Visualizing Census Data in GIS

Andrew Rowan, Ph.D.
Director, NJ Office of GIS

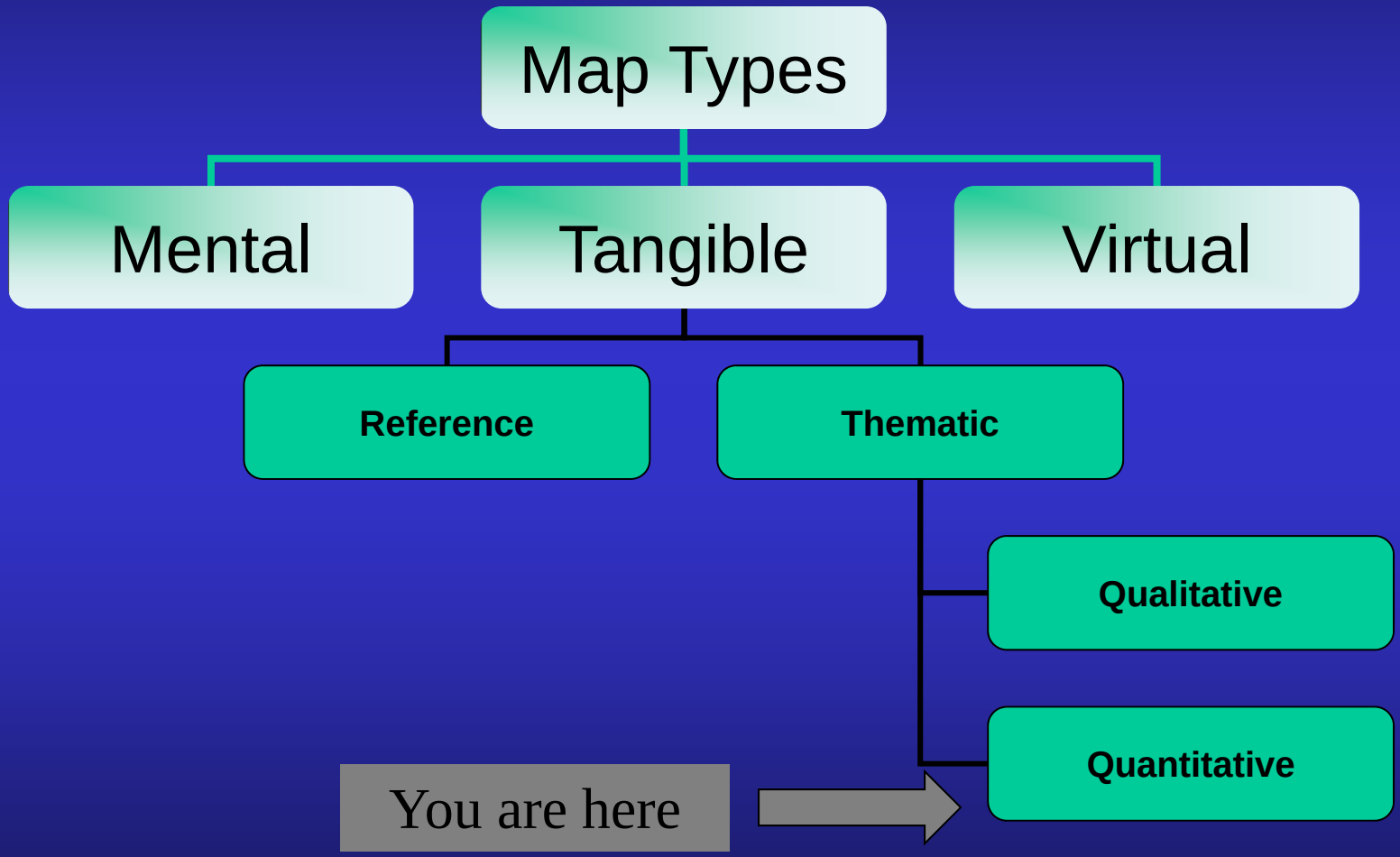
OGIS NEW JERSEY
OFFICE OF GEOGRAPHIC INFORMATION SYSTEMS
NEW JERSEY OFFICE OF INFORMATION TECHNOLOGY

