Data Exploration & Visualization

Module 9

Geospatial Visualization

Dr. ZENG Wei

DSAA 5024

The Hong Kong University of Science and Technology (Guangzhou)

Data Exploration & Visualization

Module 9: Geospatial Visualization

- Geospatial data representation
  - Map projection
  - Vector vs. raster
- Geospatial visual representation
  - Geometry-based
  - Choropleth map
  - Cartograms
  - Fields/Lines (Future Week)

Geographic data

- **Geo**graphy data attributes
  - Locations of objects on **the Earth**.
  - Location ⇒ distance, area, elevation, etc.
- Collected by location-positioning sensors
  - mobile phones
  - GPS devices
- Collected by location-based services
  - Foursquare, 美团
  - Google, Twitter, Facebook

Map projection

- The earth is a spheroid.
- The best model of the earth is a globe
  - not easy to carry
  - not scalable
  - not good for making measurement (distance, area, angle)
  - not easy to display on 2D screens

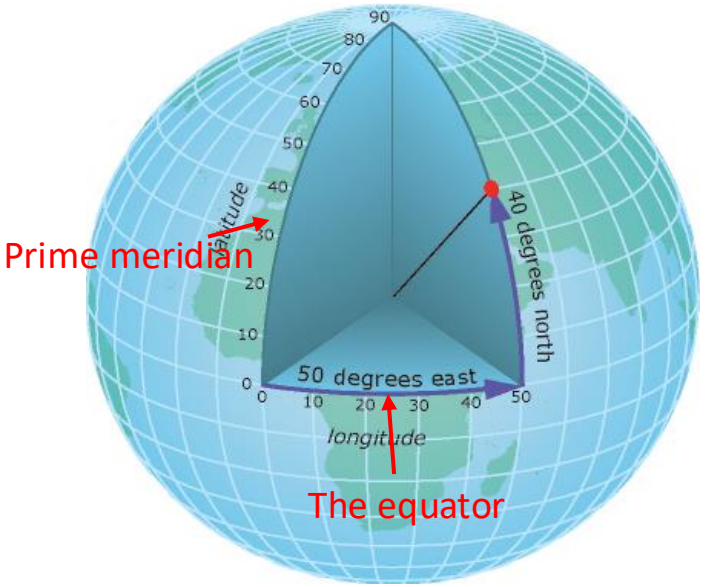
Map projection

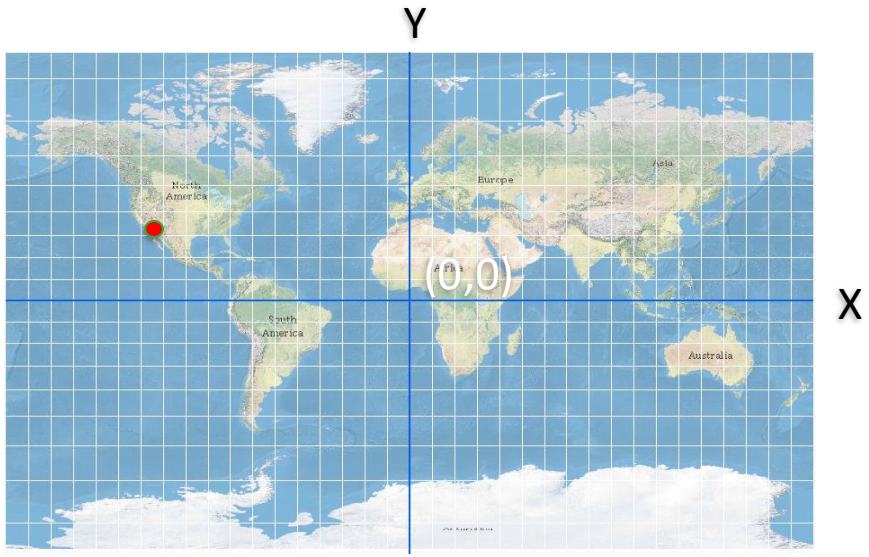
- Map projection is a method for mapping spatial patterns on a curved surface (the Earth's surface) to a flat surface.

(x, y) → $f(x, y)$

input
unprojected
polar coordinates
angles (lat/long)

output
projected
Cartesian coordinates



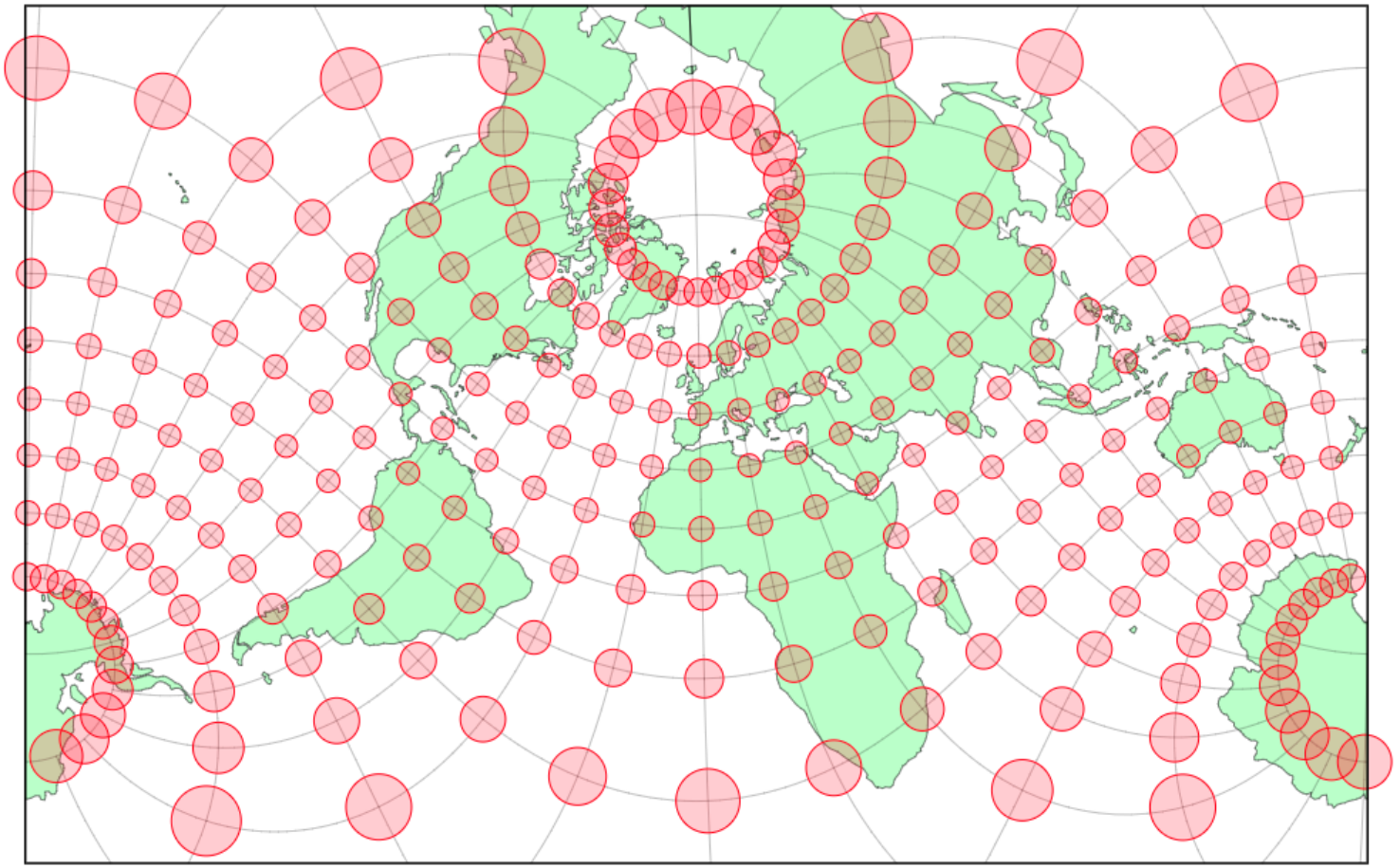


Map projection property

- **Conformal:** local shapes are preserved
- **Equal-Area:** areas are preserved
- **Equidistant:** distance from a single location to all other locations are preserved
- **Azimuthal:** directions from a single location to all other locations are preserved

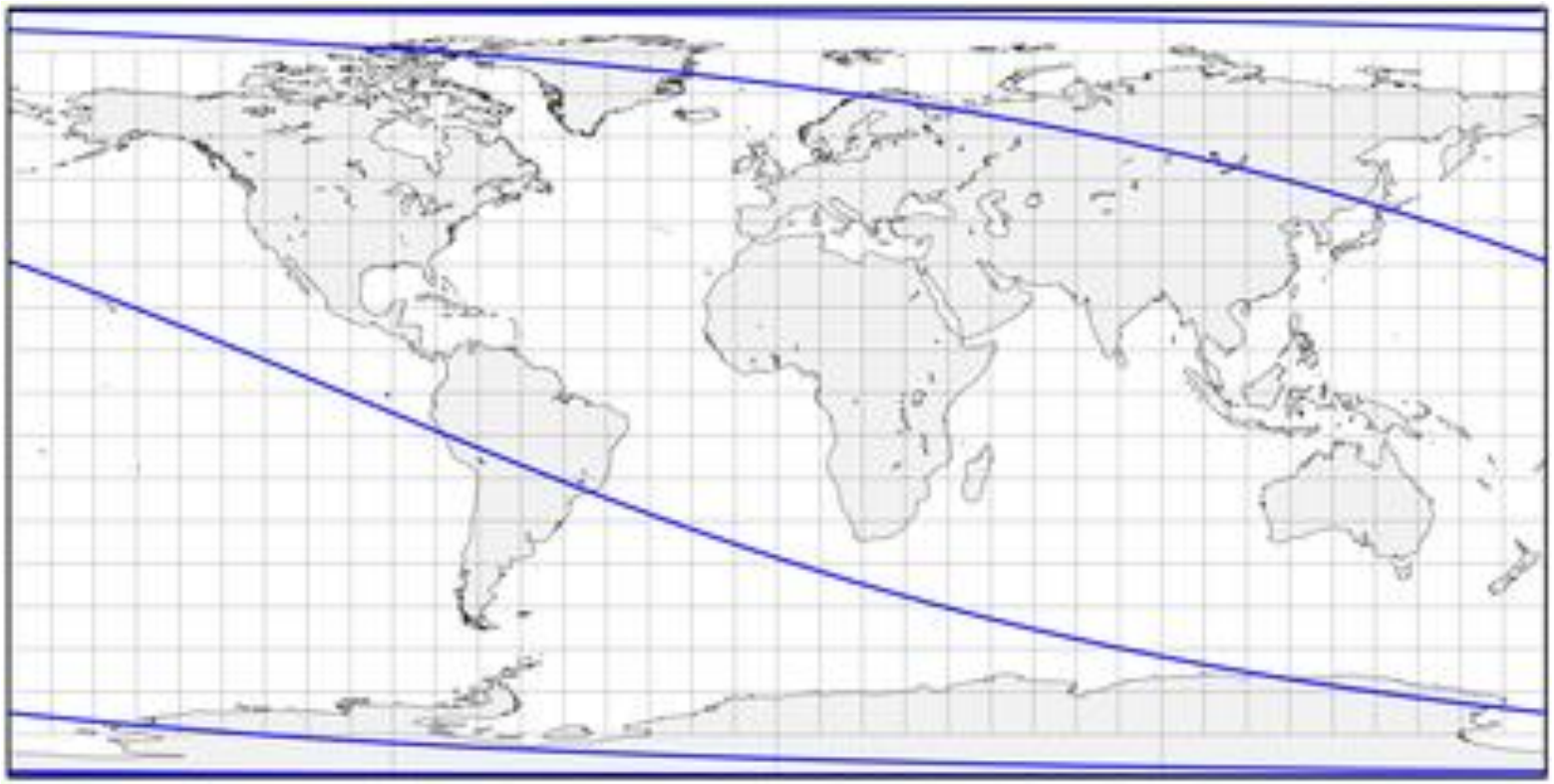
Map projection property

- Tissot's Indicatrix (https://www.jasondavies.com/maps/tissot/)

