

CRP 4080: Introduction to Geographic Information Systems for planners

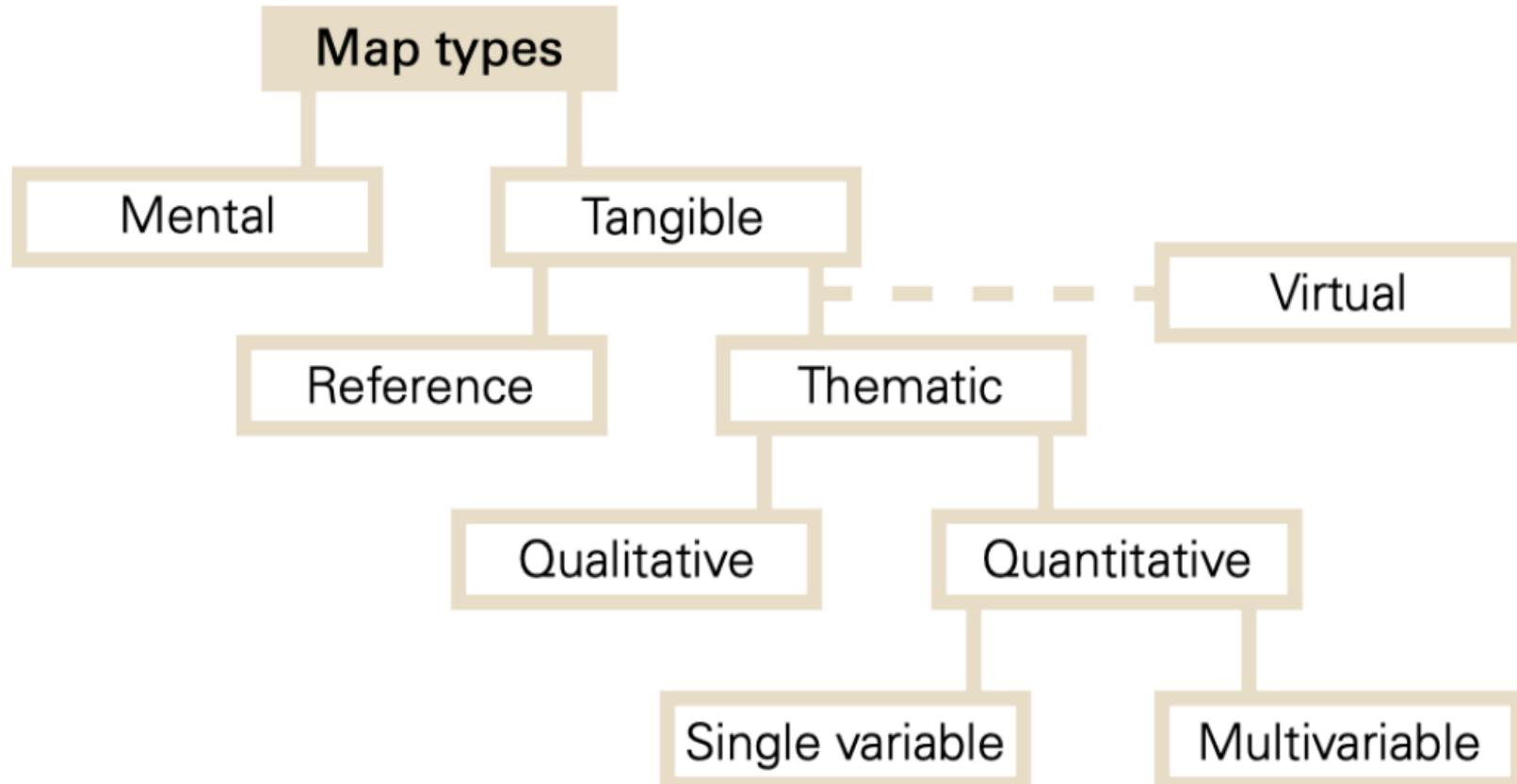
Lecture 2: Thematic Mapping

Wenzheng Li, Ph.D.
City and Regional Planning
Fall 2024

Overview

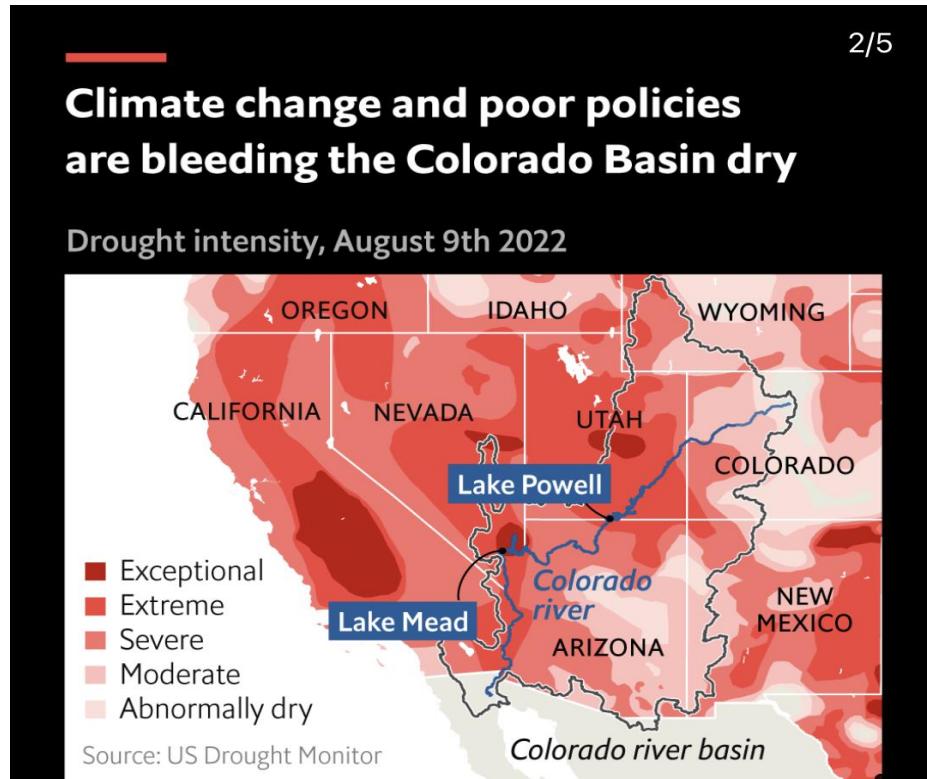
- 1) Lab 2 is released, due on Next Thursday (Sep 12th)
- 2) Access to ArcGIS Pro – final listed submitted. Please wait for the email to activate your account.
- 3) Lab access – all students in this class can access this lab, regardless programs.
- 4) Discussion section

A Typology of Maps



Source: Redrawn from Thematic Map Design, by Borden Dent.

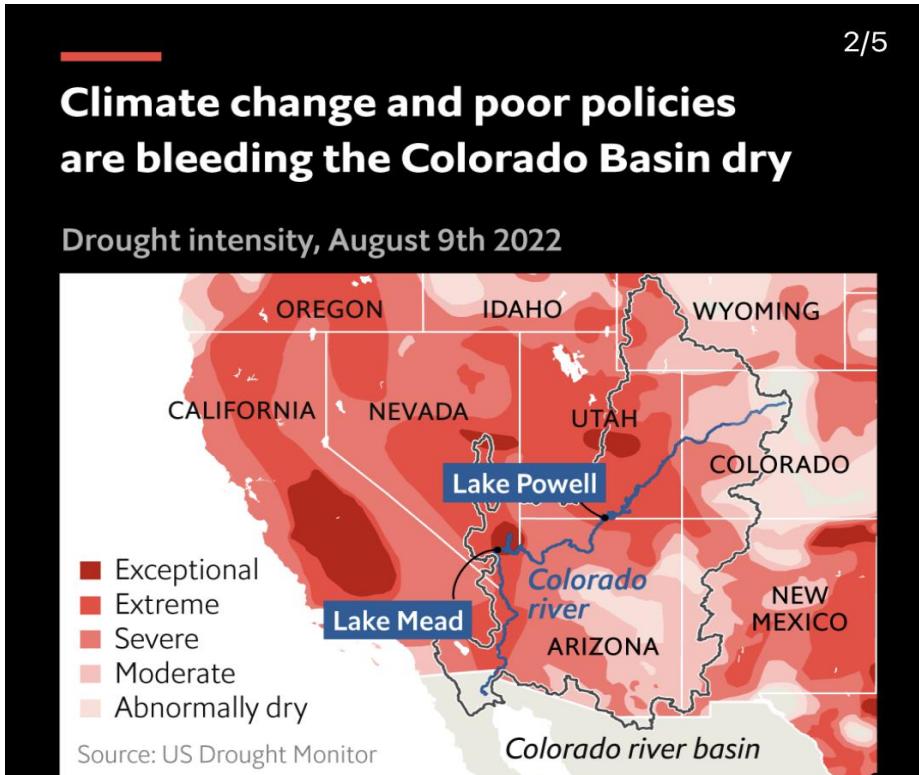
What is the difference between these two maps?



What is the difference between these two maps?



General Purpose
(Reference)



Thematic
vs.

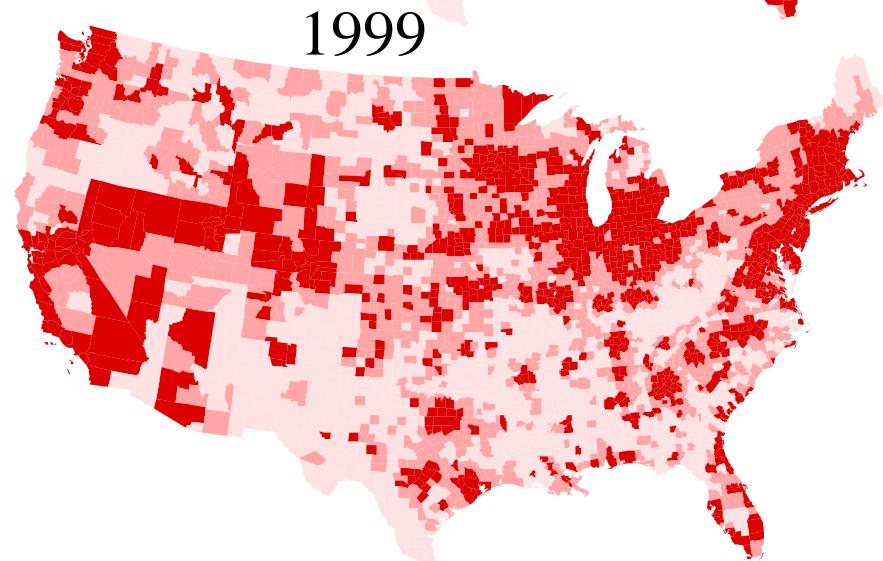
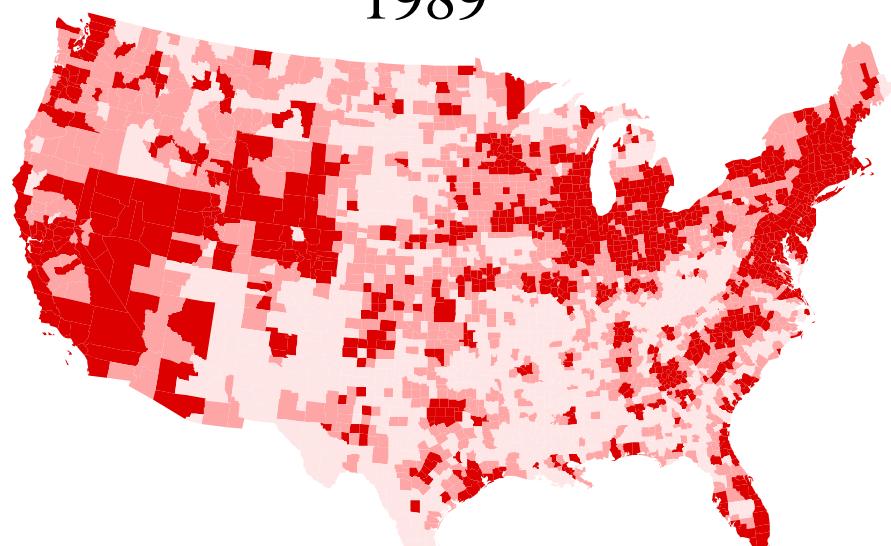
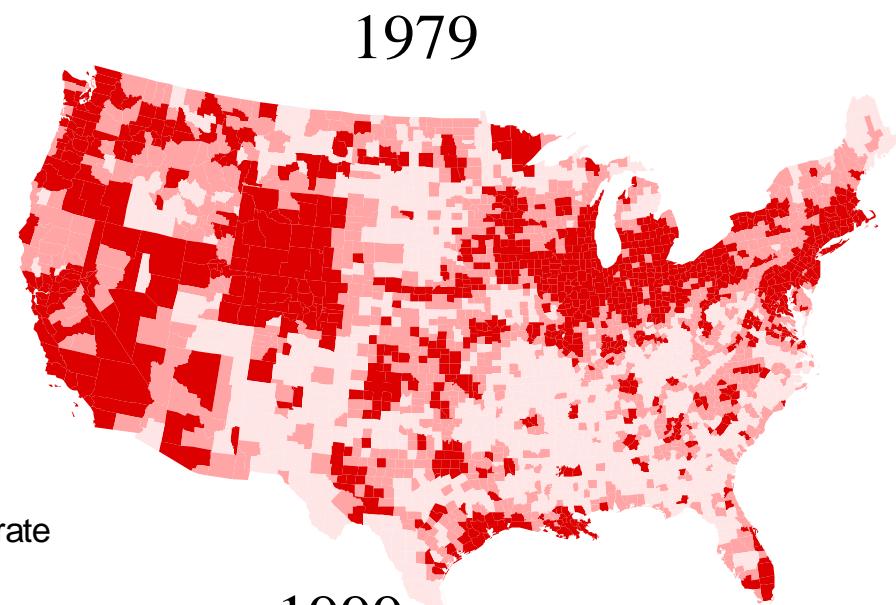
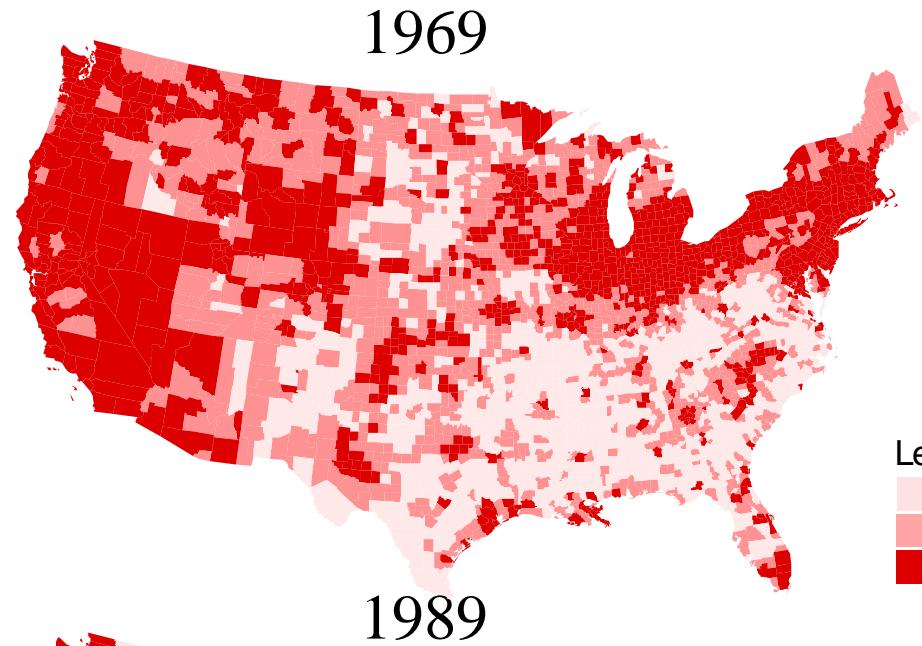
What is thematic map?

- Thematic map: a map that shows the spatial distribution of one or more geographic phenomena, which includes the key components of.
 - a geographic basemap,
 - **thematic data**.
- Your **main subject** must be distinctly visible against the base map

Why Thematic mapping?

- Describe a spatial pattern
- Explaining the pattern
- Predict the pattern
- Policy prescription/Finding solutions

Median Family Income



Legend

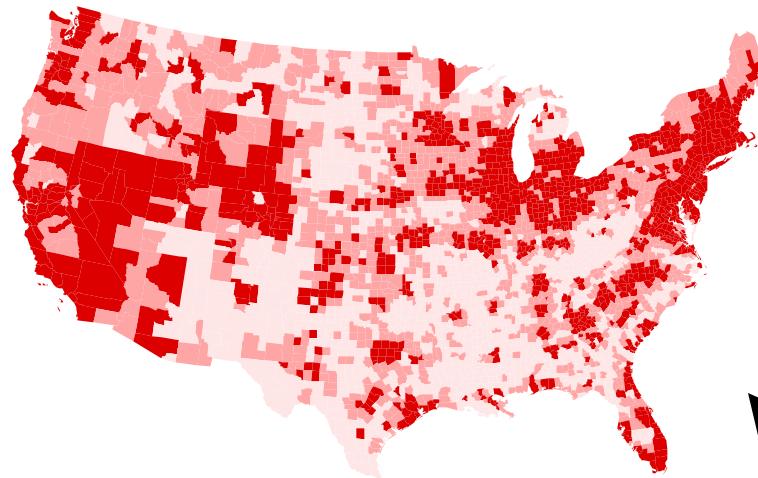
- Low
- Moderate
- High

Tobler's 1st law of geography: everything is related to everything else, but near things are more related than distant things.

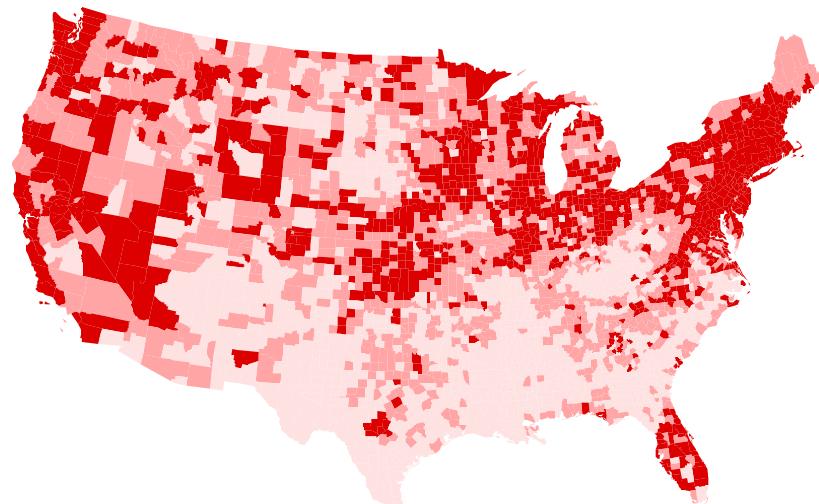
Economic Structure, Institutions, and Well Being (1990)

Median Family Income

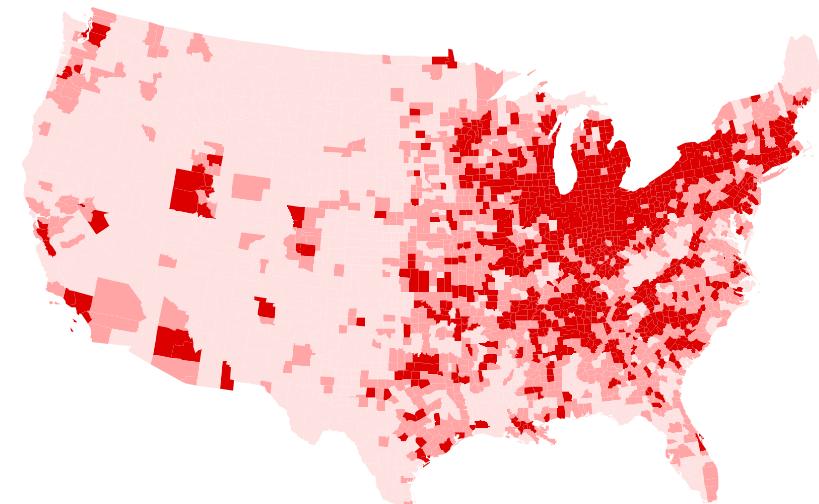
Low
Moderate
High



Institutions
(Government Social Welfare)



Economic Structure
(Manufacturing, Core Sector)

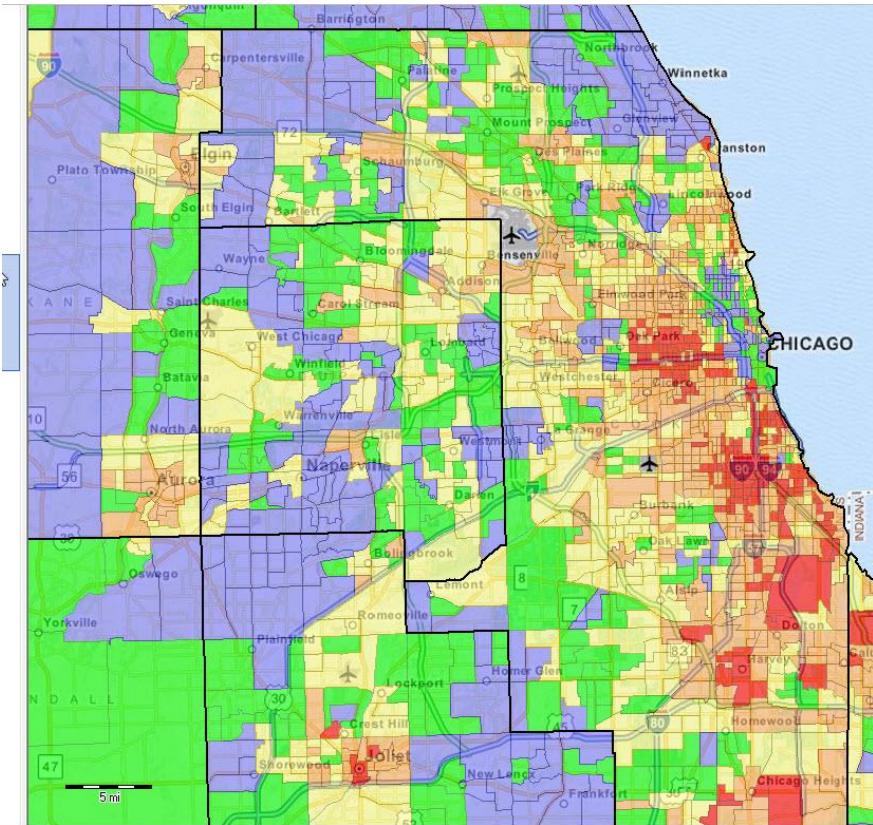


Types of Thematic maps?

