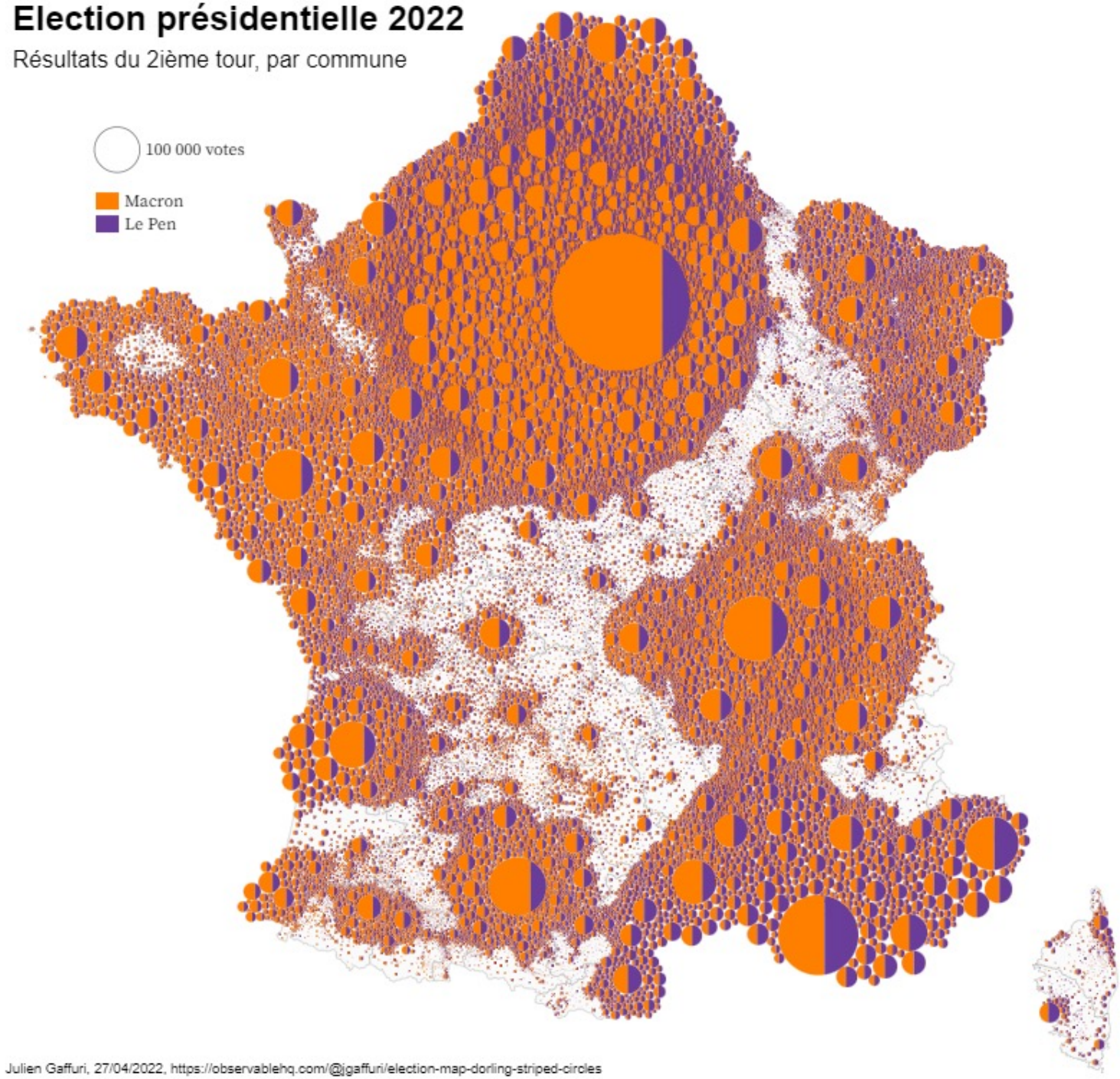
Data Visualization

Lecture 11

Election présidentielle 2022

Résultats du 2ième tour, par commune

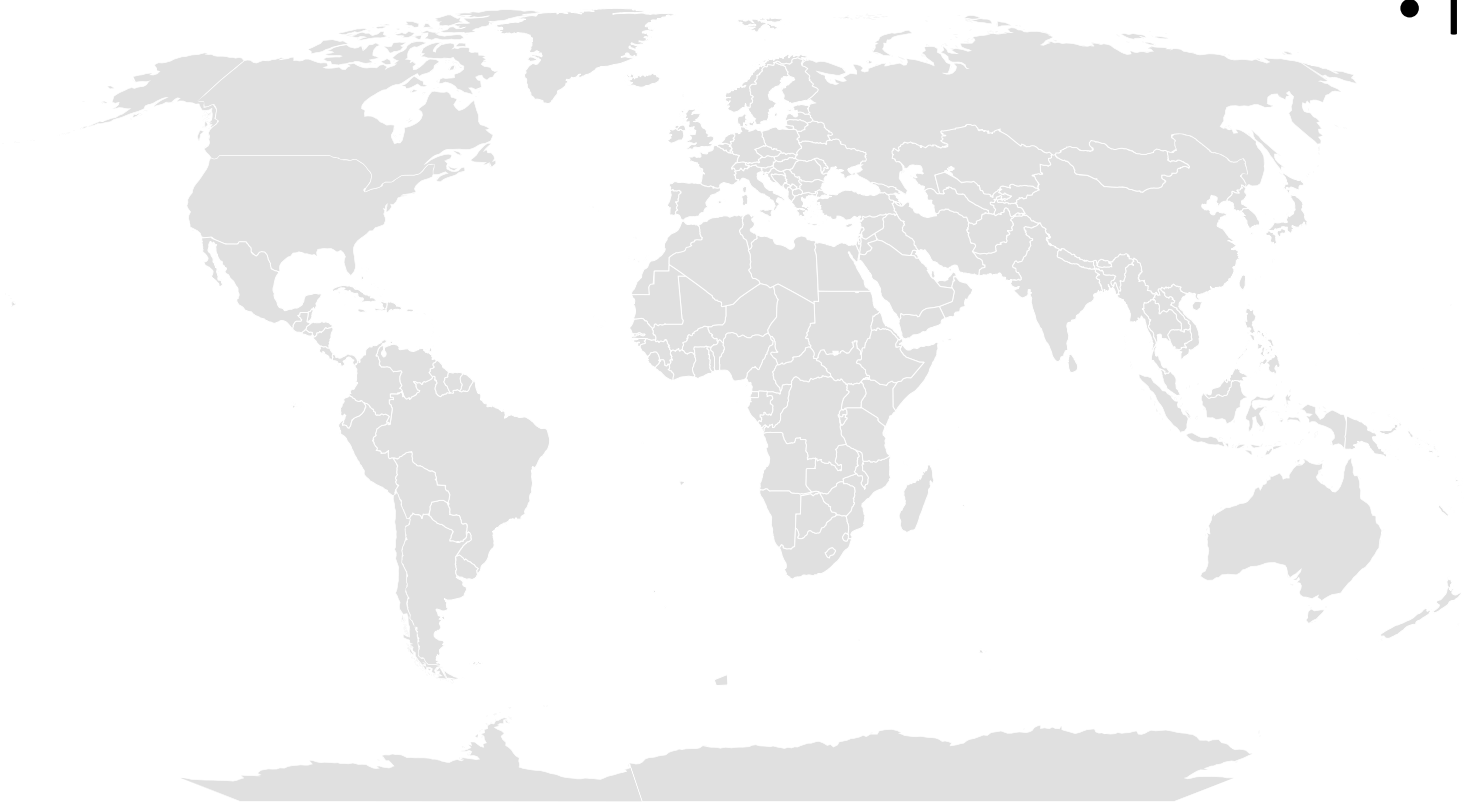


Julien Gaffuri, 27/04/2022, https://observablehq.com/@jgaffuri/election-map-dorling-striped-circles

