Visualizing Labor Migration Using Quantitative Data

Introduction to the tidycensus Package

Let's Update Our Files!

- Open RStudio
- Switch to the terminal window
- Type the following:
  cd "/YOUR/GITPATH/HERE"
- Type the following:
  git pull "https://github.com/visualizinglabormigration/workshop/"

Overview

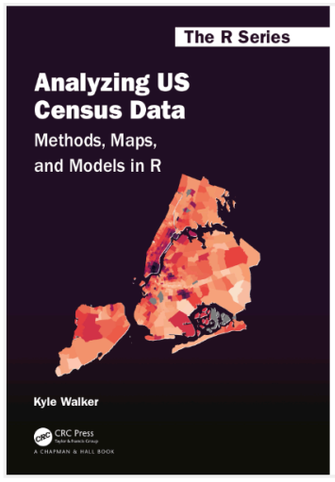
- Introduction to the American Community Survey (ACS)
  - History
  - Methodological hazards
- Accessing ACS data using the tidycensus package in R
- Mapping the distribution of migrant populations in the United States
  - Introduction to spatial data and the sf package
  - Visualizing areal data

Analyzing US Census Data: Methods, Maps, and Models in R by Kyle Walker

walker-data.com/census-r/index.html

Preface

This is the online home of *Analyzing US Census Data: Methods, Maps, and Models in R*, a book published with CRC Press in 2023.



To support the development of the book and help me maintain this free online version, please consider one of the following:

- Purchasing a hard copy of the book;
  - Buy from Routledge
  - Buy on Amazon
- Contributing to the book by filing an issue or making a pull request at the book's GitHub repository;
- Sponsoring book development and maintenance as a GitHub Sponsor;

Origins of the ACS

- The United States Constitution requires a full enumeration of the population every ten years for the purpose of congressional apportionment
- In 1940, the Census Bureau began using separate questionnaires
  - Short form —> population
  - Long form —> sample
- In 2010, the long form was replaced by the ACS
  - 1-year vs. 3-year vs. 5-year estimates

Challenges Associated with using ACS Data

- Continuous survey method affects variable definitions and comparability
  - e.g., Poverty over the previous 12 months of a multi-year period
- Smaller sample sizes increase the amount of measurement error
  - Especially problematic for less populated geographies
- The presence of measurement error
  - Requires additional work when using derived quantities
  - Violates traditional regression assumptions

Calculating Margins of Error for Derived Quantities

- We often need to calculate margins of error for the following types of derived quantities:
  - Products
  - Proportions
  - Ratios
  - Sums
- Formulas provided by the Census Bureau

Correcting for Measurement Error in Models

- Simulation exploration (SIMEX)
- Multiple overimputation
- Bayesian estimation
  - HMC vs. INLA