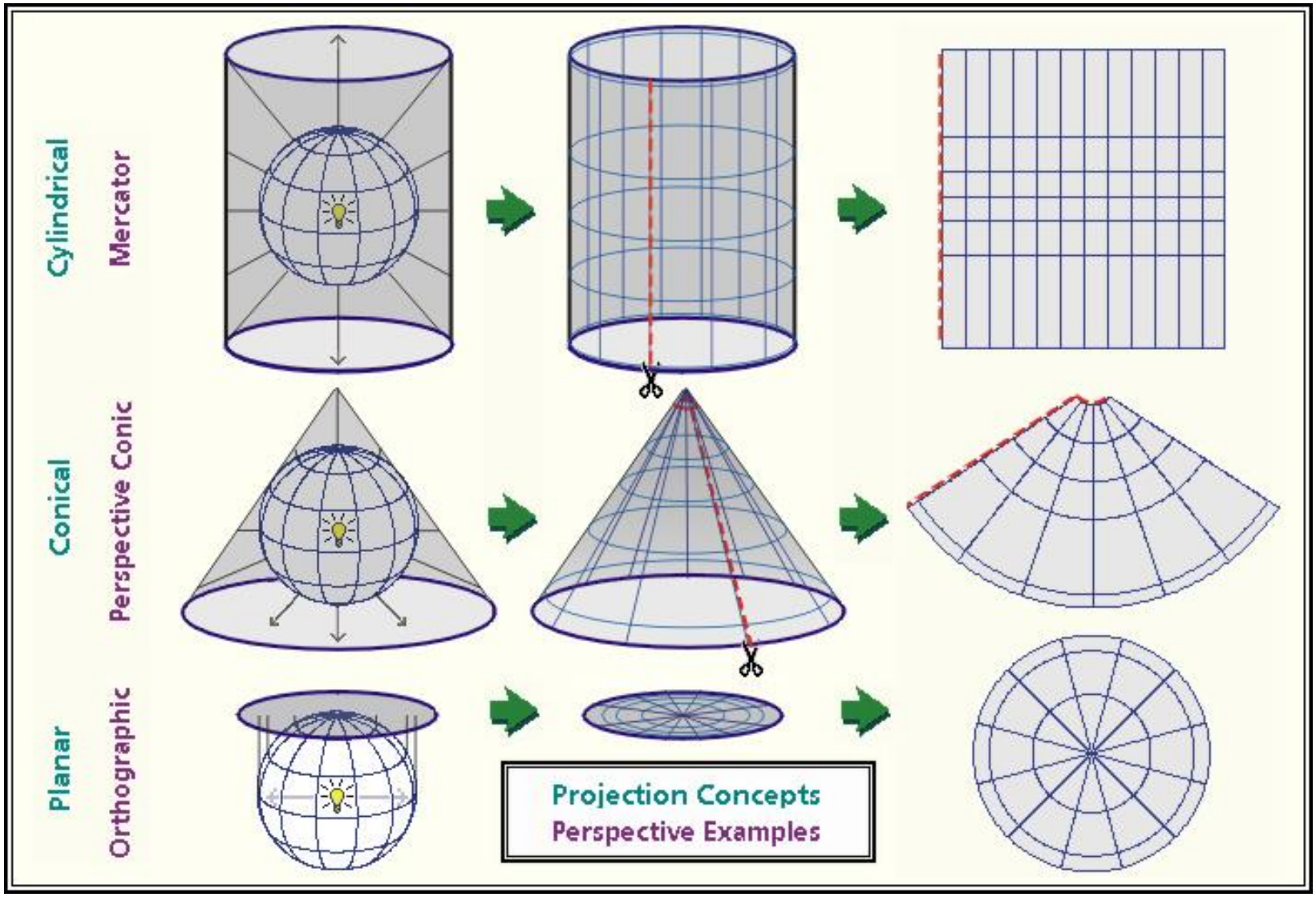


Map projection property

- No map projection can preserve all 4 properties without distortion

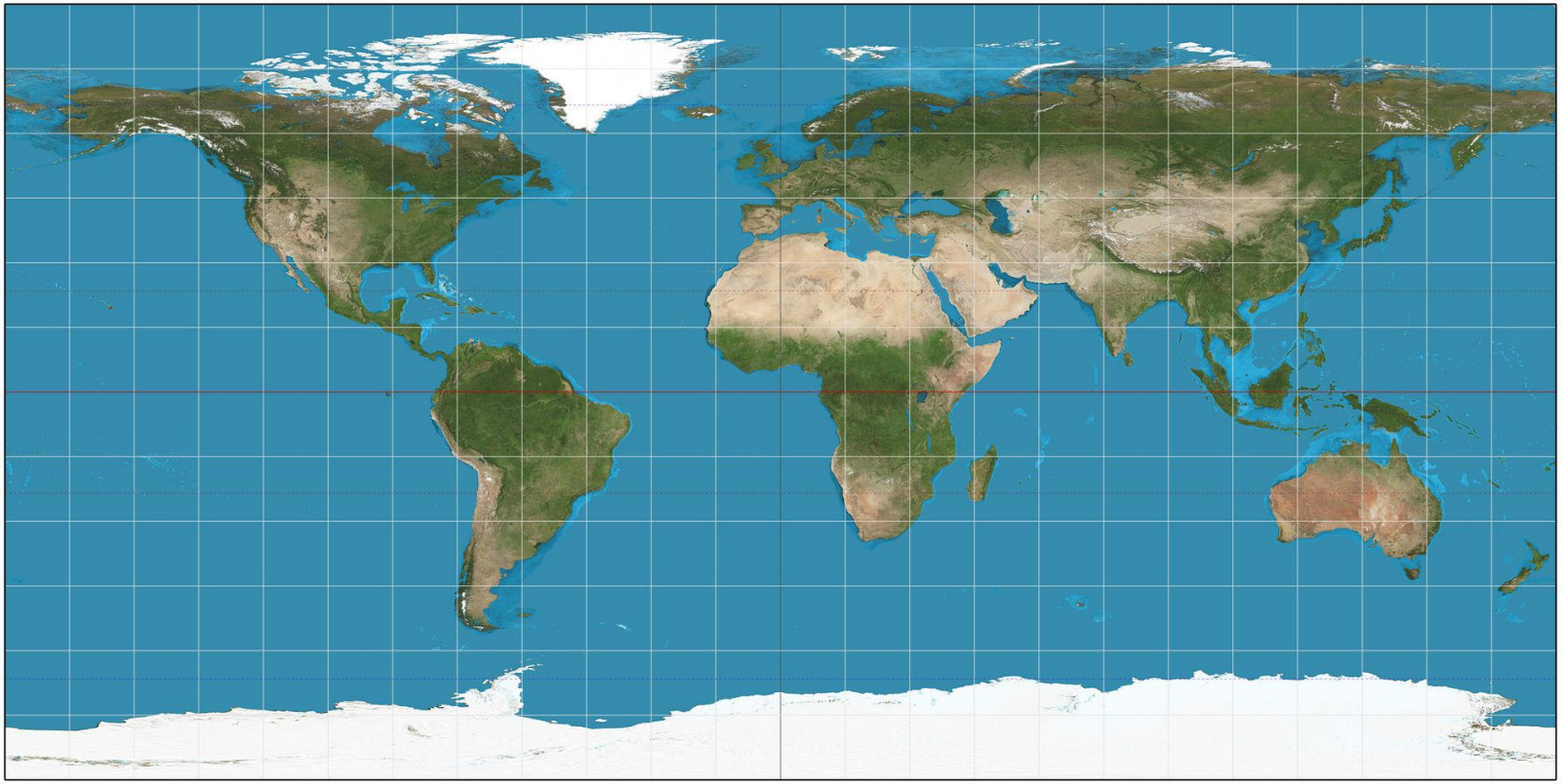


Map projection taxonomy



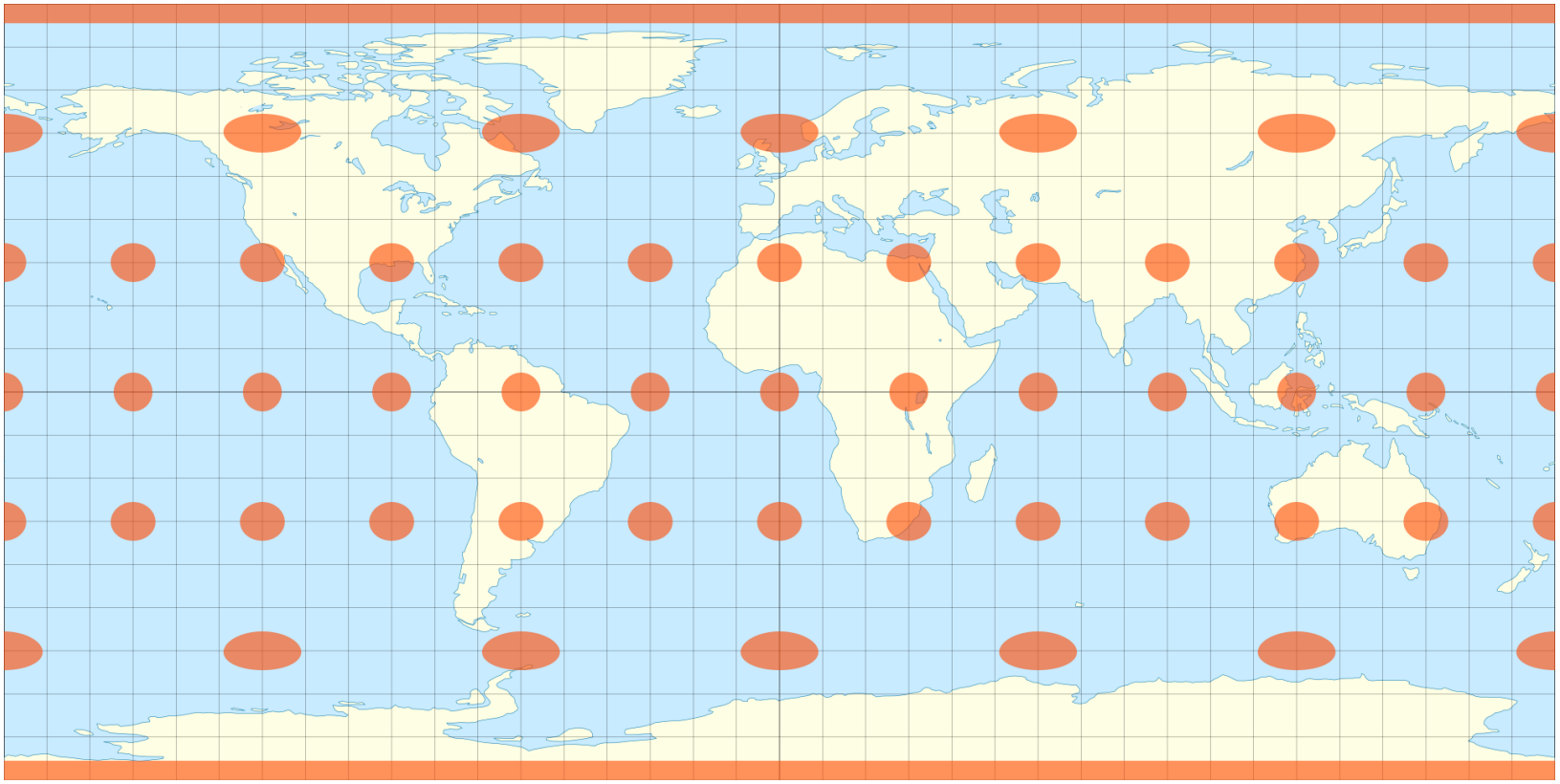
Equirectangular projection

- y = lat, x = lon
- Preserves lat and lon



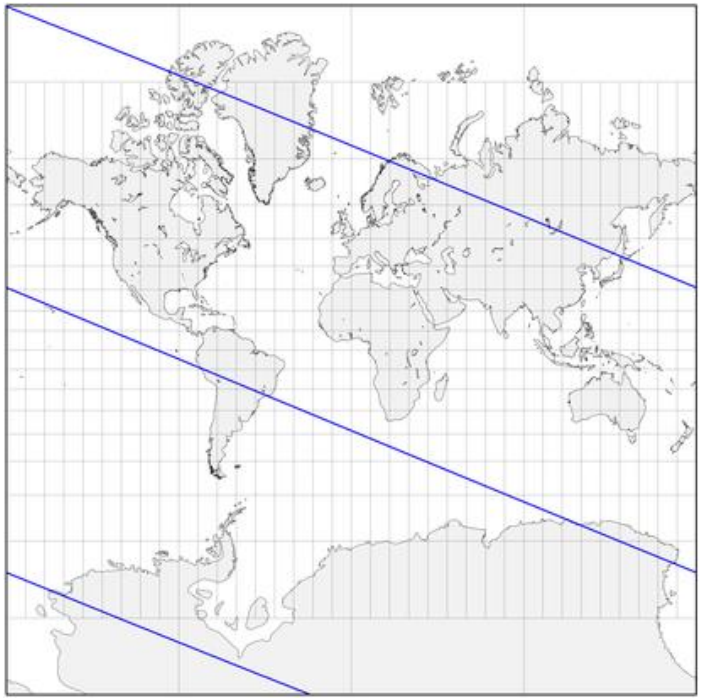
Equirectangular projection

- Neither equal area nor conformal
- Seldomly used in navigation or cadastral mapping



Mercator projection

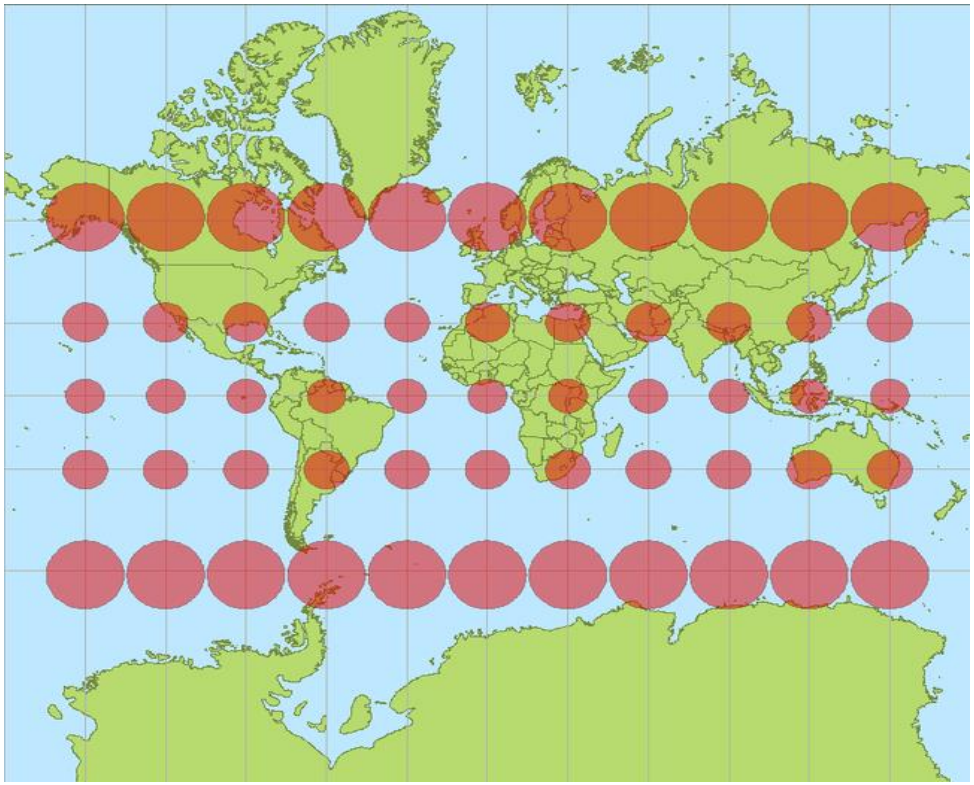
- Conformal: local shapes are reserved
- Bearing: following a compass direction makes a straight line
  - can be used for navigation

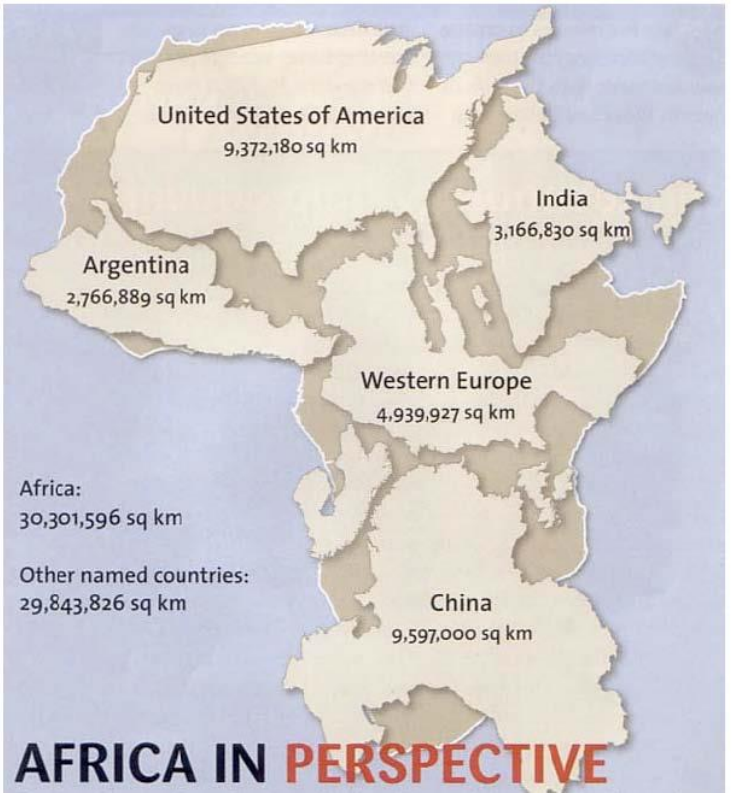


Mercator projection

- Areas are distorted
  - y = log (tan (45 + lat/2))
  - x = long

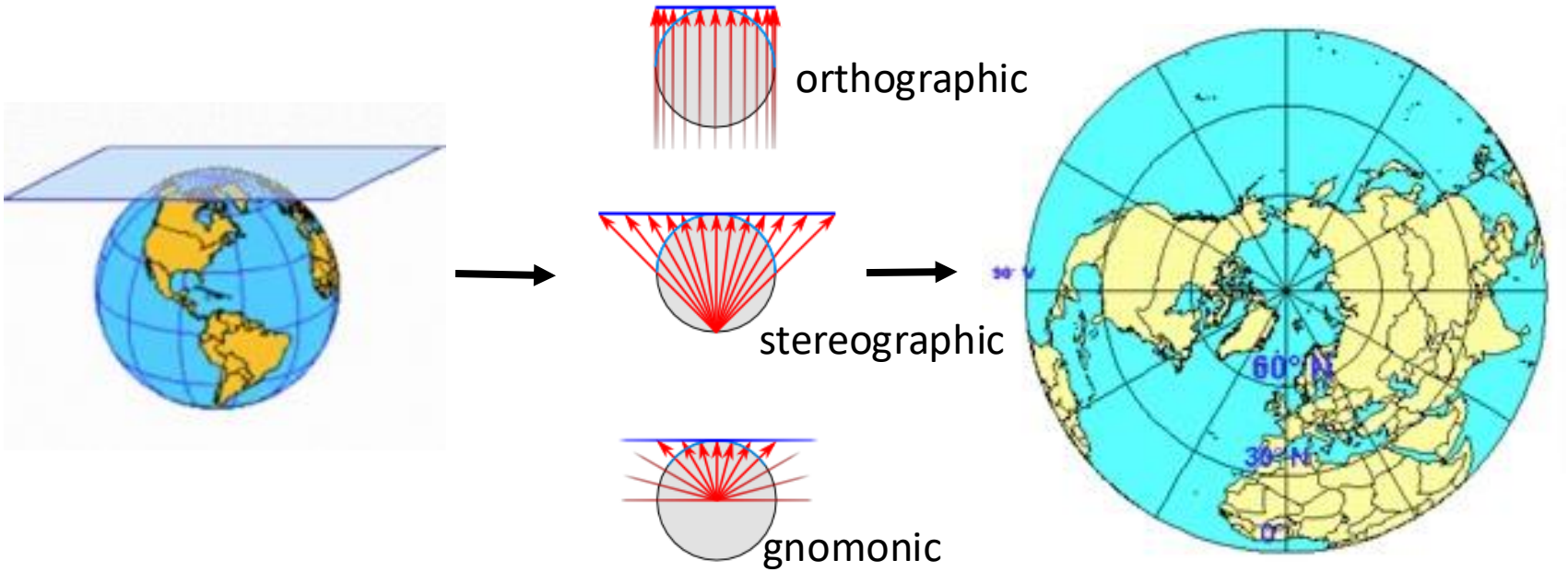
http://thetruesize.com/





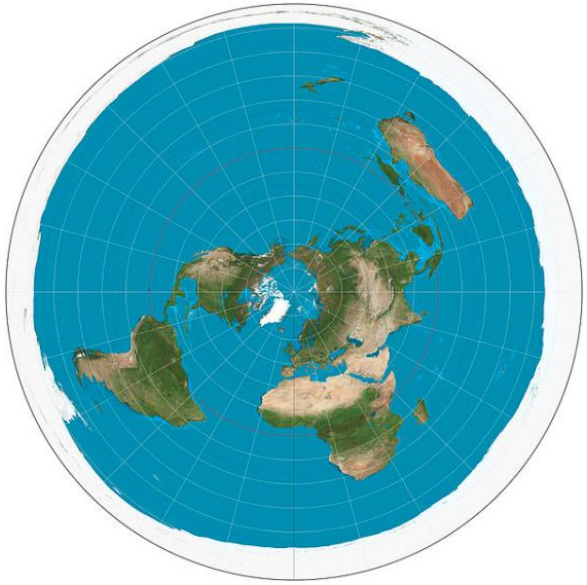
Azimuthal projection

- Directions from a central point are preserved
- Great circles through the central point are represented by straight lines on the map.



Azimuthal projection

- Directions from a central point are preserved
- Great circles through the central point are represented by straight lines on the map.