Geographic Visualizations

Maps, GIS

Geographic Map



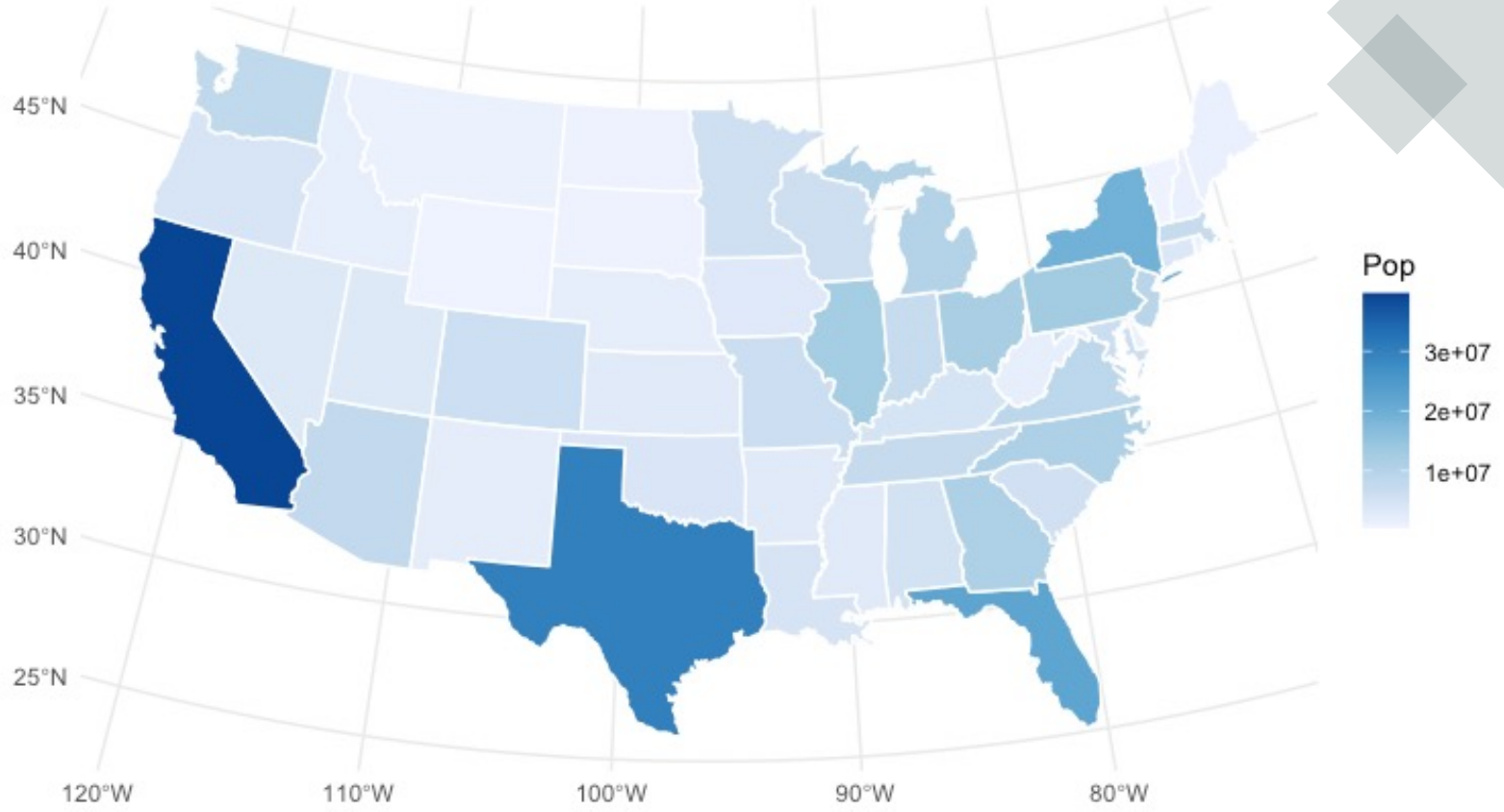
- Interlocking marks
  - Shape coded
  - Position coded
  - Area coded
  - These attributes are *taken*, unavailable for coding additional information.