Outline

- VERY accelerated cartography primer
- Software choices
- Demos/how-to
  - ArcGIS Desktop (formerly ArcView)
  - ArcGIS Online
- Notes on data prep
- Q&A

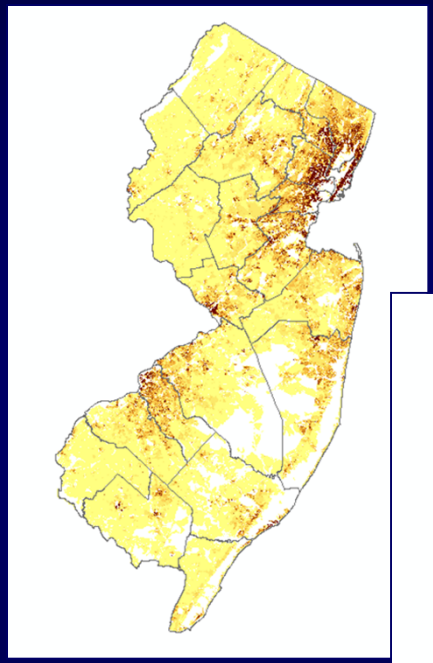
Cartography, In Brief

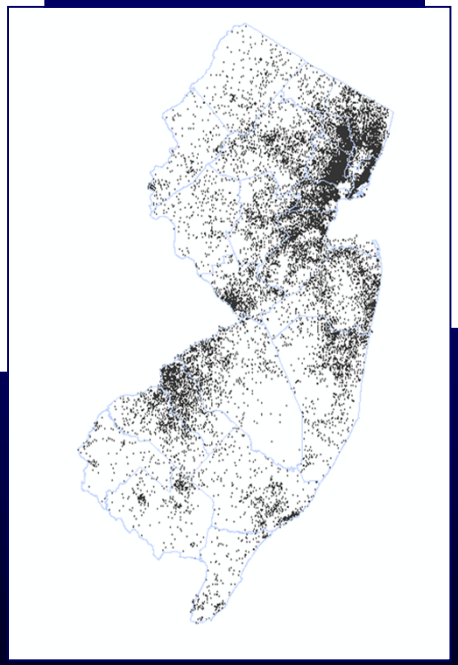
Map Types

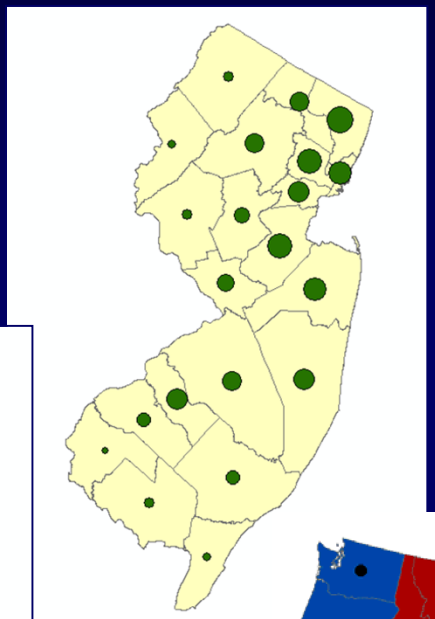


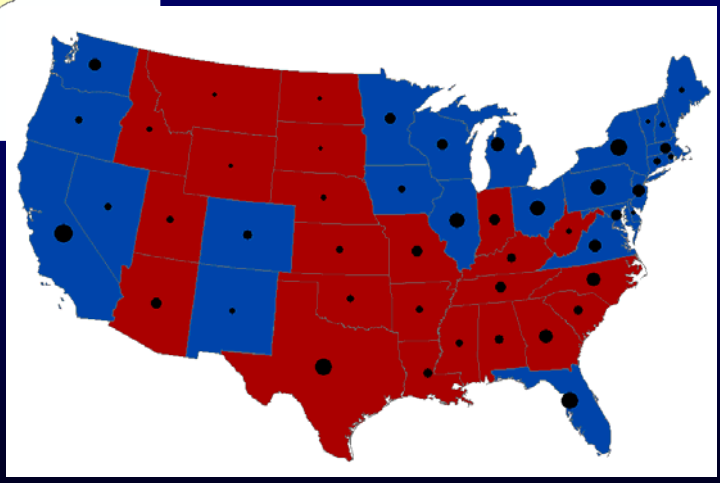
Source: Maantay and Ziegler, via Wiggins

