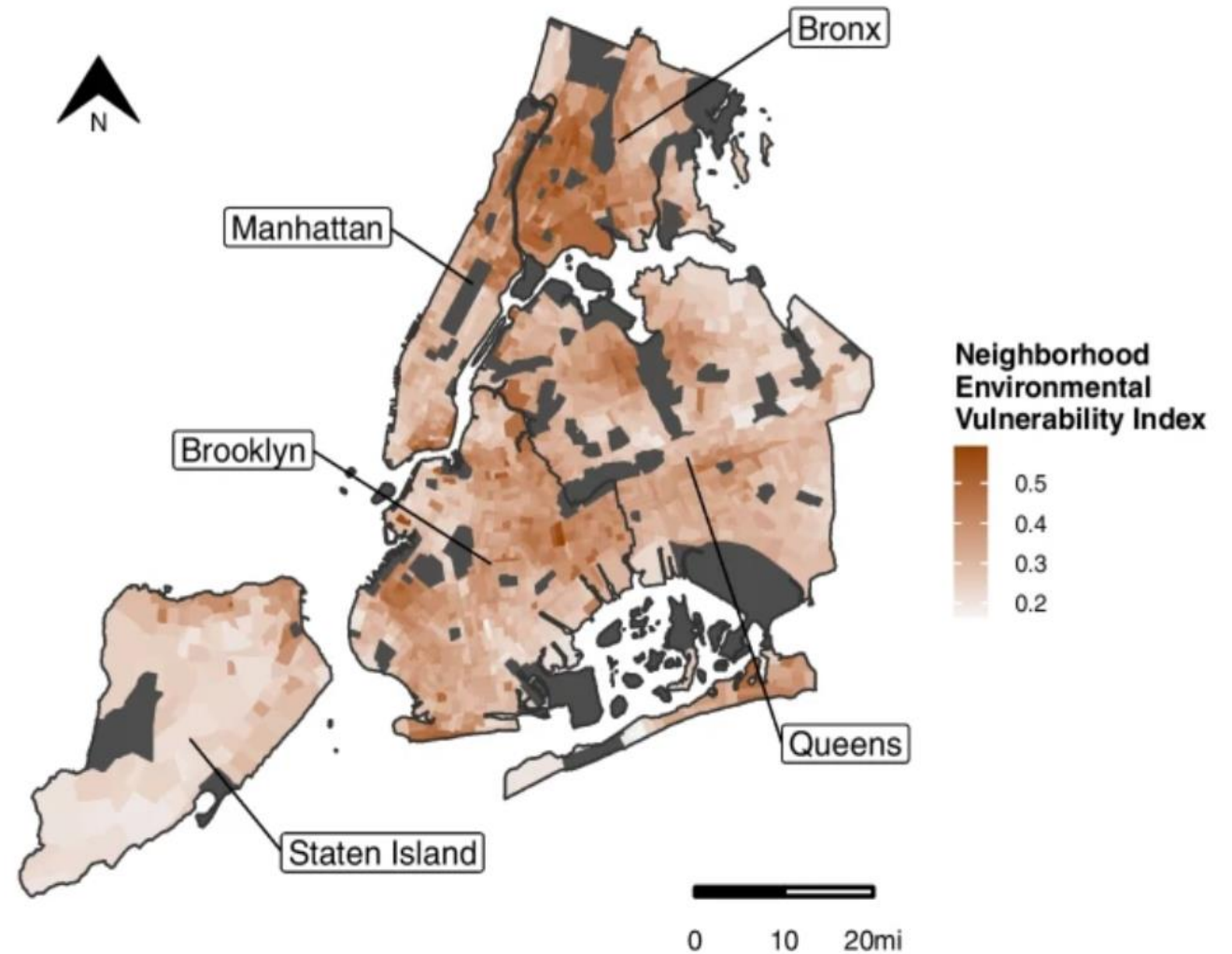
TABLE (TABULAR) JOIN

Spatial data + tabular data



Choropleth Maps