Thematic Maps

- Shows spatial variability of some attribute (theme)
  - combines geographic / reference map with tabular data
  - join together
    - region: interlocking area marks (provinces, countries, regions with outlines)
      - could also have point marks (cities, locations with lat/lon coordinates)
    - region: categorical key attribute in tabular data
      - used to look up value attributes
- major idioms
  - choropleth
  - symbol maps
  - cartograms
  - dot density maps

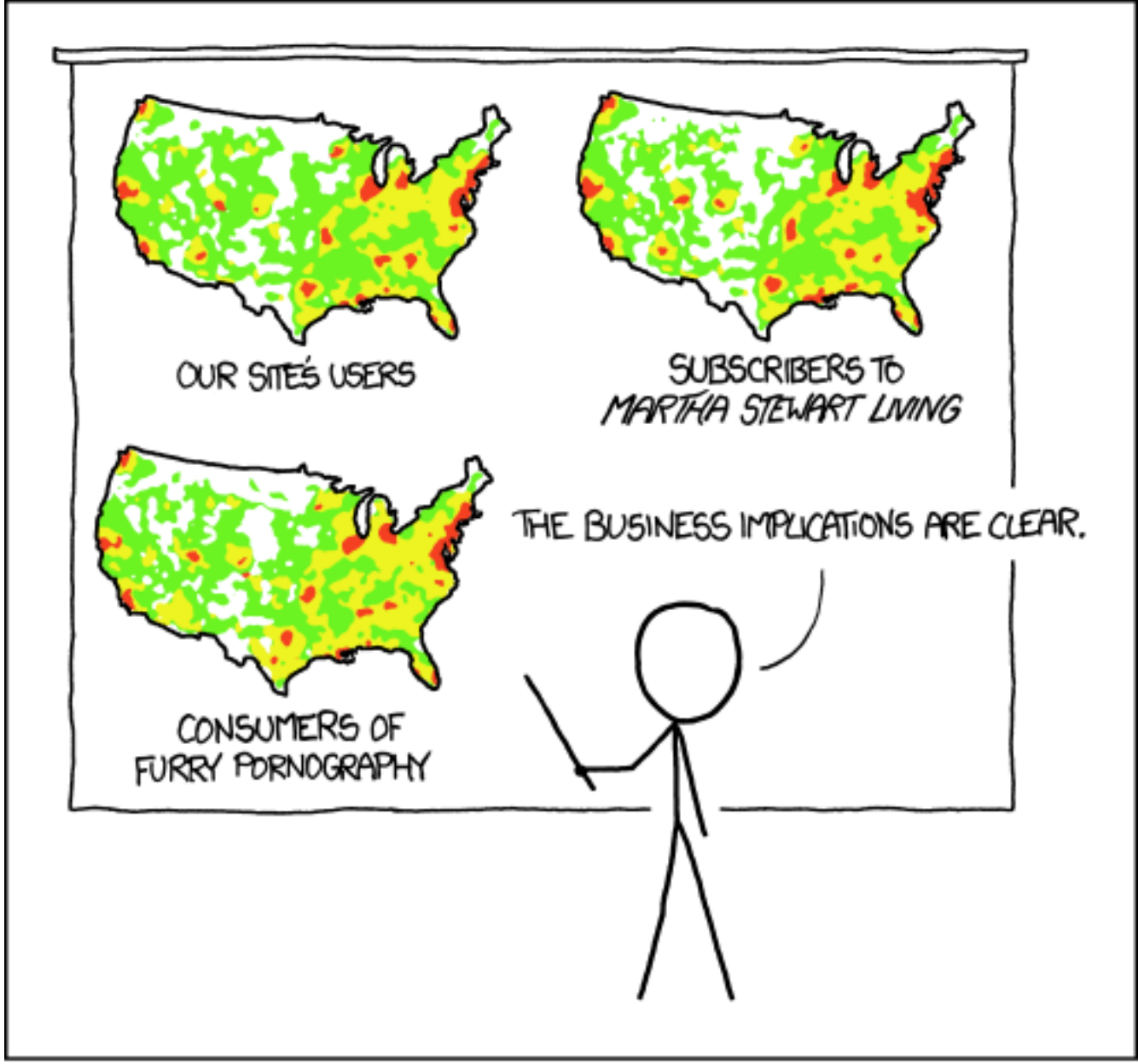
Visualization Idiom
Choropleth map

- Use given spatial data
  - when central task is understanding spatial relationships
- data
  - geographic geometry
  - table with 1 quant attribute per region
- encoding
  - position:
    use given geometry for area mark boundaries
  - color:
    sequential segmented colormap



Beware! Population Maps Trickiness!

- spurious correlations: most maps just show where people live
- consider when to normalize by population density
  - encode raw data values
    - tied to underlying population
  - but should use normalized values
    - unemployed per 100 citizens, mean family income
- general issue
  - absolute counts vs relative/normalized data
  - failure to normalize is common error
- area impacts: population vs population density
  - Trump voting maps



PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE BASICALLY JUST POPULATION MAPS