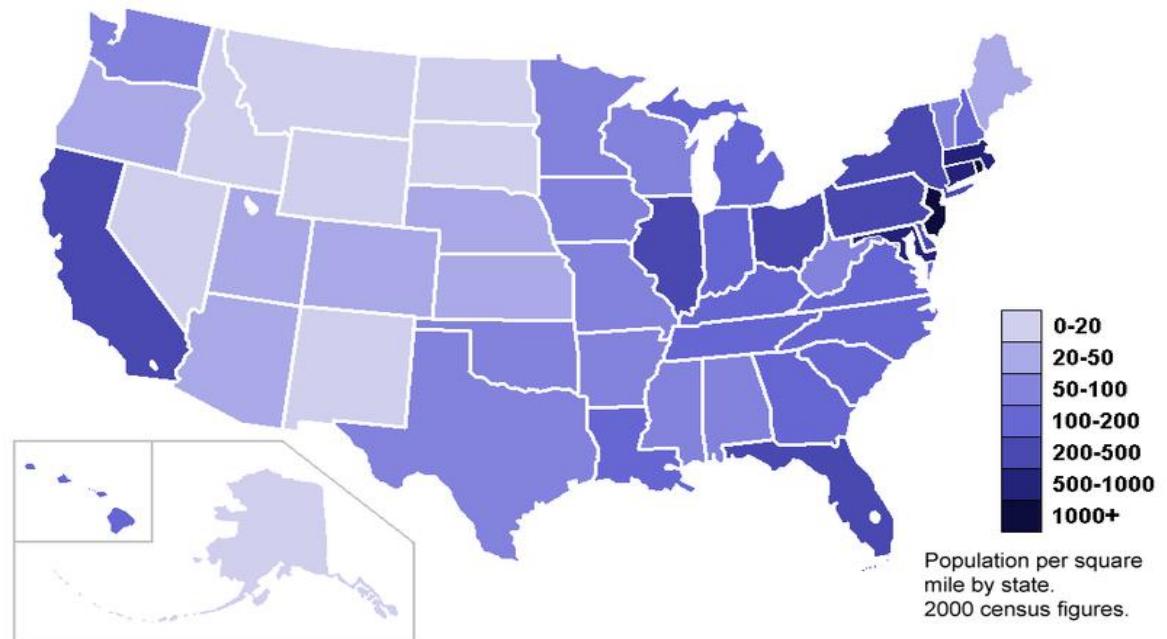
Based on how the data is mapped, the applied visual variable, and symbol dimensionality, there are different types of thematic maps:

1. Choropleth Map (Graduated Color Map): boundaries based on preexisting reporting zones (“enumerated units”) - such as a census population map
2. Proportional/Graduated Symbol Map: circle or line data, size of each circle or line represents the value of the attribute being mapped.
3. Dot Density Map: count data, number of points (randomly placed) in an area represents the value of the attribute being mapped.
4. Isoline Map
5. Other Types: Flow Map, Cartogram

Choropleth Map (graduate color map)

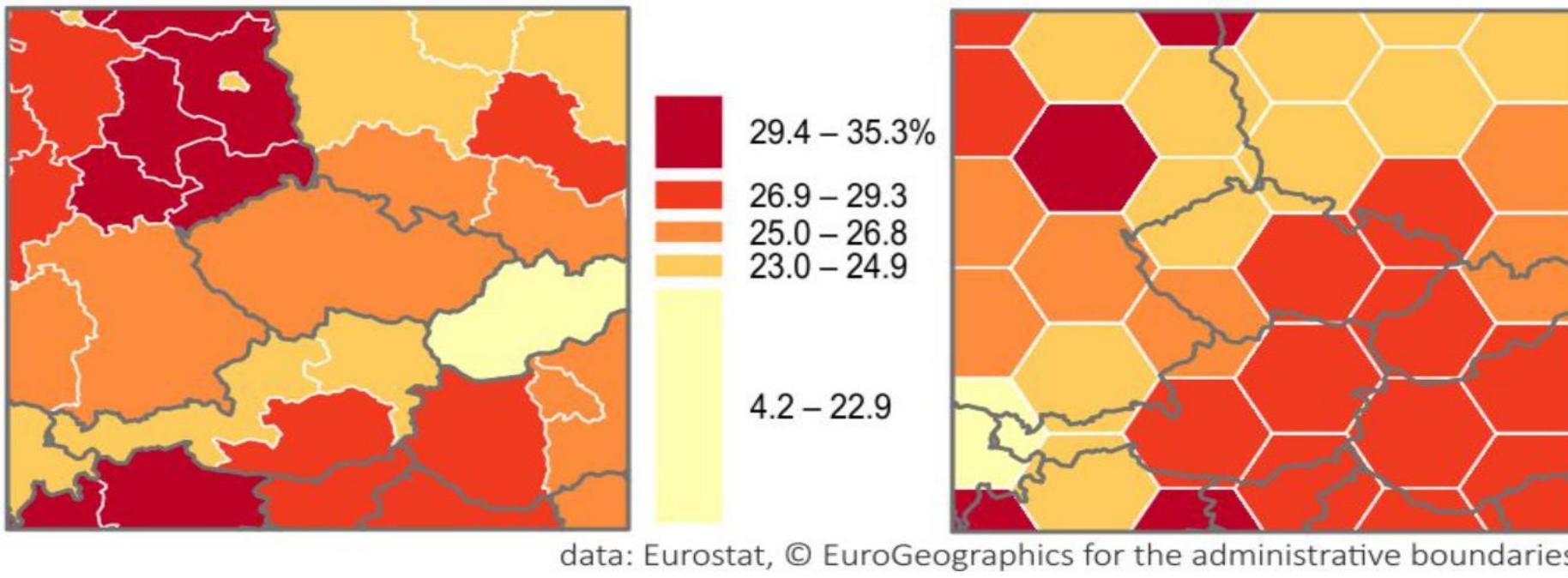


- Greek: Choros (place) Pleth (value)
- Data occur in a pre-defined “enumerated unit” (i.e. states, census tracts, regular shapes...)
- Not for continuous (field data) – boundary is discrete...
- Normalization (per area, per total population)
- Value is spread uniformly throughout the area: no variation within enumerated unit
- Graduate color: darker color – higher values; lighter color – lower values



Choropleth Map (graduate color map)

Share of population over the age of 60

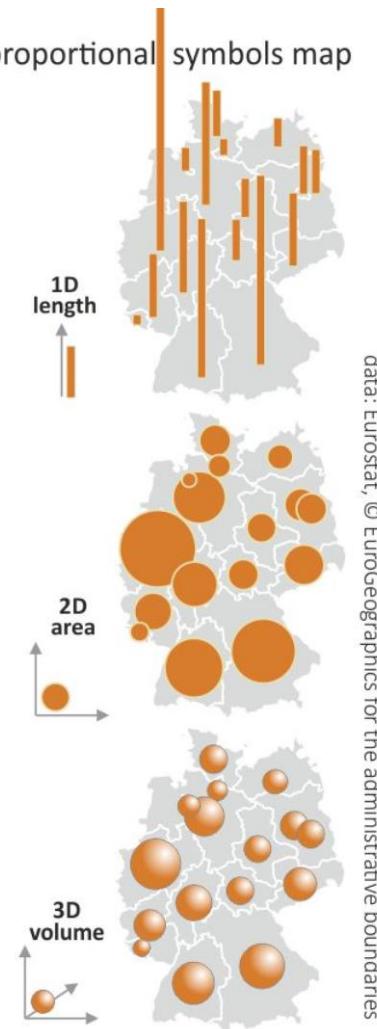
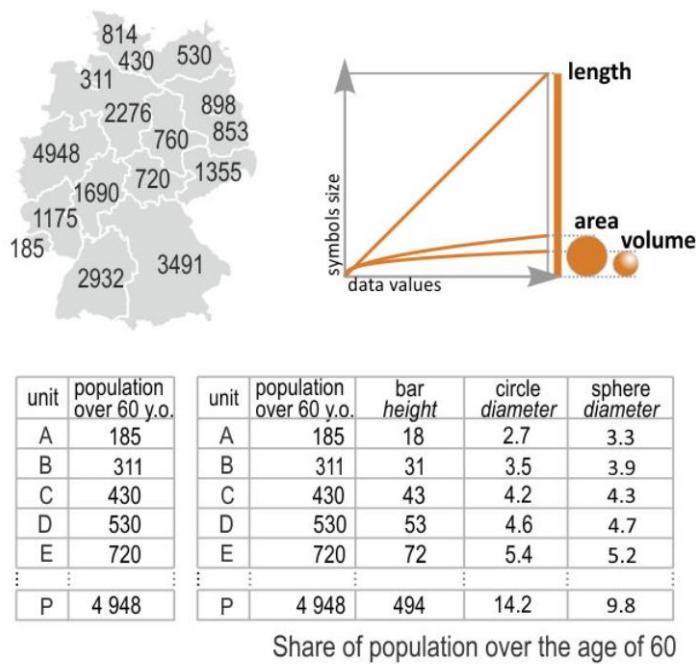


Left: Enumeration units on a choropleth map may be political boundaries (e.g., census)

Right: a tessellation of regular shapes like hexagons.

Proportional Symbol Maps

base map and data → symbols scaling → proportional symbols map



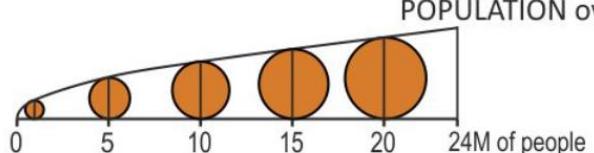
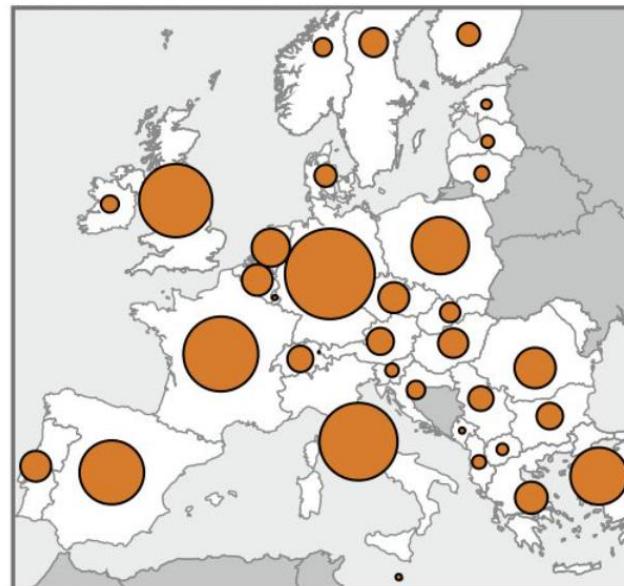
- A proportional symbol map is a thematic map that scales the size of a point symbol proportional to the represented value.
- **The scaling ratio** is the first important design consideration for proportional symbol maps. The symbol sizes should be large for the values of the phenomenon to be easily estimated and compared.
- Decide scaling ratio:
 - Linear Scaling: The size of the symbol is directly proportional to the value it represents.
 - Square Root Scaling: The size of the symbol is scaled by the square root of the value (avoid **overly large symbols**)

Graduated Symbol Maps

A graduated symbol map, sometimes called **range-graded symbols**, is a **classified proportional** symbol map.

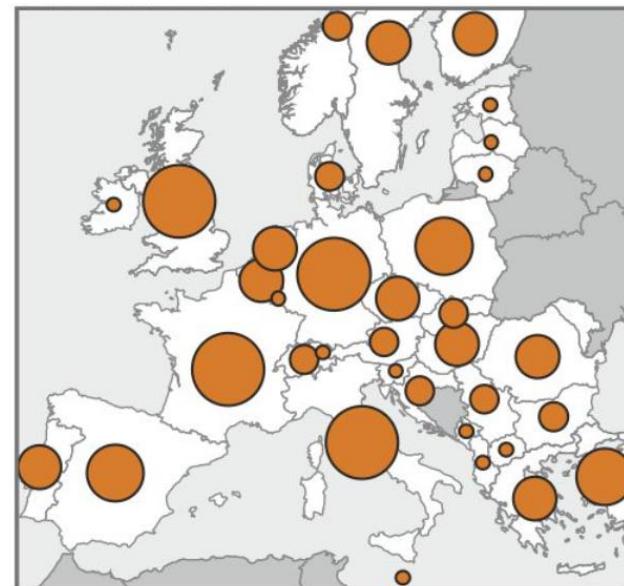
Proportional: ➔

- scales the symbol size based on the actual data value
- too many values: symbols may become indistinguishable.
- high values can become so large as to obscure other symbols.

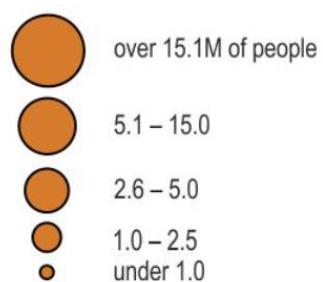


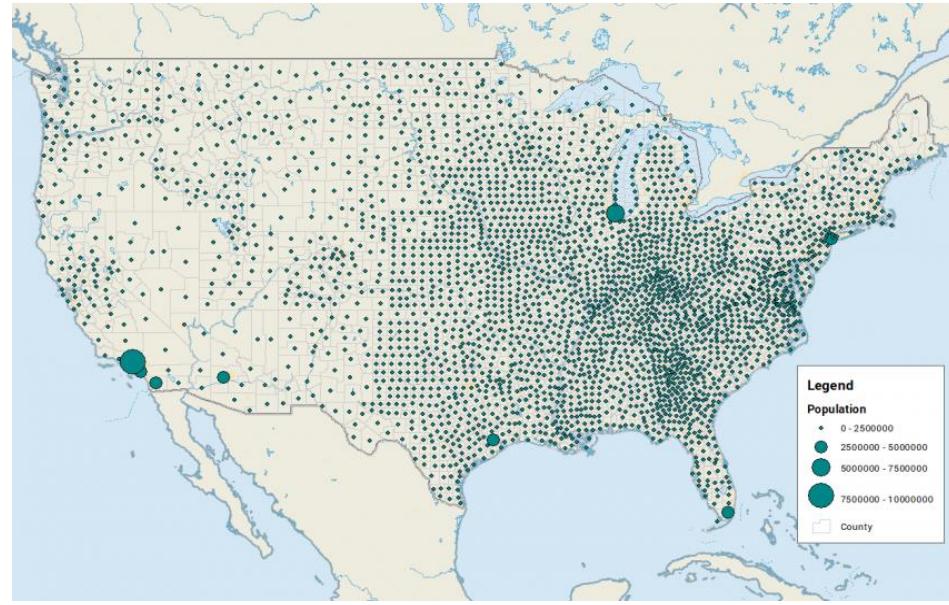
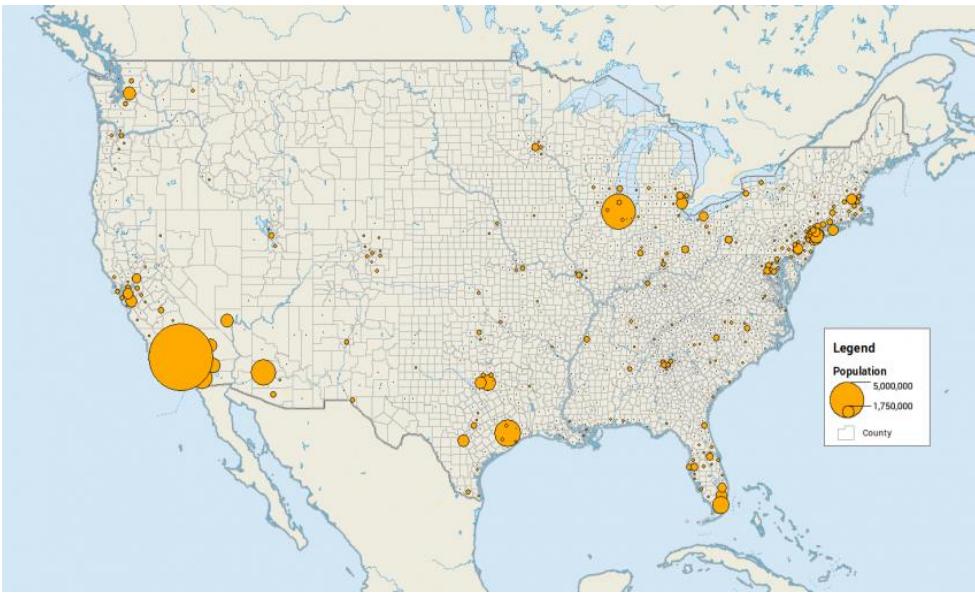
← Graduated:

- quantitative values are grouped into classes based on pre-defined symbol sizes. Within class, all features drawn with same symbol.
- You can't discern the value of individual features; you can only tell that its value is within a certain range.
- Easier to read



data: Eurostat, © EuroGeographics for the administrative boundaries





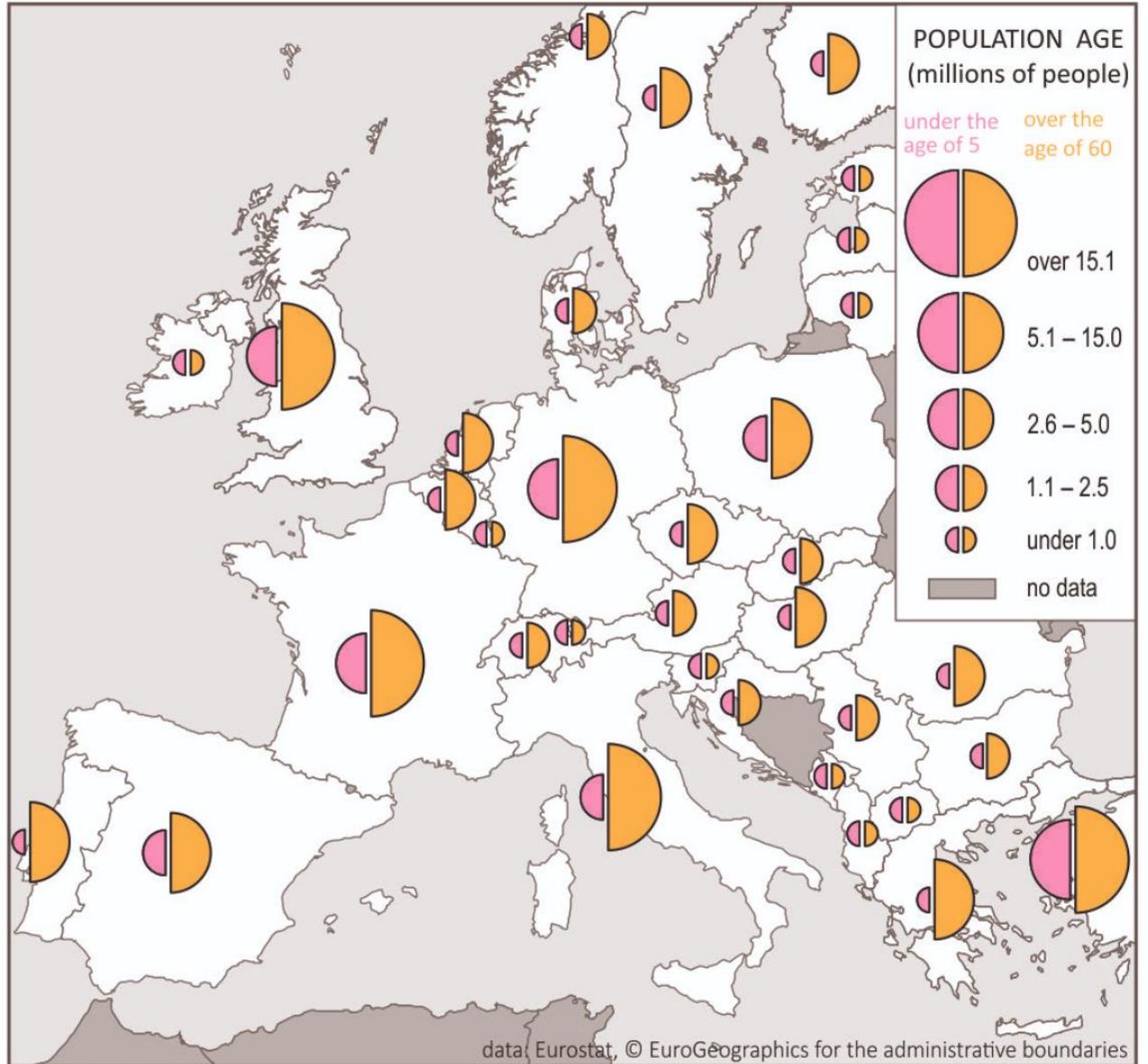
proportional symbol maps scale the size of the symbol based on value. **gradual symbols** puts each value into a class and scales it based on that “bin”.

proportional maps: symbology is unclassed: dots are scaled with absolute magnitude.

graduated symbol maps divides quantities into classes using data classification techniques. An advantage of graduated symbol maps over choropleth maps is that the **size of the geographic feature doesn't matter**.