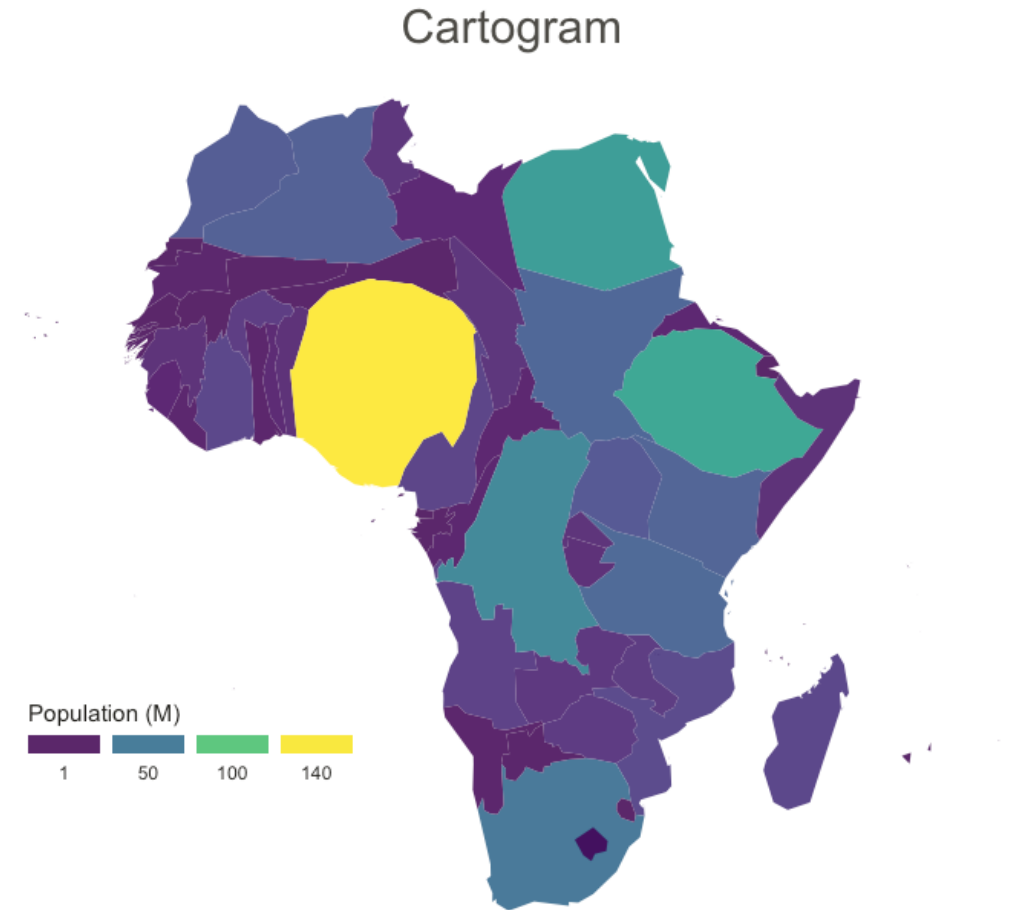
Uses **colors** or **shading** to show variation over geographic areas