Choropleth Maps Recommendations

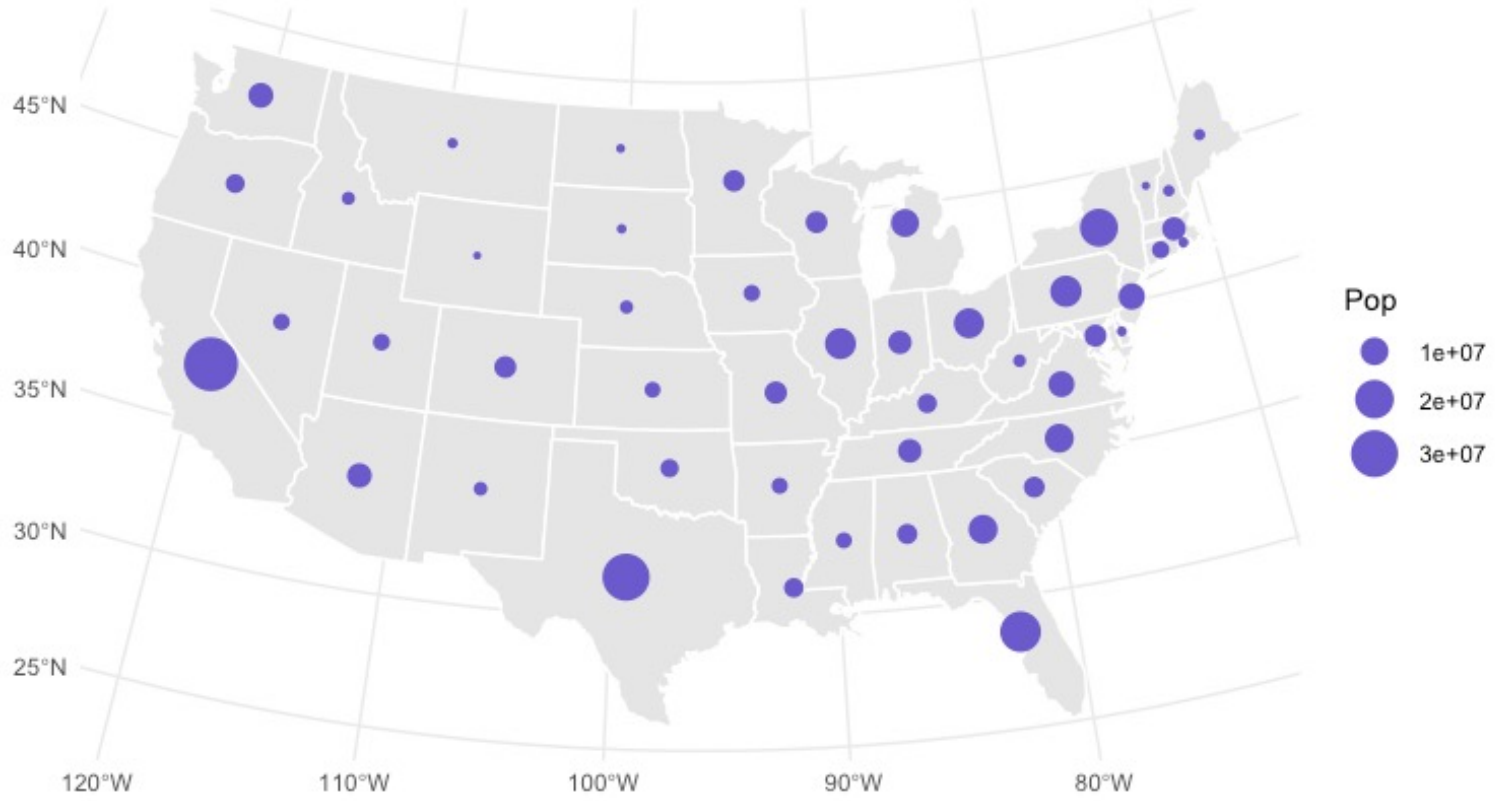
- only use when the central task is understanding spatial relationships
- show only one variable at a time
- normalize when appropriate
- be careful when choosing colors and bins
- best case: regions are roughly equal sized

Choropleth Maps
Pro vs Con

- pros
  - easy to read and understand
  - well established visualization (no learning curve)
  - data is often collected and aggregated by geographical regions
- cons
  - most effective visual variable is already in use for geographic location
  - visual salience depends on region size, not true importance *(“land does not vote” wrt voting results maps)*
    - large regions appear more important than small ones
  - color palette choice has a huge influence on the result