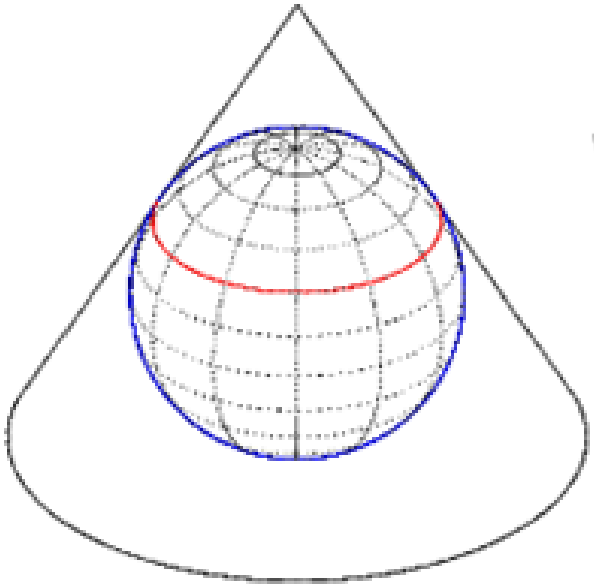


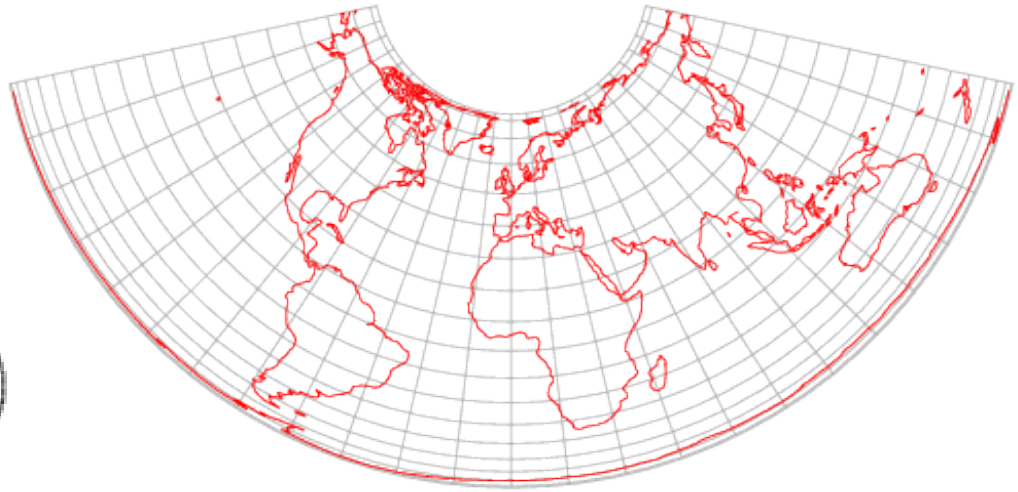
Polar azimuthal equidistant projection

The United Nations Emblem

Conic projection

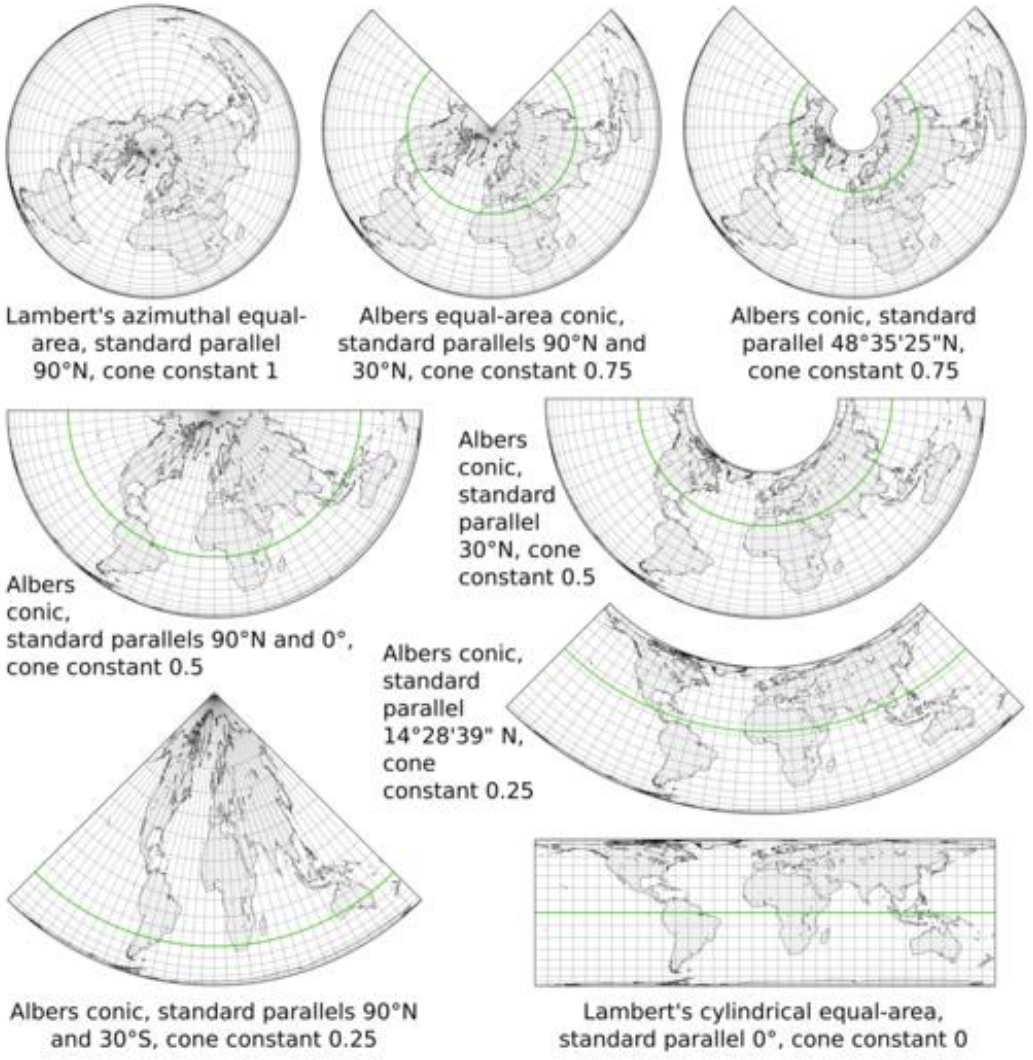
- Meridians are straightly equally-spaced lines
- Meridians coverage at a point, usually the pole
- Parallels are arcs of circles, concentric in the point





Conic projection

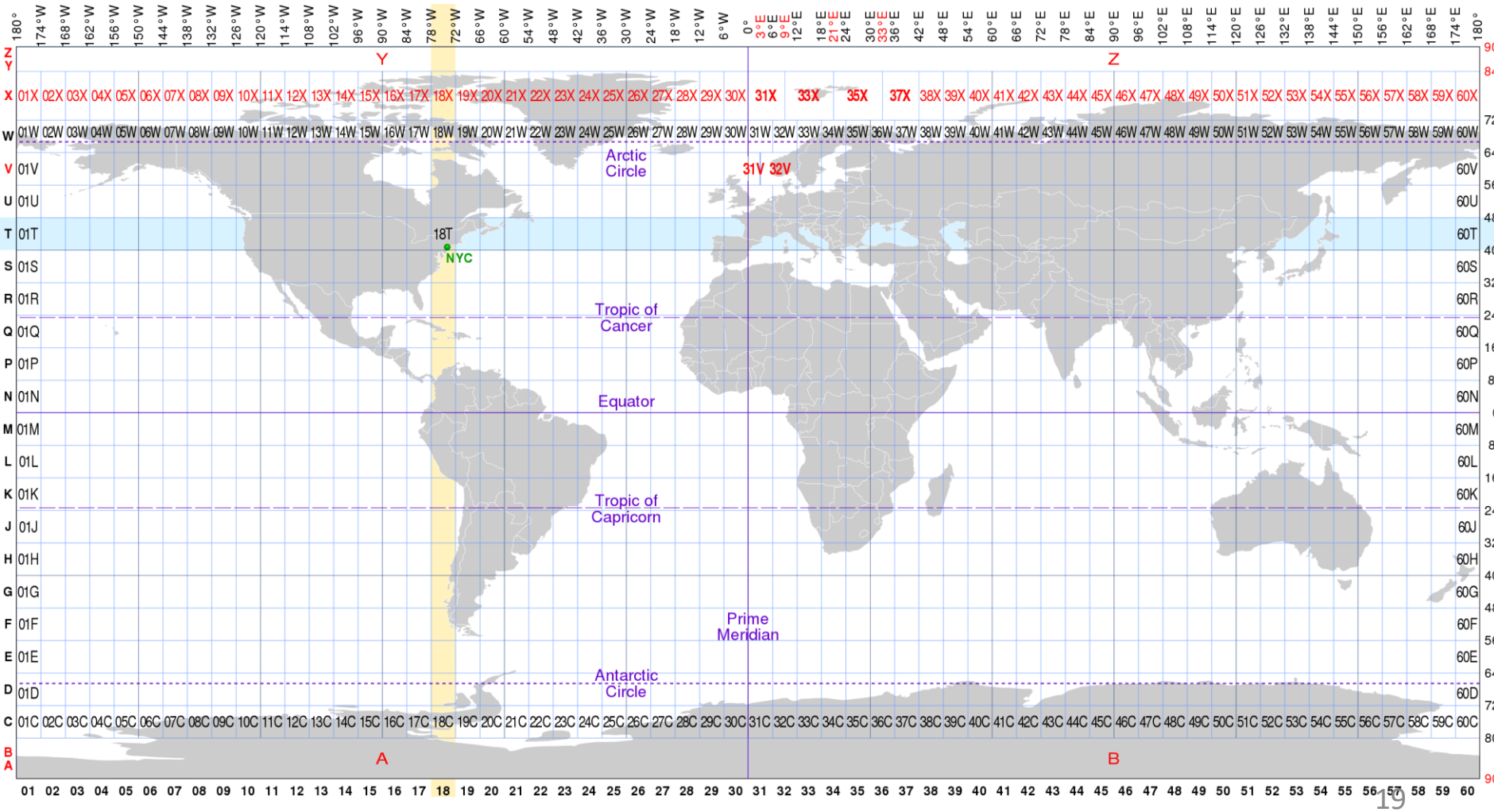
- General cases of azimuthal and cylindrical projections



In applications

- Which map projection is used in maps (Gaode, Baidu, Google)?
  - Equirectangular? Mercator? Azimuthal? Conic?
- Why it is used?
  -

Universal Transverse Mercator (UTM)

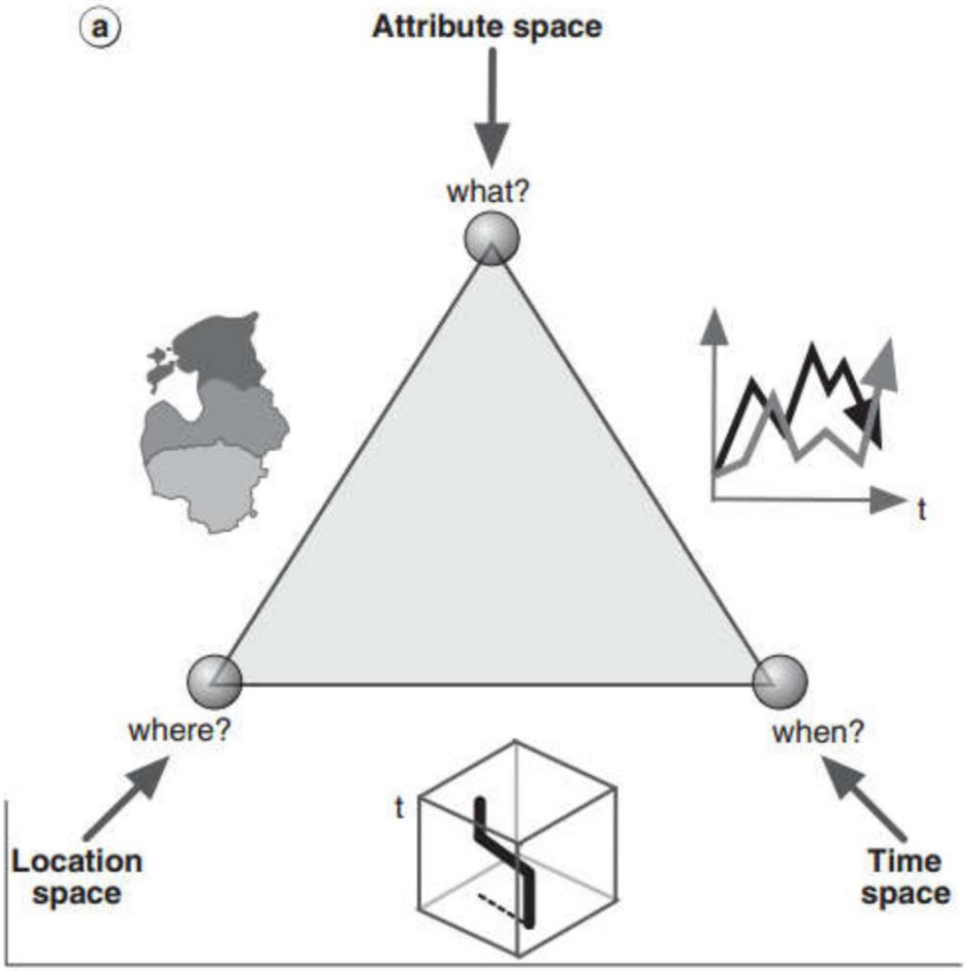


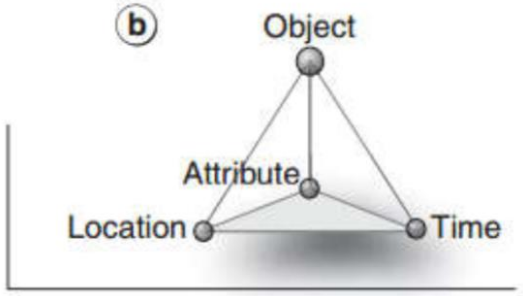
Data Exploration & Visualization

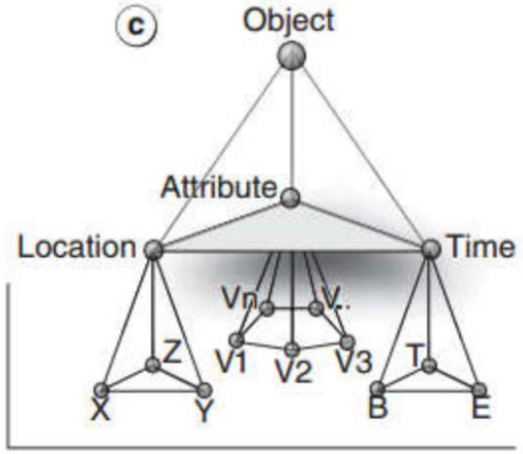
Module 9: Geospatial Visualization

- Geospatial data representation
  - Map projection
  - Vector vs. raster
- Geospatial visual representation
  - Geometry-based
  - Choropleth map
  - Cartograms
  - Fields/Lines (Future Week)

Geographic data



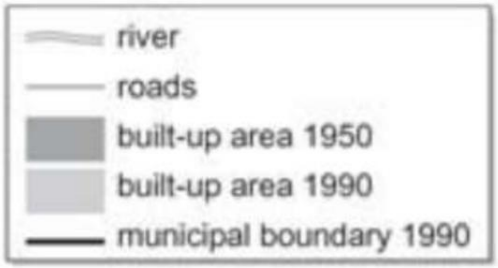




Kraak, Menno-Jan, and Ferjan Ormeling. Cartography: visualization of spatial data. Guilford Press, 2011.

Geographic data analysis

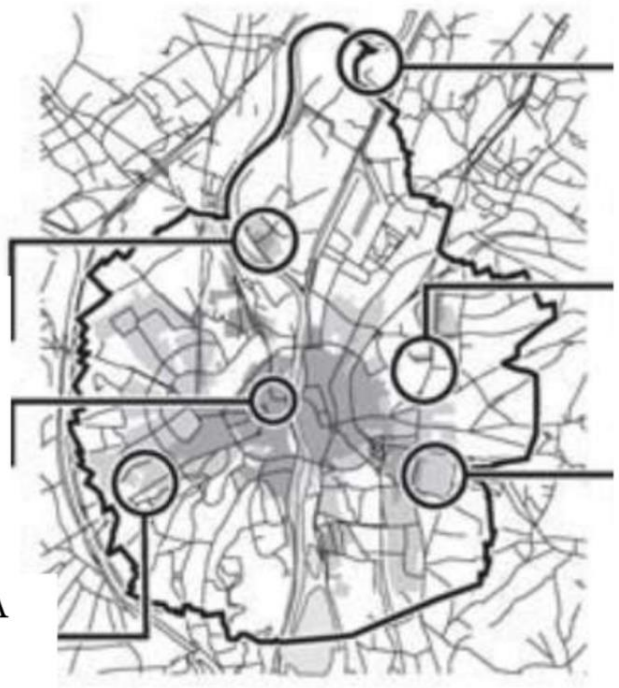
- Analysis tasks for movement data focus on Spatial (S), Temporal (T), and Attributive (A) properties.



what is the name of this village?
Identification: Borgharen

where is the city hall?
location: x,y = 1764,3180

What is the shortest route between A and B?
optimal path: start at A, go left at …



What relation exists between road network and river?
Pattern: river interrupts road network

What if: a new built-up area is created here?
models: will affect traffic intensity

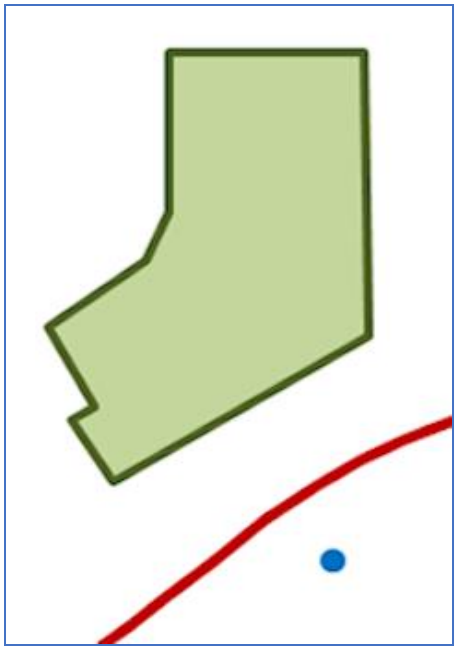
What has changed?
trends: growth urban area

Kraak, Menno-Jan, and Ferjan Ormeling. Cartography: visualization of spatial data. Guilford Press, 2011.