Types of Visualizations



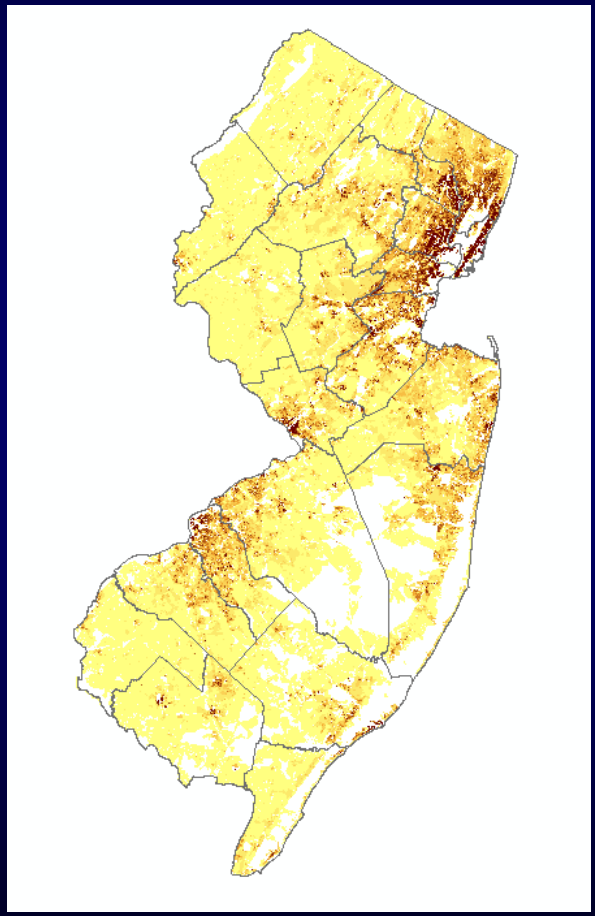






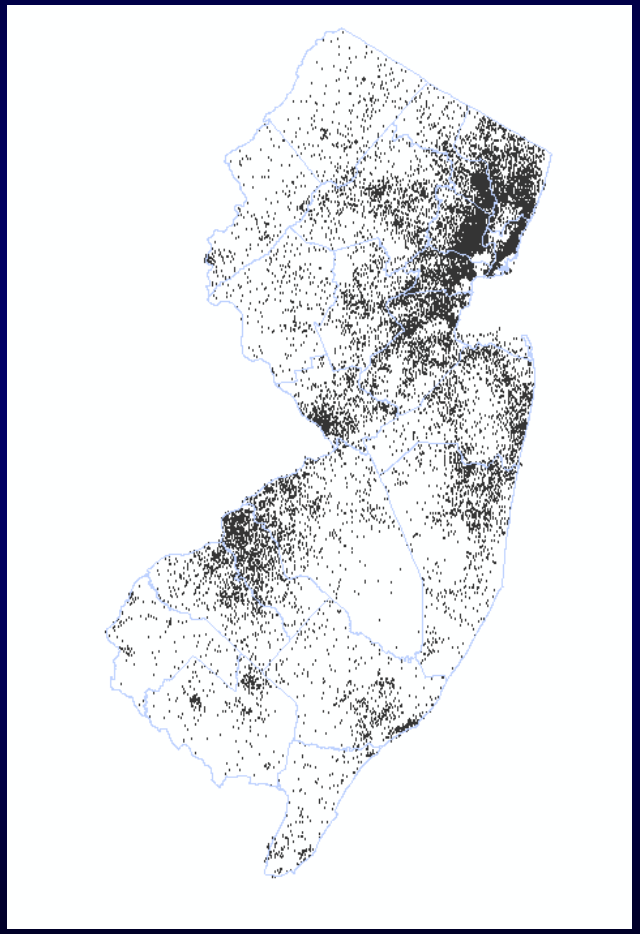
Graduated Color (Choropleth)

- Areas shaded by value
- Values MUST be independent of polygon size to produce accurate presentation – counts must be normalized by area or percentage
- Choose data classification method carefully