Graduated Symbol Maps

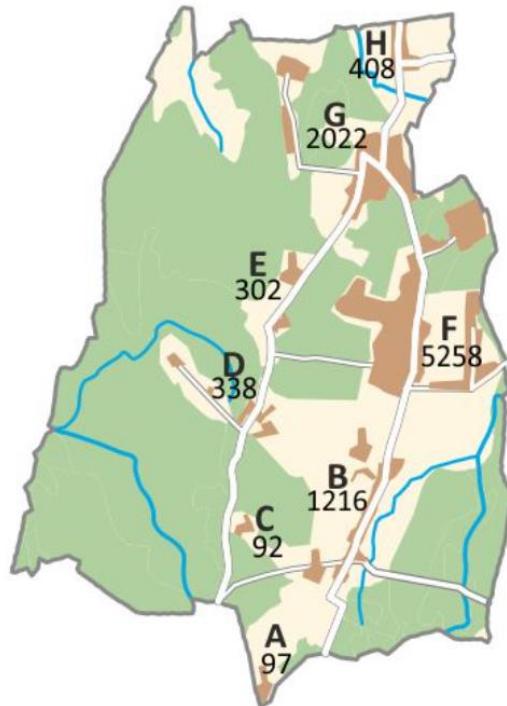
A graduated symbol map can represent more than one set of data.



Dot Density Map

- A dot density map places dots within an enumeration unit—like a district or a neighborhood—in proportion to the represented value/ feature, to preserve the **distribution and variation** of density of a phenomenon.

base map and data → dot size
dot value → dot density map



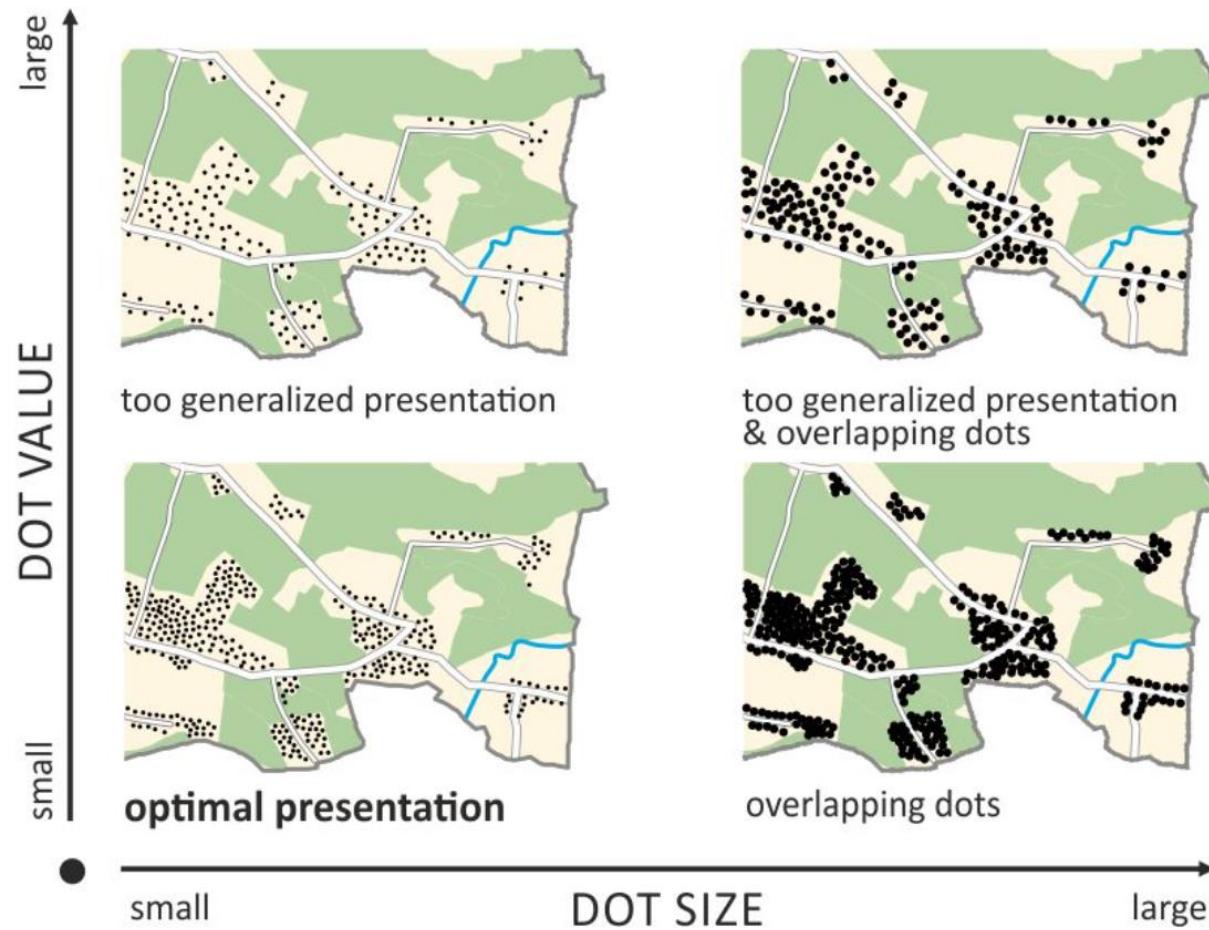
• 20 people

name	population	number of dots
A	97	5
B	1216	61
C	92	5
D	338	17
E	302	15
F	5258	263
G	2022	101
H	408	20



Dot Density Map

Y axis: number of items per dot

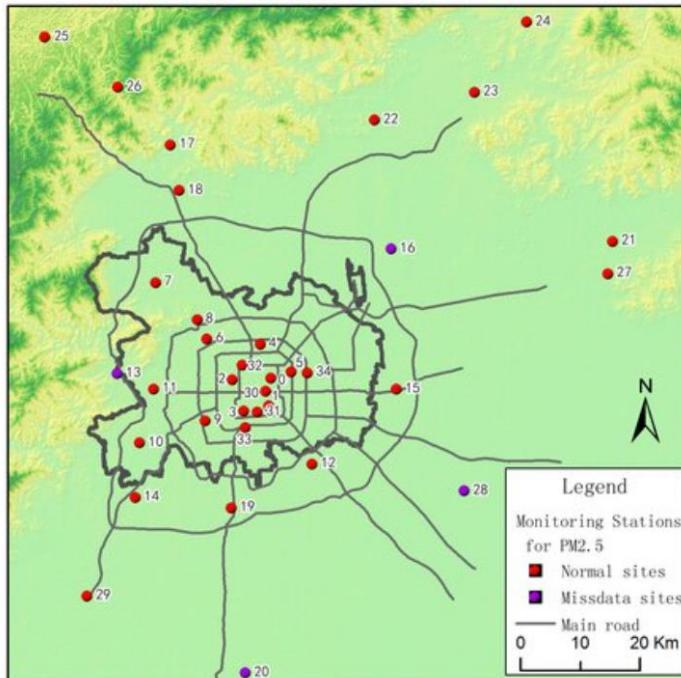


- Shows variations in spatial density
- Main concerns are dot placement and selection of dot value and size
- Simplest form is one-to-one mapping (one dot for one item) - more common to have one dot to many items
- dot placement is randomized typically - additional information (e.g., land use) should be used to improve the location of dots

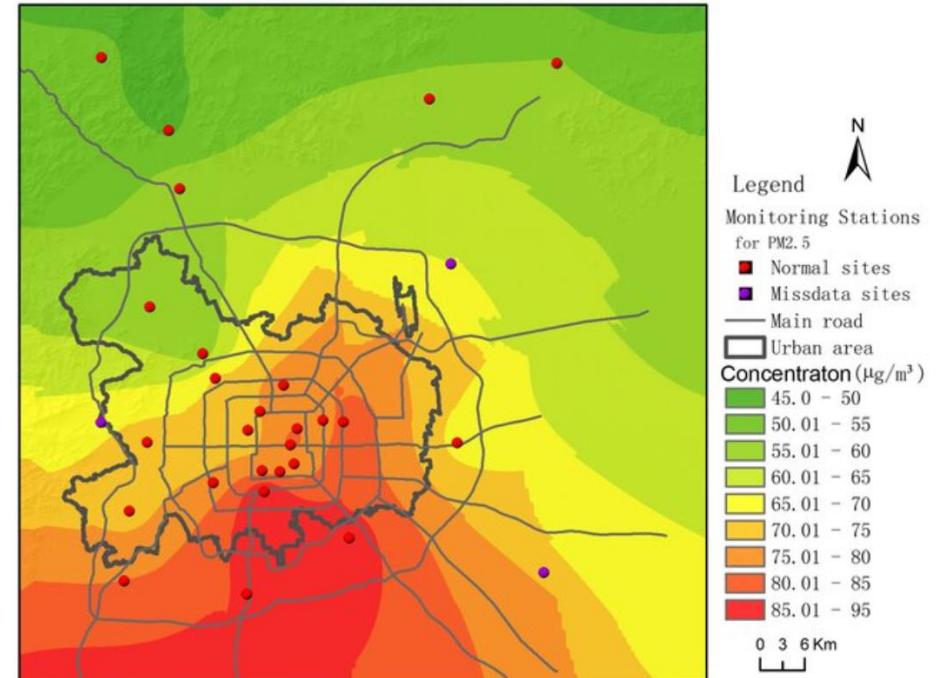
Isoleine Maps

An **isoline map (contour map)** represents data with lines derived through interpolation that connect points of equal attribute value on a map. The lines are called isolines or isarithms.

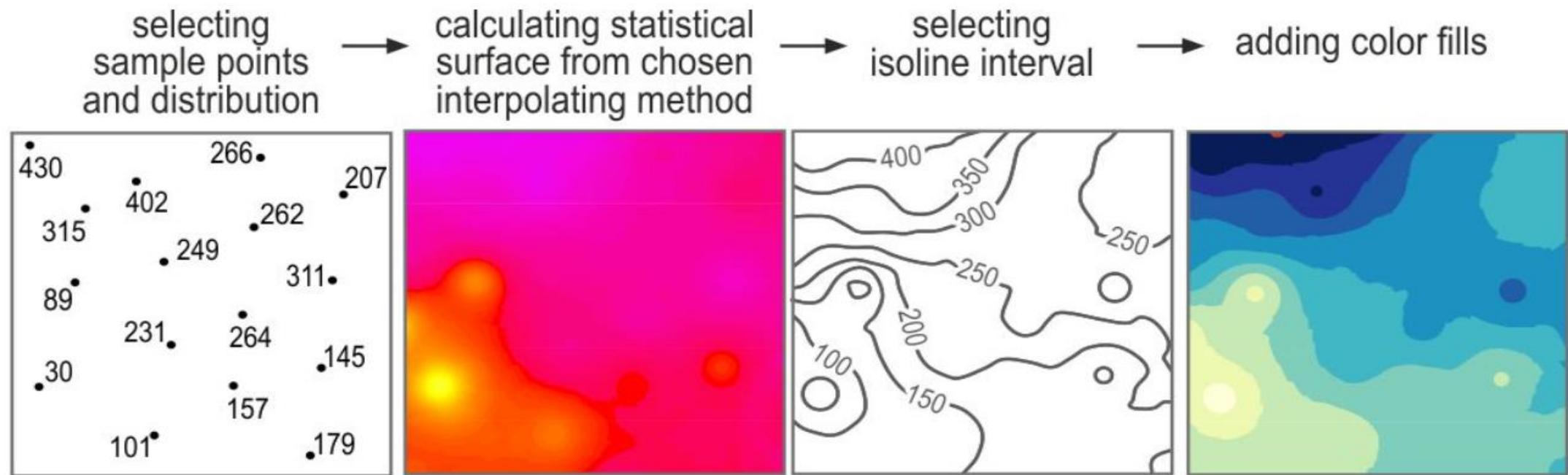
- Isometric maps use isolines resulting from interpolating values collected at **sample points** (e.g., weather stations).
- Isoplethic maps use isolines resulting from interpolating values enumerated across **areas**. (e.g., population density)



Code	Sitename
0	DongChengDongSi
1	DongChengTianTan
2	XiChengGuanYuan
3	XiChengWanShouXiGong
4	ChaoYangAoTiZhongXin
5	ChaoYangNongZhanGuan
6	HaiDianWanLiu
7	HaiDianBeiBuXinQu
8	HaiDianBeiJingZhiWuYuan
9	FengTaiHuaYuan
10	FengTaiYunGang
11	ShiJingShanGuCheng
12	YiZhuangKaiFaQu
13	MenTouGouLongQuanZhen
14	FangShanLiangXiang
15	TongZhouXinCheng
16	ShunYiXinCheng
18	ChangPingZhen
19	DaXingHuangCunZhen
21	PingGuZhen
22	HuaiRouZhen
23	MiYunZhen
25	YanQingZhen
17	ChangPingDingLing
20	JingNanYuZun
24	JingDongBeiMiYunShuiKu
26	JingXiBeiBaDaLing
27	JingDongDongGaoCun
28	JingXiNanLiuLiHe
29	QianMenDongDaJie
31	YongDingMenNeiDaJie
32	XiZhiMenBeiDaJie
33	NanSanHuanXiLu
34	DongSiHuanBeiLu



Isoline Maps



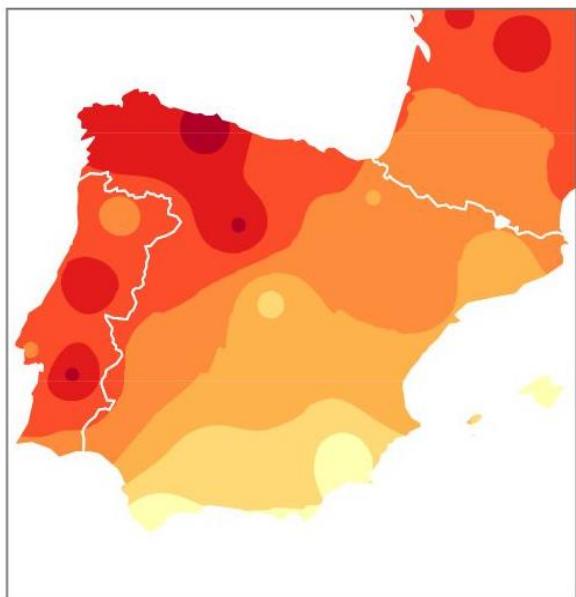
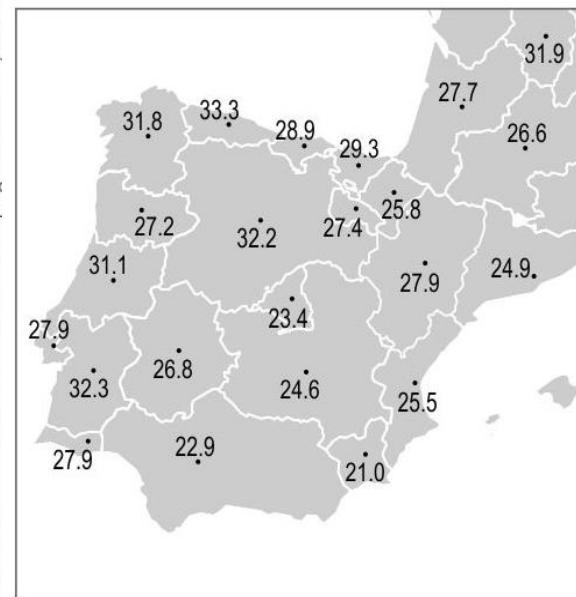
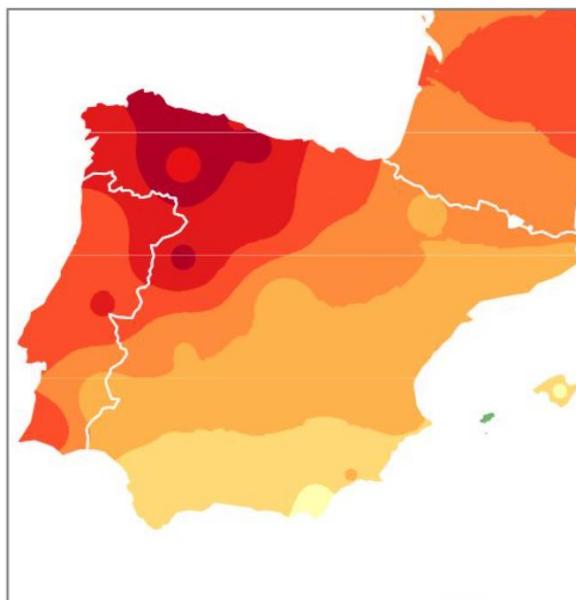
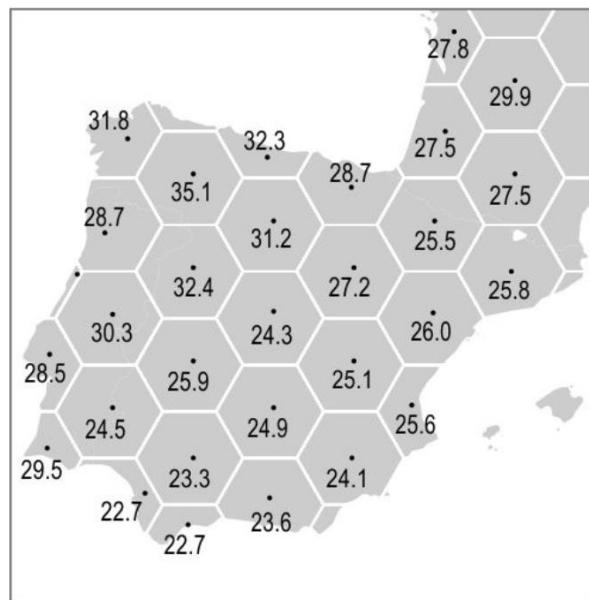
Isoline map involves several major design considerations

- **Distribution of samples, interpolation methods, isoline interval selection...**

Isoline Maps

Major concern:

- results are sensitive to the distribution of sample points

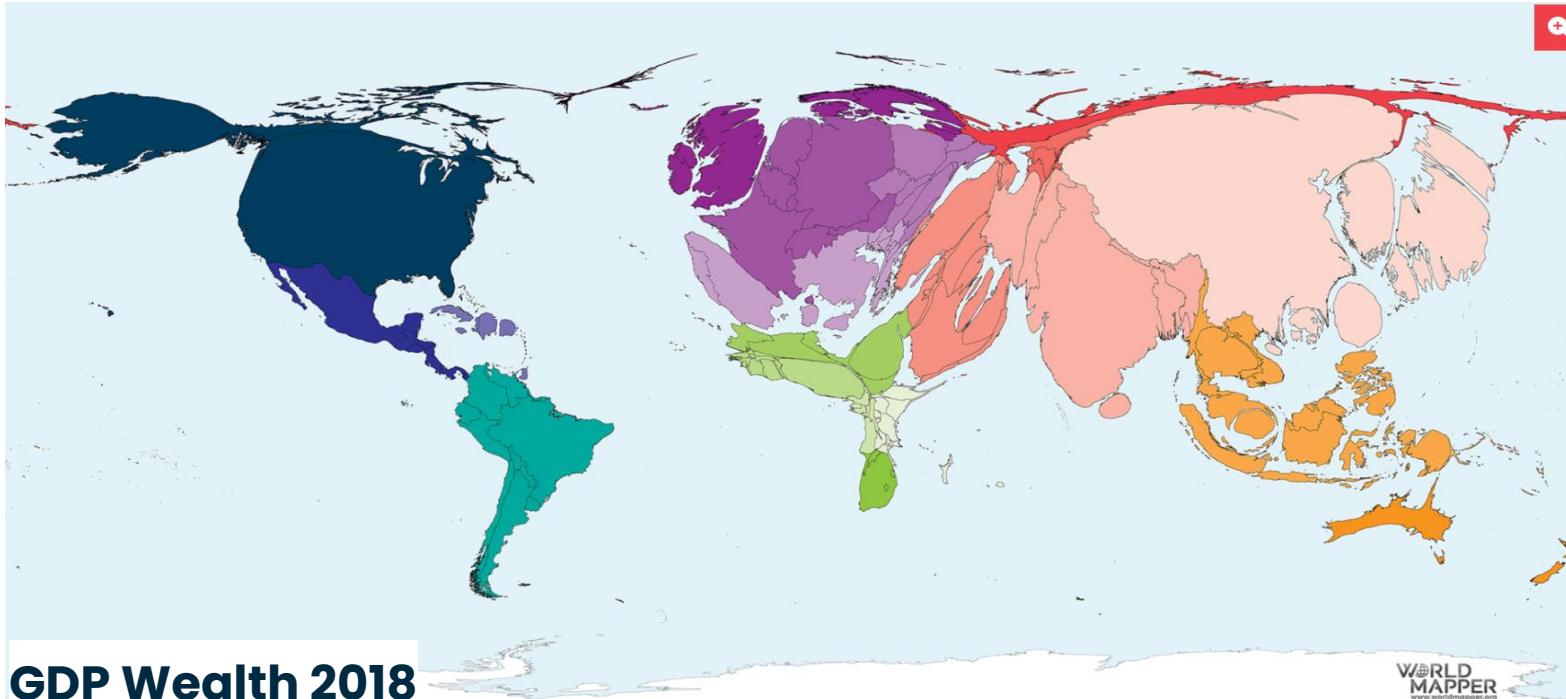


Share of population over the age of 60

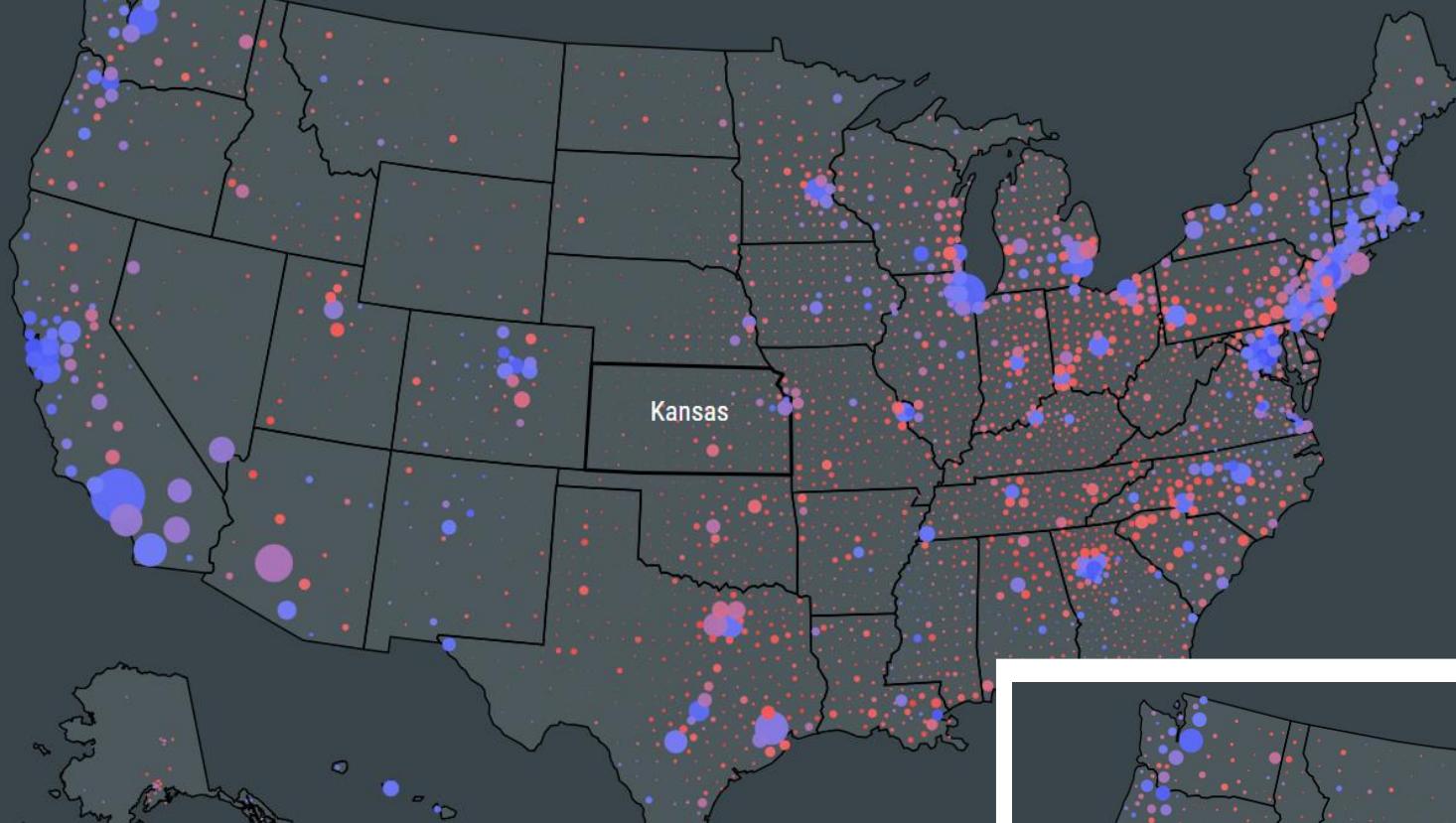


Other Maps: Cartograms

- Unique representations of geographic space intended to convey certain information
- Distort area, shape, or distance in favor of a specific objective of representation