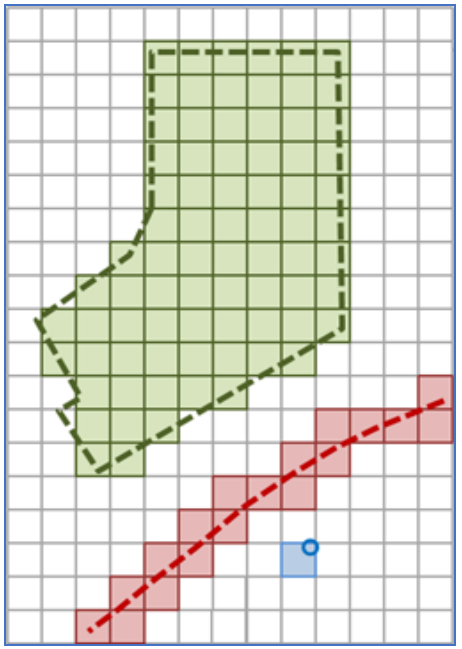
Geographic data representation

- Geographic data are typically represented as
  - Vector formats
  - Raster formats

Real world



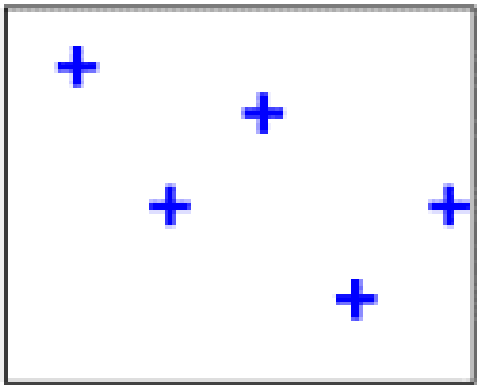
Vector format



Raster format

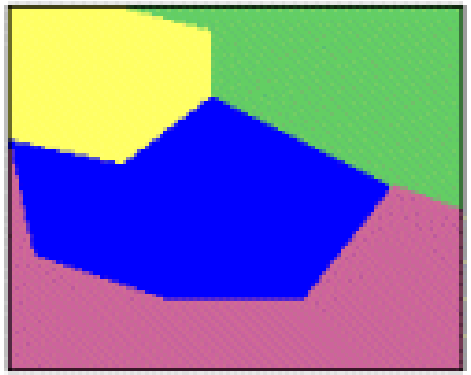
Vector-based geographic data

- Vector-based geographic data
  - Point: single vertex (x, y, and optionally z)
  - Line: two vertices with an arc between them
  - Area: a closed group of three or more arcs



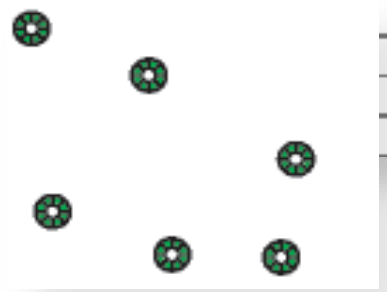
Points

Lines



Areas

Vector-based geographic data



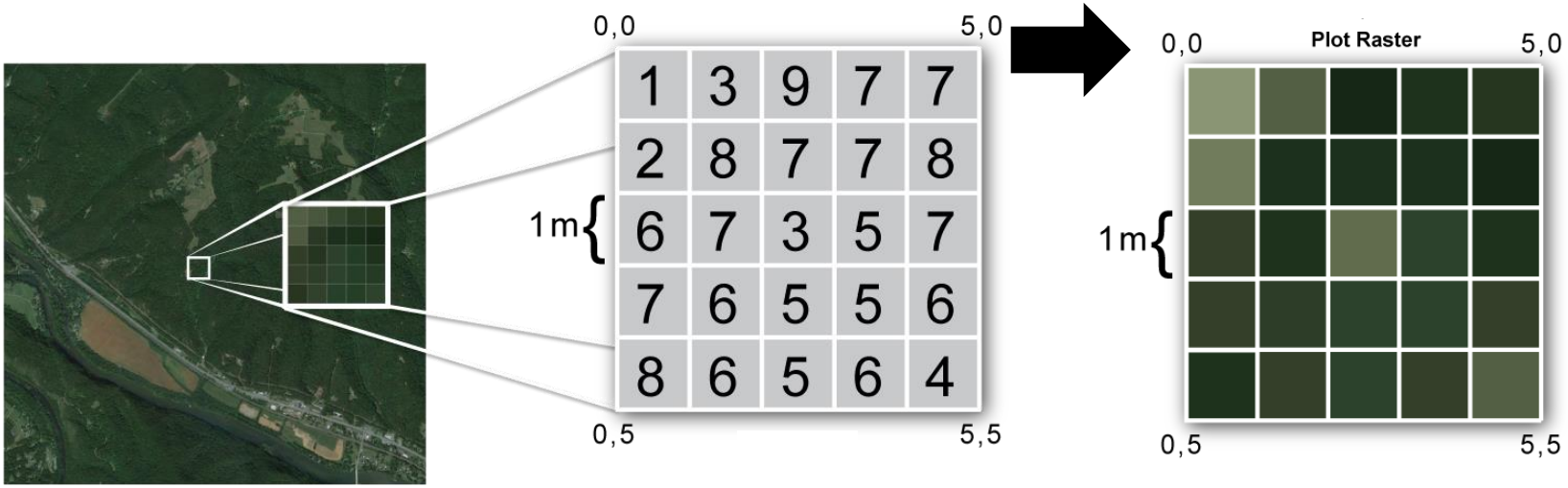
OBJECTID	SHAPE	NodeID	Type	PSI	LastInspected
1	Point	1101	39	150	7/5/2010
2	Point	1104	39	150	12/5/2009
3	Point	1107	39	150	12/5/2009
4	Point	1108	39	150	7/6/2011

OBJECTID*	SHAPE*	Street	TrafDir
1490	Polyline	DOWNING ST	W
1491	Polyline	AV OF THE AMERICAS	W
1492	Polyline	ERIVER SHORELINEW	
1493	Polyline	ERIVER SHORELINEW	

OBJECTID*	Shape*	MBL	DBLDVAL	DLNDVL
4	Polygon	275B0100044	140830	26000
5	Polygon	275B0100045	132820	26000
6	Polygon	275B0100043	188080	26000
7	Polygon	275B0020013	149030	20000

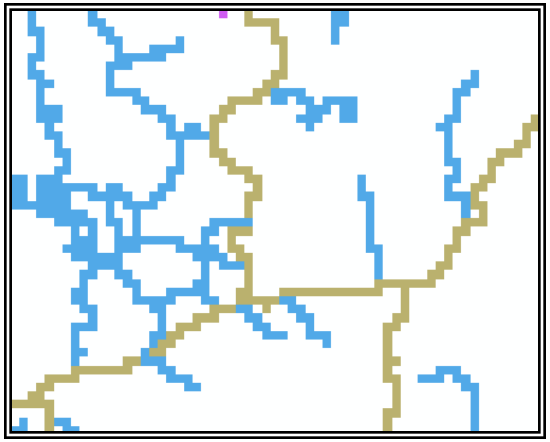
Raster-based geographic data

- **Raster data**: represents geography via ***grid cells***
  - Images
  - Digital elevation models (DEMs)
  - Point clouds



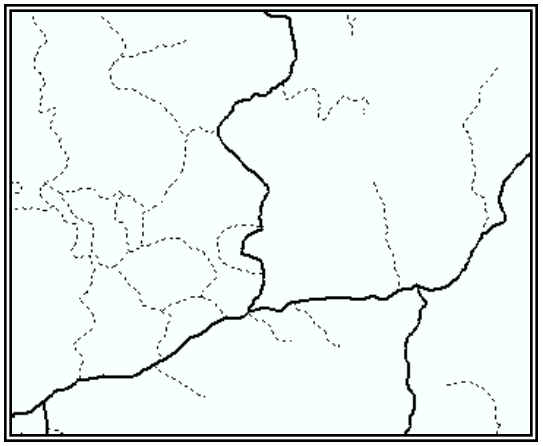
Raster vs. Vector

- Raster Is Faster, but Vector Is Corrector [Joseph Berry, 1996]



Advantages
- ✓ Store continuously changing values, e.g., elevation
- ✓ Fast and flexible for computer programs

Disadvantages
- × Topology hard to reserve
- × Do not conform to high-quality maps

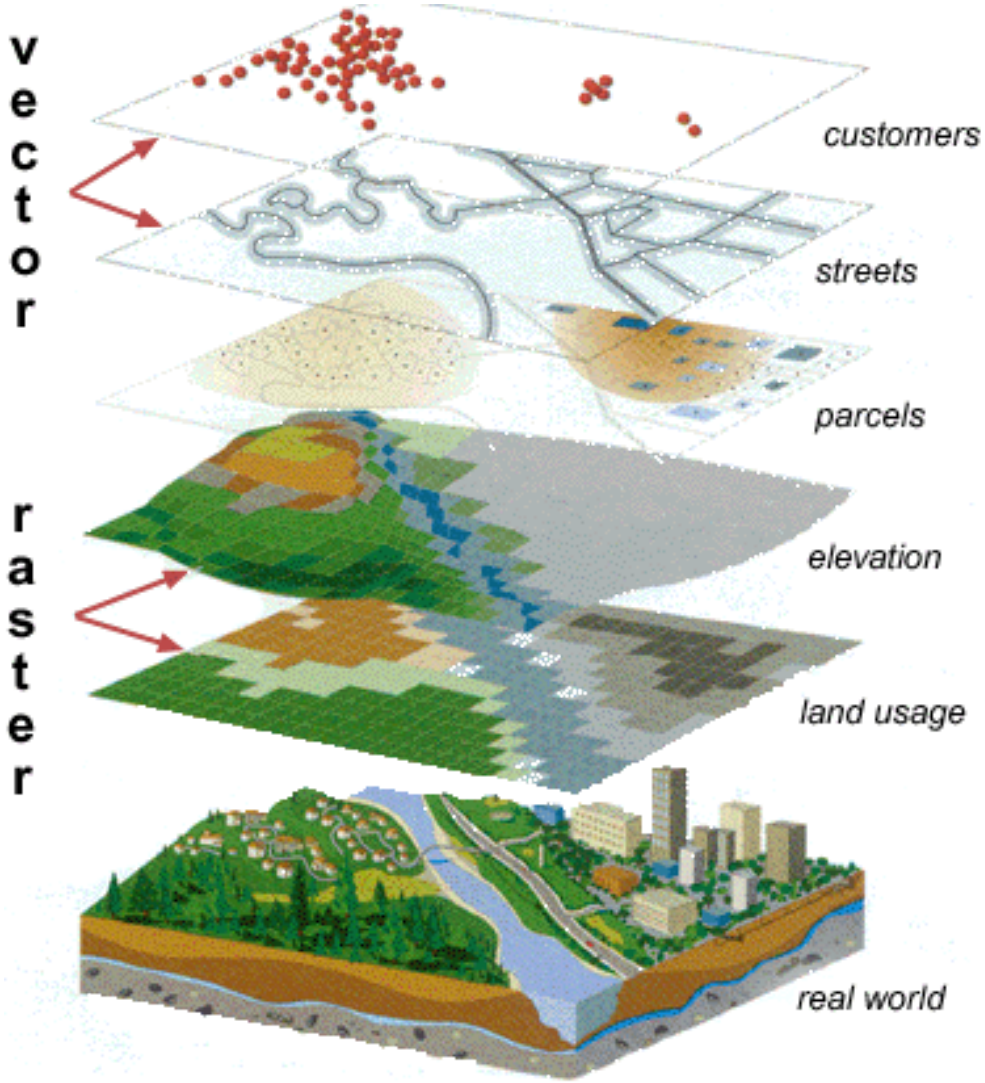


Advantages
- ✓ Store topology information, e.g. proximity
- ✓ True shape and size

Disadvantages
- × Need to store each vertex explicitly
- × Processing intensive and data cleaning extensive

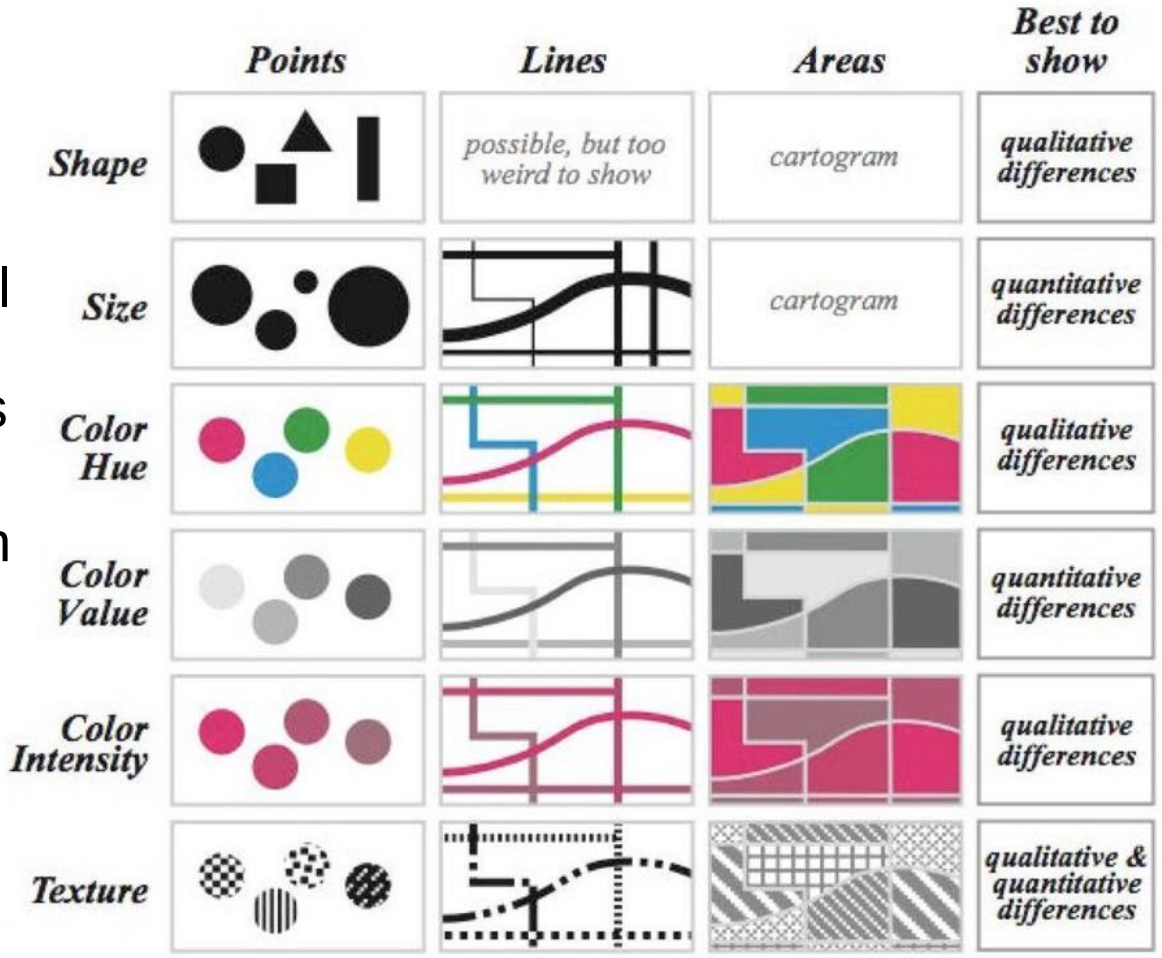
Raster vs. Vector

- Real-world usage scenario usually combines raster with vector geographic data.



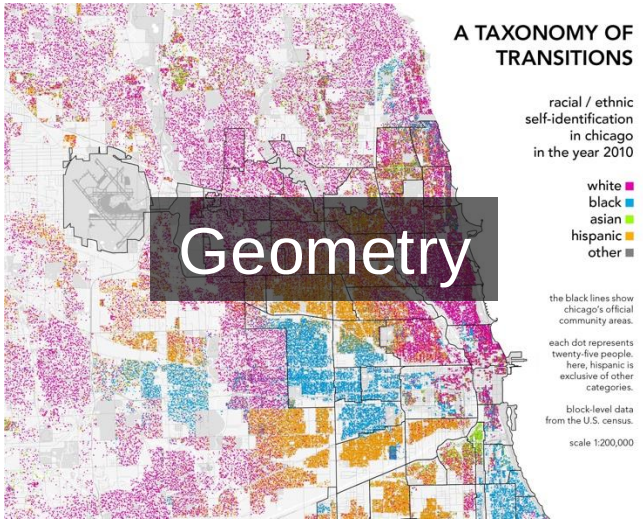
Visual mapping

- Visualization for vector-based geographic data
  - Mapping to visual channels depend on data attributes
  - Focus on high-level visualization techniques

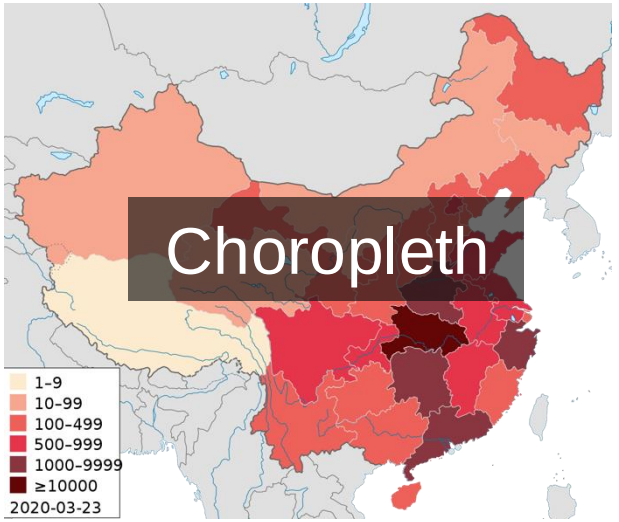


John Krygier and Denis Wood, Making maps: a visual guide to map design for GIS, Edition: 3rd Publisher: Guilford Press 2016