Overview

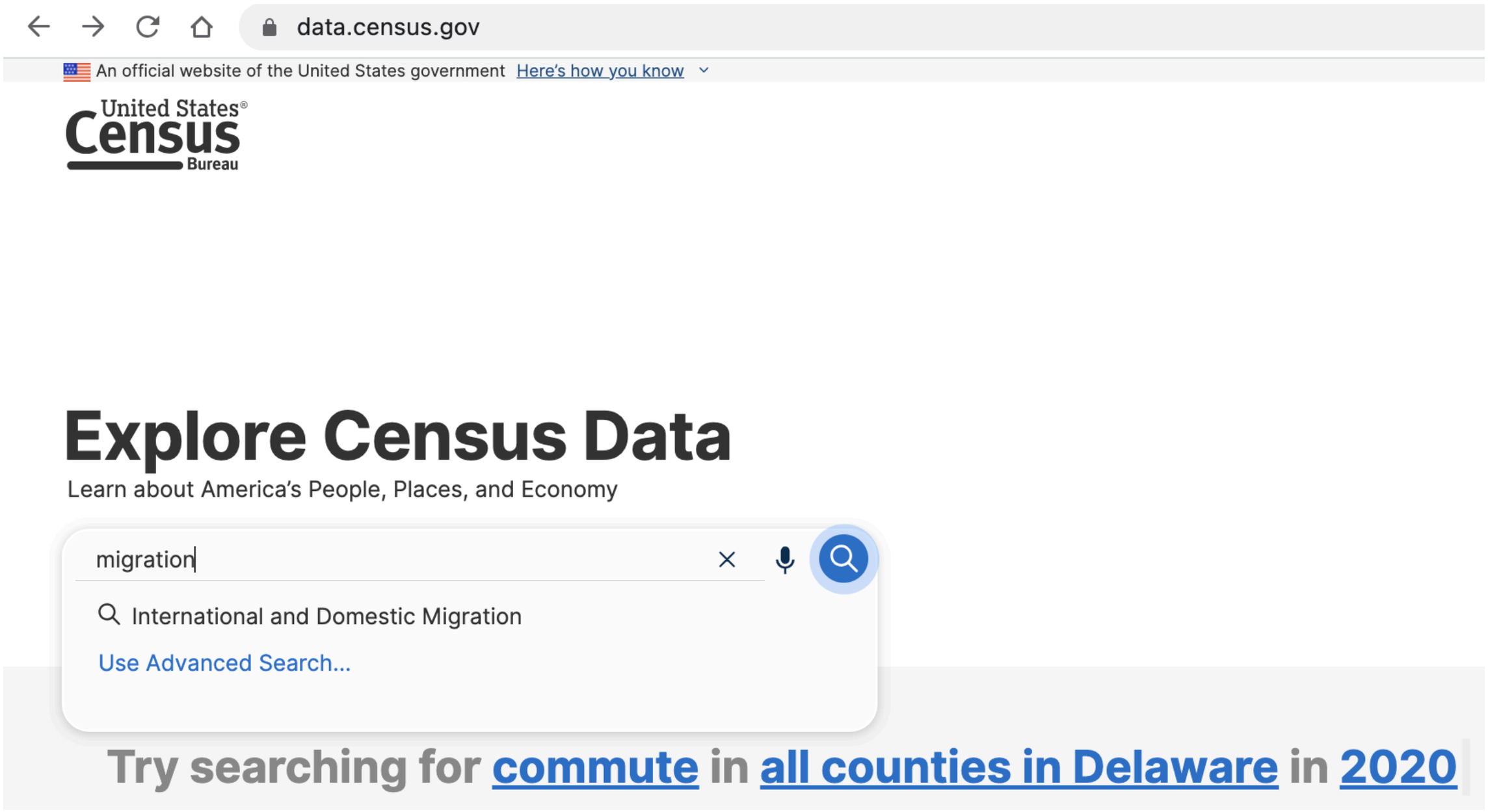
- Introduction to the American Community Survey (ACS)
  - History
  - Methodological hazards
- Accessing ACS data using the tidycensus package in R
- Mapping the distribution of migrant populations in the United States
  - Introduction to spatial data and the sf package
  - Visualizing areal data

