Data Science II

- Introduction to Data Visualization -

Visualizing Geospatial Data

Prof. Dr. Eduard Kromer
Summer Semester 2024
University of Applied Sciences Landshut

Visualizing Geospatial Data

Geospatial

- datasets often contain information linked to locations in the physical world
  - e.g. where do people with specific attributes (income, age etc.) live
  - e.g. where have specific plants been found
- plotting geographic data involves adding color to geographic areas or adding symbols (circles, squares, lines) on top of a geographic map

Challenges of Maps

- the size of a geographic area sometimes does not correspond to the importance of the data value
    - the map can distort our perception of the important values being visualized
- you should always question if a map is truly the best way to present geograhic data

Choropleth Map

Choropleth Map

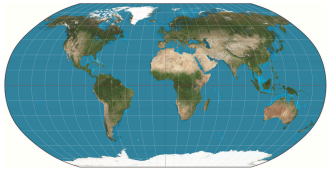
- choropleth maps use colors, shades, or patterns on geographic units to show proportionate quantities and magnitutes
- smaller numbers usually correspond to lighter colors and larger numbers to darker colors (a “color ramp”)

Projections

- the world is a globe, but maps are flat
  - a map maker must choose a map projection to transform the world's sphere into a two-dimensional plane
  - Which projection does the best job depicting the earth in two dimensions?

Projections

Mercator projection

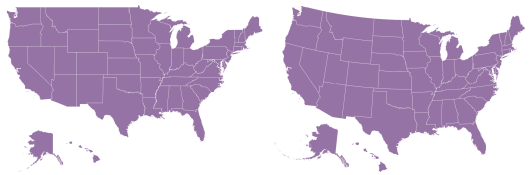


Robinson projection

Projections

- Mercator projection:
  - distorts the size of objects as the latitude increases from the equator to the north and south poles
  - countries closer to poles (e.g. Greenland) appear much larger than they actually are
    - Greenland appears to be about the same size as South America, when it is about one-eight the actual size
  - can work well for small areas

Projections



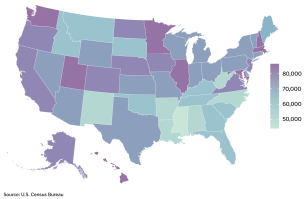
Mercator projection

Albers projection

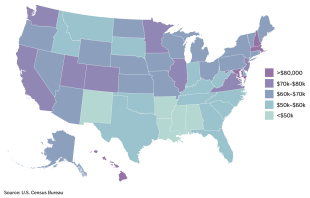
Choosing the Bins

- the choice of intervals that will shade the geographic units plays an important role for our visualization
- binning methods:
  1. no bins: choose a continuous color palette
     - each data value receives its own unique color tone
  2. equal interval bins: divide the data range into an equal number of groups
  3. data distribution bins: arrange bins to hold the same number of observations (or use variance or standard deviation)
  4. arbitrary bins: use an arbitrary criterion (e.g. round numbers) for binning

Binning Methods

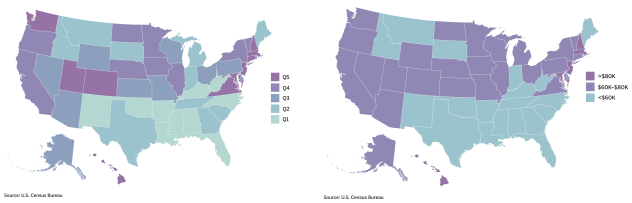


No bins



Equal interval bins

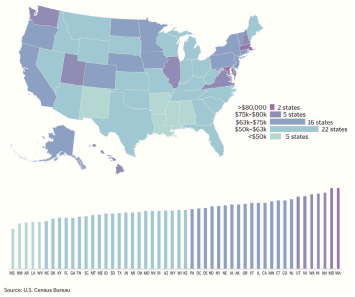
Binning Methods



Data distribution bins

Arbitrary bins

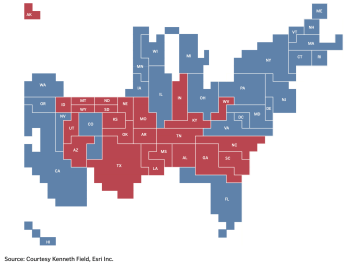
Binning Methods



Cartograms

- a cartogram reshapes geographic areas based on their values
- these adjustments more accurately visualize the data because cartograms correlate the data and the geographic size
- but they are less intuitive for us
- there are four primary types of cartograms: contiguous, noncontiguous, graphical and grided

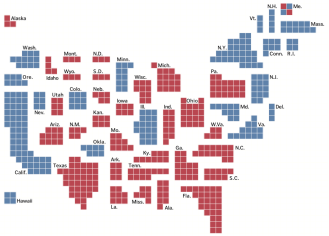
Contiguous Cartogram



Source: Courtesy Kenneth Field, Esri Inc.