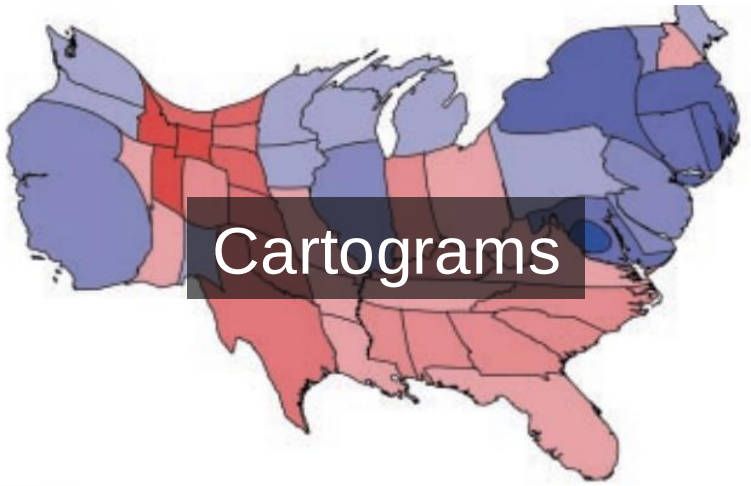
Geographic visualization

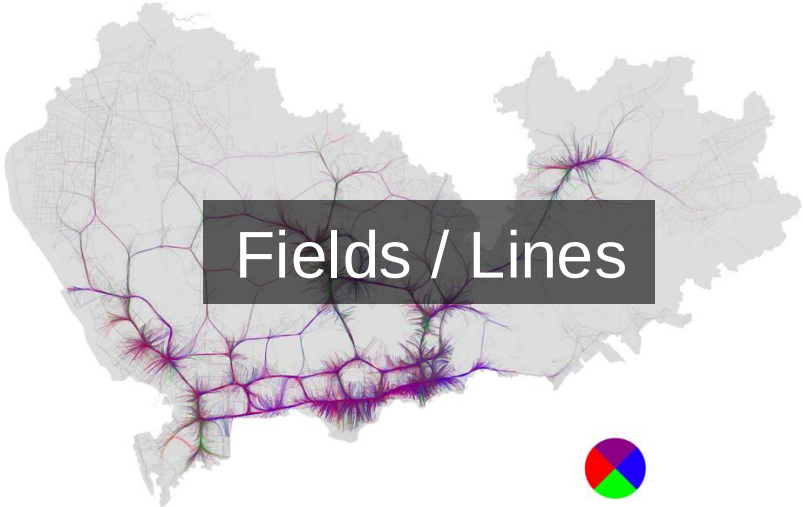


Geometry



Choropleth


Cartograms


Fields / Lines

Data Exploration & Visualization

Module 9: Geospatial Visualization

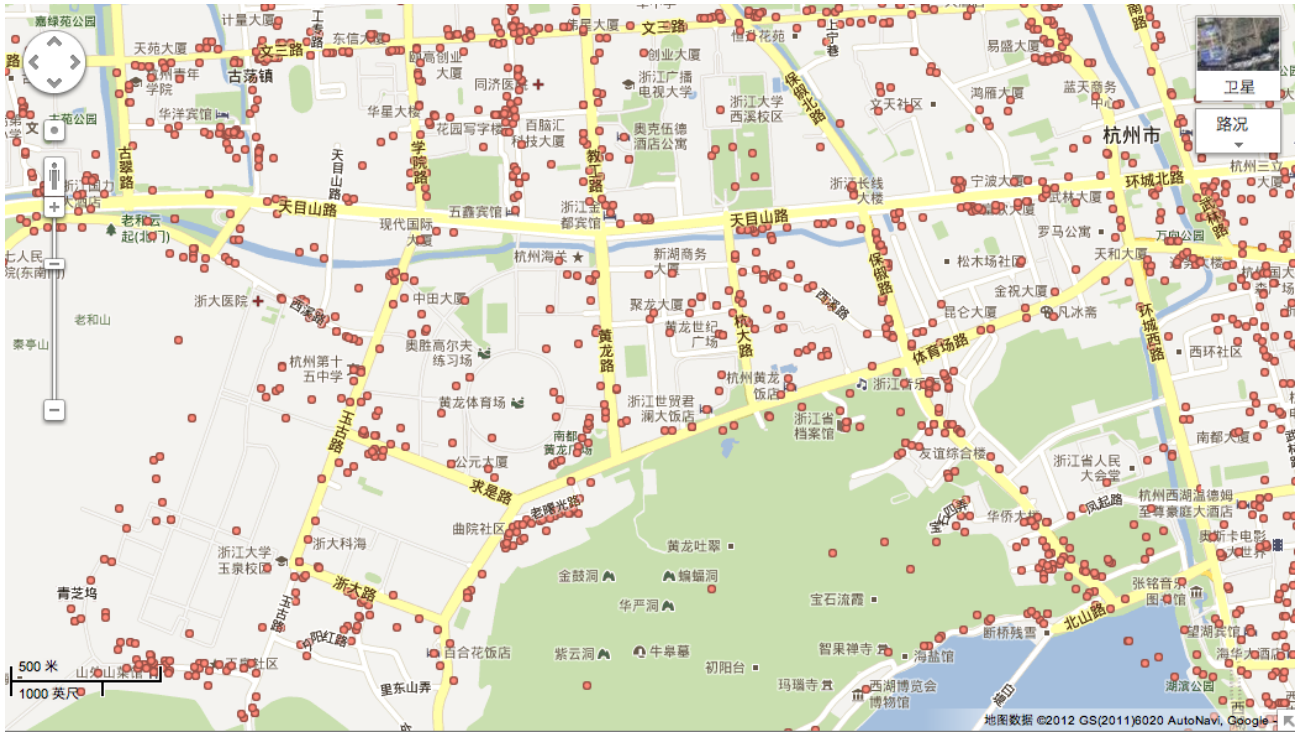
- Geospatial data representation
  - Map projection
  - Vector vs. raster
- Geospatial visual representation
  - Geometry-based
  - Choropleth map
  - Cartograms
  - Fields/Lines (Future Week)

Point-based geographic data

- Properties of point-based geographic data
  - **Position**: latitude and longitude, not necessarily with altitude.
  - **Nominal**: building types
  - **Ordinal**: hotel rating, e.g., 5-star, 4-star
  - **Quantitative**: number of people visits per day
  - **High-dimensional**
- Examples: Points-of-interest (POIs)
  - Landmarks
  - Restaurants

Geometry visualization

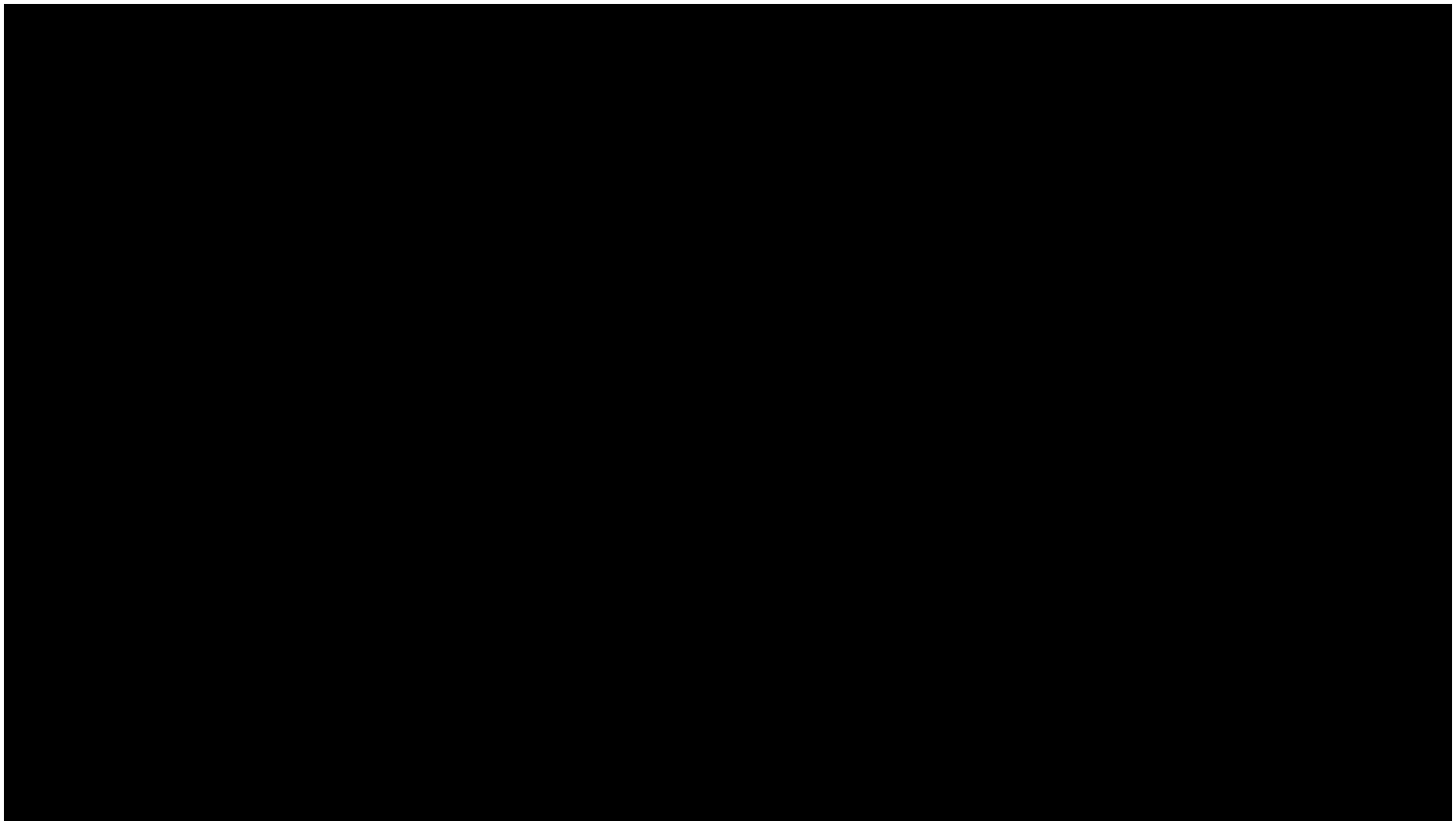
- Marking the data on the map would be the most direct and simplest way to visualize geographic data
  - Points scattered on the geographical space, with longitude and latitude information



Gas stations in Xihu District, Hangzhou, China

Geometry visualization

- Marking the data on the map would be the most direct and simplest way to visualize geographic data
  - Points scattered on the geographical space, with longitude and latitude information

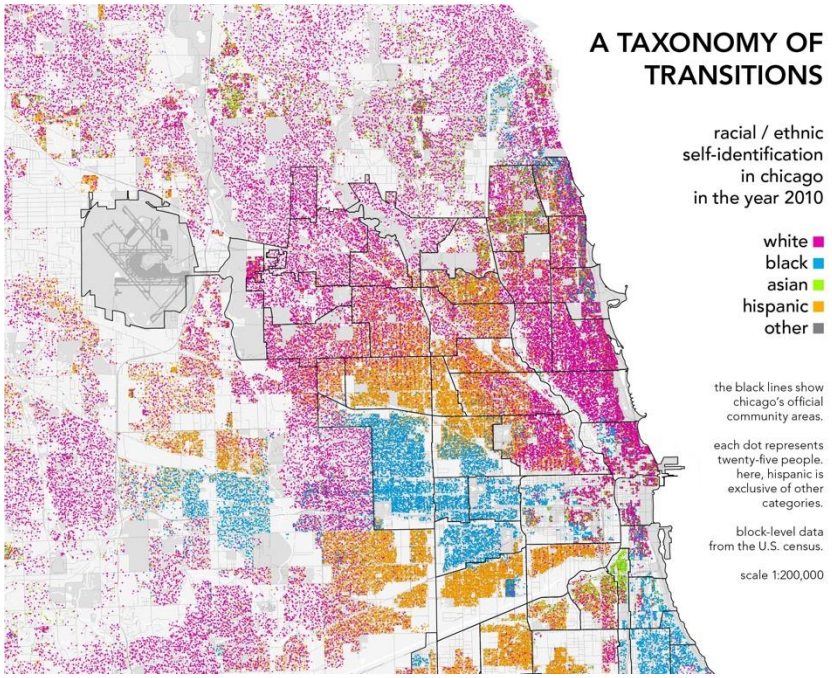


Every death on every road in Great Britain 1999 - 2010

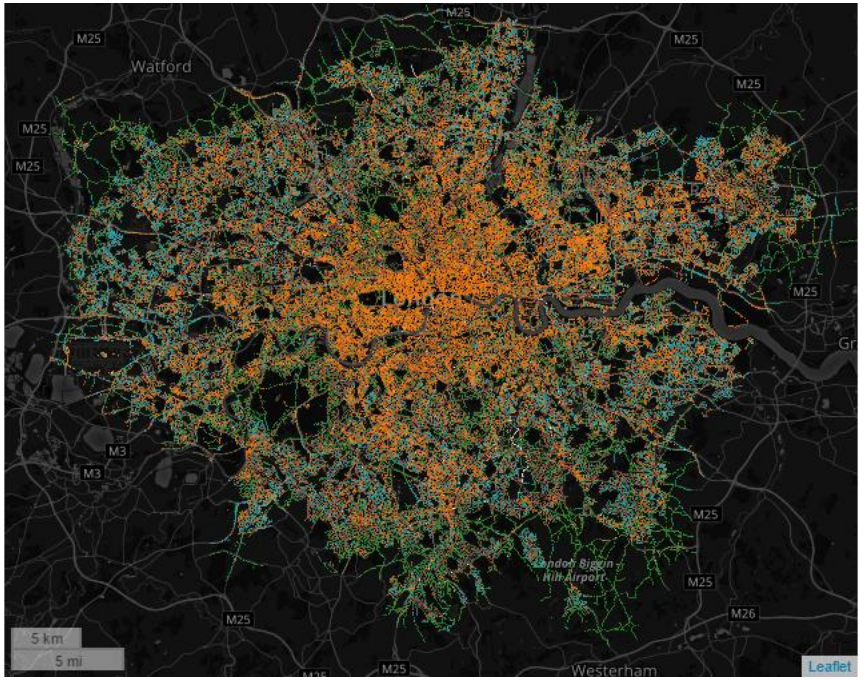
Source: BBC

Geometry visualization

- Marking the data on the map would be the most direct and simplest way to visualize geographic data
  - Encode nominal attributes with hues



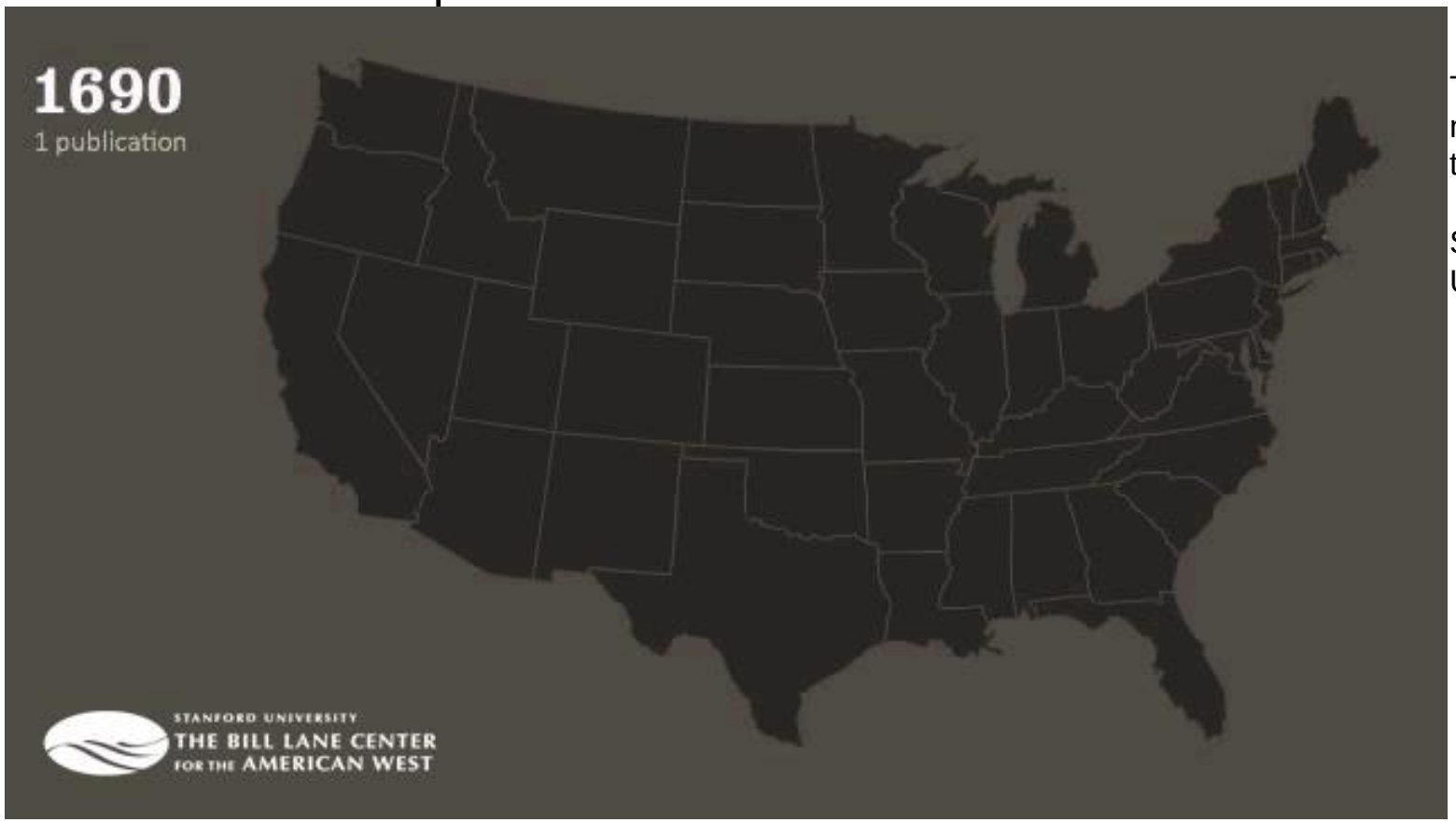
A taxonomy of transitions in Chicago: 2010
Source: Radical Cartography
http://www.radicalcartography.net/index.html?chicagodots vs
http://www.encyclopedia.chicagohistory.org/pages/3889.html



Street view categories in London, 2017

Geometry visualization

- Marking the data on the map would be the most direct and simplest way to visualize geographic data
  - Encode quantitative attributes with sizes