Election of 2020

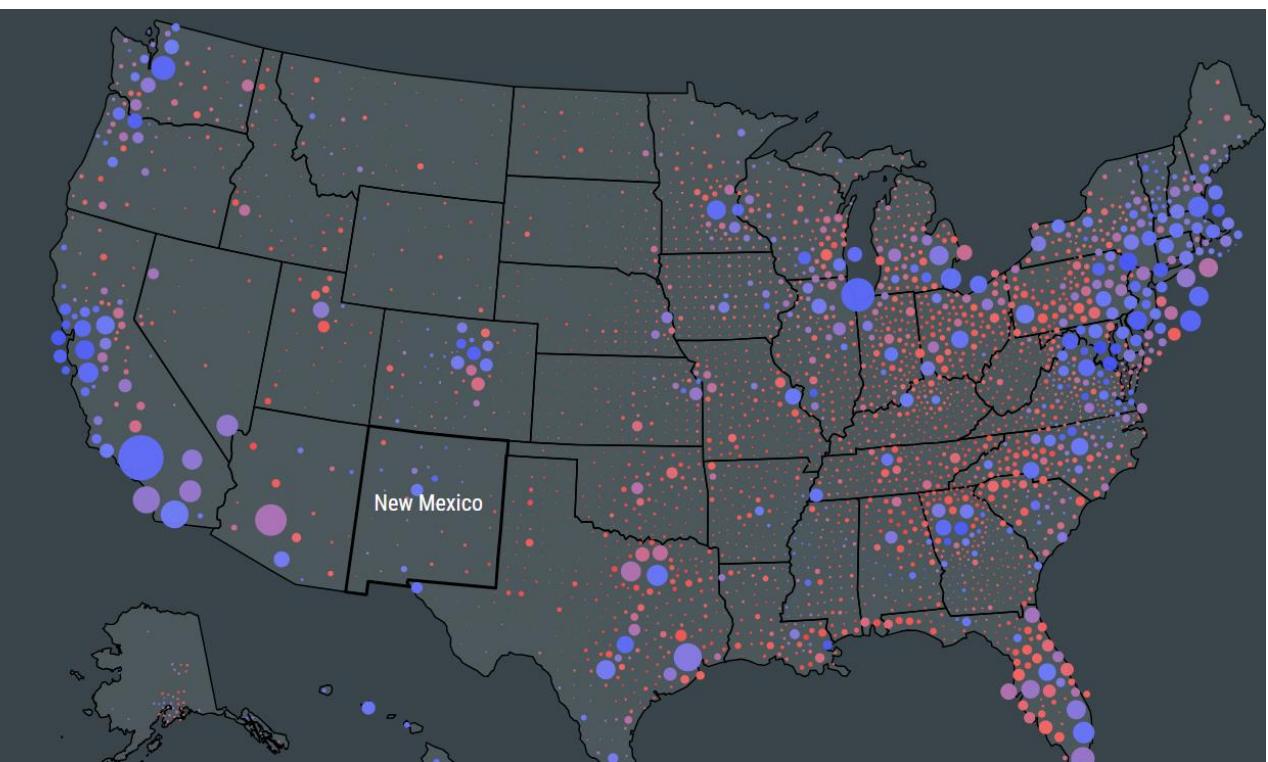
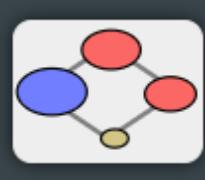


Democratic Joe Biden	81,152,510
Republican Donald Trump	74,192,165
Other Parties Libertarian Green	2,991,383 1,808,705 380,372

BUBBLES



CARTOGRAM

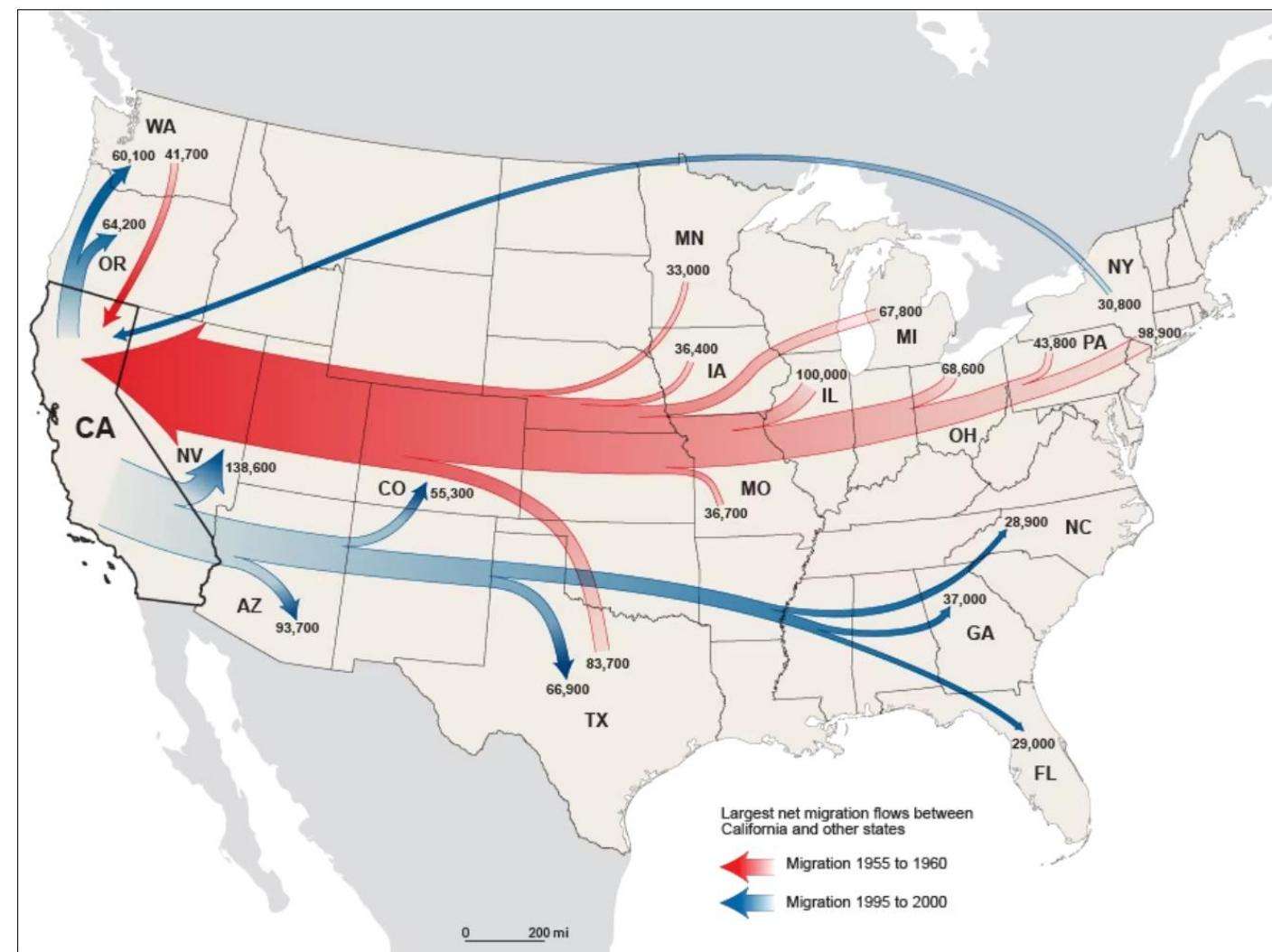


Other Maps: Flow map

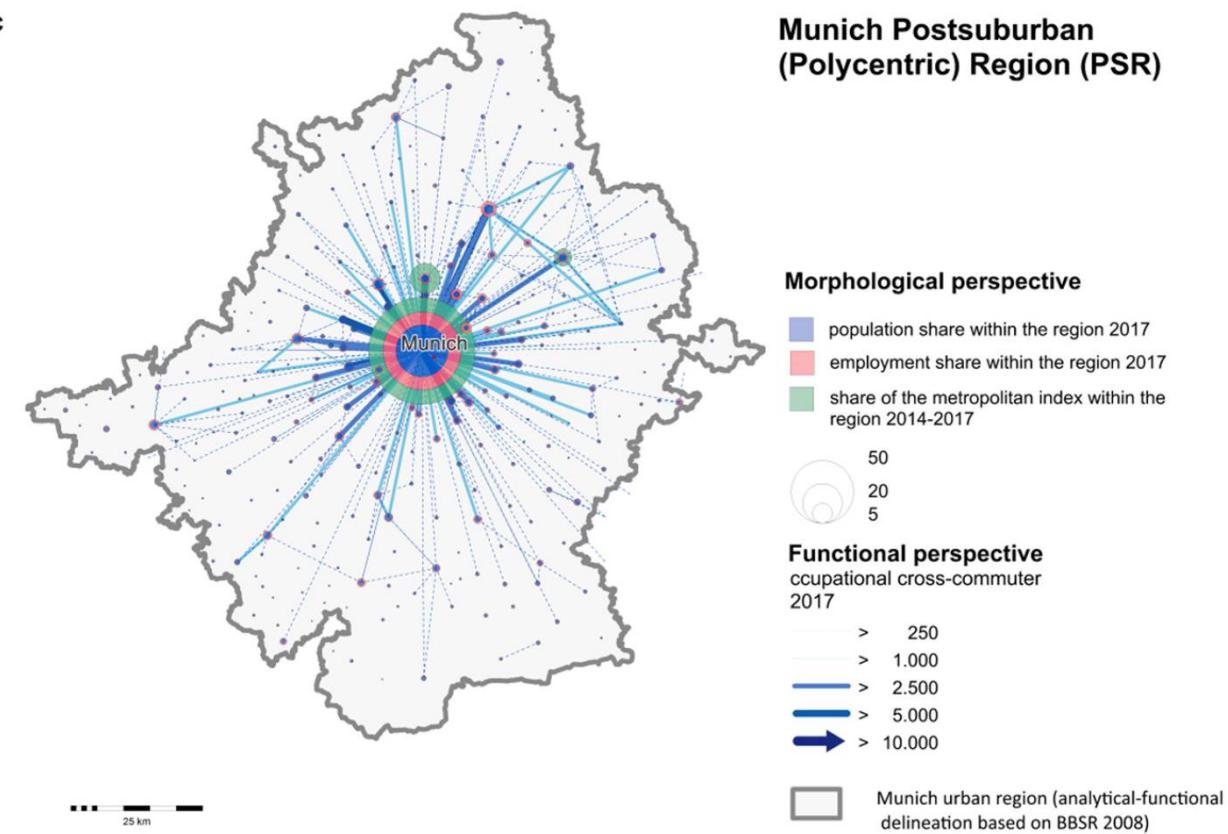
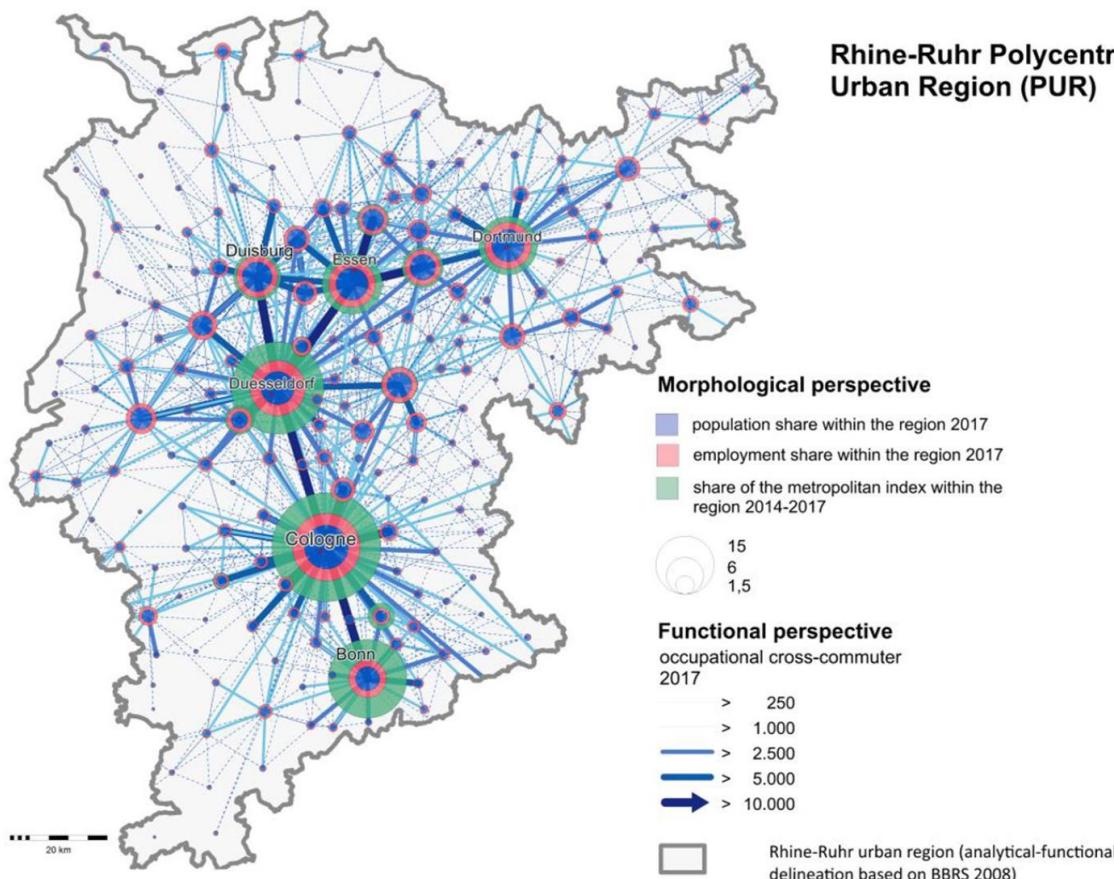
- A *flow map* represents the direction and/or magnitude of a phenomenon along linear objects or between locations.

Net Migration Between California and Other States: 1955-1960 and 1995-2000

March 7, 2013



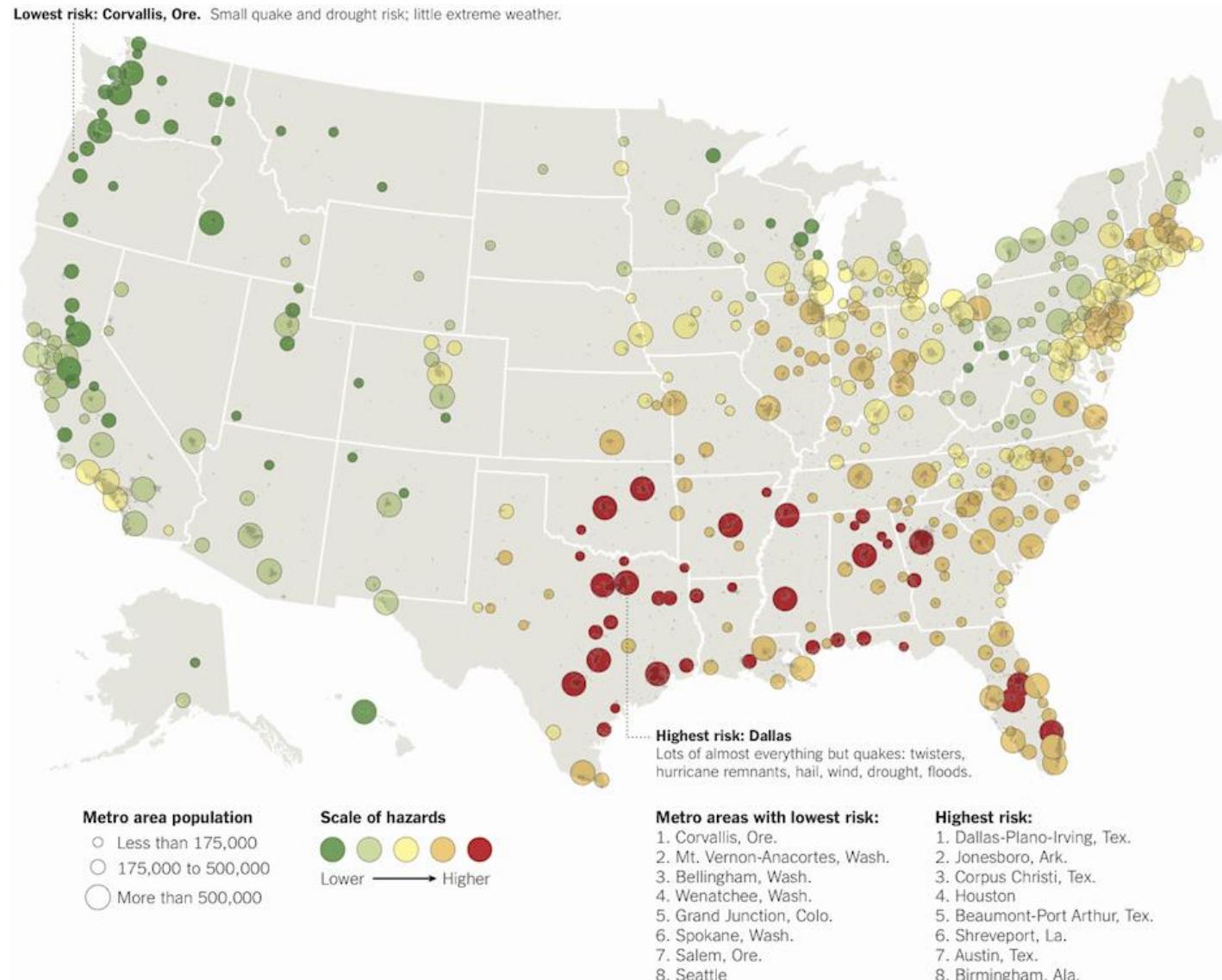
Other Maps: Flow map



Multivariate maps

It is common to plot two variable on a map by using two thematic map types, and these are referred to as **bivariate maps**.