Finding Aggregate Census Data

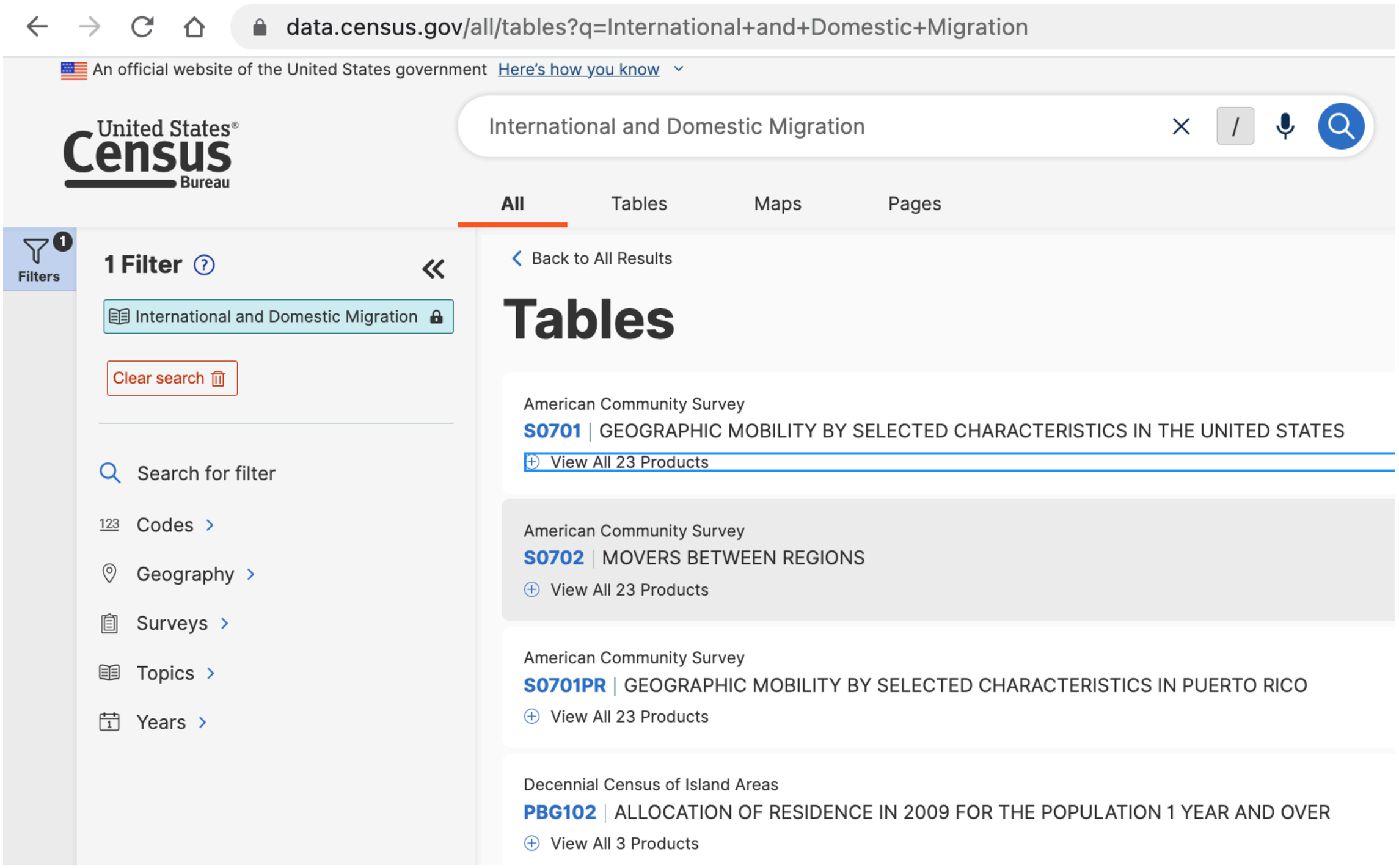
- United States Census Bureau
  - Tabular data stored as census data tables
  - TIGER/Line files for geographic data and select demographic and economic data
    - ASCII files vs. shapefiles
- National Historical Geographic Information System (NHGIS)
  - Integrated source for both census and non-census data
- tidycensus package in R
  - Uses the Census API to return tabular and spatial data stored as a simple features (sf) object

According to the ACS five-year estimates, what percentage of the adult population in each county in Virginia migrated from another state in 2021?