Source: Uong et al., 2023

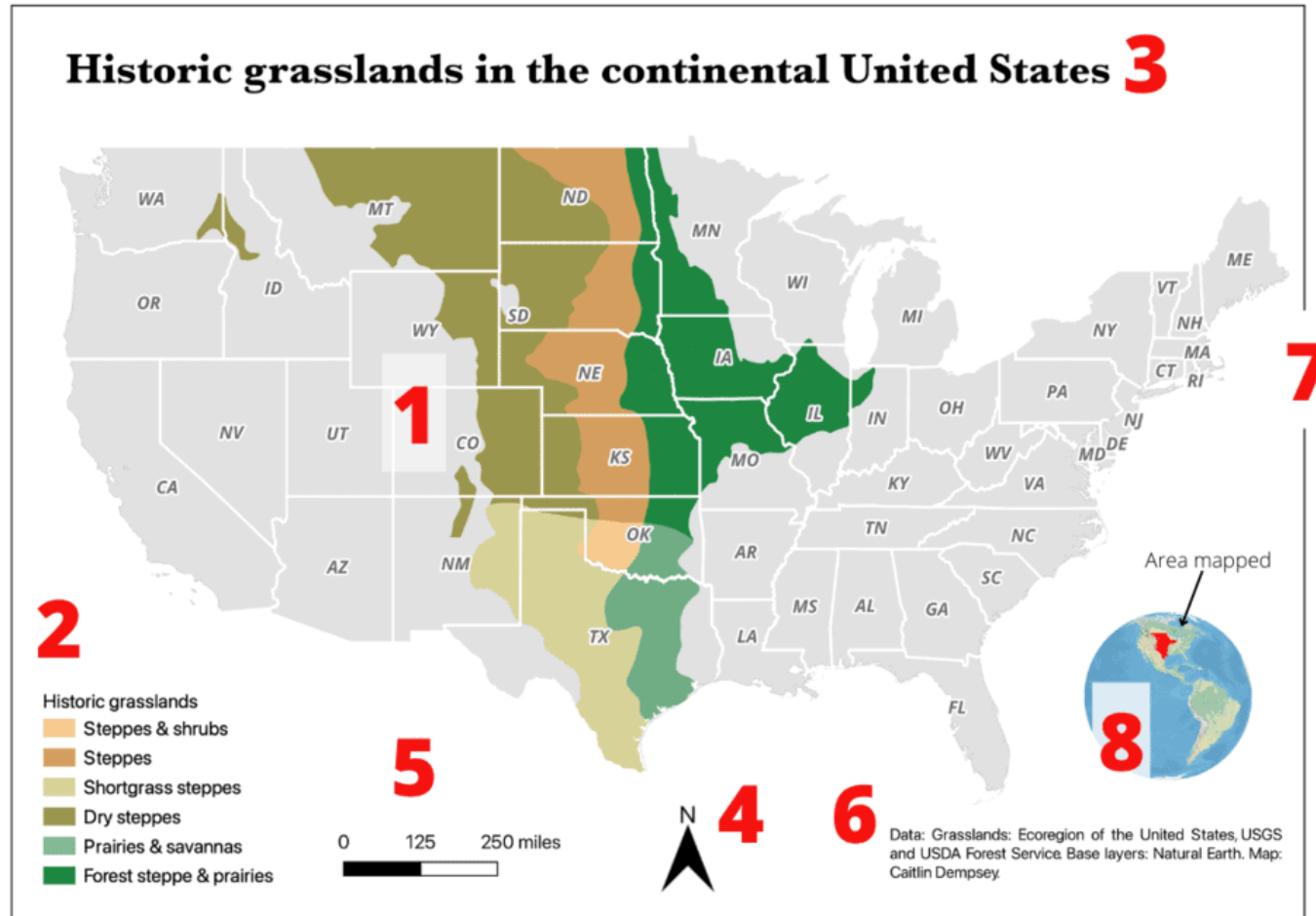
Cartogram

Uses **shapes/size** to show variation over geographic areas



Source: <https://www.data-to-viz.com/graph/cartogram.html>

Anatomy of a Map



Source: <https://www.geographyrealm.com/whats-in-a-map/>

Code Demo & Exercise

Closing Notes

Acknowledgements

Special thanks to **Christina Mehranbod** for her contributions to the original presentation on which parts of this workshop are based! She is an excellent spatial and injury epidemiologist, also looking for postdocs and jobs next year.



Thank You!

Please reach out if you have any questions!

- Email: spu2105@cumc.columbia.edu
- Github: @phispu
- Twitter: @phispu