Map of U.S. risk to natural disasters using graduated symbols to represent metro area populations (and hue to represent risk)

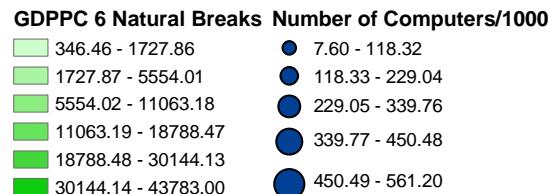
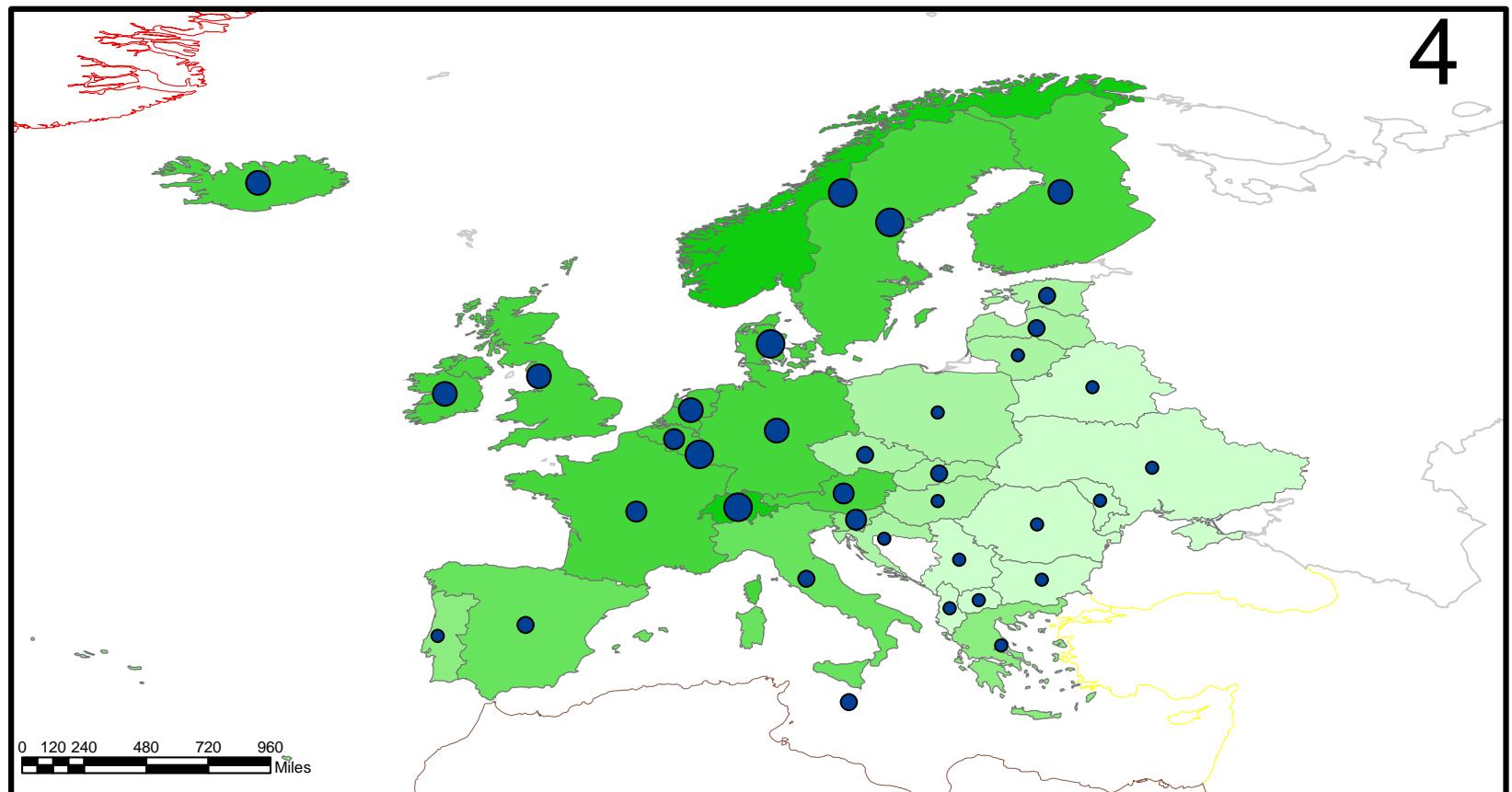


Multivariate maps

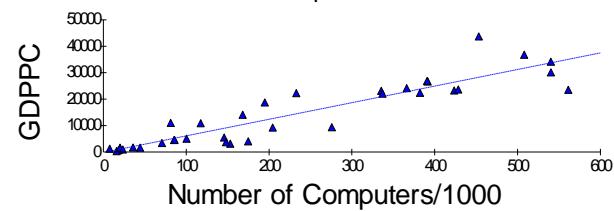
Multivariate map:
Bivariate

European Economic Development 2000

4



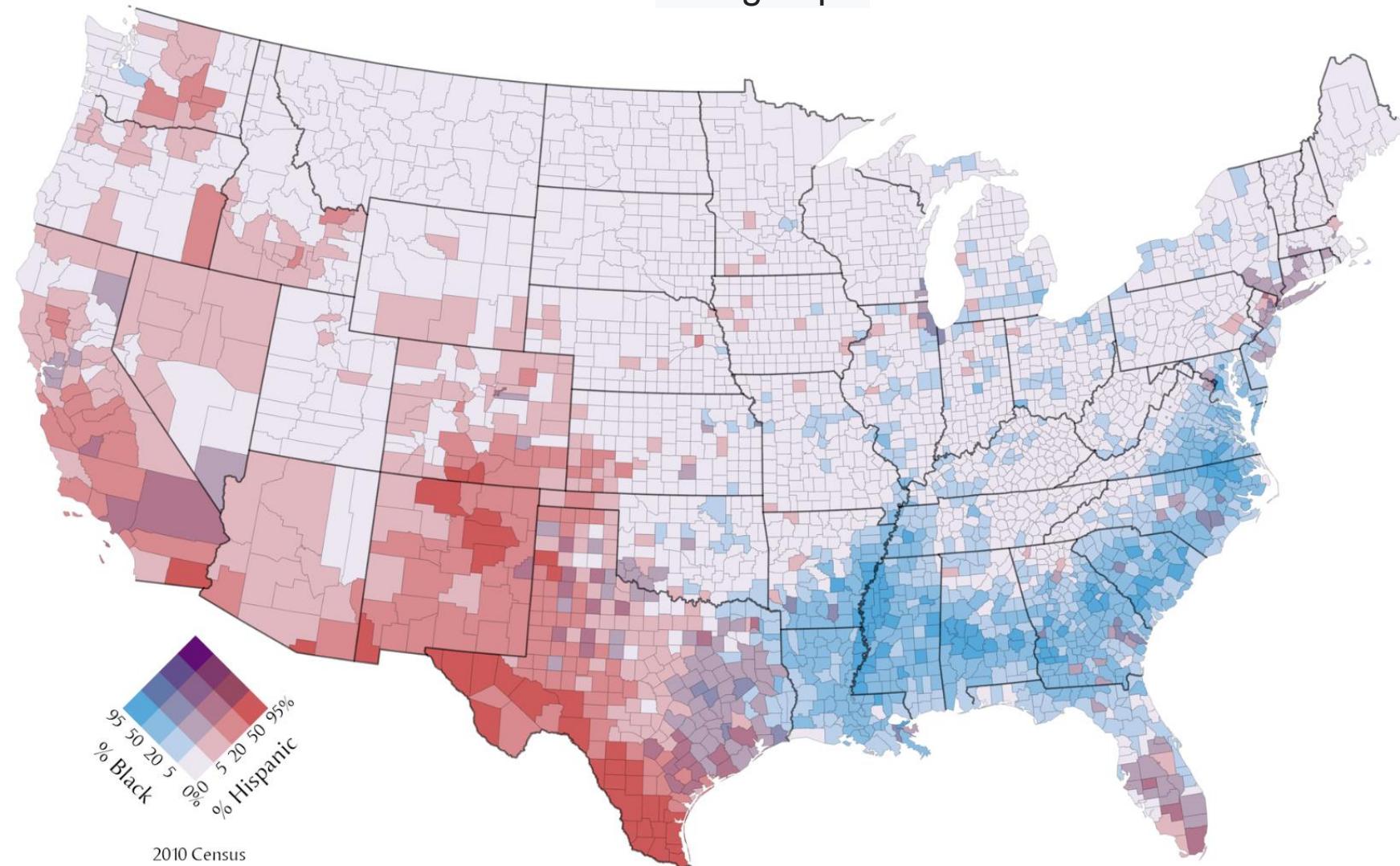
Relationship between GDPPC and Number of Computers



Created by Joe D. Francis
September 2004

Multivariate maps

Bivariate choropleth map comparing the Black (blue) and Hispanic (red) populations in the United States, 2010 census; shades of purple show significant proportions of both groups.



Type of Thematic Maps — A Summary

Choropleth map

Strengths:

- allows good representation of a theme, tied to well-defined enumeration units
- user/reader friendly

Limitations:

- creates the risk of unintentionally masking important details due to improper choice of the number of classes and classification methods
- boundaries of the enumeration units (e.g., census units) often do not associate with the distribution of the phenomenon (e.g., misleadingly suggest homogeneity within the enumeration unit)
- often has enumeration units of different sizes, and the largest ones visually dominate
- does not allow exact values to be derived

Type of Thematic Maps — A Summary

Proportional / graduated symbol map

Strengths:

- effectively represents datasets with a large range of values
- can represent individual point data and enumerated data
- represents discrete, abrupt surfaces

Limitations: not very reader/user friendly

- may cause problems with overlapping symbols
- does not allow exact values to be read (graduated symbol maps only)
- ...

Type of Thematic Maps — A Summary

Dot density map

Strengths:

- represents distribution and variations in pattern, such as clustered or dispersed distribution
- represents smooth, discrete surfaces

Limitations:

- does not allow quantities to be read
- ...

Type of Thematic Maps — A Summary

Isoline map

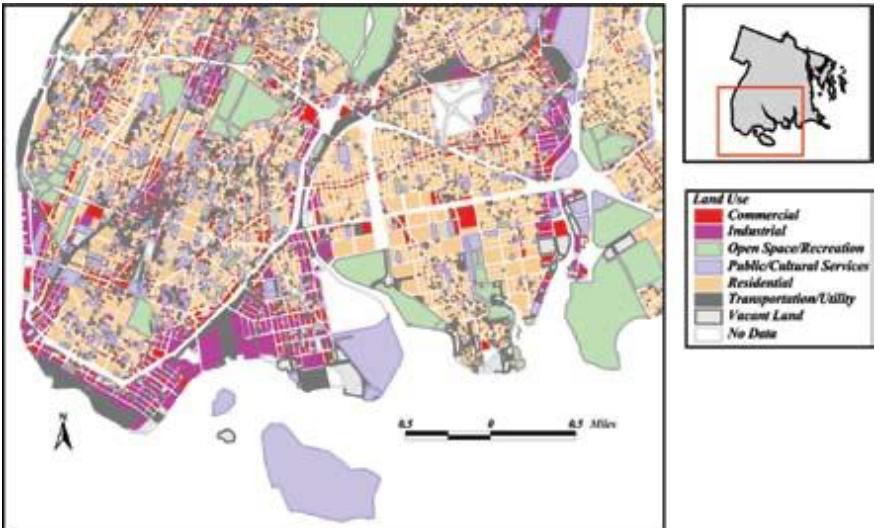
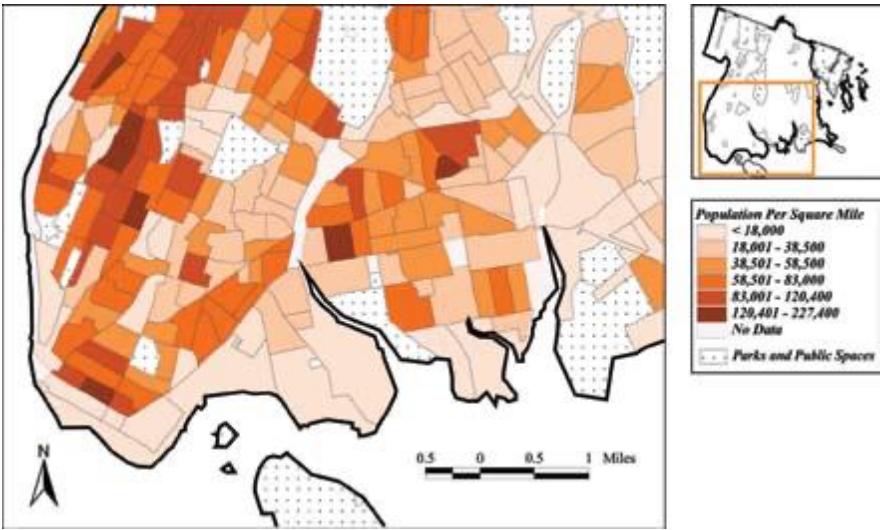
Strengths:

- effectively represents the arrangement of magnitudes, the orientation of surface gradients, and the main lines of distribution
- represents smooth, continuous surfaces

Limitations:

- does not allow exact values to be read

Data Mapped: Quantitative vs. Qualitative Map

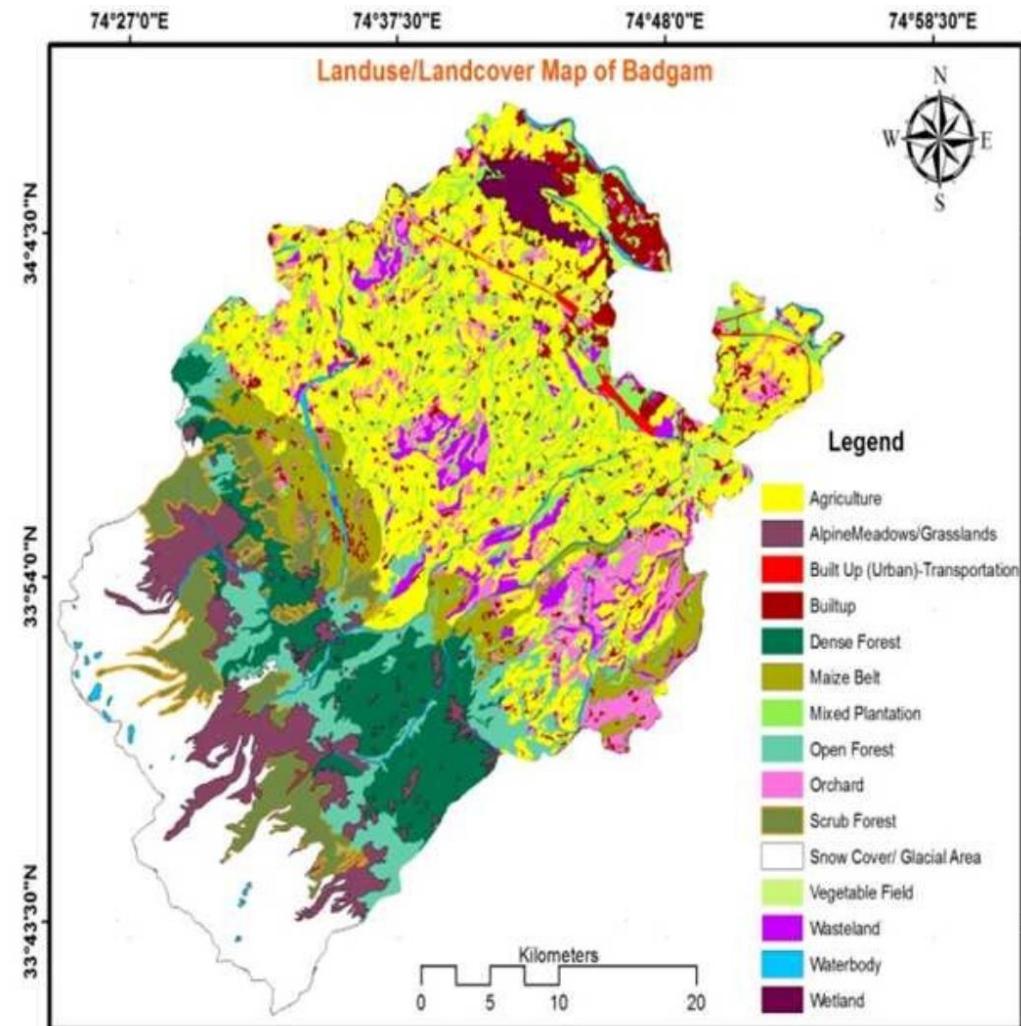


- Quantitative maps show numerical relationships of the variables being mapped
- Qualitative maps are based on descriptive information, and show location and boundaries of differences of kind or type

- Qualitative maps make use of **nominal** (e.g. land cover class) or **ordinal** (e.g. a ranking) measurement scales,
- while quantitative maps make use of **interval** (e.g. Celsius temperature) or **ratio** (e.g. density, income, etc..) measurement scales.
- Is the attribute being mapped nominal, ordinal, interval, or ratio?

Qualitative Thematic Maps — Measurement scales: Nominal

- Objects are classified into groups. The groups have names, not numeric values. There is no ordering implied. Also called *categorical* or *qualitative* data.
 - Examples: soils, land use, land cover, zoning, forest types, ethnicity
- The thematic map of a nominal attribute field is often called a “unique value” map.



Qualitative Thematic Maps — Measurement scales: **Ordinal**

Implies a hierarchy of rank—a ranking of classes.

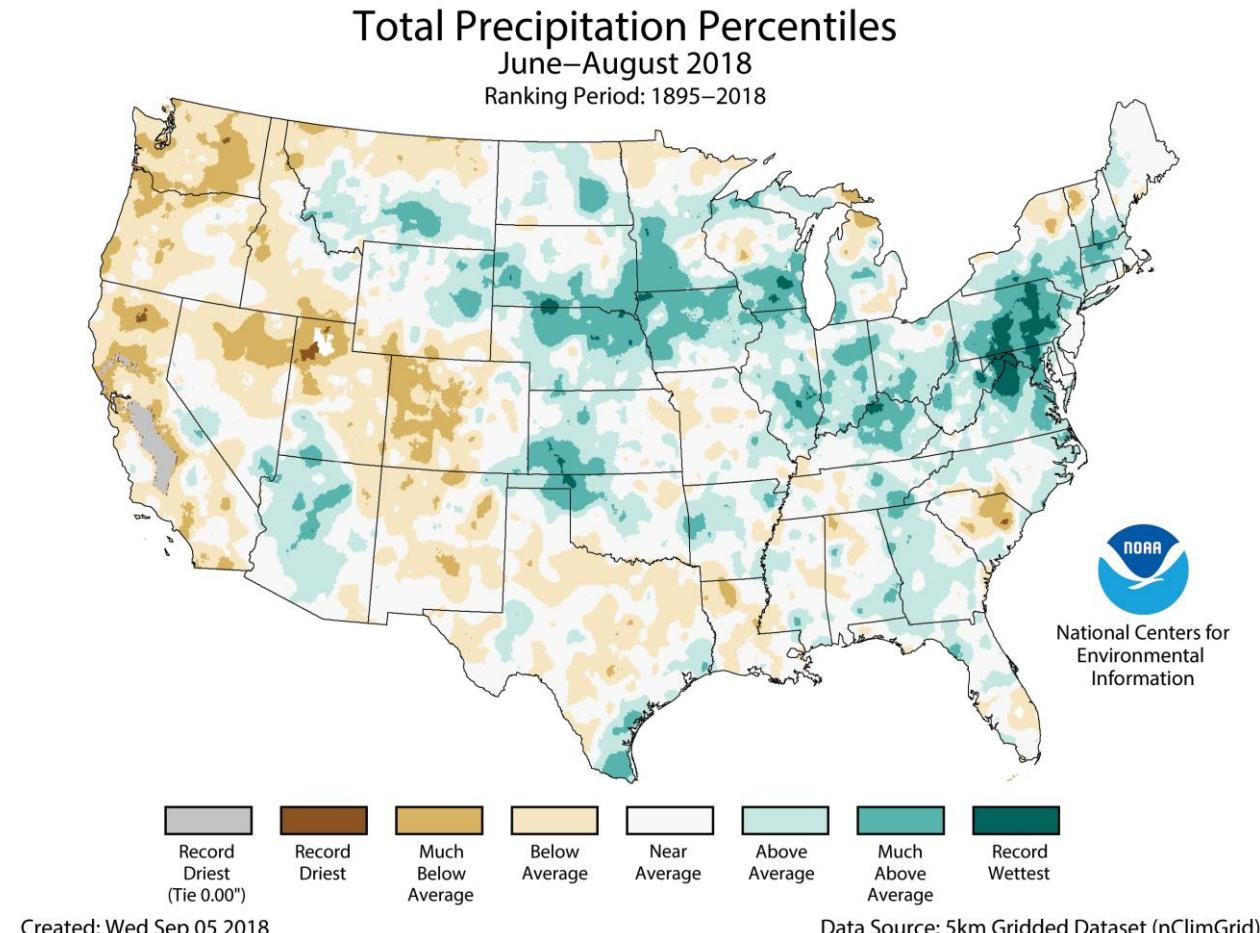
Opinion poll response of “strongly agree,” “agree,” “disagree”

Soils can be ordered from “poorly drained” to “somewhat poorly drained” to “well-drained” to “excessively drained”

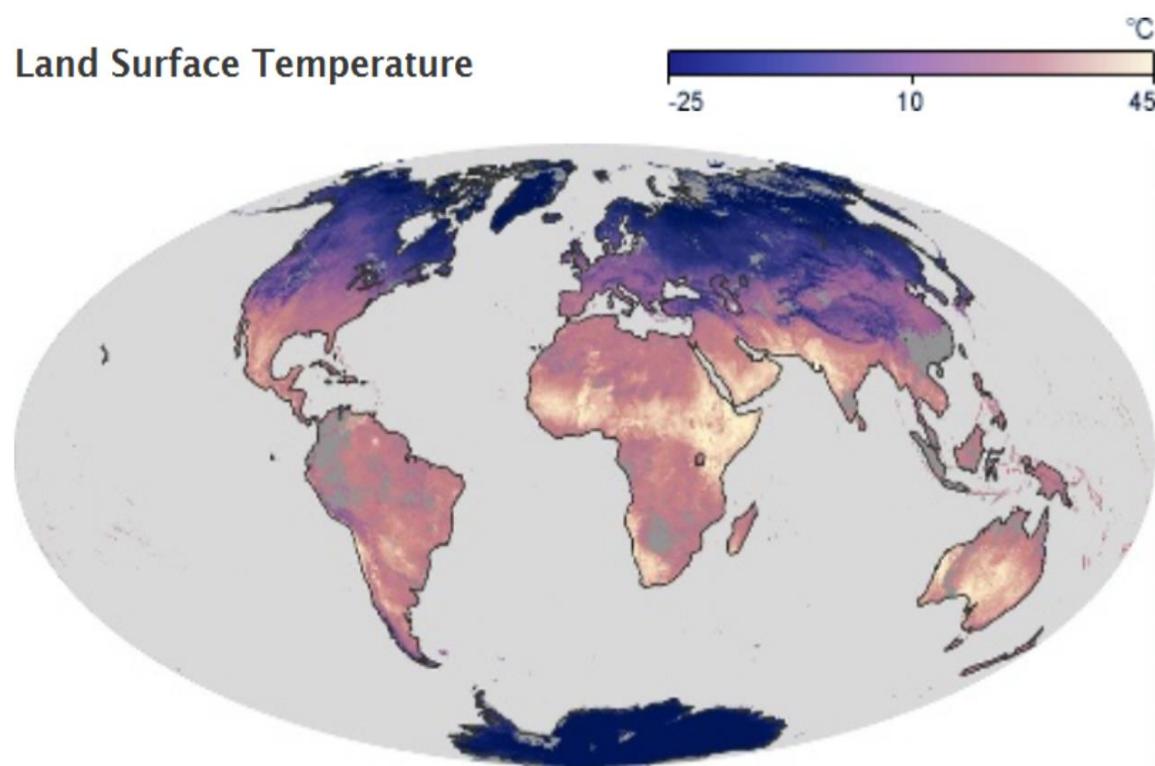
*building condition (poor to excellent)
Roads are assigned congestion levels.*

- We can assign numbers to these categories, but it doesn't automatically imply that we can use arithmetic relationships.