Finding Tables Online



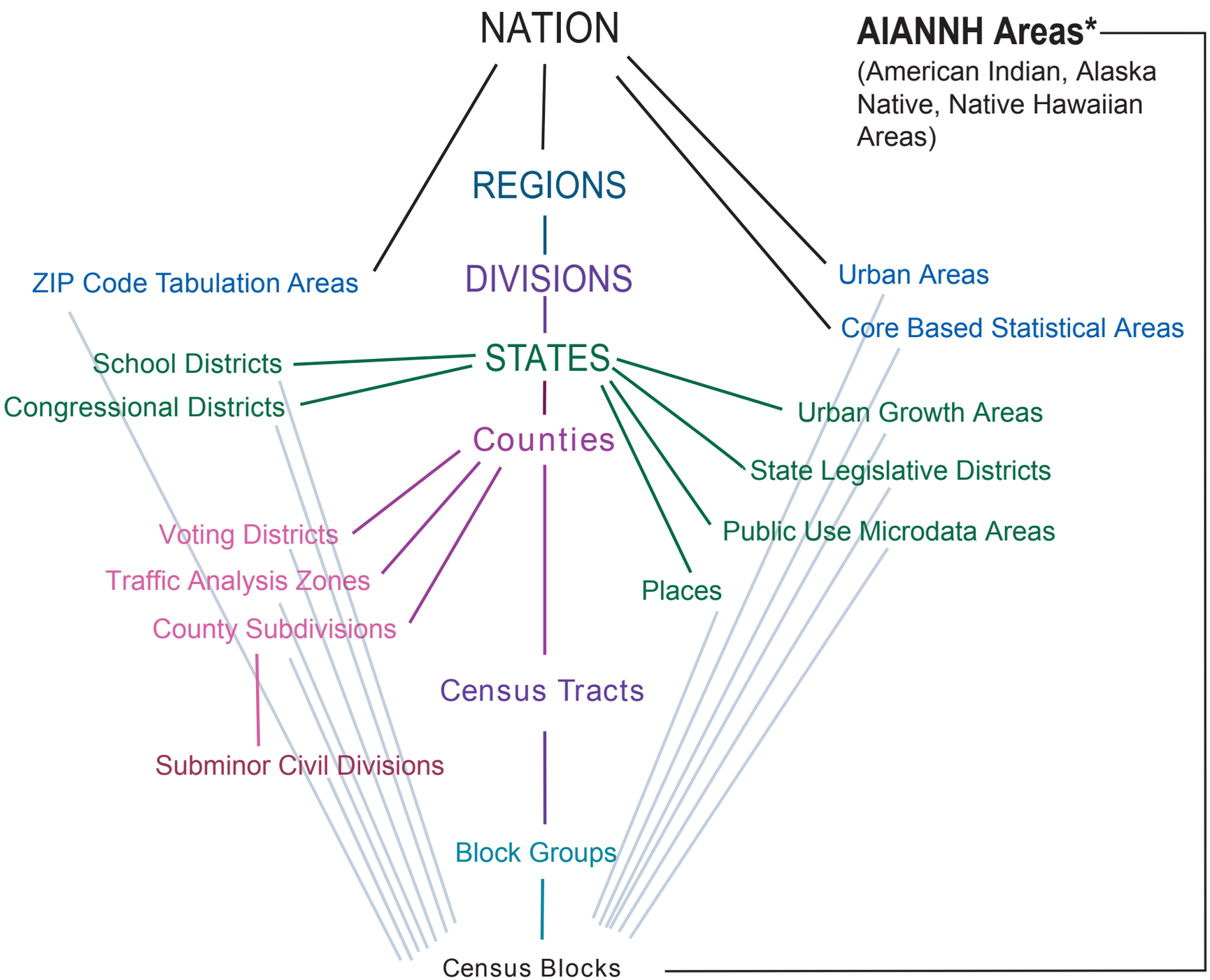
Finding Tables Online



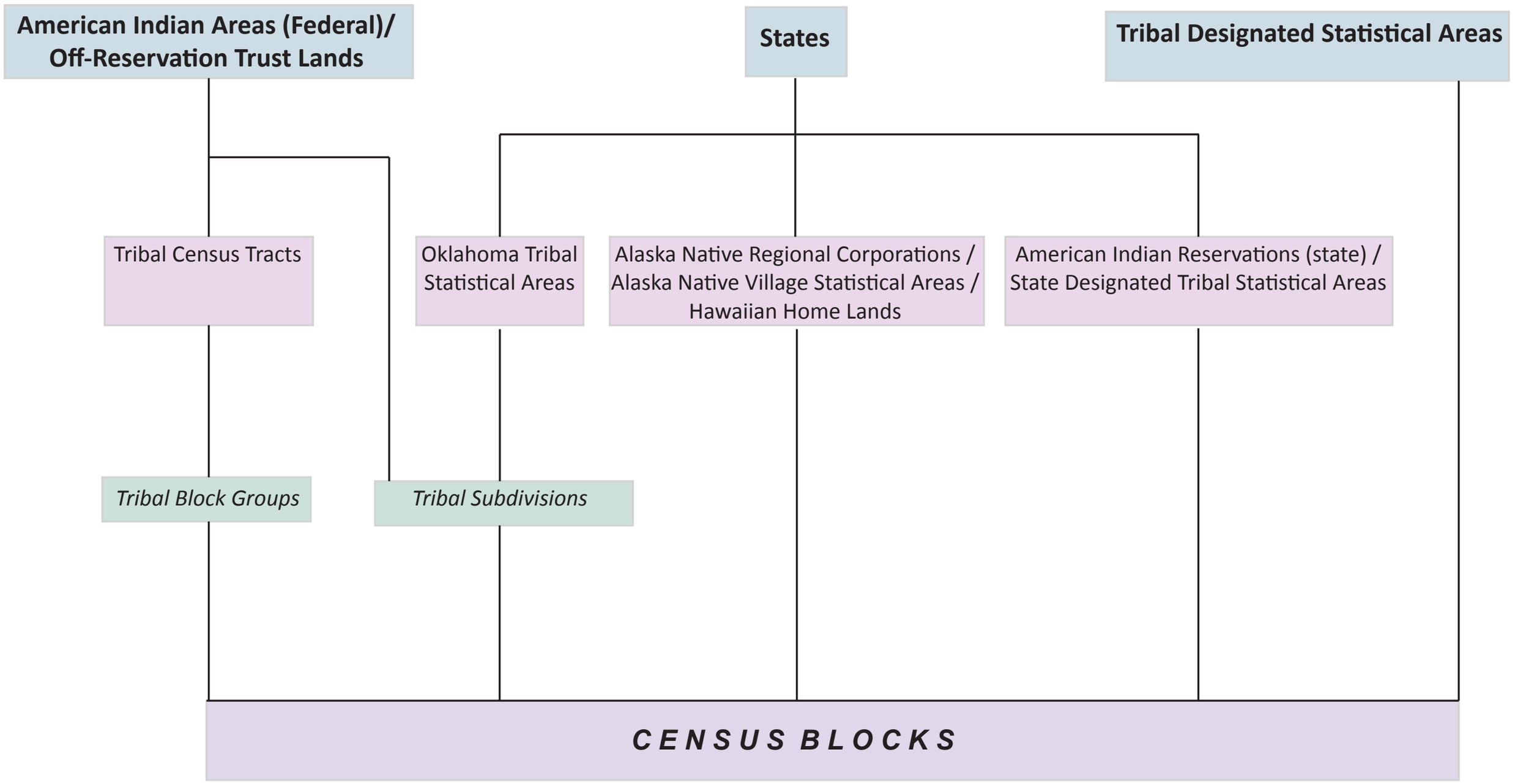
A Guide to ACS Tables

- Time
  - 1-, 3-, and 5-year estimates
- Geography
  - Varied levels of aggregation ranging from the block-level to the national-level
- Table
  - Detailed Tables (B or C)
  - Data Profiles (DP)
  - Subject Tables (S)
  - Comparison Profiles (CP)

Standard Hierarchy of Census Geographic Entities



Hierarchy of American Indian, Alaska Native, and Native Hawaiian Areas



A Guide to ACS Tables

- Time
  - 1-, 3-, and 5-year estimates
- Geography
  - Varied levels of aggregation ranging from the block-level to the national-level
- Table
  - Detailed Tables (B or C)
  - Data Profiles (DP)
  - Subject Tables (S)
  - Comparison Profiles (CP)

Let’s find out how to do this all in R!

Overview

- Introduction to the American Community Survey (ACS)
  - History
  - Methodological hazards
- Accessing ACS data using the tidycensus package in R
- Mapping the distribution of migrant populations in the United States
  - Introduction to spatial data and the sf package
  - Visualizing areal data