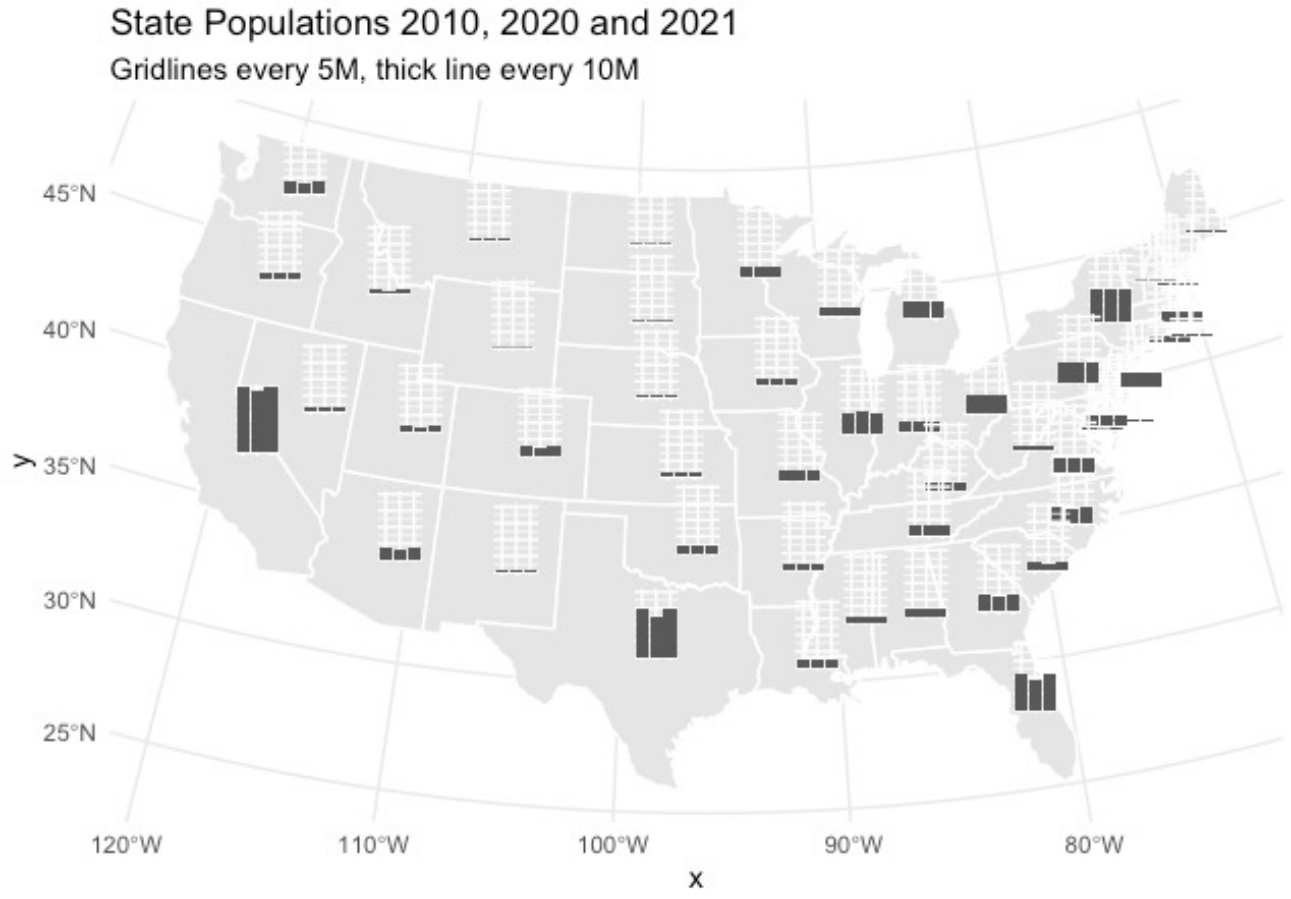
Visualization Idiom
Symbol Map

- symbol is used to represent aggregated data
  - allows use of size and shape as visual channels
    - aka proportional symbol maps, graduated symbol maps
- keep original spatial geometry in the background
- often a good alternative to choropleth maps



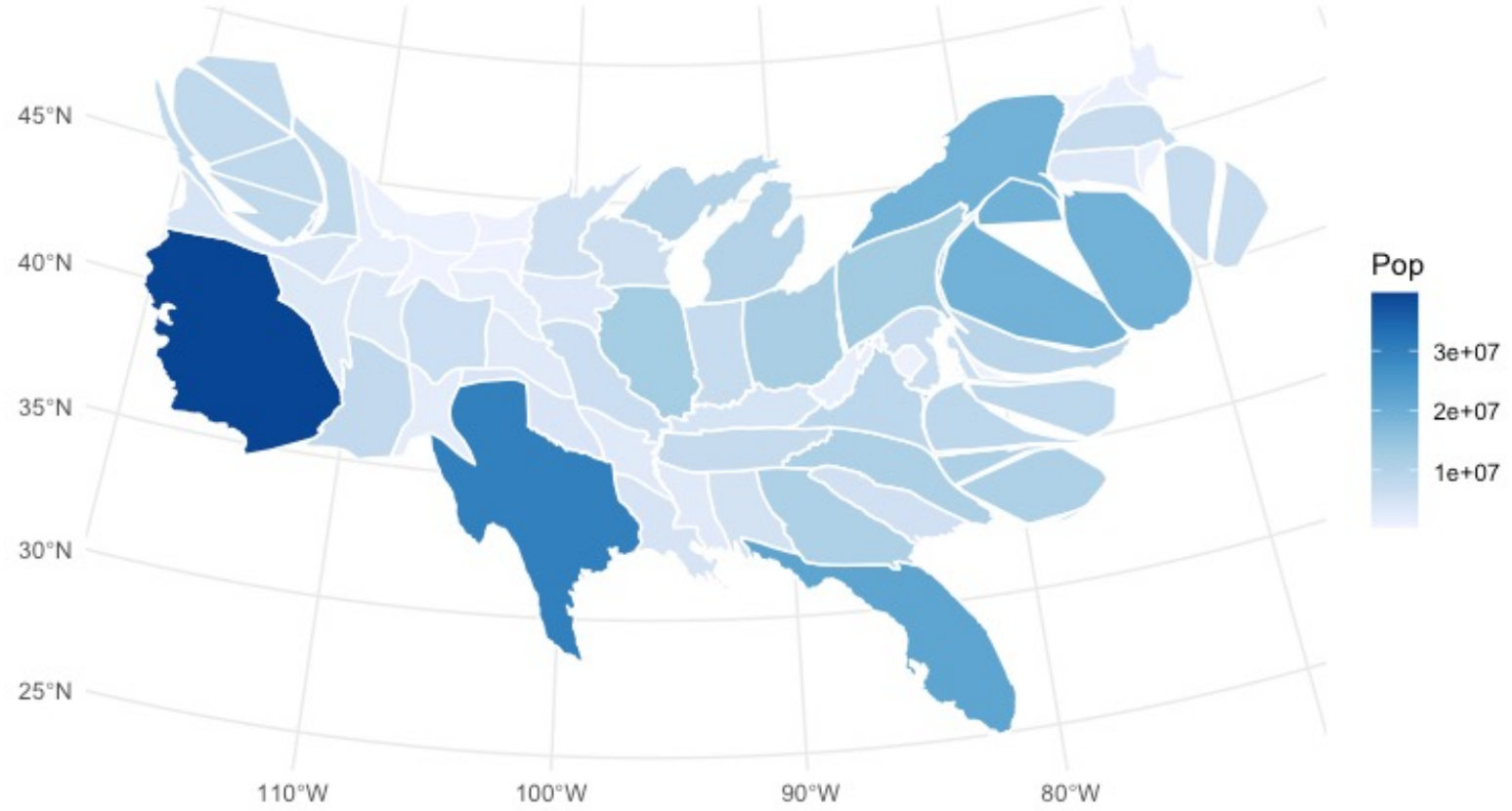
Symbol Maps with Glyphs
Vary the symbol to encode more

- pros
  - somewhat intuitive to read and understand
  - mitigates problems with region size vs data salience
    - marks: symbol size follows attribute value
    - glyphs: symbol size can be uniform
- cons
  - possible occlusion / overlap
    - symbols could overlap each other
    - symbols could occlude region boundaries
  - complex glyphs may require explanation / training