The map of an ordinal attribute field should use “graduated” symbols or shade patterns.



Quantitative Thematic Maps — Measurement scales: Interval



- represent quantitative data
- the difference between values is meaningful and consistent across the scale.
- no true zero point, meaning zero does not indicate the absence of the quantity being measured.
Example: *years, Temperature (in Celsius or Fahrenheit)*
- can't make statements like "twice as much"

Quantitative Thematic Maps — Measurement scales: **Ratio**

- Ratio: Ratio-scale magnitudes are absolute, and have a known **starting point of zero**.
- Ratio, proportion, percentage (proportion \times 100)
- These attributes support **arithmetic operations** (multiplication and division, e.g., "twice as much" or "half as much").
-Examples: age of structure, square footage of a parcel, assessed value of a parcel of land, road width in feet, air quality measurement.

Quantitative Thematic Maps — Measurement scales: Quiz

Each scale is represented once in the list below.

- Favorite candy bar
- Weight of luggage
- Year of your birth
- Egg size (small, medium, large, extra large, jumbo)
- Military rank
- Number of children in a family
- Jersey numbers for a football team

Quantitative Thematic Maps — Measurement scales: Quiz

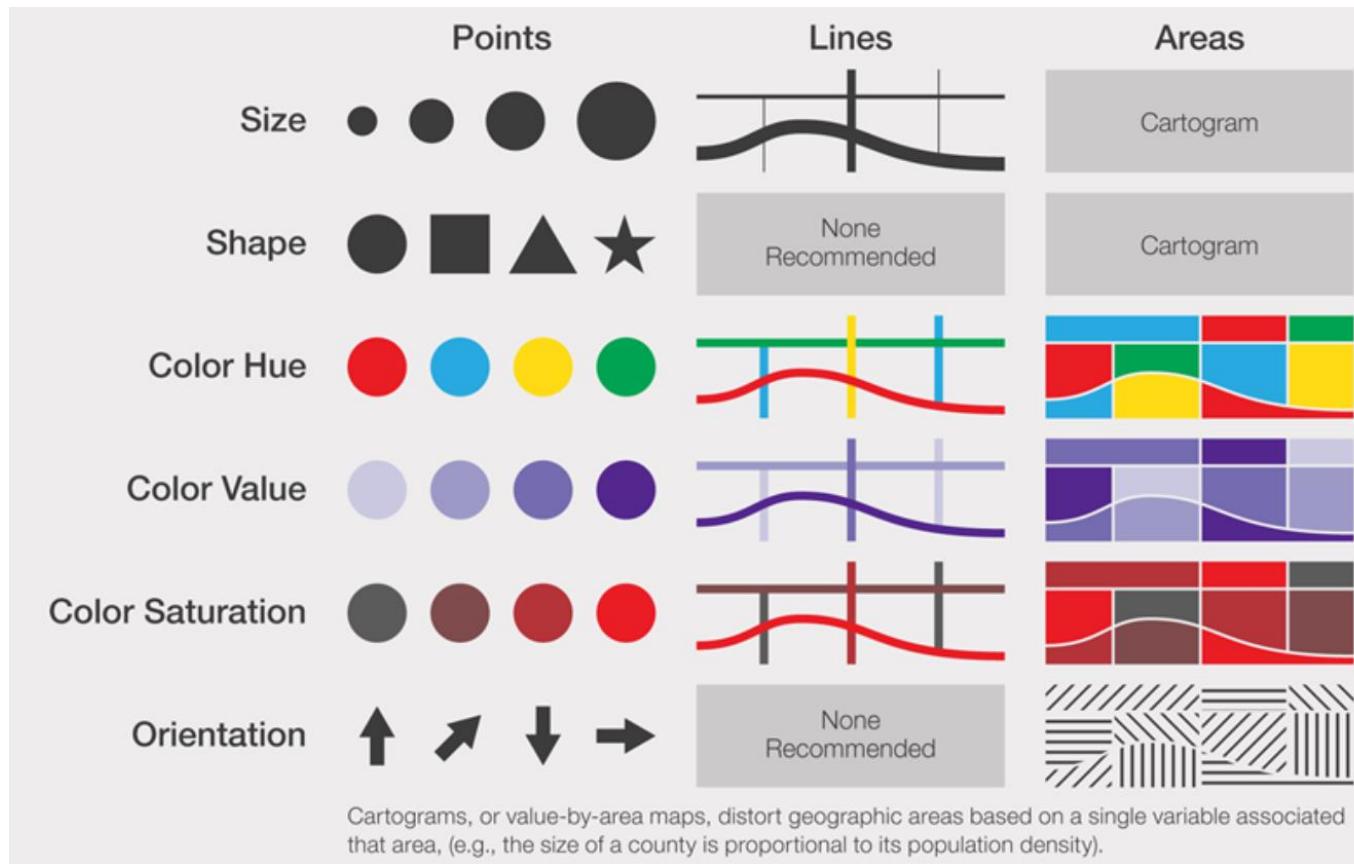
Each scale is represented once in the list below.

- Favorite candy bar - Nominal (N) Different candy bars are just categories with no ranking.
- Weight of luggage - Ratio (R). quantitative data with a true zero point
- Year of your birth - Interval (I). the differences between values are meaningful, but there is no true zero (e.g., 0 AD is not the absence of time)
- Egg size (small, medium, large, extra large, jumbo) - Ordinal (O).
- Military rank - Ordinal (O). the differences between ranks are not quantifiable
- Number of children in a family - Ratio (R). quantitative data with a true zero
- Jersey numbers for a football team - Nominal (N). numbers are just labels

Symbology

Common Parameters (visual variables) of Symbols

Color Theory



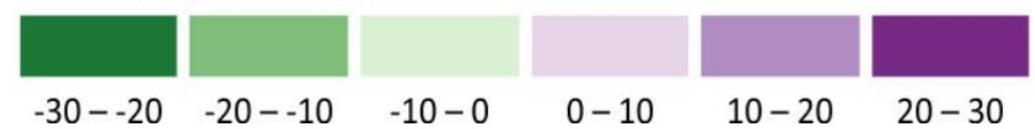
Symbology - Color Scheme

- *Sequential values of a single hue:* continuous variable such as mean height, population, etc. Conventionally, darker colors are used to display higher values.
- *Diverging hues:* good way of displaying bi-polar datasets (good or bad, negative or positive).

Urban population (% of total population)



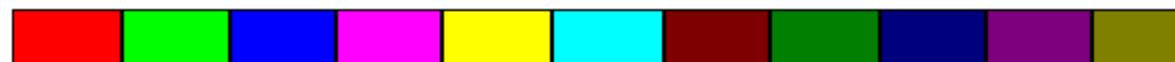
Urban population growth (in %)



- *Binary:* present or absent, above or below a threshold, forest or farmland

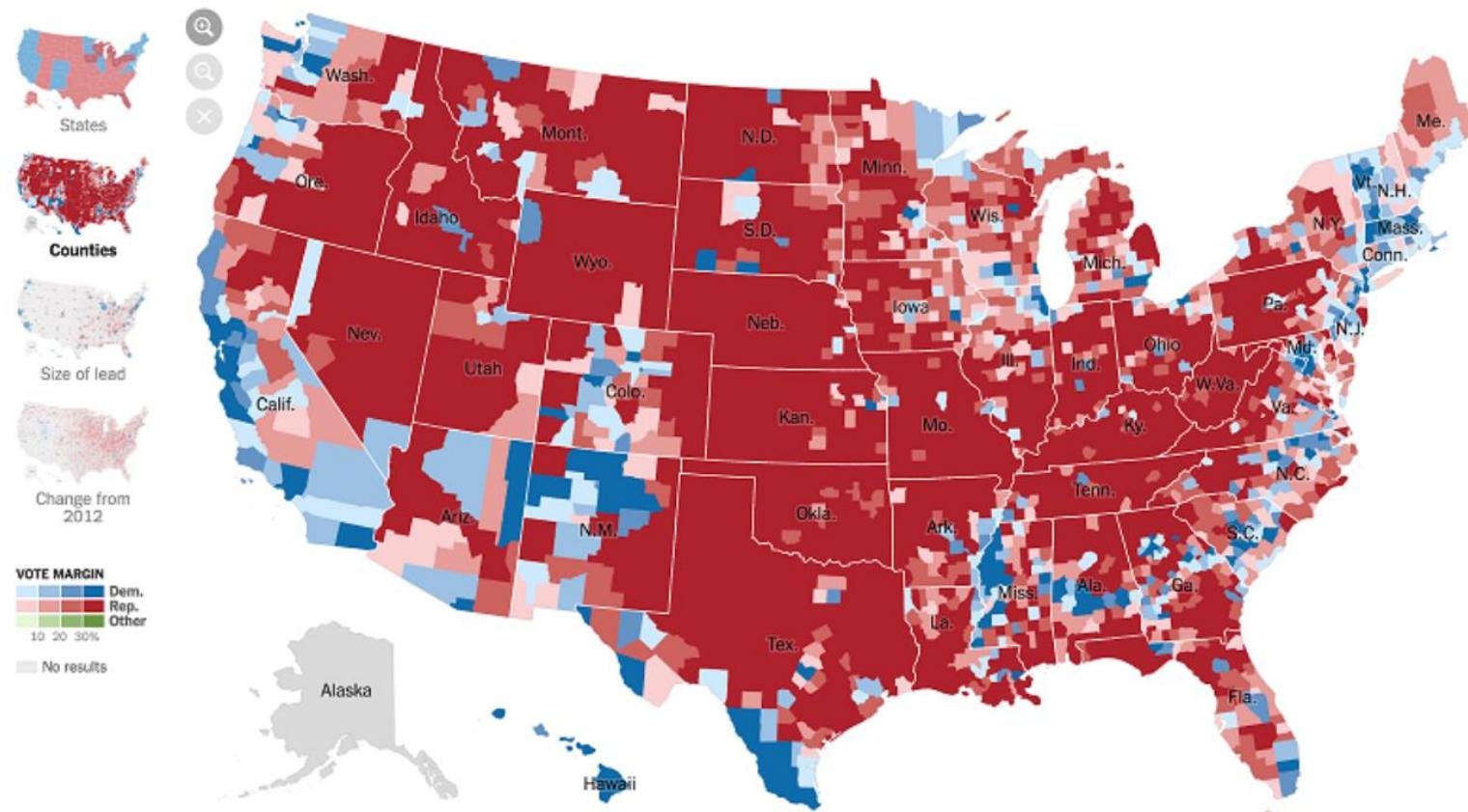


- *Individual values*



Data Normalization

Data Normalization and classification should be considered in all types of aggregated mapping



Choropleth Map - 2016 U.S. Presidential Election. Source: New York Times, 2016.

Data Normalization

- Data normalization is the process of taking enumerated data and attempting to remove biases and misleading messages that are founded in differences between areas (enumeration units).
- Two general types of normalization: **statistical normalization** and visual normalization (e.g., cartogram: purposefully exaggeration).

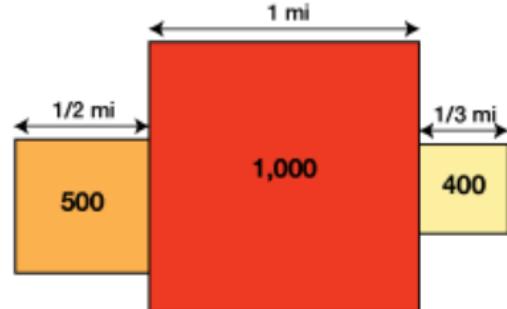
Data Normalization — Statistical normalization

Some common types of normalization:

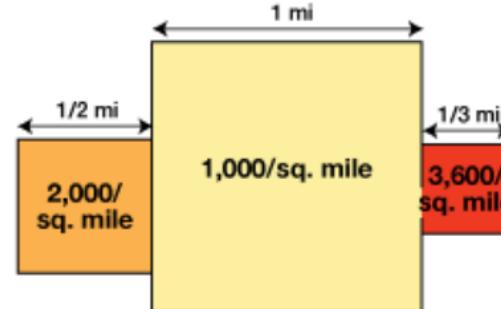
- Normalization by unit area
- Normalization by relevant population
 - Remember: data are normalized against the same universe of values from which the enumerated phenomena was measured.
- Normalization by summary value across units
- Normalization by a Prior Timestamp (i.e. Percent Change)

Data Normalization — Statistical normalization

- Normalization by unit area (Density)



Total Number of Persons
(Total Persons)

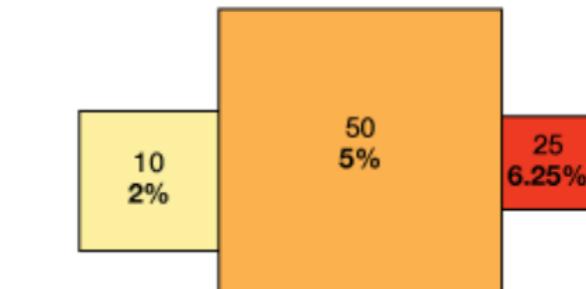


Population Density
(Persons/Square Mile)

- Normalization by relevant population (Rate)



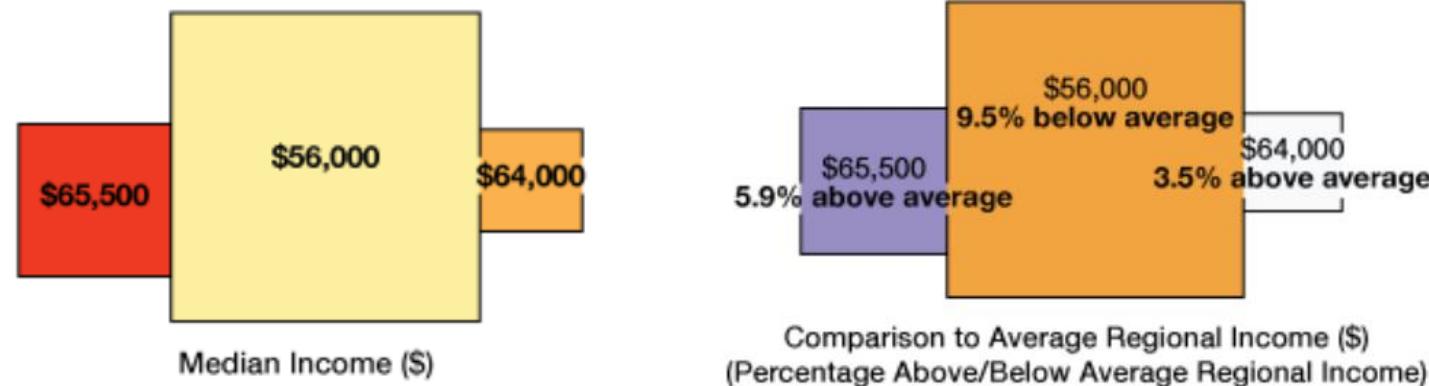
Number of Persons in Labor Force
(Total Persons in Labor Force)



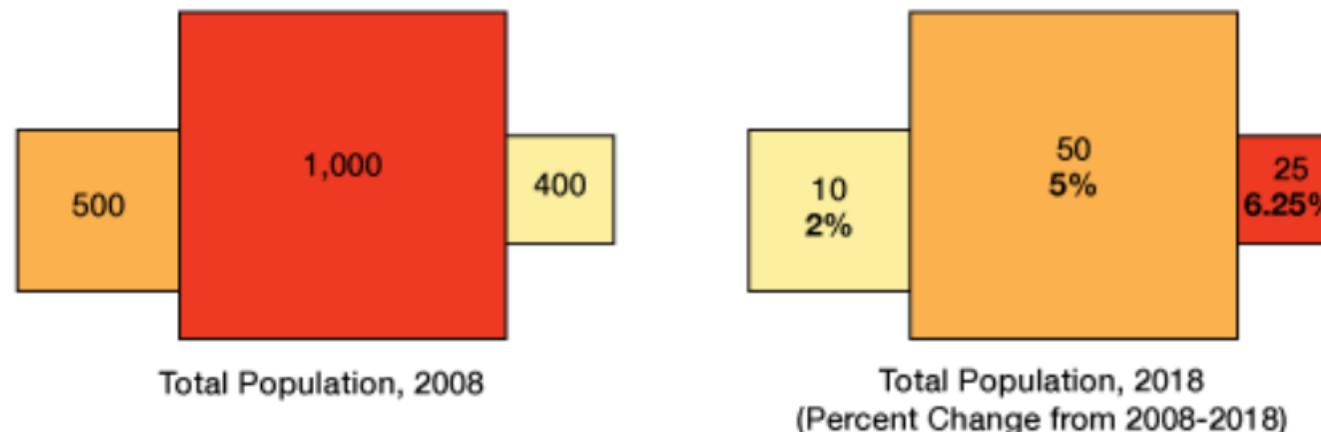
Persons Unemployed and Percentage Unemployed
(Persons in Labor Force Seeking Work/
Total Persons in Labor Force)

Data Normalization — Statistical normalization

- Normalization by Summary Value across all Units (i.e. Above/Below Average, Standard Deviation).

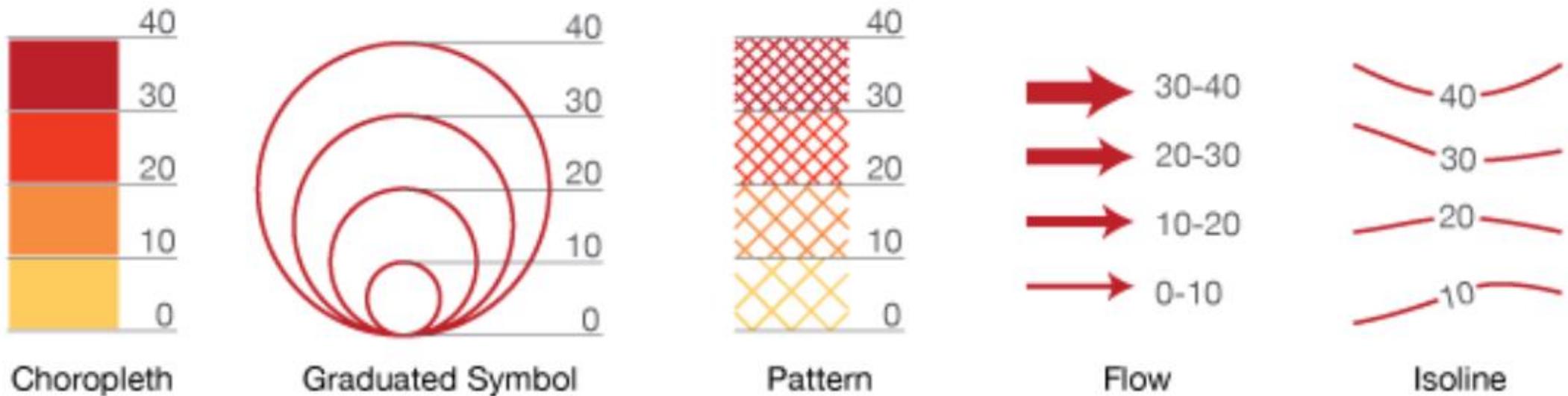


- Normalization by a Prior Timestamp (i.e. Percent Change).