Working with Spatial Data

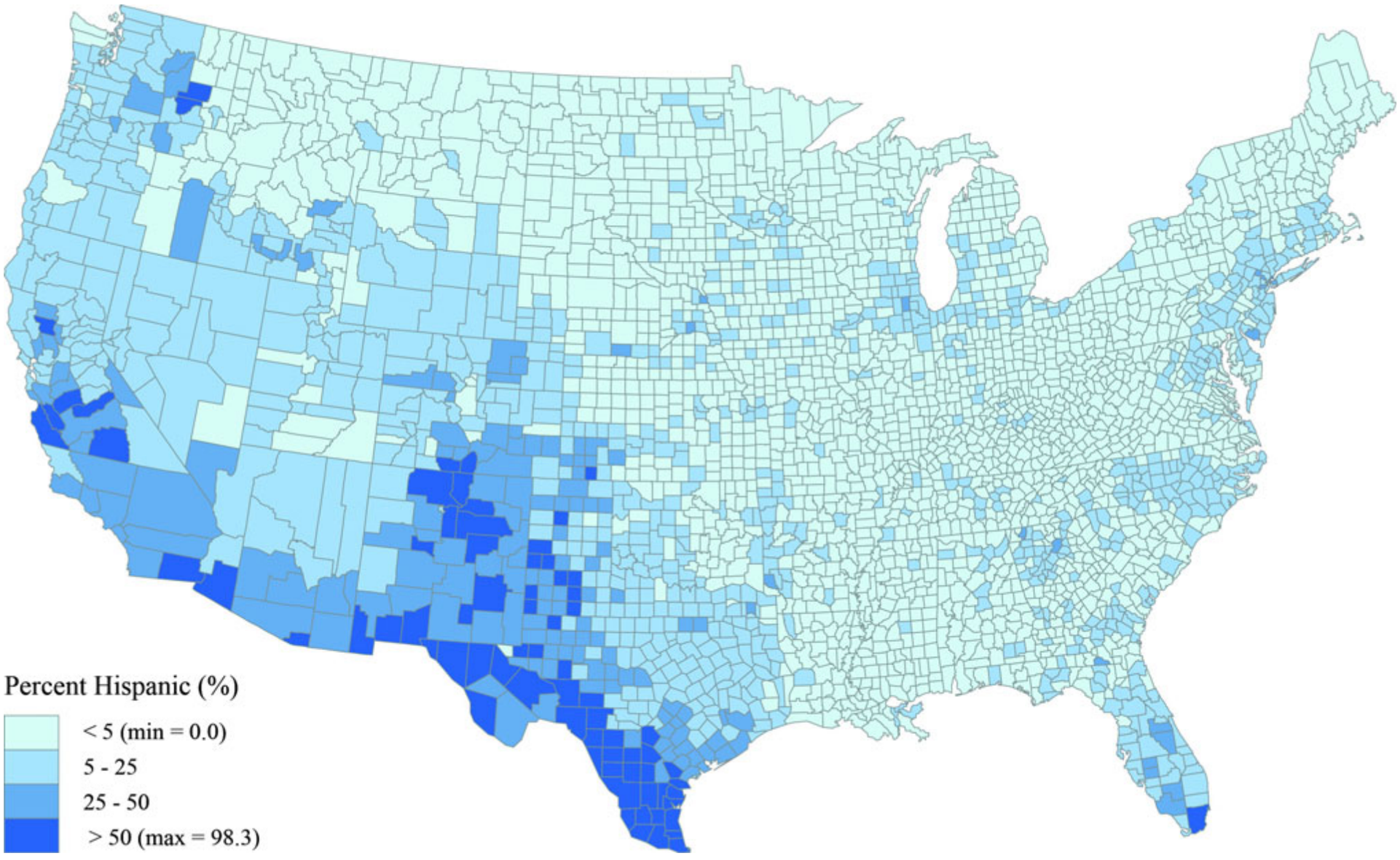
- Spatial data
  - Vector vs. raster
    - Vector data represent points, lines, and polygons in terms of discrete locations
    - Raster data describes the world in terms of a grid
  - We can't talk about location without a coordinate reference system (CRS)
    - Projected (X-Y) vs. unprojected (Lon-Lat)
- Tabular data
  - Traditional data file with IDs that allow it to be merged to the identities represented in the spatial data

Let’s find out how to do this all in R!

Visualizing Areal Data Using Choropleth Maps

- Choropleth maps are thematic maps that use color to represent variation in the value of a variable across a given geography
- Your choice of colors should align with the type of variable you are working with
  - May require breaking data into classes
- Your choice of color should align with readability demands

Figure 1. Hispanic Concentration, ACS 2006-2010



O'Connell, Heather A. and Carla Show. 2014. "Spatial Variation in the Relationship between Hispanic Concentration and Country Poverty: A Migration Perspective." *Spatial Demography 2: 30–54.*

Visualizing Areal Data Using Choropleth Maps

- Choropleth maps are thematic maps that use color to represent variation in the value of a variable across a given geography
- Your choice of colors should align with the type of variable you are working with
  - May require breaking data into classes
- Your choice of color should align with readability demands

Let’s check out ColorBrewer!

Let’s find out how to do this all in R!