GIS & spatial data

Librarian Panel on Data and Databases

Jessica Benner | jbenner@andrew.cmu.edu

GIS & Spatial Data Librarian

Presented at SICSS (Summer Institute in Computational Social Science)

May 19, 2025

Joining non-spatial information to spatial data

Spatial information

- Country
- State / province
- County
- City
- Zip code
- Neighborhood
- Census geographies
- Addresses
- … more

Non-spatial information

- Everything else!
- The things you are interested in!

Performed with a common Identifier using attribute join operation

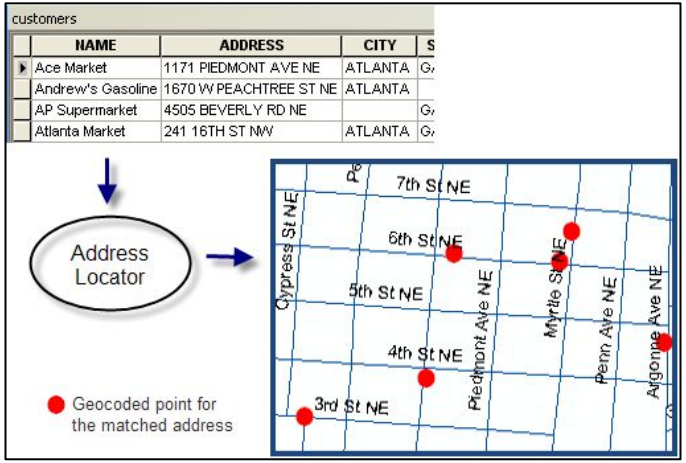
Data with Addresses

Geocoding in R from National Socio-Environmental Synthesis Center

MMQGIS Plugin in QGIS

Publish hosted layer in ArcGIS Online: 40 credits / 1000 addresses

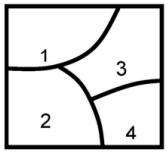
Guide with a variety of geocoding options (Columbia)



ArcGIS description of geocoding

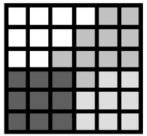
Common models to represent spatial data

VECTOR



Land-use

- Objects
- Object-based Model
- Discrete and definite
- Good for precise, well-known features
    - Building
    - Street
    - Park
- Shapefile, CSV, CAD drawings, GeoJSON,



RASTER

- Phenomena
- Field-based Model
- Distributed continuously
- Good for physical and environmental systems
    - Terrain
    - Temperature
    - Noise
- JPEG, PNG, TIFF

Admin Boundaries

State GIS Clearinghouses for a variety of data on individual states including boundaries. List of Clearinghouses from University of Pittsburgh.

Administrative boundary data guide from Brandeis

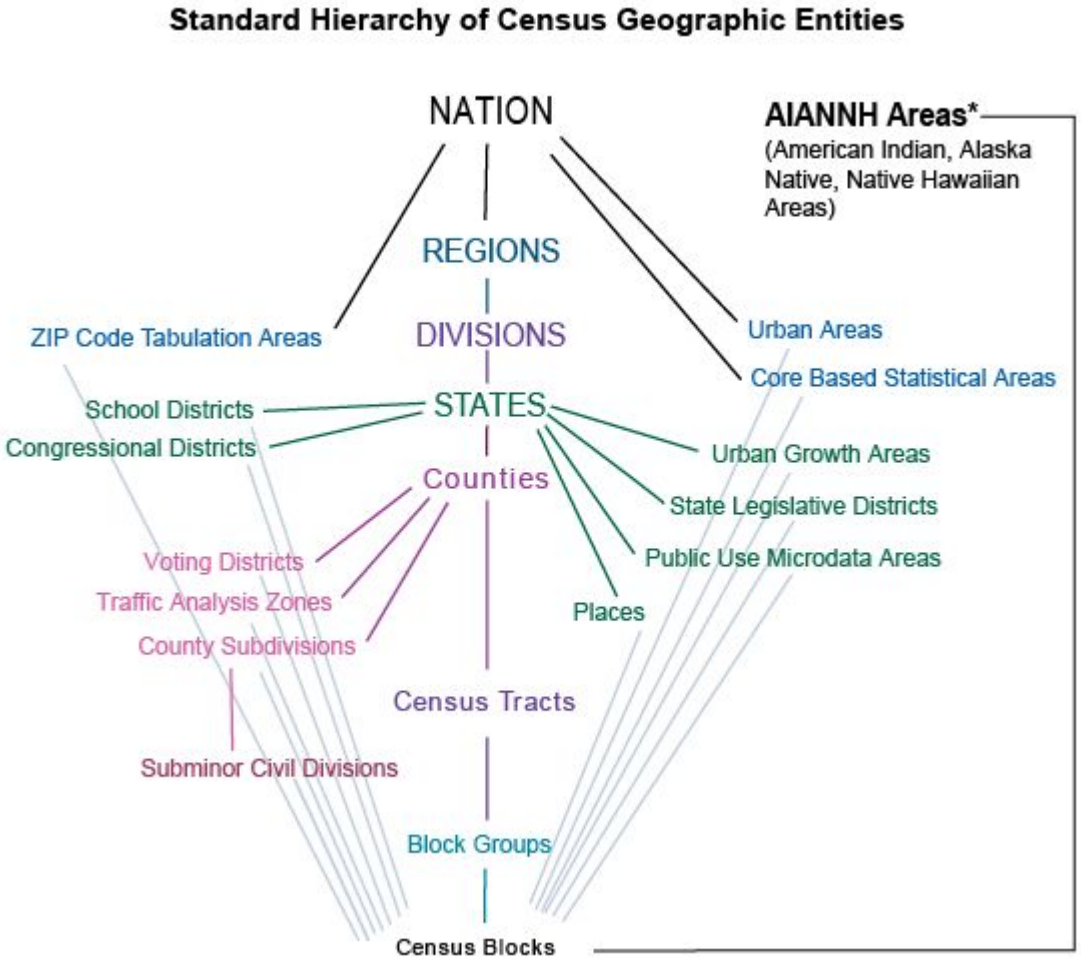
Natural Earth includes cultural, physical, and raster layers at a few scales