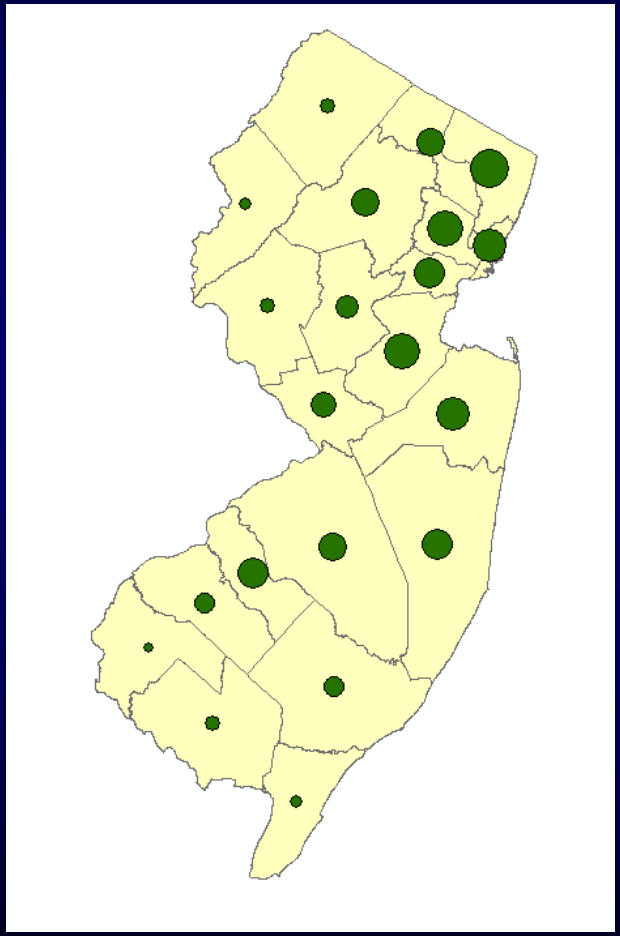
Dot Density

- Appropriate for count data, not ratios
- Each dot represents a fixed number of people
- Easy to understand, portrays spatial pattern, but provides no concrete numbers to the reader