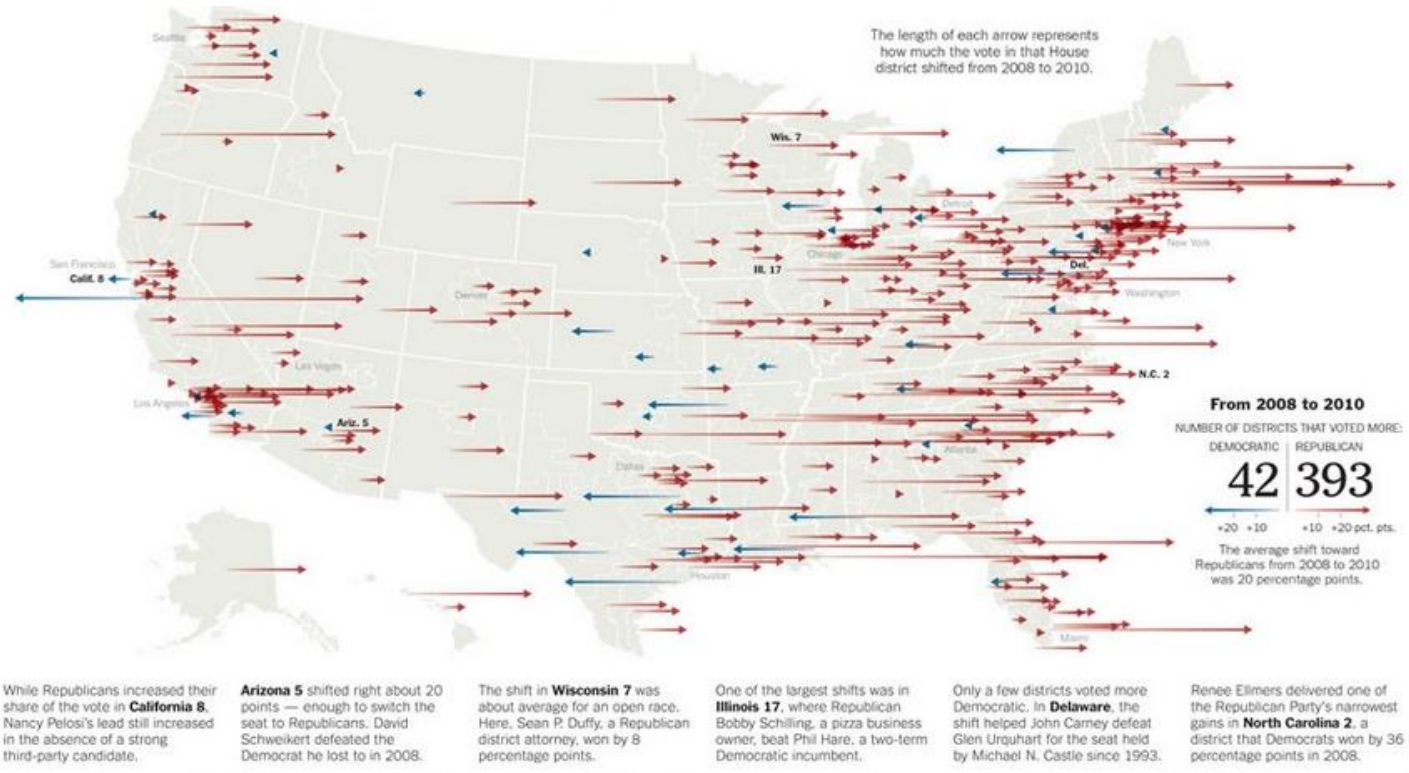


The growth of newspapers across the US: 1690 - 2011.

Source: Stanford University

Geometry visualization

- Marking the data on the map would be the most direct and simplest way to visualize geographic data
  - Encode nominal attributes with orientation
  - Encode quantitative attributes with length

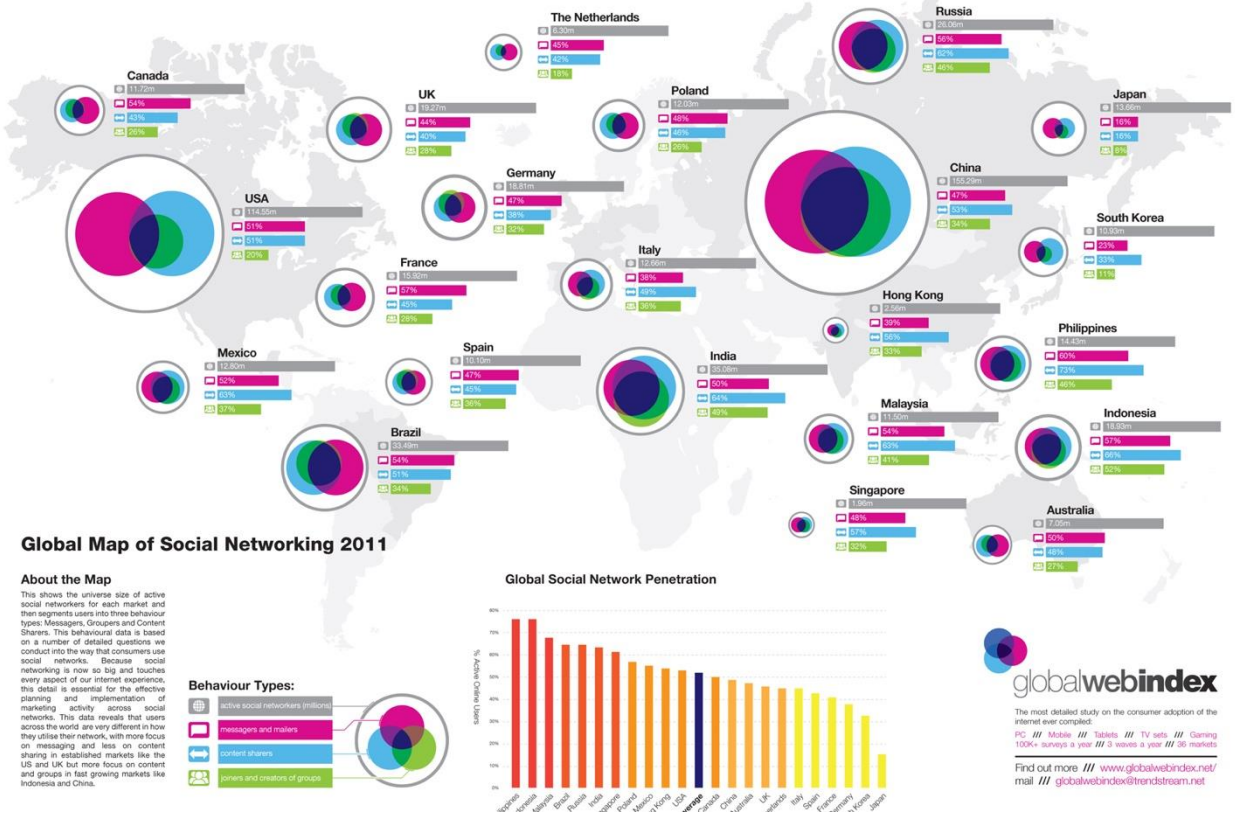


After the Vote: districts across the country shift to right.

Source: New York Times

Geometry visualization

- Marking the data on the map would be the most direct and simplest way to visualize geographic data
  - Encode quantitative attributes with size
  - Encode nominal attributes with hue

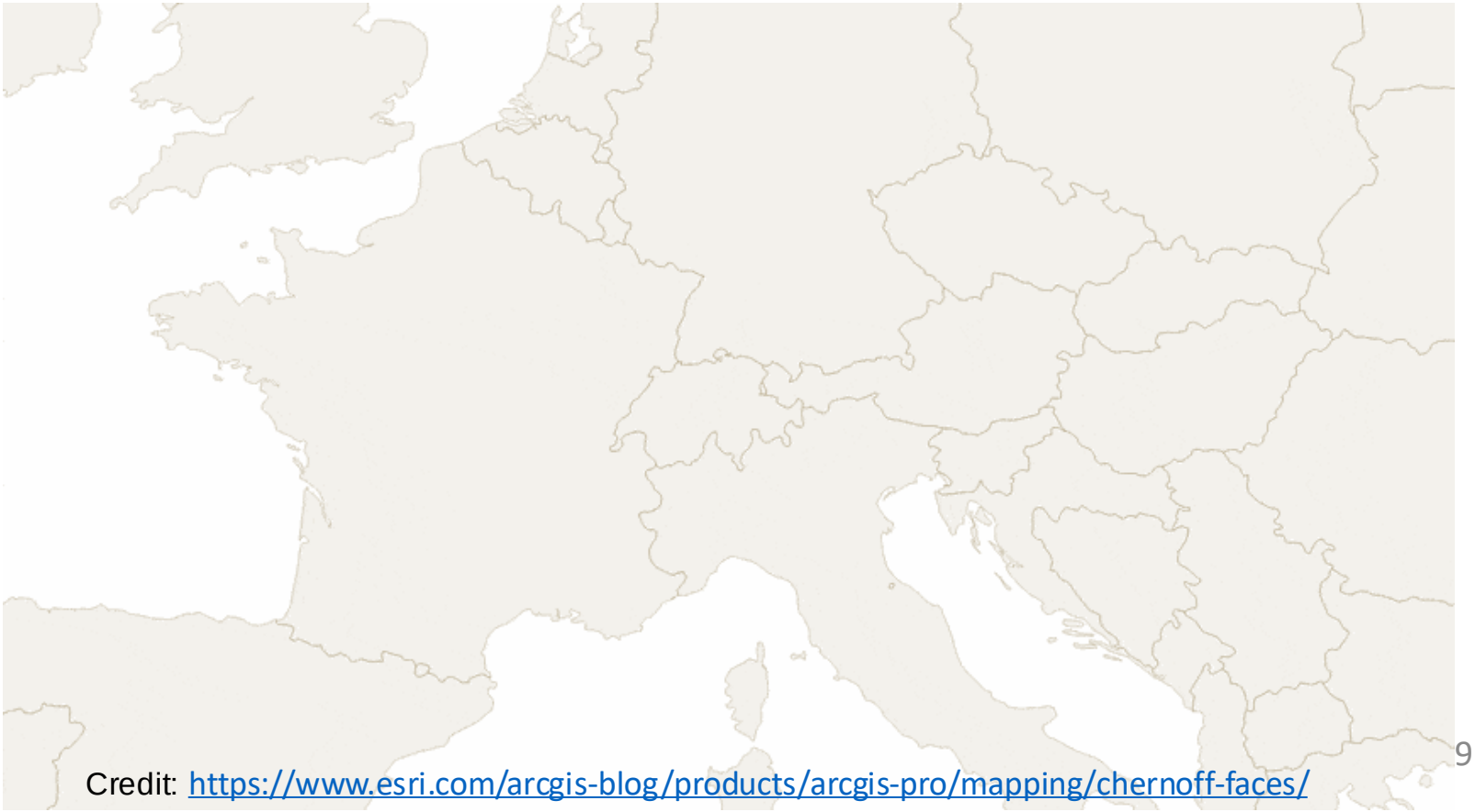


Global Map of Social Networking 2011

Source: Global Web Index

Geometry visualization

- Marking the data on the map would be the most direct and simplest way to visualize geographic data
  - Encode high-dimensional attributes with Chernoff faces



Credit: https://www.esri.com/arcgis-blog/products/arcgis-pro/mapping/chernoff-faces/

Data Exploration & Visualization

Module 9: Geospatial Visualization

- Geospatial data representation
  - Map projection
  - Vector vs. raster
- Geospatial visual representation
  - Geometry-based
  - Choropleth map
  - Cartograms
  - Fields/Lines (Future Week)

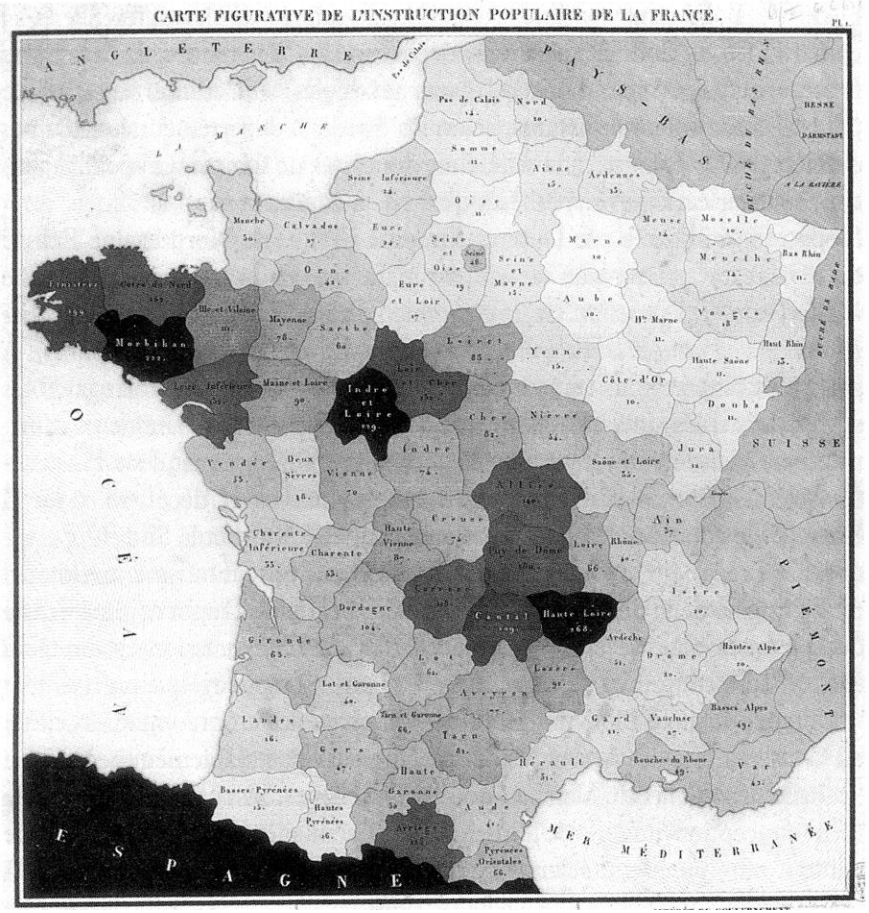
Choropleth map

- A **choropleth map** is a thematic map in which areas are **shaded or patterned in proportion** to the measurement of the statistical variable being displayed on the map,
  - population density
  - per-capita income

- The term comes from the Greek for "area/region“ and "multitude“.

Choropleth map

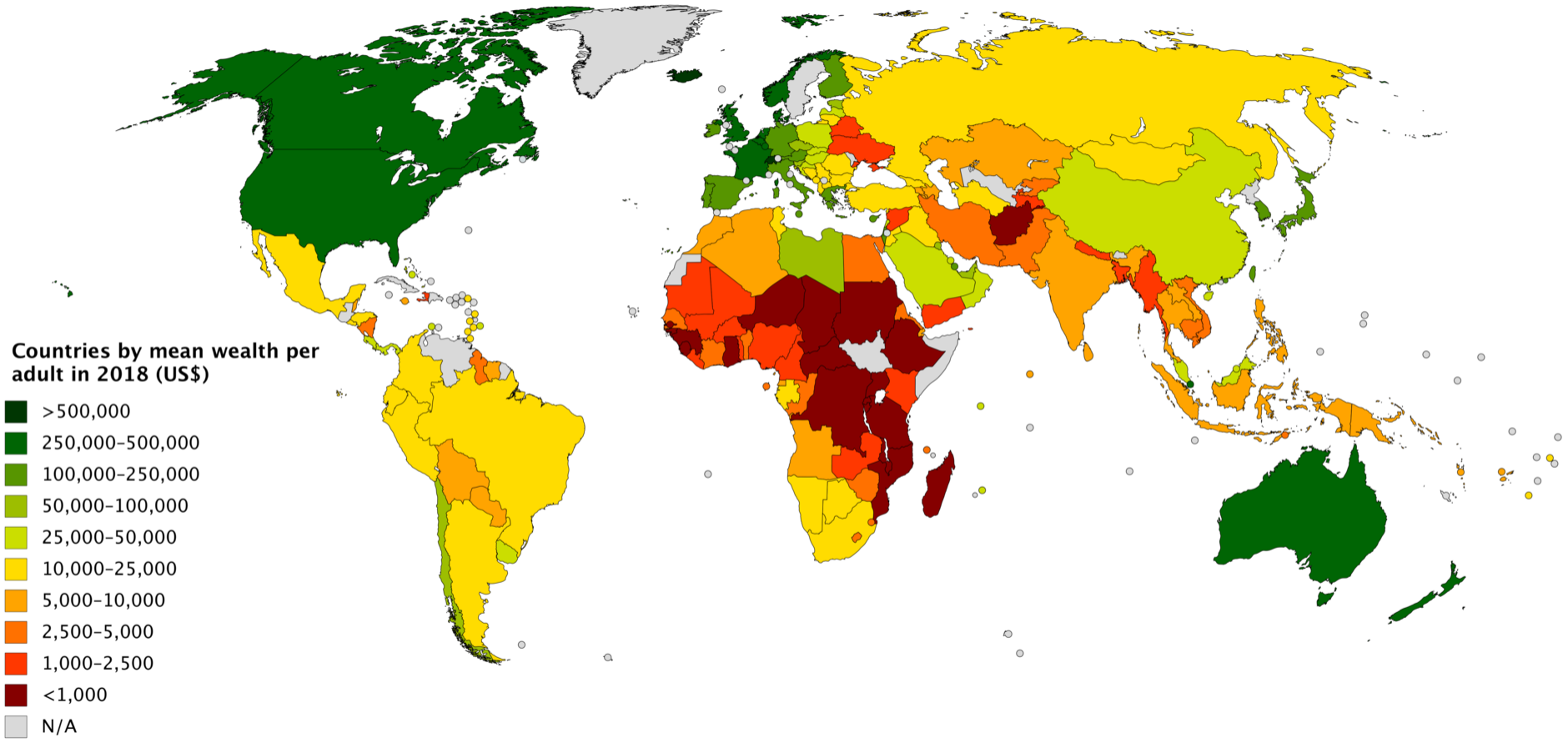
- The earliest known choropleth map was created in 1826 by Charles Dupin.
  - "cartes teintées" (coloured map in French)
- The term "choroplethe map" was introduced in 1938 by the geographer John Kirtland Wright.



The distribution of illiteracy in France, 1826.