Data Classification

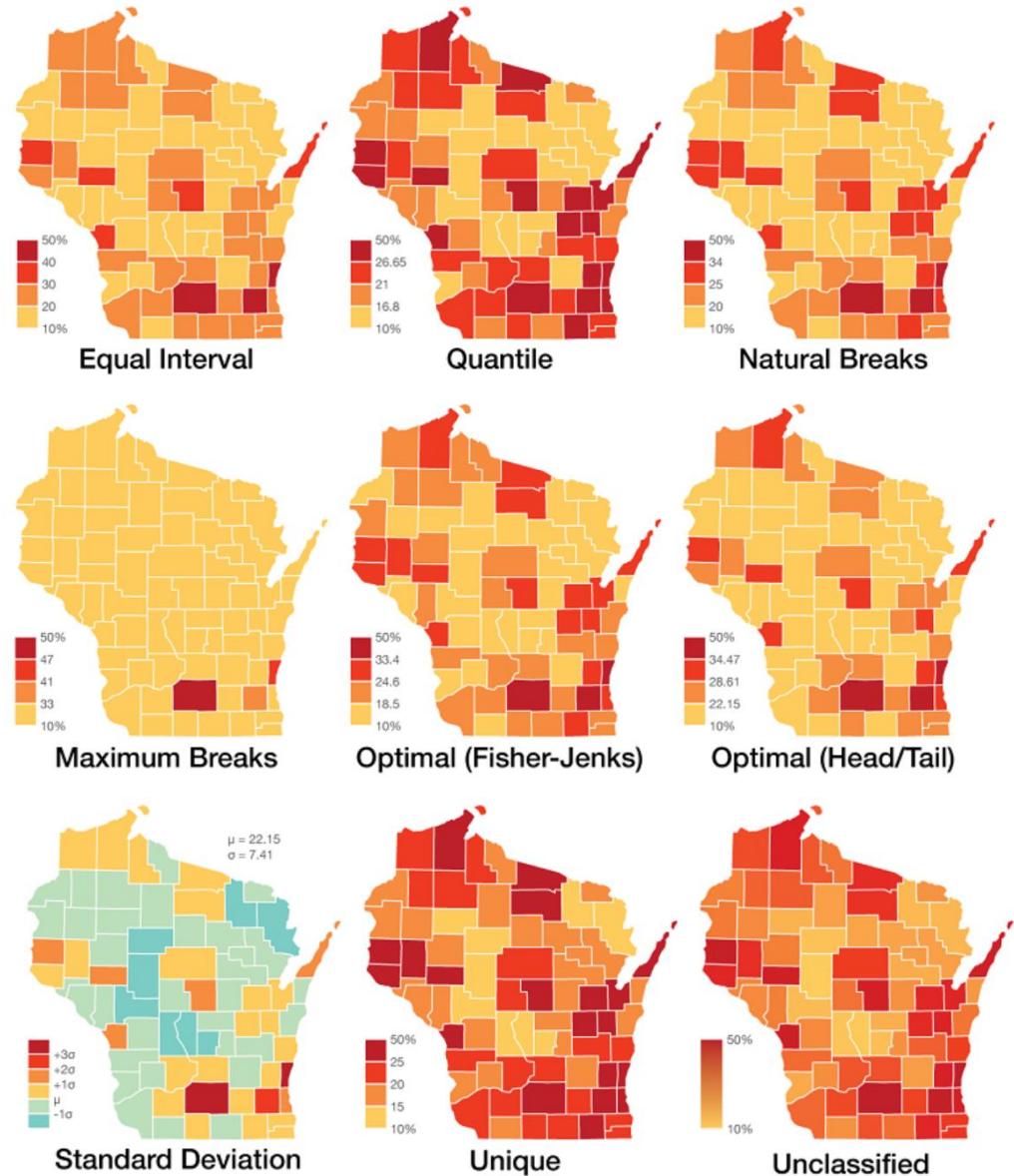
Classification is an intellectual process that groups similar phenomena to gain relative simplicity in communication and user interpretation (Robinson et al., 1995; Longley et al., 2015).



Data classification across different types of maps.

Data Classification

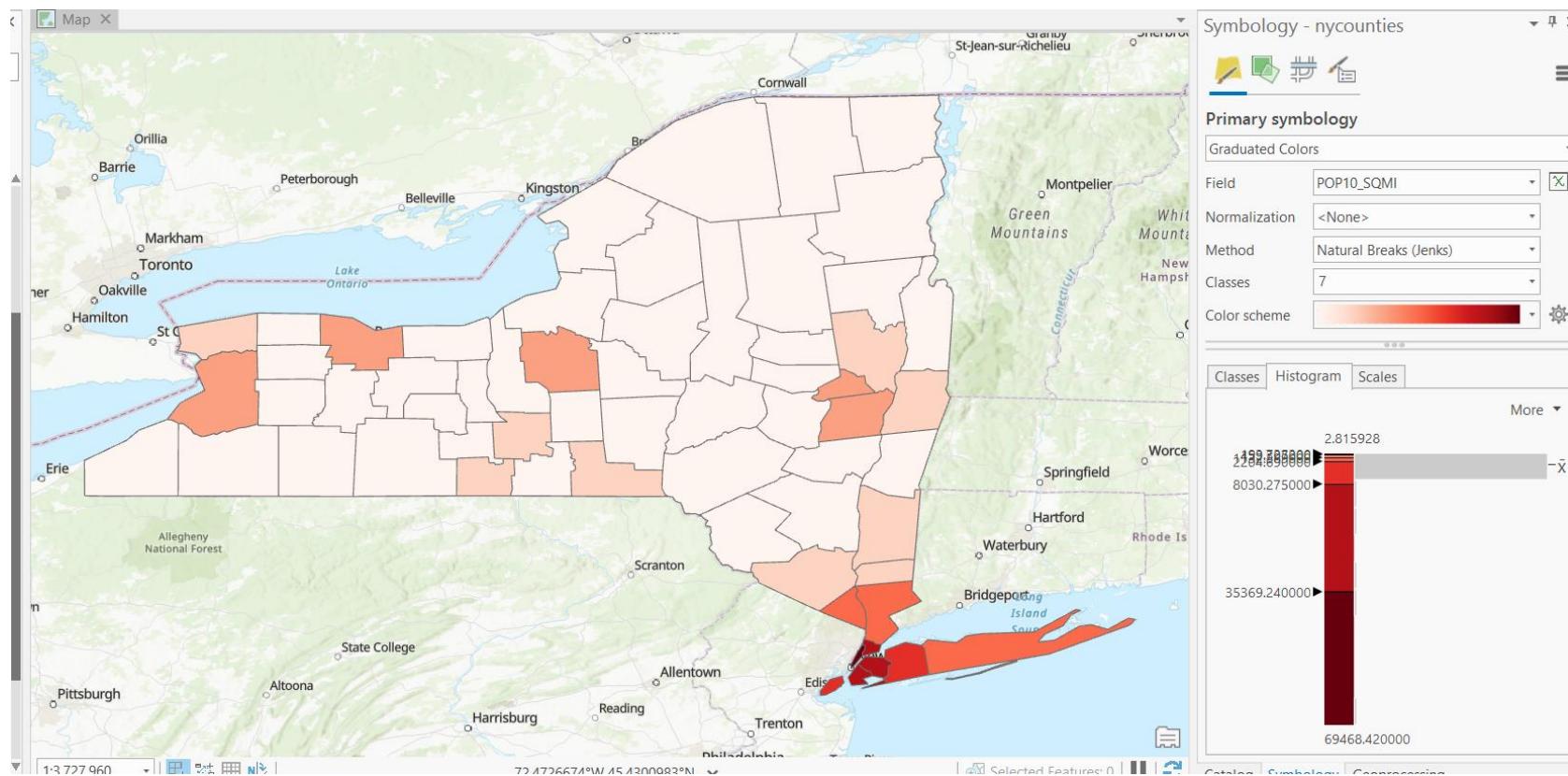
- Changing a classification scheme can modify the message of the map entirely
- Some practical guidance:
 - Try different classification scheme to see which fit the underlying data distribution and/or your purpose.
 - thematic map could be accompanied by an explanatory note stating that the data classification depicted on the map.



Common classification methods (Percentage of residents over the age of 25 in Wisconsin that possess a Bachelor's degree or higher in Wisconsin in 2016 by county).

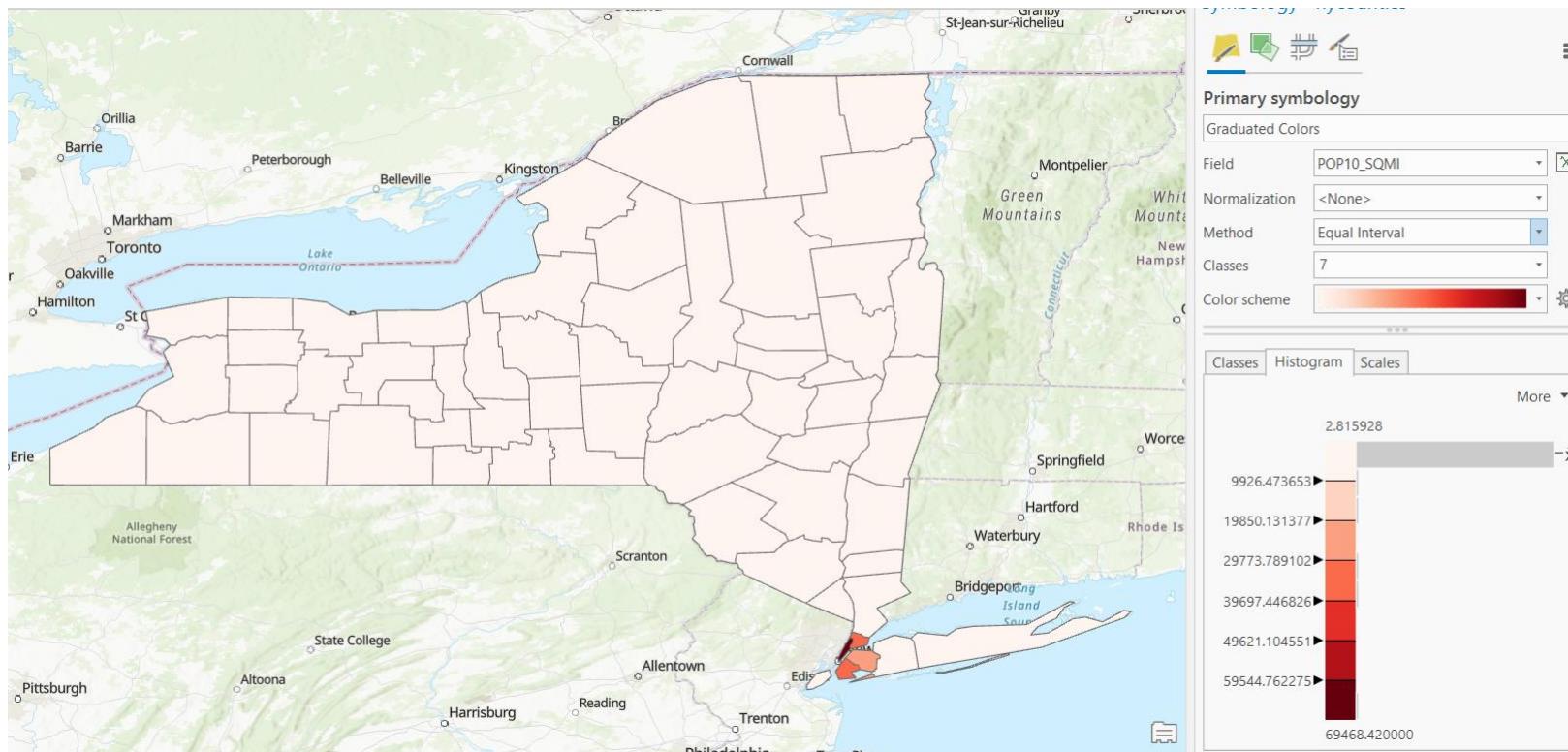
Data Classification — Natural Breaks/Jenks

- Uses Jenk's optimization algorithm - minimizes the variance within the groups while maximizing the variance between the groups.
- Identifies break points by looking for groupings and patterns inherent in the data.
- Suggested use: Clustered and Skewed data..



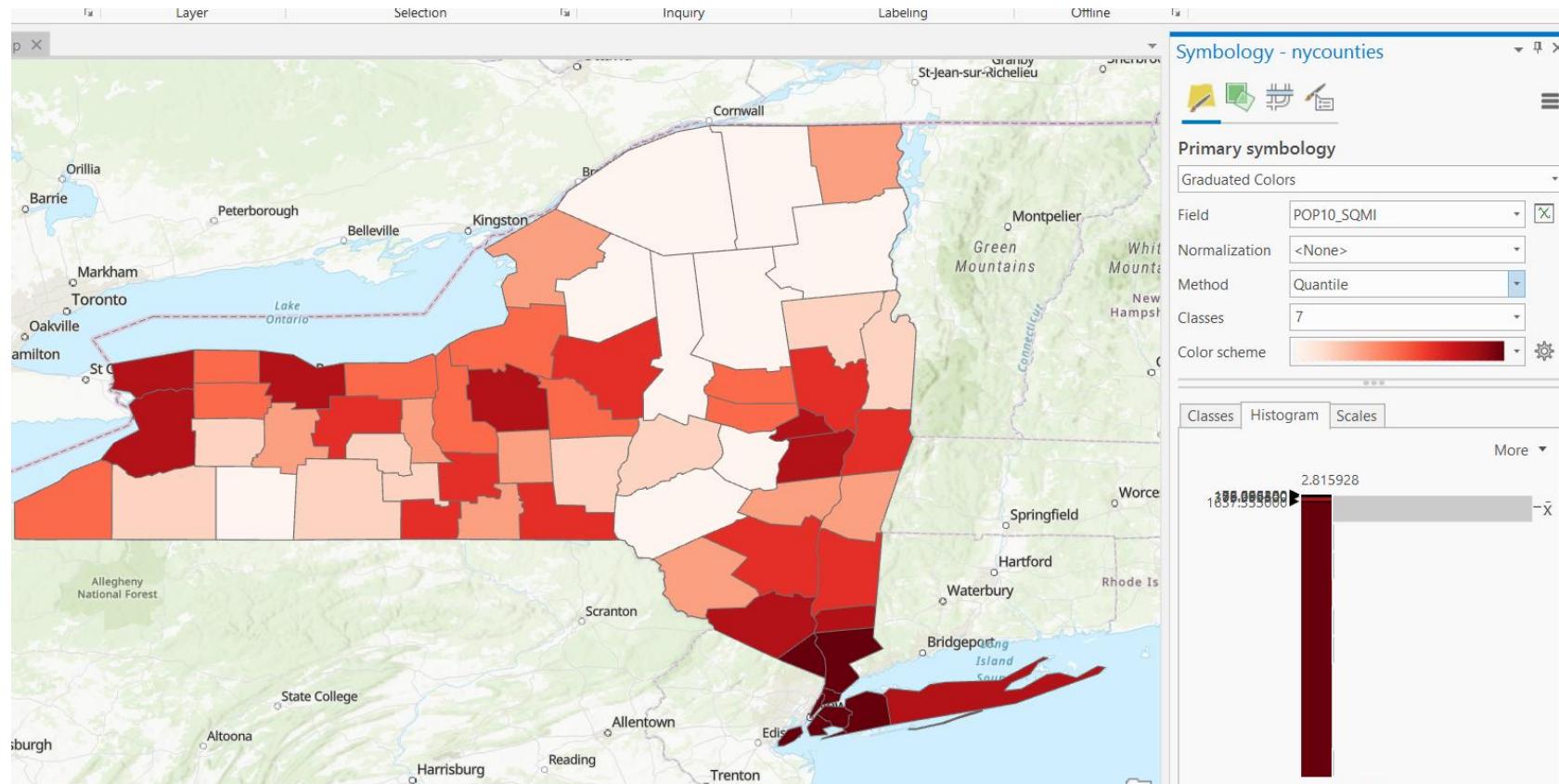
Data Classification — Equal Interval

- Creates classes of equal value ranges. The entire range of data values ($\max - \min$) is divided equally into predetermined number of classes. The value ranges in each class are equal in size.
- For example, if the range was 100, each class size has a value of 20 for the 5 classes. e.g. 0-20, 21-40, 41-60, 61-80, 81-100).
- There is no consideration for how many records go into each class.
- Suggested use:** Uniformly distributed data with familiar data ranges.



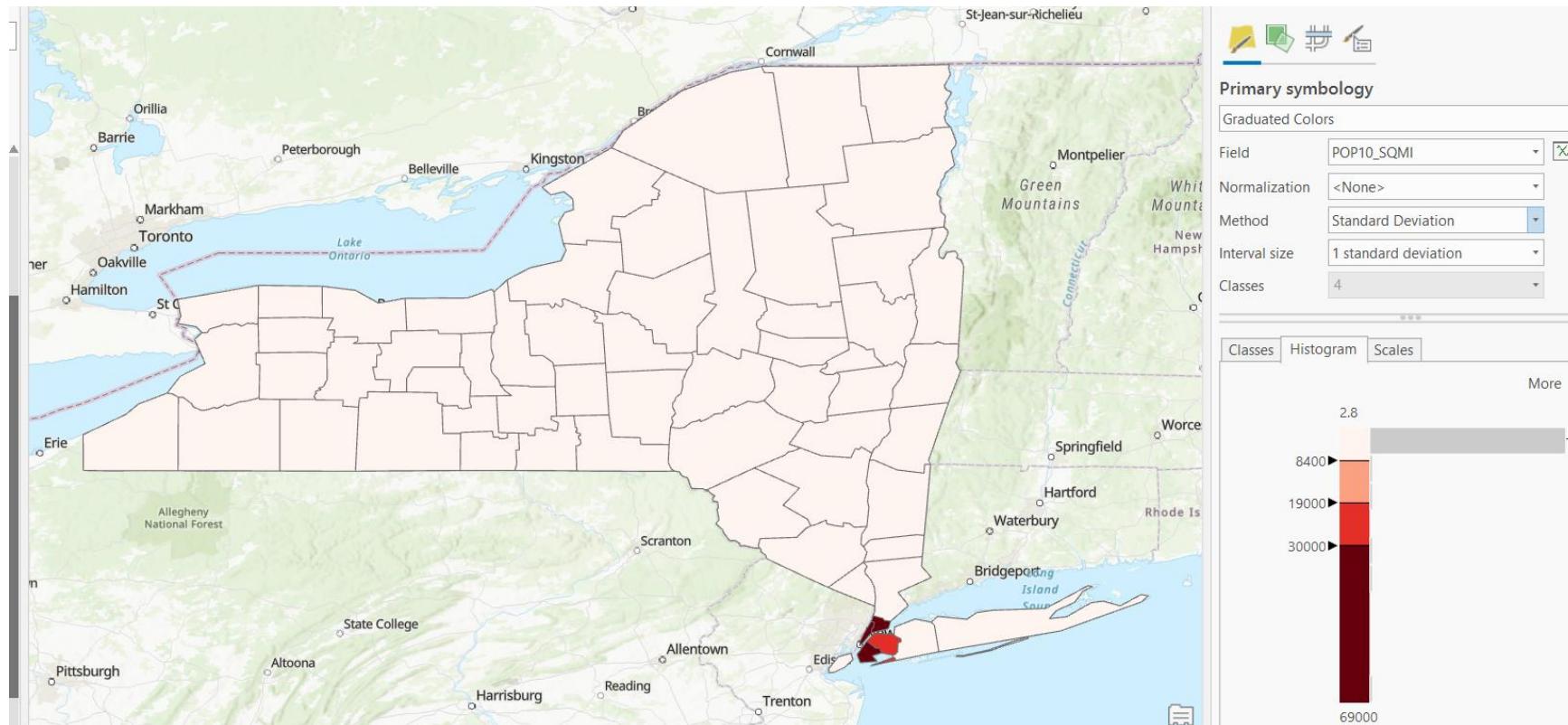
Data Classification — Quantile

- Each class contains the same number of features.
- Within class variations can be large, observations of similar value can be placed in different classes – misleading visualization
- Suggested use: Evenly(Close to uniformly) distributed data



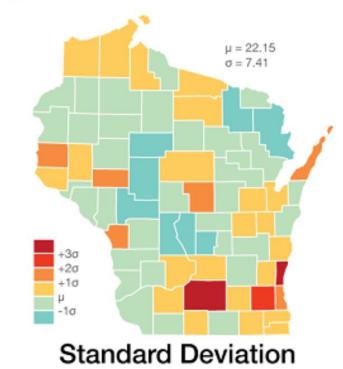
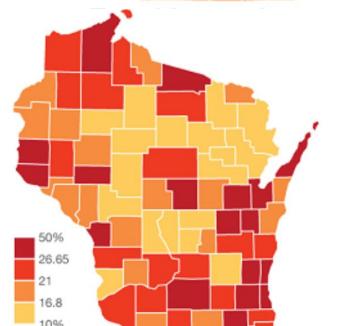
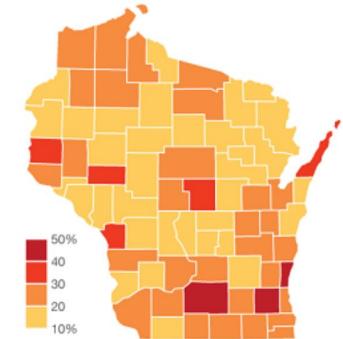
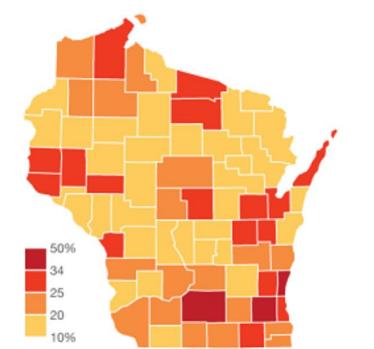
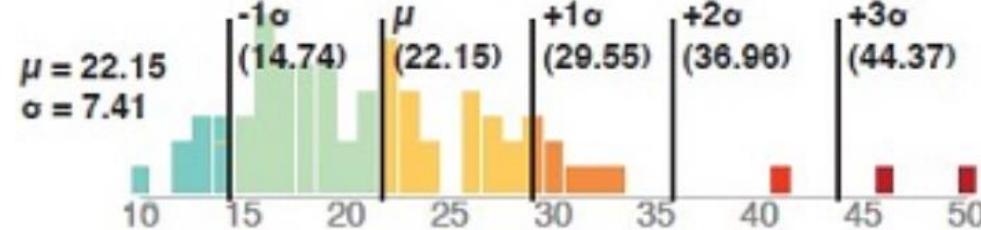
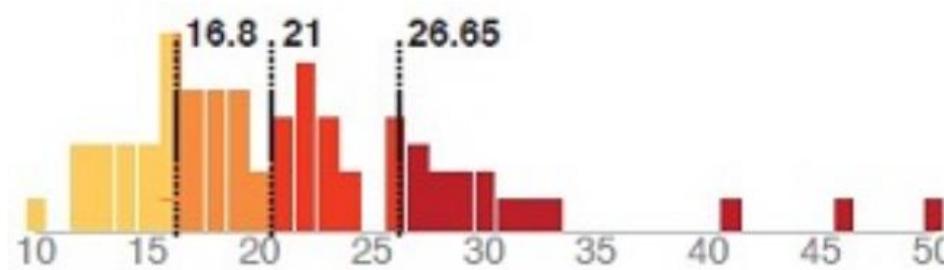
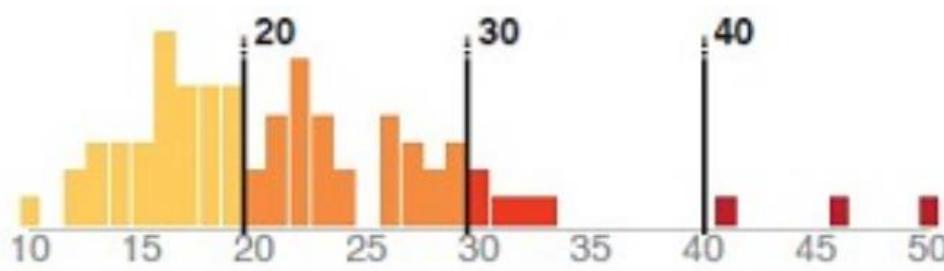
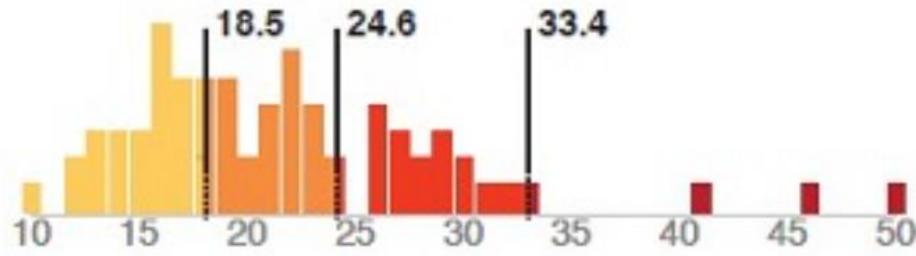
Data Classification — Standard Deviation

- Assignment of classes is based on the distance of an observation from the mean of all the values.
- ArcGIS first finds the mean value, calculates the standard deviation, and then places the class breaks above and below the mean at e.g., 1, .5, or .25 standard deviations.
- ArcGIS will aggregate outliers (beyond three standard deviations from the mean) into two classes: '>3 Std Dev' and '<3 Std Dev'.
- **Suggested use:** Normally distributed data



Exercise - Data Classification?