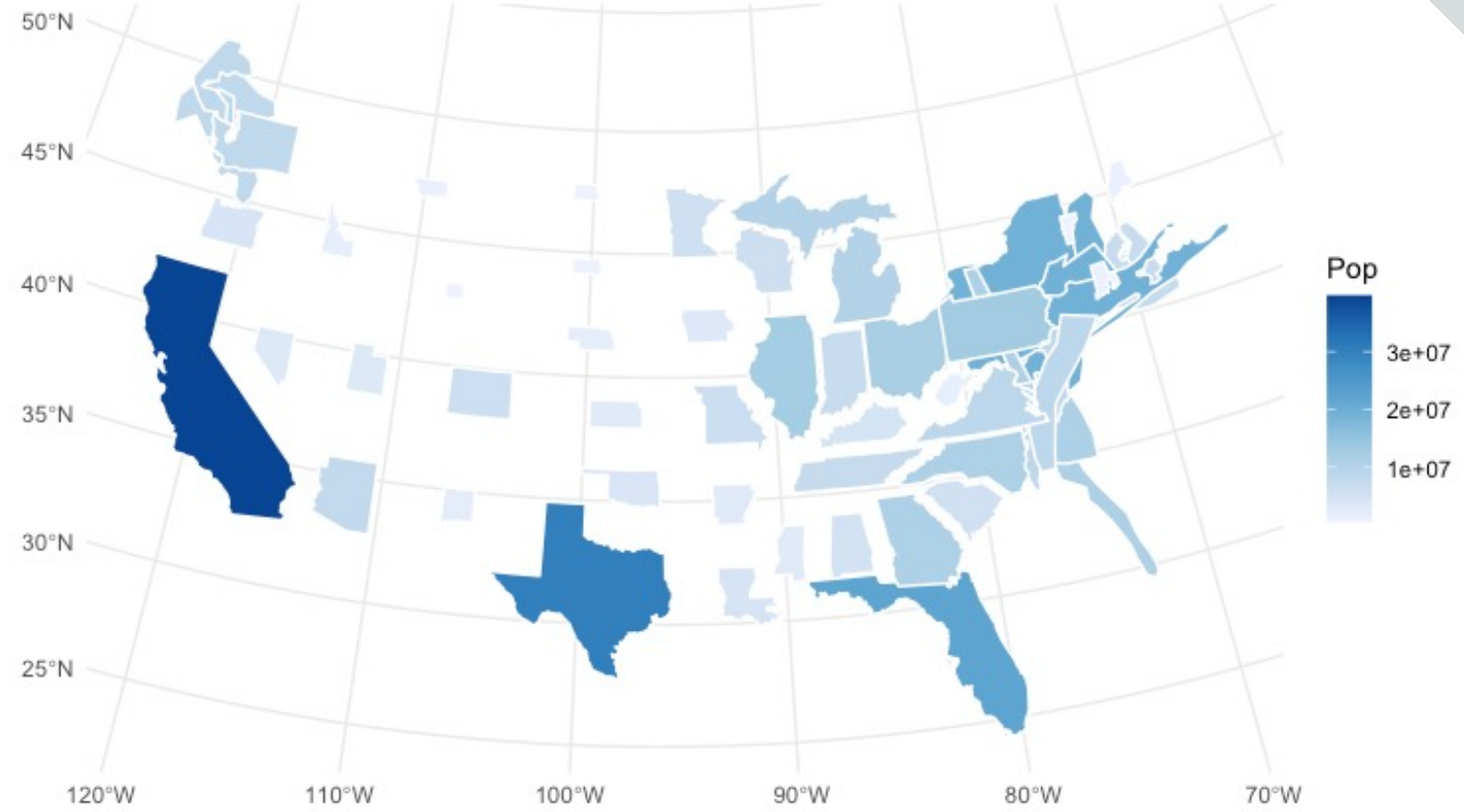
Visualization Idiom
Contiguous Cartogram

- interlocking marks:
  shape, area and position coded
- derive new interlocking marks
  - combine original interlocking marks with new quantitative attribute
- algorithm for new marks
  - input: target sizes
  - goal: shape as close to original as possible
  - requirement: maintain constraints
    - relative position
    - contiguous boundaries with neighbors



Visualization Idiom
Non-contiguous Cartogram

- interlocking marks:
  shape, area and position coded
- derive new interlocking marks
  - combine original interlocking marks with new quantitative attribute
  - resize new marks according to quant attribute
  - maintain centroid position