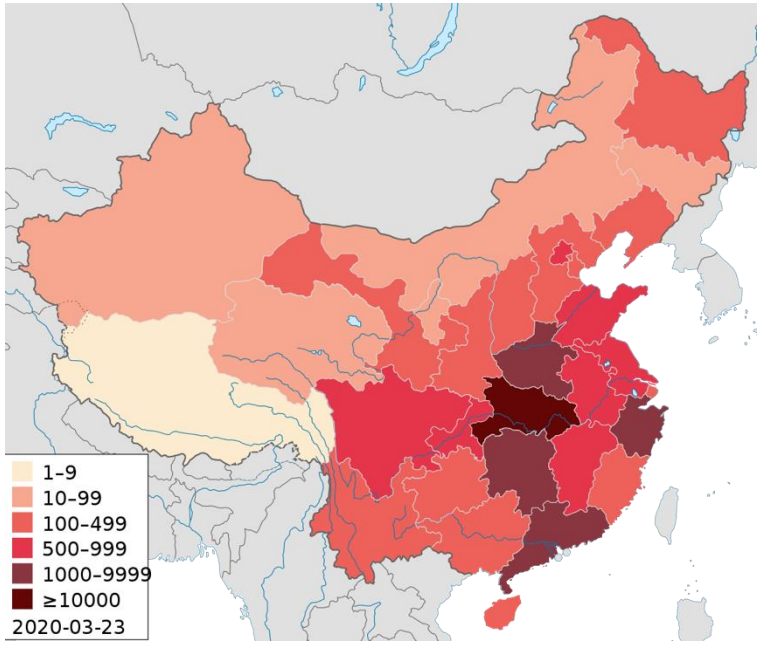
Example

- Regions by mean wealth per adult in 2018



When to use choropleth map?

- Appropriateness of Data – political boundary (Census Bureau units)
  - Not appropriate – continuous phenomena should not be mapped by choropleth maps
- Theme in well-defined enumeration units
  - No more than 5 to 7 classes
  - Data must work as ratios or proportions



COVID-19 cases in mainland China by provinces as of 7 March 2020

Data classification

- Six common methods
  - Equal intervals
  - Standard deviation
  - Geometric progression
  - Quantile
  - Natural breaks
  - Optimal

Equal intervals

1. Calculate the range of the data (R)

$$R = H - L$$

2. Obtain the common difference (CD)

$$CD = R / (\#\ \text{of Classes})$$

3. Obtain the class limits by calculating
   - First class limits: [L, L + 1 x CD)
   - Second class limits: [L + CD, L + 2 x CD)
   - …
   - Last class limits: [L + (n-1) x CD, L + n x CD]

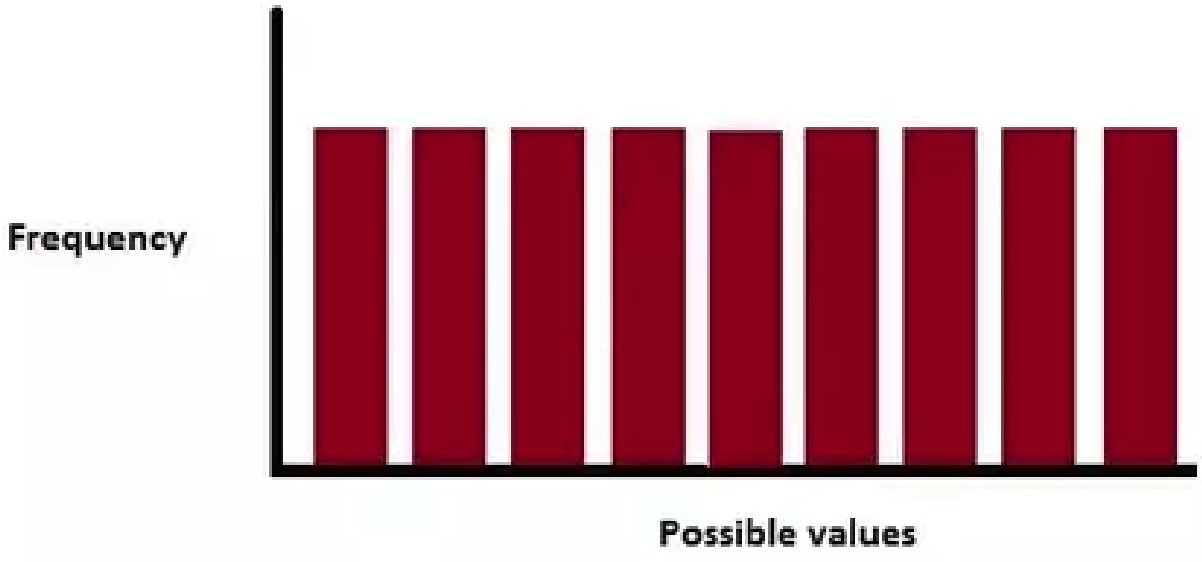
Equal intervals

- Advantages
  - easy to compute the intervals
  - easy to interpret the resulting intervals
  - no gap in the legend display
  - only lowest limits can be shown in legend

- Disadvantage
  - skewed data is not appropriate.

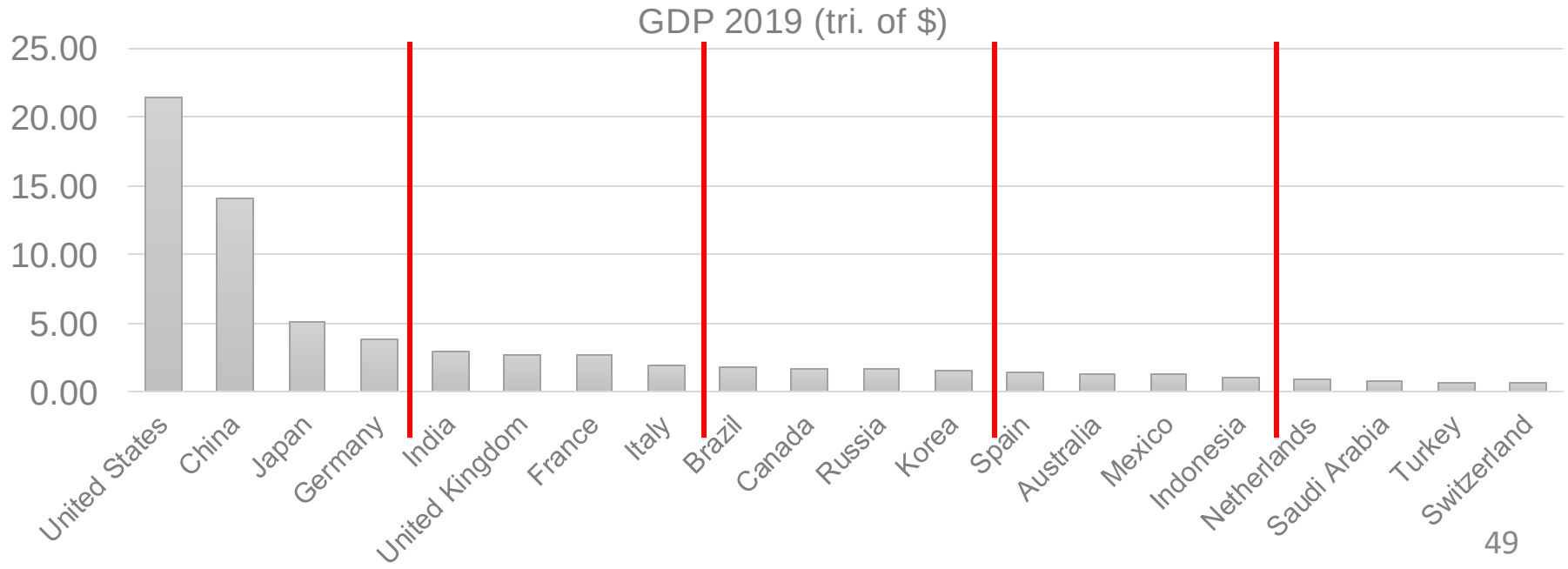
Equal intervals

- Useful when histogram of data array is uniform
  - rare in geographic phenomena



Quantile

- Divide classes so that the total number of items in each class are approximately the same
  - Ordered data are placed in classes.
  - K = # of enumeration units / number of classes



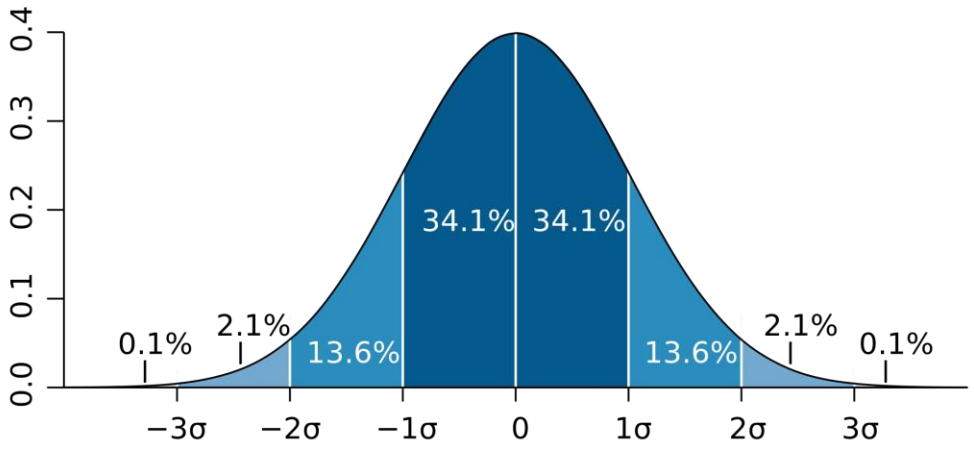
Quantile

- Advantages
  - class limits can be computed manually
  - if enumeration units are same, each class will have the same map area
  - quantile are useful for ordinal-level data, no numeric information would be necessary to create the classification.
- Disadvantage
  - gap result may vary
  - fails to consider data distribution.

Standard deviation

- Computing steps
  - Compute mean μ and standard deviation σ of the data
  - Obtain class limits

    $\ldots, [\mu - 2\sigma, \mu - \sigma),\ [\mu - \sigma, \mu),\ [\mu, \mu + \sigma),\ [\mu + \sigma, \mu + 2\sigma)$ ...
- Useful when the data array approximates a normal distribution



Standard deviation

- Advantages
  - Distribution of data is taken into account
  - if normal distributed data is used, the mean is a good divider
  - no gap in the legend

- Disadvantage:
  - only work with normal-distributed data
  - negative values may be in the range

Geometric progression

- Useful technique when frequency of data declines continuously with increasing magnitude

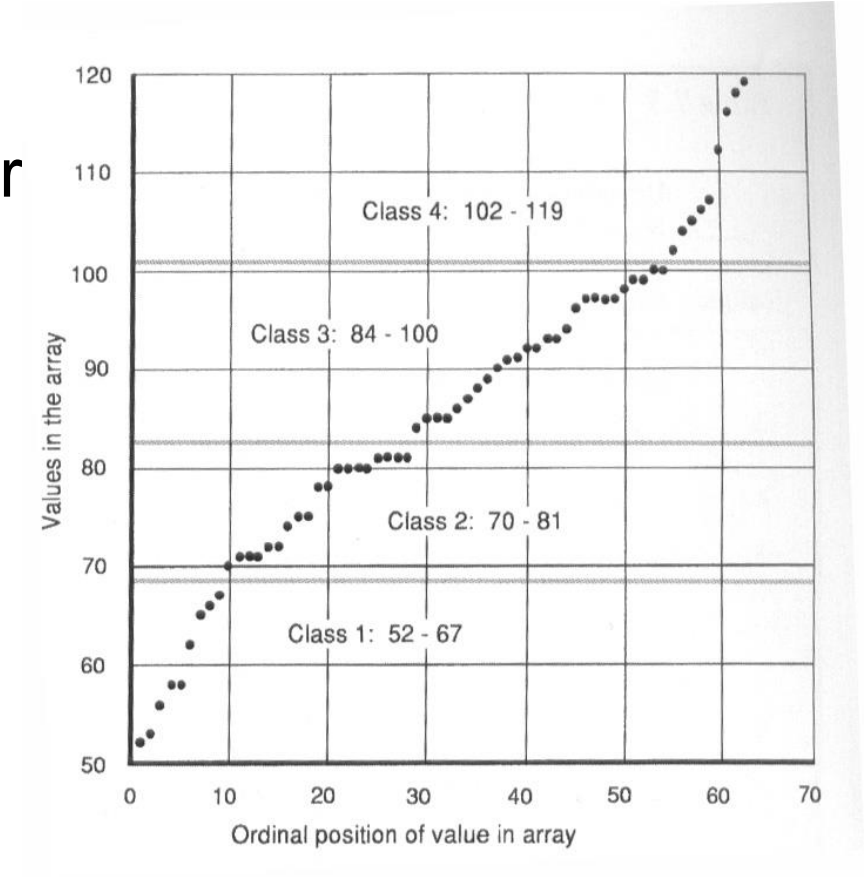
$$a, ar^1, ar^2, \ldots ar^n$$

- Computing steps:
  - compute common multiplier (a is the lowest value, r is the common multiplier and n is the number of classes)
  - use "$X_{min} \times r^n = X_{max}$" to obtain r
  - e.g. $118 \times r^5 = 790 \rightarrow r = 1.46$
  - So the interval goes from 118, 118×1.46, 118×1.46^2, 118×1.46^3, 118×1.46^4

Graphic array

- Class boundaries are identified at places wher slopes change remarkably

- Disadvantage
  - not suitable for large amount of data



Jenks optimization

- Forming groups that are internally homogenous while assuring heterogeneity among classes
- Groups are created based on gaps.
- Minimize differences within class and maximize differences between classes.
- Based on GVF (Goodness of Variance Fit) - an optimization techniques to minimize the sum of the variance within each of the class.

GVF (Fisher-Jenkins Algorithms)

1. Compute the squared deviation of all data

$$SDAM = \sum_{i=1}^{n} (Xi - \overline{X})^2$$

2. Compute SDCM (Squared Deviation, Class Means)

$$SDCM = \sum_{j=1}^{m} \sum_{i=1}^{n} (X_{ij} - X_{cj})^2 \quad .....for \;\; j = 1....m, \;\; m = number\; of \;\; classes$$

$where \;\; SDCM \; is \; Squared \; Deviation \; Class \; Mean, \; Xij \;\; is \; the \; data \;\; i$
$and \; assigned \;\; to \; Class \;\; j, \; X_{cj} \; is \; the \; mean \; of \;\; class \; j.$

3. Compute GVF = (SDAM - SDCM) / SDAM
4. The goal is to maximize the value of GVF (closer to 1.0 is the better value)

GVF (Fisher-Jenkins Algorithms)

- Input: [4, 5, 9, 10]
- Step 1: compute SDAM
  - mean = (4 + 5 + 9 + 10) / 4 = 7
  - SDAM = $(4-7)^2+(5-7)^2+(9-7)^2+(10-7)^2=26$
- Step 2: Compute SDCM
  - For [4] [5,9,10]:
    - SDCM = $\{(4-4)^2\}+\{(5-8)^2+(9-8)^2+(10-8)^2\}=0+9+1+4=14$
  - For [4,5] [9,10]:
    - SDCM = $\{(4-4.5)^2+(5-4.5)^2\}+\{(9-9.5)^2+(10-9.5)^2\}=0.25+0.25+0.25+0.25=1$
  - For [4,5,9] [10]:
    - SDCM = $\{(4-6)^2+(5-6)^2+(8-6)^2\}+\{(10-10)^2\}=4+1+9+0=14$

GVF (Fisher-Jenkins Algorithms)

- Step 3: Compute GVF
  - For [4] [5,9,10]:
    - GVF = (26 - 14) / 26 = 12 / 26 = 0.46
  - For [4,5] [9,10]:
    - GVF = (26 - 1) / 26 = 25 / 26 = 0.96
  - For [4,5,9] [10]:
    - GVF = (26 - 14) / 26 = 12 / 26 = 0.46
- GVF for [4,5] [9,10] is highest.
- Limitation: It's a data-intensive algorithm.

Data Exploration & Visualization

Module 9: Geospatial Visualization

- Geospatial data representation
  - Map projection
  - Vector vs. raster
- Geospatial visual representation
  - Geometry-based
  - Choropleth map
  - Cartograms
  - Fields/Lines (Future Week)

Cartograms

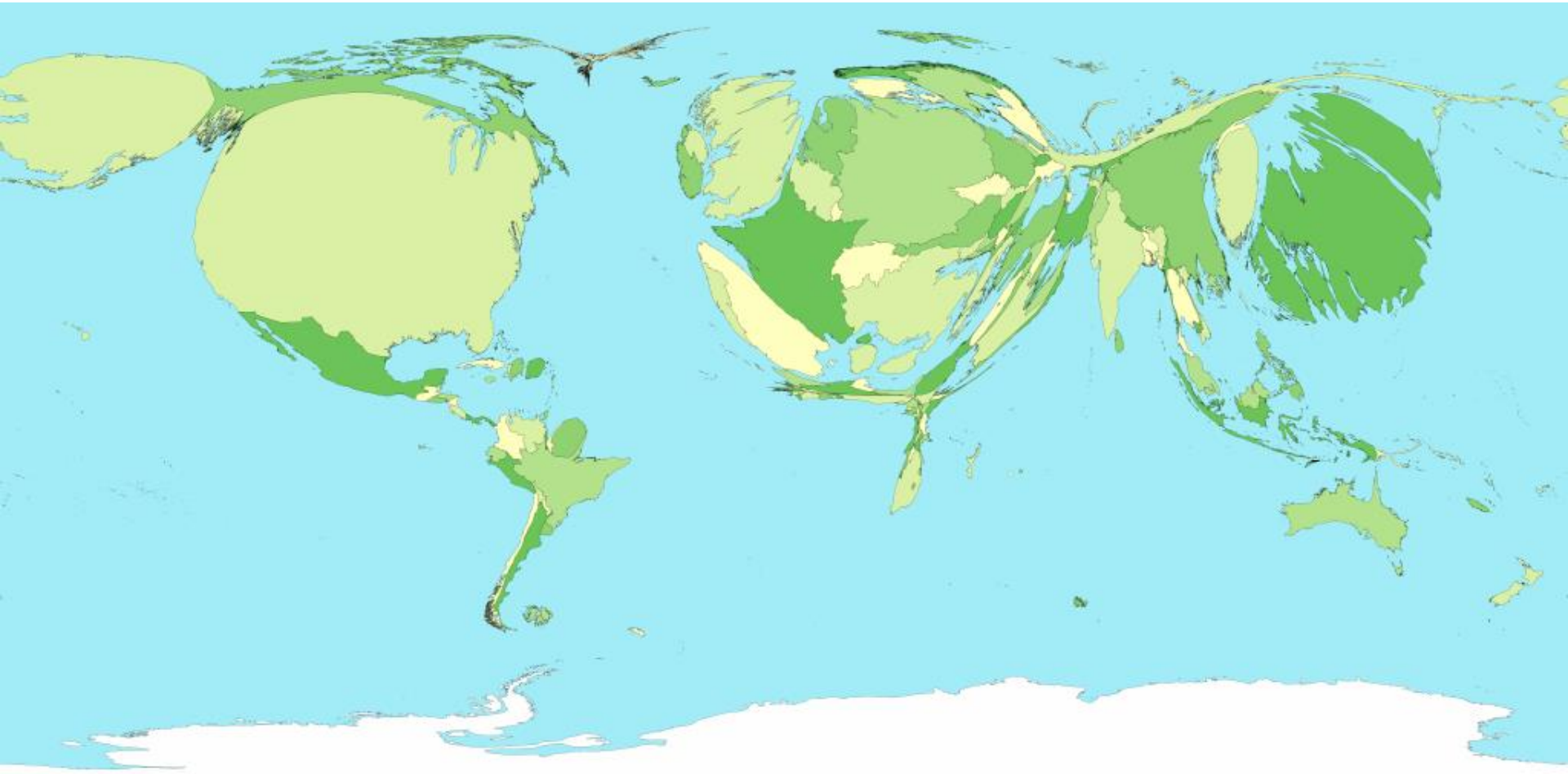
- A cartogram is a map in which some thematic mapping variable is substituted for land area or distance.
  - Area cartogram, distance cartogram
- Examples:
  - Travel time
  - Population
  - GNP
- The geometry or space of the map is *distorted*, sometimes extremely.