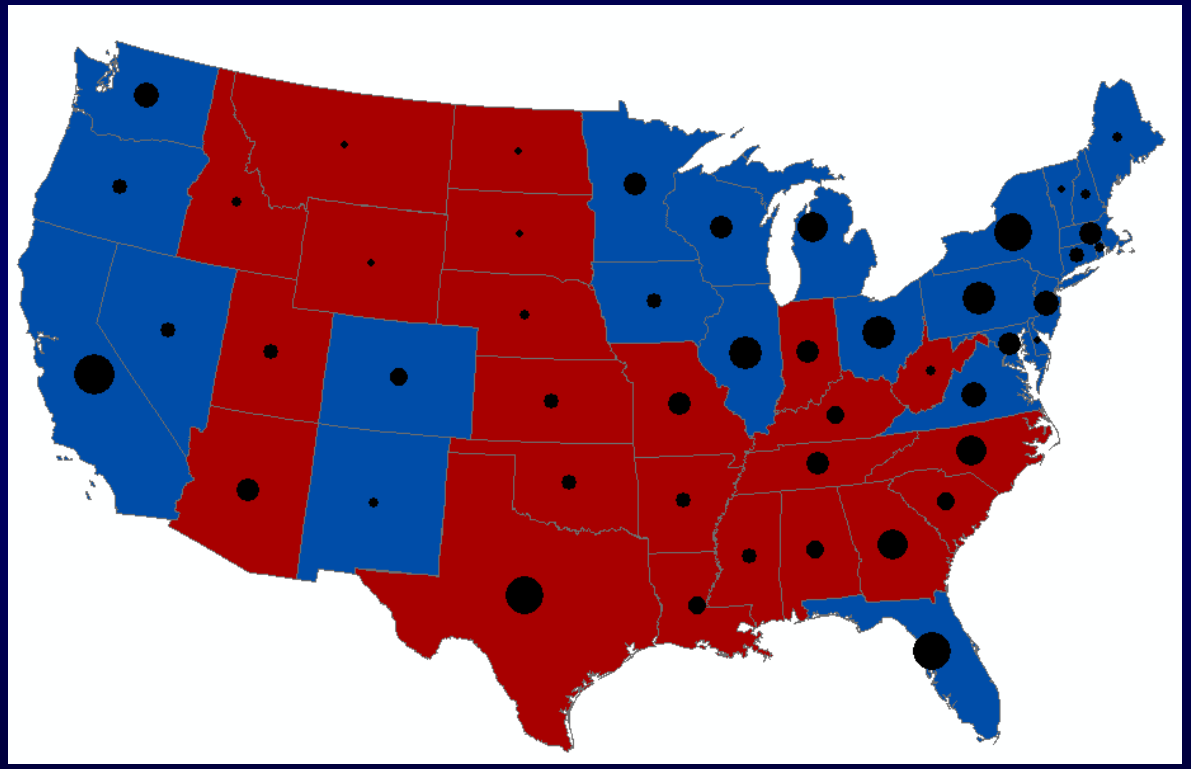
Proportional/Graduated Symbol

- Can be suitable for both count and ratio
- Use care not to overwhelm the reader's ability to perceive the message
- Best for relatively few values



Multiple Variables (Quantity, Category)

- Be careful not to get too complex
- Consider multiple maps instead



Classification Methods

Common methods for grouping values into classes (colors or symbol sizes on map)

- Natural breaks: algorithm looks for clustering in the data
- Equal interval: 0-10, 10-20, etc.
- Quantile: equal number of polygons in each class

Software Options

Desktop GIS

- ArcGIS Desktop
  - By far the most common choice
  - “Basic” version more than adequate
  - Takes some time investment to learn
- MapInfo
- GeoMedia
- QGIS and other free software

Free, Lightweight?

- Google Earth not meant for thematic maps. Add-ons exist that enable thematic maps, but core product does not do them.
- ArcGIS Explorer – free desktop s/w from Esri, can only do thematic maps if layers are prepared first in ArcGIS.

ArcGIS Online

- Cloud-based, software as a service from Esri
- Rapidly growing library of data and maps shared by users
- Subscription version starts at $2500/yr for 5 users
- Free accounts exist, some limitations on capabilities (click “sign in” and then click “create public account”)
- Some data prep in ArcGIS will be necessary

Demos/How-To

Data Preparation

Obtaining Data

- Geographic data (shapefiles) and statistical data published separately
- Geographic data: www.census.gov > geography > TIGER, or njgin.nj.gov
- For statistical data, many and varied choices
- Must use the geography that corresponds to the summary level of your statistical data
- You will need “GEOID” field in both files to define one-to-one match to join the two together
- GEOID always in TIGER files, but you may have to create it for statistical tables

Creating a GEOID

- Unique ID for each polygon
- Cobination of codes for state and lower divisions
- Content varies depending on what geography you use

Area	fields needed to create join id
County subdivision (NJ - Municipalities, aka Minor Civil Divisions)	STATE + COUNTY + COUSUB
Cenus Tract	STATE + COUNTY +TRACT
Block Group	STATE + COUNTY + TRACT + BLKGRP
Block	STATE + COUNTY + TRACT + BLKGRP + BLOCK
Congressional District (111th Congress)	STATE + CD
State Legislative District (Upper Chamber, same as lower in NJ)	STATE + SLDU
School District (Elementary)	STATE + SDELM
School District (Secondary)	STATE + SDSEC
School District (Unified)	STATE + SDUNI
Voting District	STATE + COUNTY + VTD
Place	STATE + PLACE
American Indian Area	AIANNH + AIHHTLI

Data Prep for QGIS

- QGIS reads shapefiles (and others) for spatial data
- For tabular data, Comma-Separated Values in a text file (.csv, can export from excel)
- Need to prepare extra .csvt file to define field types or numbers will be treated as text
- For more details, see http://qgis.spatialthoughts.com/2012/03/using-tabular-data-in-qgis.html

Using Excel Files in ArcGIS

- Access is a better suited tool, but many people find Excel easier to use/learn
- Add a single worksheet to your map, not a whole workbook
- First row: field names
- Data begins on row 2
- Field types determined by values in row 2
- No extra blank lines