Visualization Idiom
Grid Cartogram

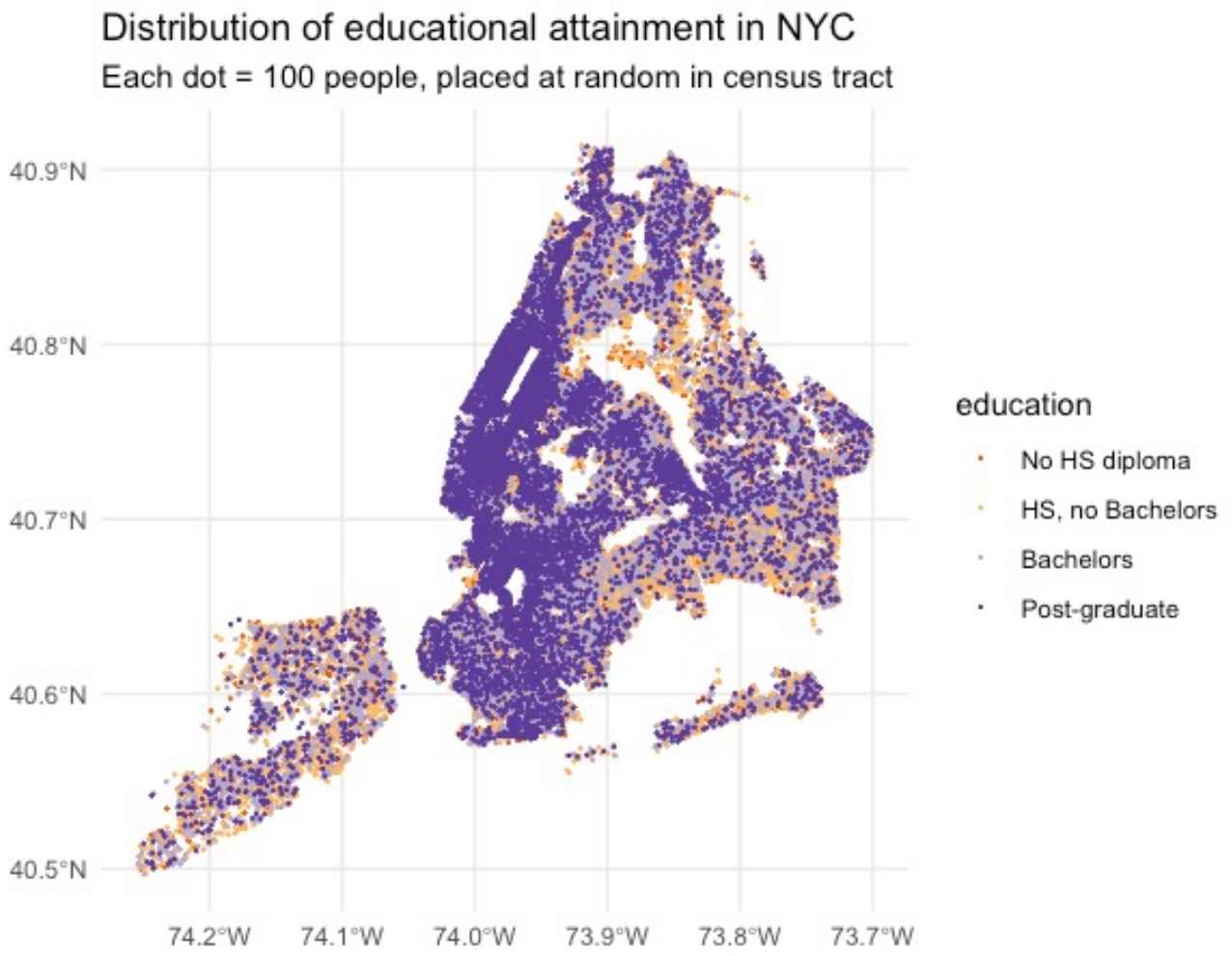
- uniform-sized shapes arranged in a rectilinear grid
- maintain approximate spatial position and arrangement

Cartogram
Pro & Con

- pros
  - Can be intriguing and engaging
  - best case: strong and surprising size disparities
  - non-contiguous cartograms often easier to understand
- cons
  - require substantial familiarity with original dataset & use of memory
    - compare distorted marks to memory of original marks
    - mitigation strategies: animated transitions or side-by-side views
  - major distortion is problematic
    - may be aesthetically displeasing
    - may result in unrecognizable marks
  - difficult to extract exact quantities

Visualization Idiom: Dot density maps

- visualize distribution of a phenomenon by placing dots
- one symbol represents a constant number of items
  - dots have uniform size and shape
  - allows use of color channel
- task: show spatial patterns, clusters



Dot Density Maps
Pro vs Con

- pros
  - straightforward to understand
  - avoids choropleth non-uniform region size problem
- cons
  - challenge: normalization, just like choropleths
    - shows population density (correlated with attributes), not effect of interest
  - perceptual disadvantage:
    difficult to extract quantities
  - performance disadvantage:
    rendering many dots can be slow
    - the plot on the preceding slide has over 60k individual dots, and took almost an hour to process and render

Geographic Coordinates and Projections

Curvature

- The Earth surface has curvature; a sheet of paper, or a computer screen does not
  - Representing one on the other introduces distortion
  - Many ways to put coordinates on the surface of the Earth
  - Many possibly interesting invariants:
    - Area? Angles? Geodesics to be straight lines? No discontinuity in areas of interest? Minimal distortion in areas of interest?
- **Coordinate Reference System** (CRS) is a framework to precisely represent Earth locations as coordinates
  - Choose: Earth ellipsoid, coordinate representation (horizontal datum), map projection (or choose the GCS – geographic coordinate system – longitude/latitude)
  - Standardized in EPSG codes, ISO standards
- Choice of ellipsoid and datum can give the same location different coordinates
  - And any high-precision choice has to be continuously updated: tectonic movements, tidal effects, effects from the last ice age: Scandinavia rises with 1cm/year after the last ice age

CRS Types

- Geographic (geodetic) coordinate system
  - Spherical coordinates, longitude and latitude. Discontinuous at poles, and opposite the *prime meridian*.
    - WGS84 – World Geodetic System; NAD83 (North Americn Datum); ED50 (European Datum, NATO); OSGB36 (UK Datum);
- Geocentric coordinate system
  - 3D system, used by satellites and GPS
- Projected coordinate system (planar, grid)
  - Map projections
- Engineering coordinate system
  - Custom-made for small area by explicitly surveying

Map Projection Types and Properties

Types

- Cylindrical
  - Meridians map to vertical lines, parallels map to horizontal lines
- Pseudocylindrical
  - Some meridians map to regularly spaced curves instead of straight lines
- Conic
  - Meridians as straight lines, parallels as circle arcs
- Pseudoconical
  - Non-central meridians may be complex curves
- Azimuthal
  - Meridians as straight lines, parallels as complete concentric circles
- Pseudoazimuthal
  - Equator and central meridian to perpendicular intersecting straight lines. Parallels bow away from the equator, meridians bow toward the central meridian.
- Other
  - Not based on a particular projection
- Polyhedral
  - Can be folded up to a polyhedral approximation of the sphere, usually maps each face with low distortion, but has discontinuous cut-lines

Properties

- Conformal
  - Local angles preserved, local shapes undistorted, local scale constant in all directions
- Equal-area
  - Area preserved everywhere
- Compromise
  - Balance between conformality and equal-area, to reduce overall distortion
- Equidistant
  - All distances from one (or two) specific points are correct
- Gnomonic
  - All great circles are straight lines
- Retroazimuthal
  - Shortest rout direction to a fixed location corresponds to map direction to that location