Census Geographies

[Guidance on the Census Geography levels](#) from Missouri Census Data Center

[TIGER/Line shapefiles](#)

[Social explorer Geodata](#)



US Census Data (Licensed Data)

PolicyMap is a U.S. national data and mapping tool and analytics platform with multidisciplinary applications. Useful for social sciences, urban studies, real estate and housing analysis, community and economic development, public administration, public health, policy and political science, education, business, economics, statistics, and geography, among others. Includes thousands of U.S. data indicators for demographic and socioeconomic analysis, from a neighborhood census block group in many cases, up to a national level. **Contemporary datasets, good amount of other data.**

Social Explorer is a visual and numerical presentation of U.S. census data and a wide range of demographic information, 1790 – present. Creates maps and reports at levels from national to block group. **Has historical census data, good for visualization.**

US Census Data (Open Data)

Data.census.gov is the data dissemination platform to access demographic and economic data from the U.S. Census Bureau. This help page has a video and other resources to help you use the site. **All US Census data, can be hard to use. The 'Profile' feature can be useful for aggregate demographics.** General note: Decennial are counts, ACS & PEP are estimates.

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. **Probably the best for desktop GIS.**

International Census Data

Living Atlas in ArcGIS. Search term Census then choose Country.

IPUMS International is dedicated to collecting and distributing census data from around the world. The world's largest collection of publicly available individual-level census data. If you use their data, they require that you cite them and agree to data security. (need to apply for an account)

US Census International Database, population estimates for over 227 countries - mostly country level data

Notes about working with US Census data

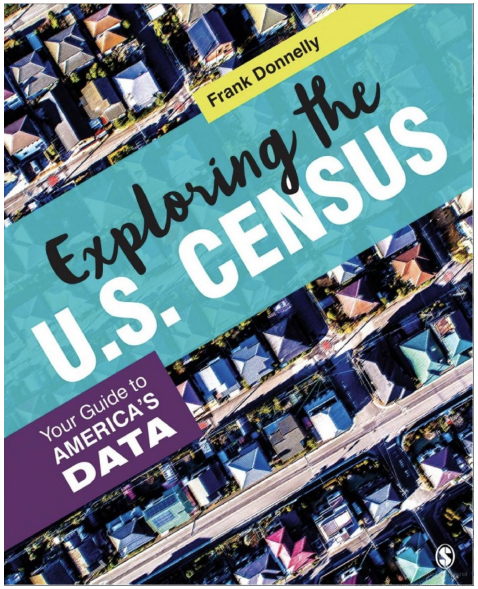
Frank Donnelly at Brown is a good source.

He wrote a book >>

His guide to Census Data & article about the datasets

Chapter from the Census about working with ACS (American Community Survey) data in rural areas.

Historical analysis is best done with NHGIS



Access the book online via CMU

Basemaps

Basemaps R package https://jakob.schwalb-willmann.de/basemaps/

OpenStreetMaps in R https://jcoliver.github.io/learn-r/017-open-street-map.html

OSMapiR https://wiki.openstreetmap.org/wiki/OsmapiR_(R_package)

Bbox finder: http://bboxfinder.com/#0.000000,0.000000,0.000000,0.000000 Use a rectangle to find bounding box coordinates

R spatial tools

Data Sources

tigris provides cartographic files from the US Census

tidycensus access to the US Census APIs for the American Community Survey and then Decennial Census

rnaturalearth for access to **Natural Earth** data (basemap)

RSocrata for access to Socrata Open Data portals (including **NYC Open Data**) Socrata Portal

Full listing of spatial tools for R from Columbia University

Specific Countries and Regions

India Urban Data Exchange (IUDX), Smart City data, might be API access only

Maps of India, lots of topics, only maps area available and some may not be fully accurate

China Data Institute, basic administrative and environmental data, may be outdated

China Dimensions (via CIESIN, Columbia University)– a variety of socioeconomic data, including Geographic Information System (GIS) databases that cover the administrative regions of China

Open Africa, open repository maintained by grassroots initiative, GIS and other data

US Data.gov, clearinghouse for data from US Federal Government

Country/state/province name + “spatial data clearinghouse”, “spatial data infrastructure” “open data’

Urban Datasets

Atlas of Urban Expansion (Worldwide data), available from Lincoln Institute of Land Policy who have some other US datasets.