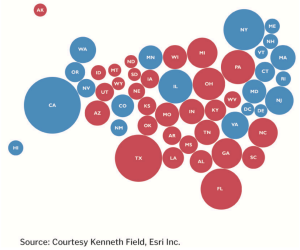
- the contiguous cartogram adjusts the size of each geographic unit according to the data
- here: each state is sized to its number of electoral votes

Noncontiguos Cartogram



- the noncontiguous cartogram adjusts the size of each geographic unit according to the data, but the units are broken apart and not kept adjacent
- this way we maintain the shape of the individual units but distort the overall view

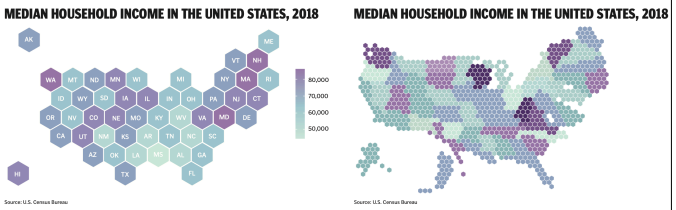
Graphical Cartogram



Source: Courtesy Kenneth Field, Esri Inc.

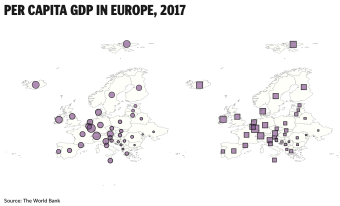
- graphical cartograms do not maintain the original shape of the geographic units and instead use other shapes sized to the data values

Gridded Cartogram



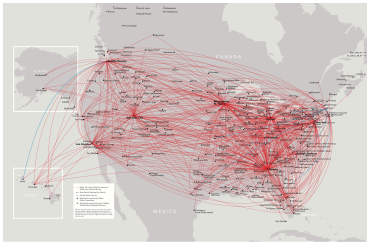
Different shapes are scaled to the data and arranged so they maintain the general shape of the major geography (hexagons are most often used).

Proportional Symbol Maps



- proportional symbol maps: symbols are sized proportionate to the data
- the two maps show per capita GDP for European countries encoded with circles and squares

Flow Map



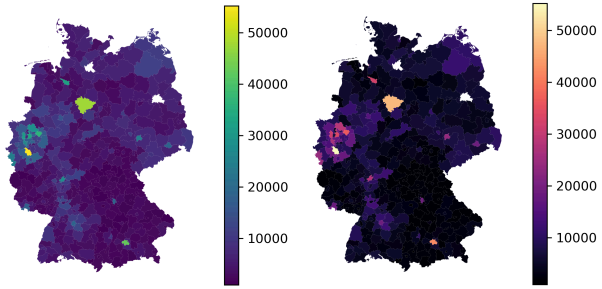
- flow maps show movement between places
- arrows and lines denote the direction of the flow
- the width of the line can correspond to the data value

Choropleth Maps with Python

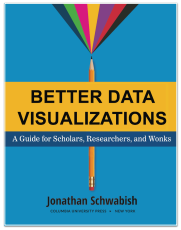
- there exist several Python libraries for choropleth map visualizations, e.g.
  - `geoplot` (see also `geopandas`)
  - `folium`
  - `plotly.choropleth` (see tutorial)
- download German geo data from "Bundesamt für Kartografie und Geodäsie"
- use geopandas to read in vg2500_krs.shp
- download German unemployment data from Statistisches Bundesamt:

  "Arbeitslose nach ausgewählten Personengruppen - Jahresdurchschnitt - regionale Tiefe: Kreise und krfr. Städte - Jahr 2020"
- from both datasets create the corresponding choropleth map

Choropleth Maps with Python



Literature



References

- Slide 5,8,10,12-14,16-21; Image Source: J. Schwabisch - Better Data Visualizations, Columbia University Press