Significant Map Projections

- Cylindrical
  - Plate Carrée
    - 120 CE, distances along meridians conserved
  - Mercator
    - 1569 CE, lines of constant bearing are straight (easy for navigation); areas distort as map approaches poles
  - Web Mercator
    - 2005 CE, simplification of Mercator for fast calculation, clips latitudes at ±85.05° for square presentation. De facto standard for web-based maps.
  - Gall-Peters
    - 1855 CE, equal-area, with standard parallels at ±45°
  - Cassini
    - 1745 CE, distances along central meridian are conserved, distances perpendicular to central meridian are preserved.
- Conical
  - Albers
    - 1805 CE, two standard parallels, low distortion between them
  - Lambert conformal
    - 1772 CE, used in aviation charts
- Polyhedral
  - Dymaxion
    - 1943 CE, Introduced by Buckminster Fuller
  - Waterman Butterfly
    - 1996 CE
- Other
  - Gnomonic
    - 580 BC, All great circles map to straight lines, shows less than half the globe
  - Two-point equidistant
    - 1919 CE, Any two control points – all straight line distances to those two points are accurate
  - GS50
    - 1982 CE, particularly low distortion of Continental US

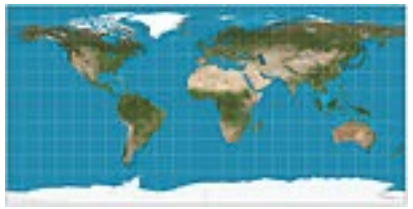
Significant Map Projections

Mercator

Web Mercator

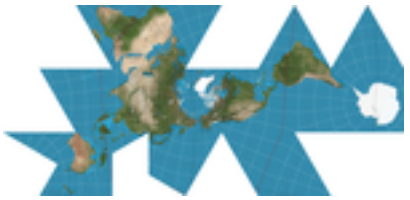
Gall-Peters

Plate Carrée



Cassini

Dymaxion



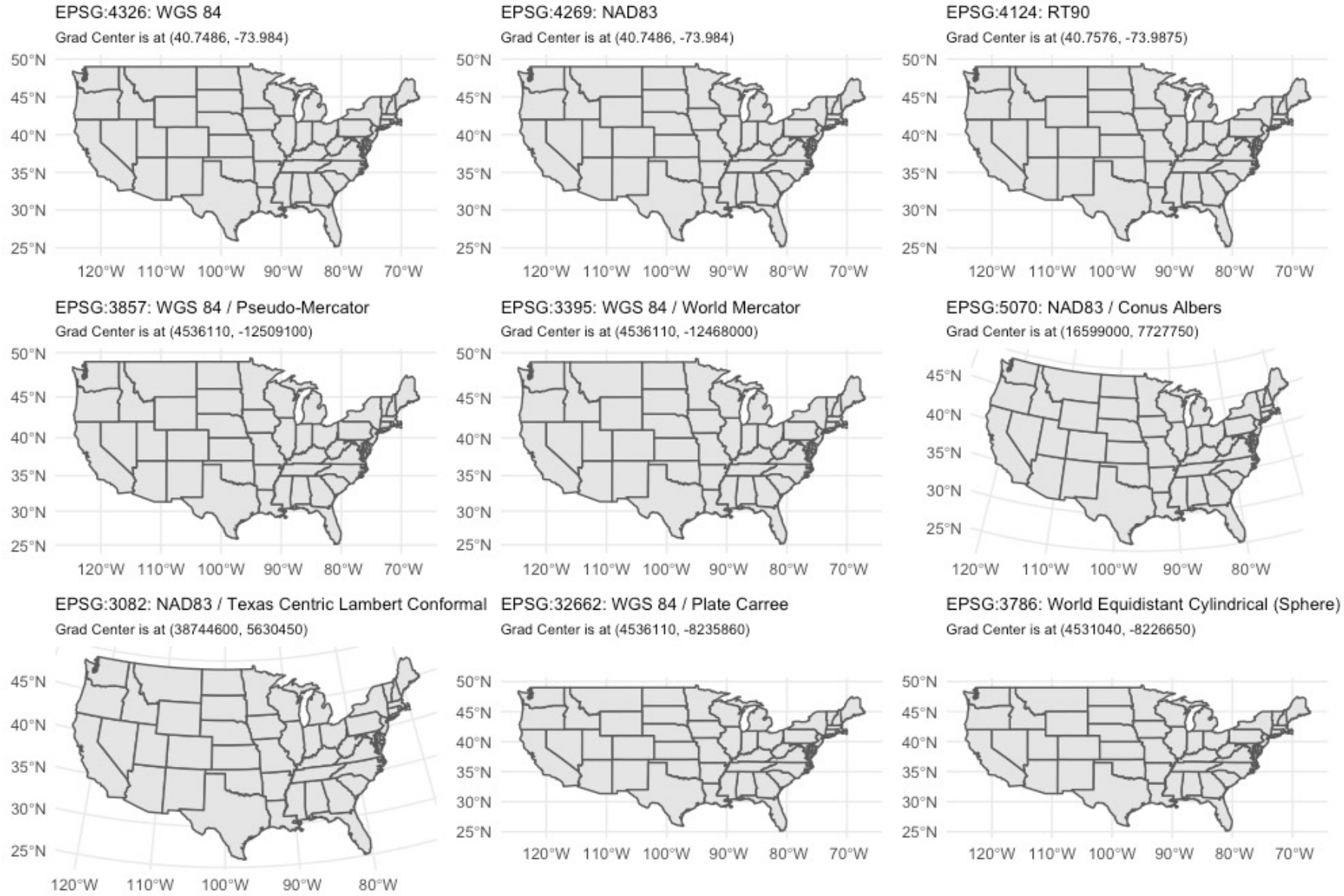
Lambert Conformal Conic

Albers Conic

Gnomonic

GS50

Significant or Interesting CRS Choices



BAD MAP PROJECTION #248:

MADAGASCATOR

MERCATOR PROJECTION BUT WITH THE NORTH POLE IN THE INDIAN OCEAN
SO IT EXAGGERATES THE SIZE OF MADAGASCAR INSTEAD OF GREENLAND