Urban Institute, think tank, economic and health related data, mainly US

Urban Indicators Database, Data from UN , including Urban Observatories

Urban Data Platform Plus, economic data for European Union and some global cities, mostly tabular data

World Datasets

World Resources Institute GIS data on Population & poverty, Agriculture, Land cover & land form, Base data, Elevation, Rainfall, Biodiversity & wildlife, Tourism, and Water

The Planetary Computer Data Catalog includes petabytes of environmental monitoring data, in consistent, analysis-ready formats, API/web access only

Relief Web, humanitarian emergencies and natural disasters, maps only

Socioeconomic Data and Applications Center, socioeconomic data, NASA data center (includes China Dimensions)

Global Earth Observation System of Systems (GEOSS), heterogeneous collections of Earth observations from satellites, airplanes, drones and in-situ sensors at global, regional and local scales

Specific data searching for individual cities

For individual cities, ask who governs, uses, or advocates for (or against) these resources?

- National agencies
- State and provincial agencies
- County and city agencies
- Regional Data Centers
- University labs and advocacy groups
- Companies operating in the area

Some specific data terms

Water data: hydrography, hydrology

Transportation data: GTFS, route

Demographic data: microdata

Land data: land use vs land cover, cadastral, parcel

Built environment data: building footprint

Imagery: bands, satellite vs instrument, resolution, DEM

ArcGIS: feature, geodatabase, attribute table

…

GIS Dictionary , Dictionary of Geography

Transportation Data

Pittsburgh example

US: Bureau of Transportation Statistics including National Transportation Atlas Database, multiple modes of transport and multiple categories

State: Pennsylvania Spatial Data Access Transportation Data

County: Allegheny County GIS Portal Transportation Data

County Service: Pittsburgh Regional Transit Open Data Portal

Regional: Western Pennsylvania Regional Data Center Transportation Data

Private Service: POGOH bicycle rides (formerly Healthy Ride Pittsburgh) https://pogoh.com/data/

Historic Map Collections

David Rumsey Map Collection, 119,000 maps from across the world, Use the Luna Viewer to browse the collection, choose by Location (Where) or Topic (What)

Library of Congress Map Collections, mostly US with some other countries.

Historic Topographic Maps - USGS (1884 - 2006), US only

Old Maps Online, collaboration to make old maps available for free, multiple countries

Digital Sanborn Maps, 1867-1970 Produced for over a century, more than 660,000 Sanborn maps chart the growth and development of more than 12,000 American towns and cities.

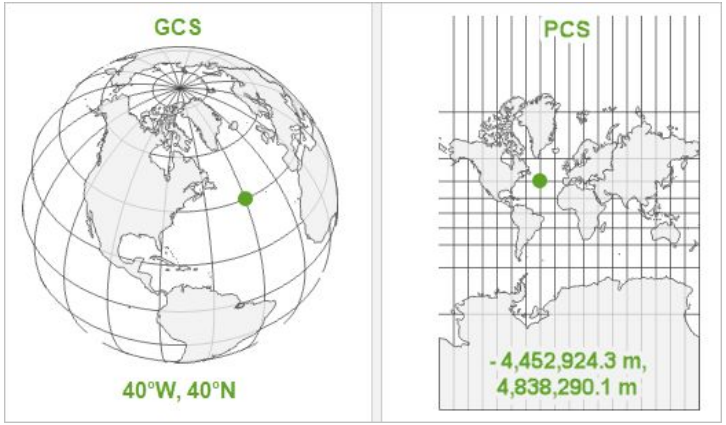
Note about coordinate reference systems and projections

If datasets aren't lining up in the visualization, you may need to check the reference system of the data and adjust it

If you want to measure things, you need a projection for best accuracy

Workshop on Projected Coordinate Systems in R by Maptime UC Davis (~2 hours, starts at 9:30 minutes into the video)

- Chapter of GIS Fundamental book they mention Chapter 3 in the list



Working with GIS and spatial data

- ➔ Joining is a common tool to add data together and add non-spatial data to geographic datasets
- ➔ No spaces in your file names or file path
- ➔ Inspect the datasets you find to make sure they have the measures/attributes you need
- ➔ Depending on what you want to do, you may need to attend to projections
- ➔ Spatial is nice but not always needed to make your point

Questions?

Jessica Benner
4406 Wean Hall
jbenner@andrew.cmu.edu
Schedule a meeting