Cartograms

Source: Our World in Data (OWID)

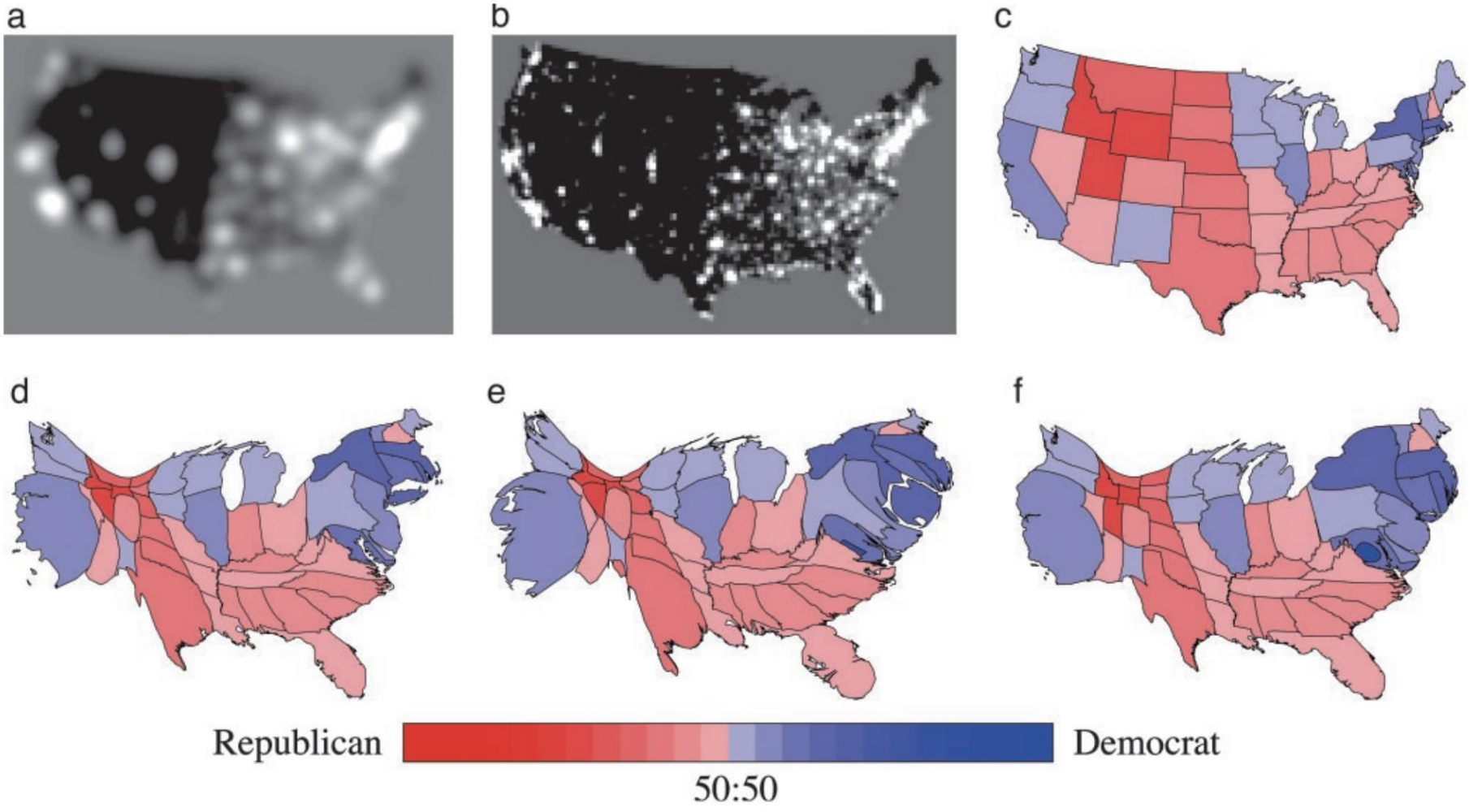
Cartograms

Gross domestic product



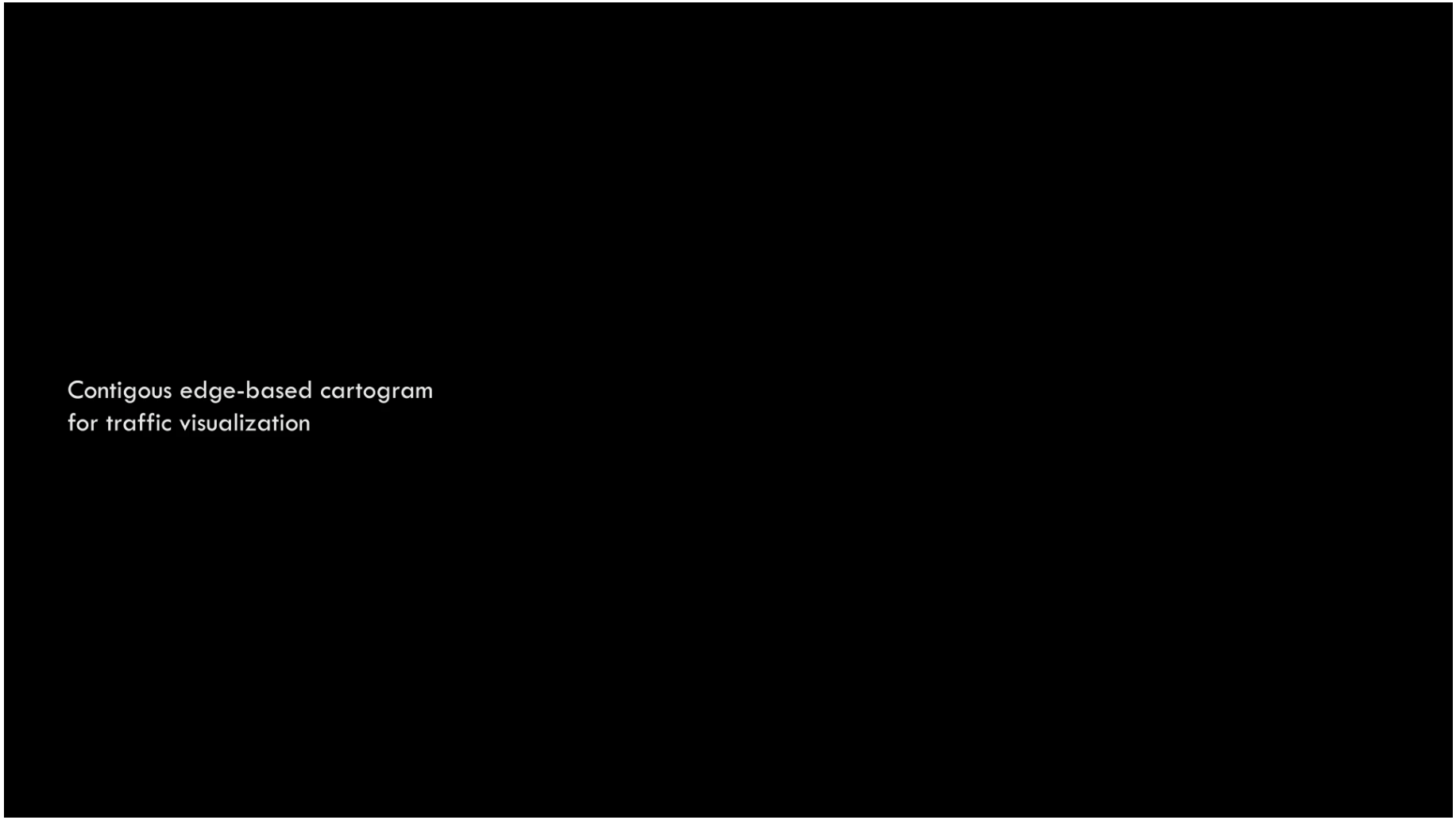
Source: http://www-personal.umich.edu/~mejn/cartograms/

Cartograms



Source: Gastner & Newman, Diffusion-based method for producing density-equalizing maps, PNAS 2004
https://www.pnas.org/content/pnas/101/20/7499.full.pdf

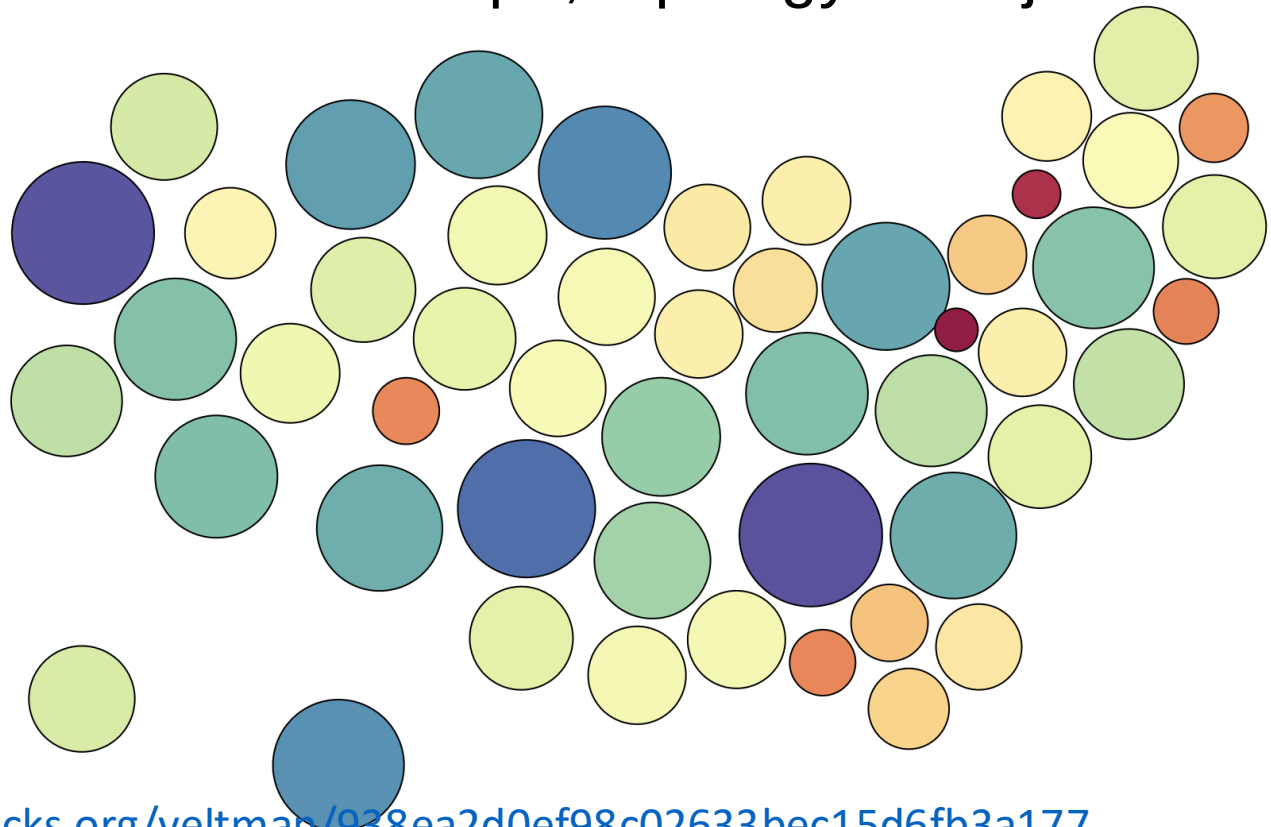
Cartograms



Pulsing Blood Vessels: A Figurative Approach to Traffic Visualization [Cruz and Machado, 2016]

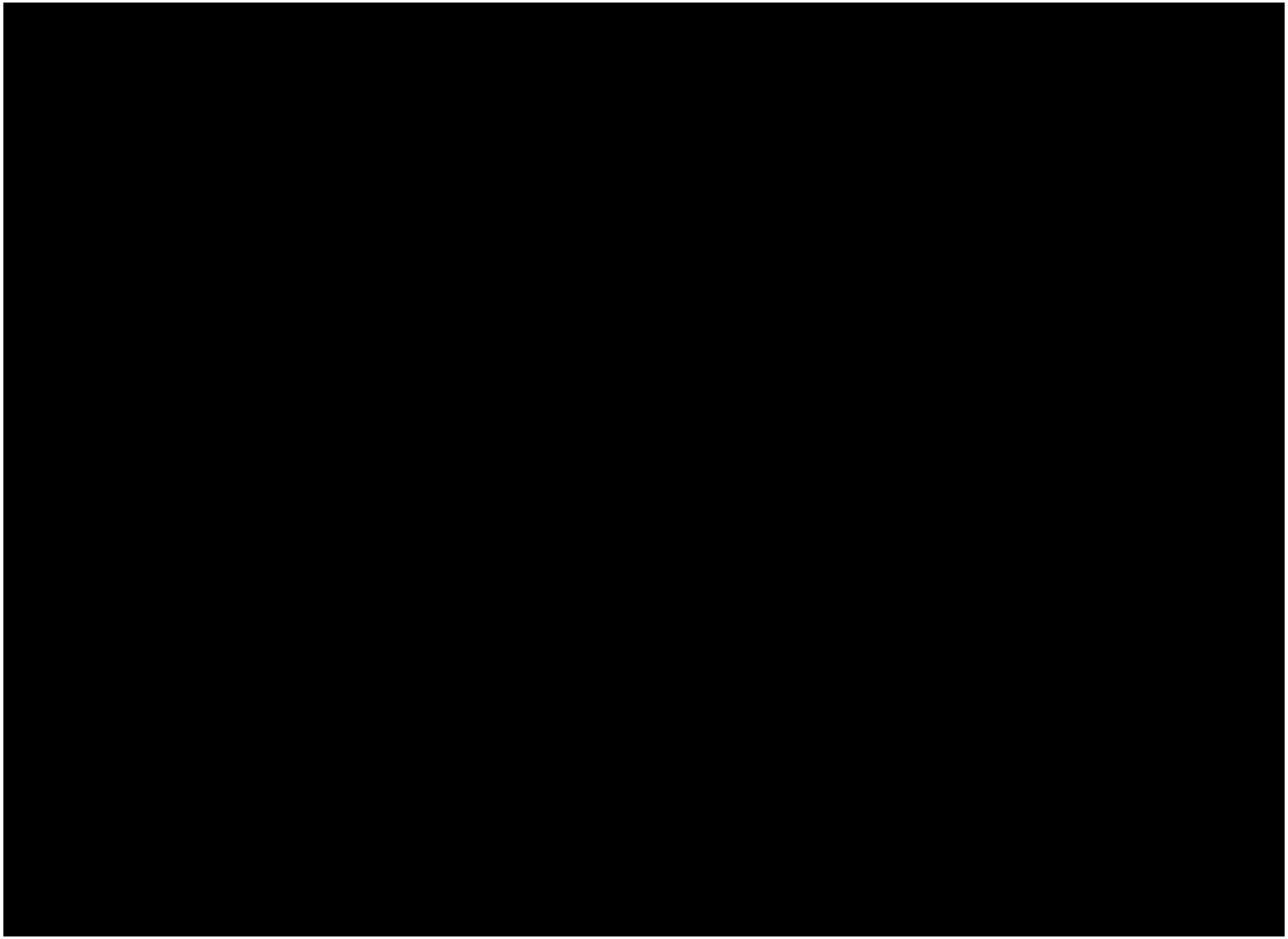
Dorling cartograms

- Circles are usually the shape of choice - more neatly positioned.
- Maintains neither shape, topology or object centroids



https://bl.ocks.org/veltman/938ea2d0ef98c02633bec15d6fb3a177

Applications in smart city



Live Singapore

Source:
Senseable City Lab

Summary

- Map projection: map the curved Earth's surface to a flat surface
  - Necessary for 2D geographic visualization
  - Cylindrical, conical, planar
  - Conformal, equal-area, equidistant, azimuthal
- Vector (point, line, area) vs. raster (grid cells)
- Geographical visualization
  - Geometry-based: overlay on maps
  - Choropleth maps: color coding
  - Cartograms: area/distance coding