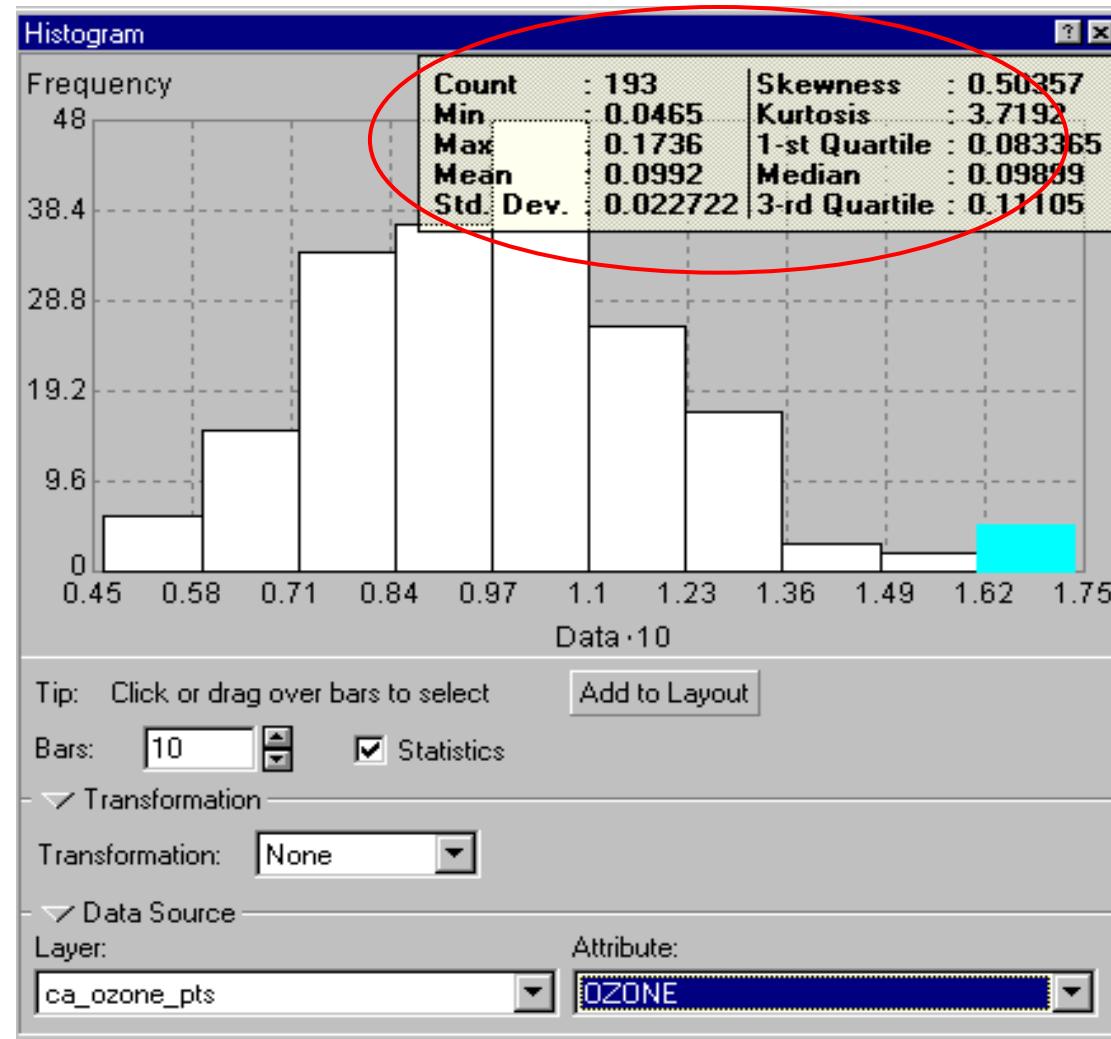
Quantile?
Natural Break?
Equal interval?
Standard Deviation?



Data Classification — How many classes?

- Thematic maps use at least 5 and as many as 7 groupings, or classes, of data
- Studies have shown that the human eye cannot differentiate among more than eleven classes, and even maps with seven classes can lead to confusion.

Exploratory spatial data analysis



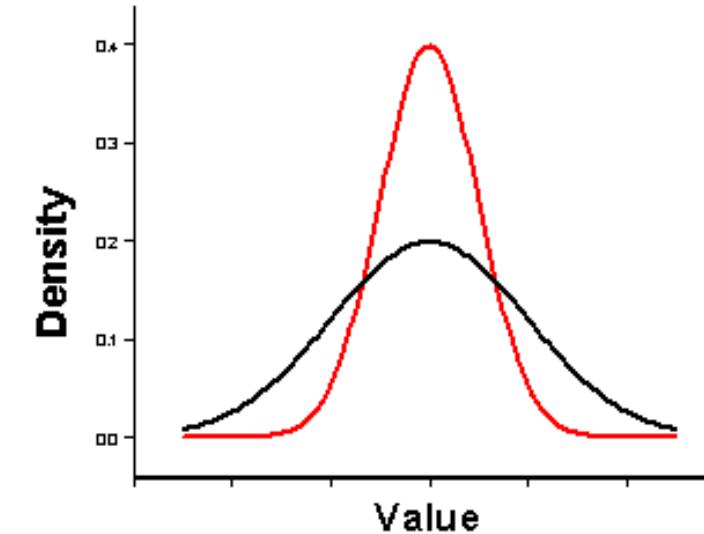
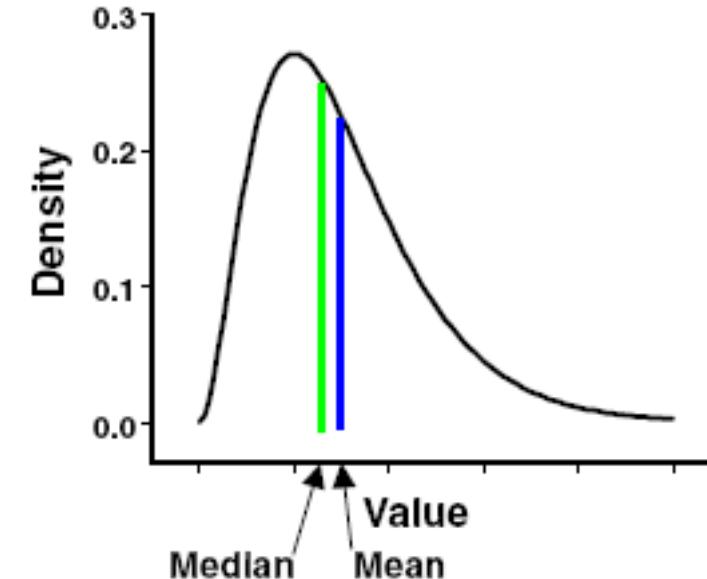
Exploratory spatial data analysis: Summary statistics

The important features of a distribution can be summarized by statistics that describe its **location, spread, and shape**.

Measures of location: where center and other parts of the distribution lie (mean, median)

Measures of spread: The spread of points around the mean value

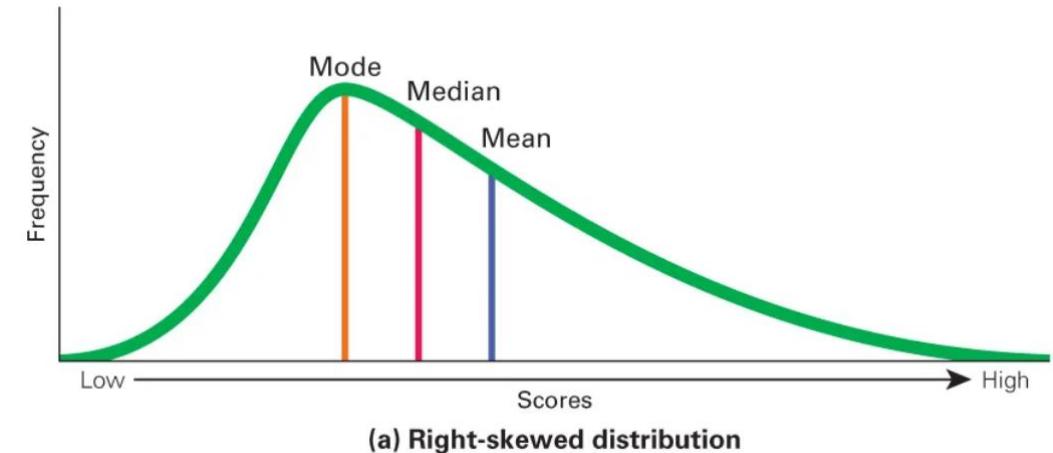
- Variance - average squared deviation of all values from the mean,
- Standard deviation - square root of the variance.



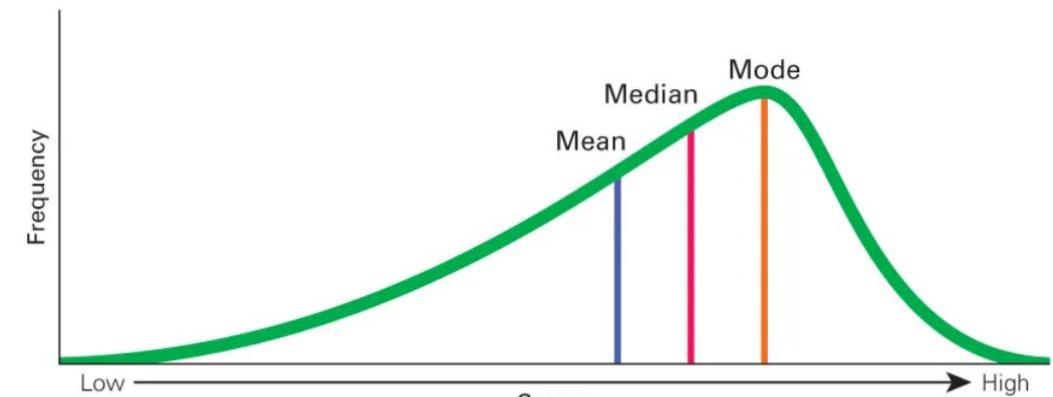
Exploratory spatial data analysis: Summary statistics

Measures of shape

- **Skewness** is a measure of the symmetry of a distribution. For symmetric distributions, the coefficient of skewness is zero.
- If a distribution has a long right tail of large values, it is positively skewed (right-skewed),
- if it has a long-left tail of small values, it is negatively skewed (left-skewed).



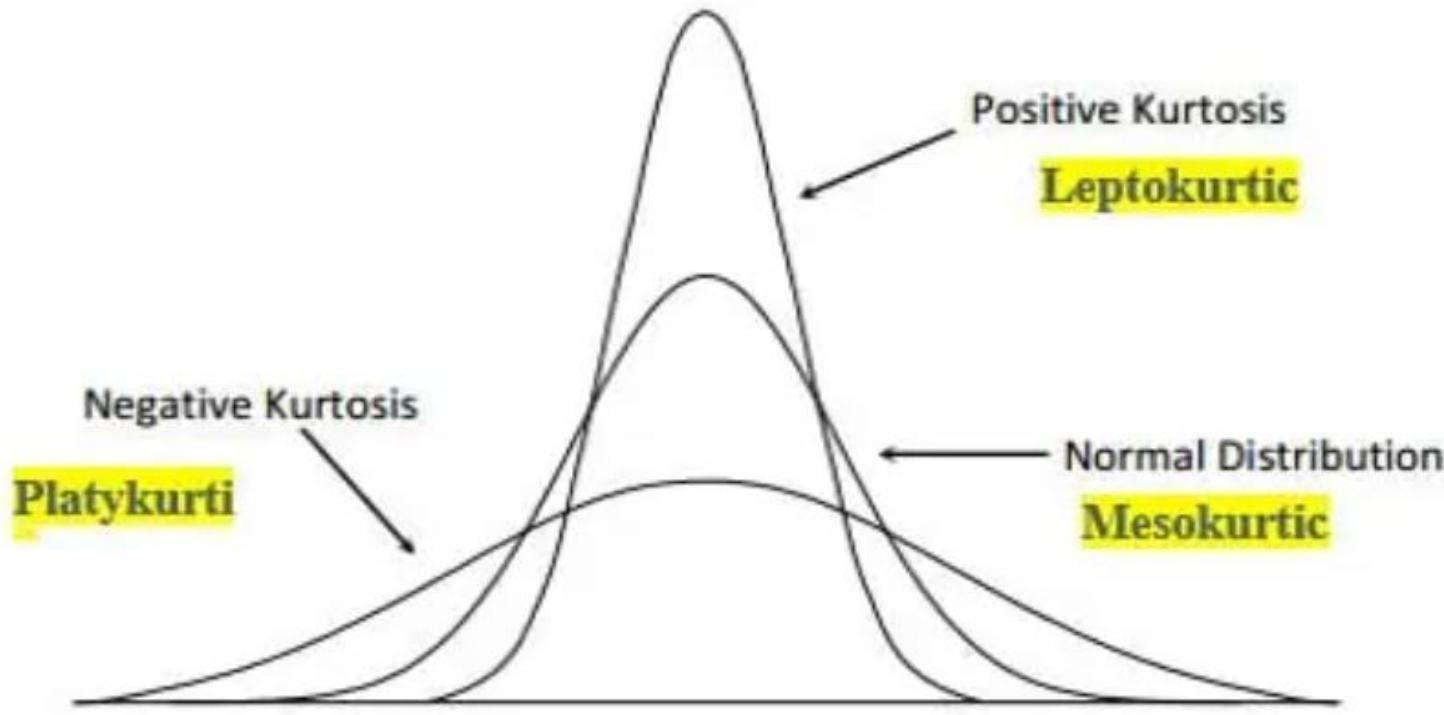
(a) Right-skewed distribution



(b) Left-skewed distribution

https://www.macmillanhighered.com/BrainHoney/Resource/22292/digital_first_content/trunk/test/griggs4e/asset/ch01/c01_fig05.jpg

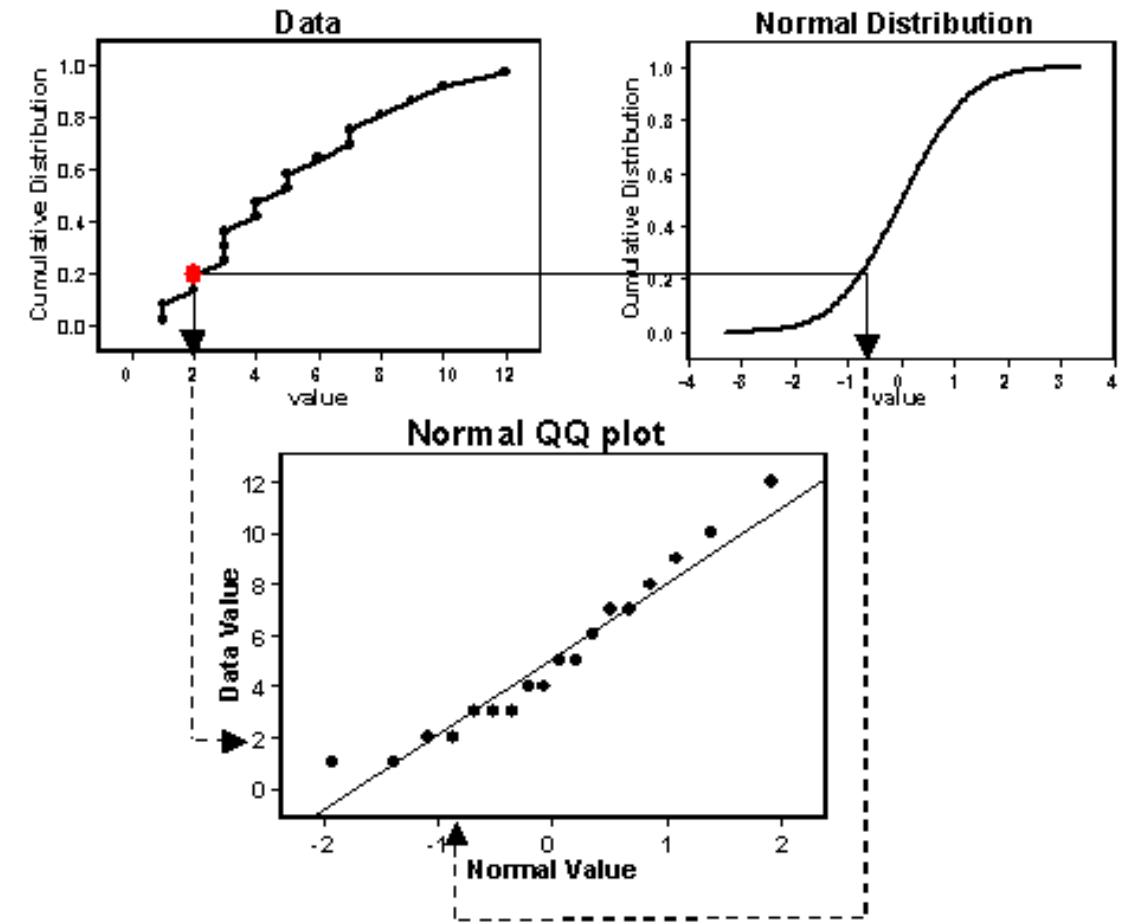
Exploratory spatial data analysis: Summary statistics



Kurtosis characterizes the relative peakedness or flatness of a distribution compared to the normal distribution. Positive kurtosis indicates a relatively peaked distribution. Negative kurtosis indicates a relatively flat distribution. Normal distributions produce a kurtosis statistic of about zero.

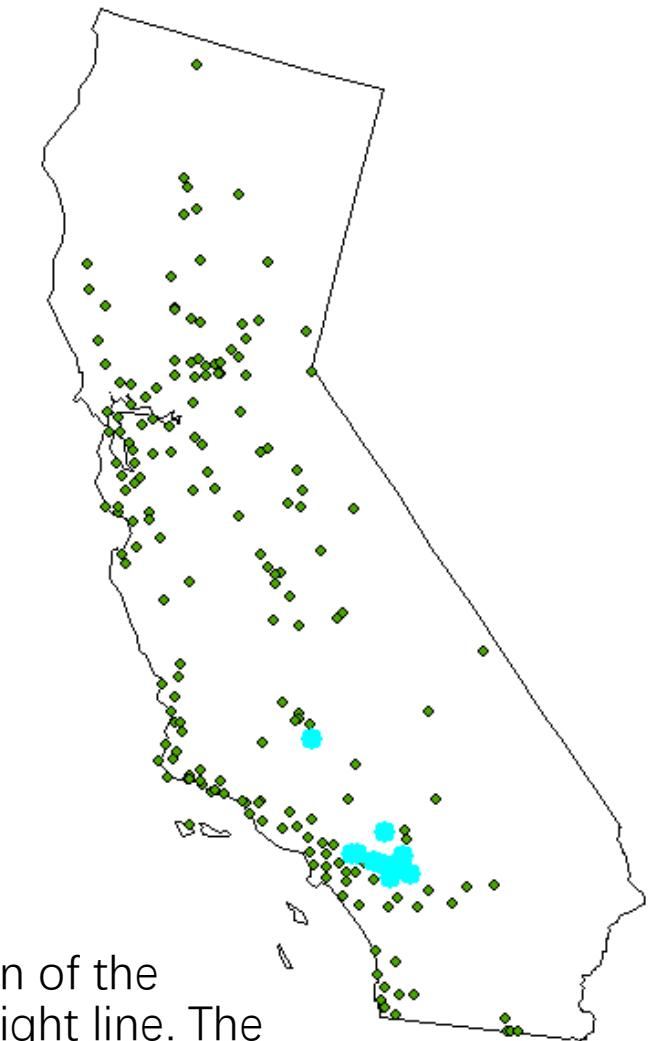
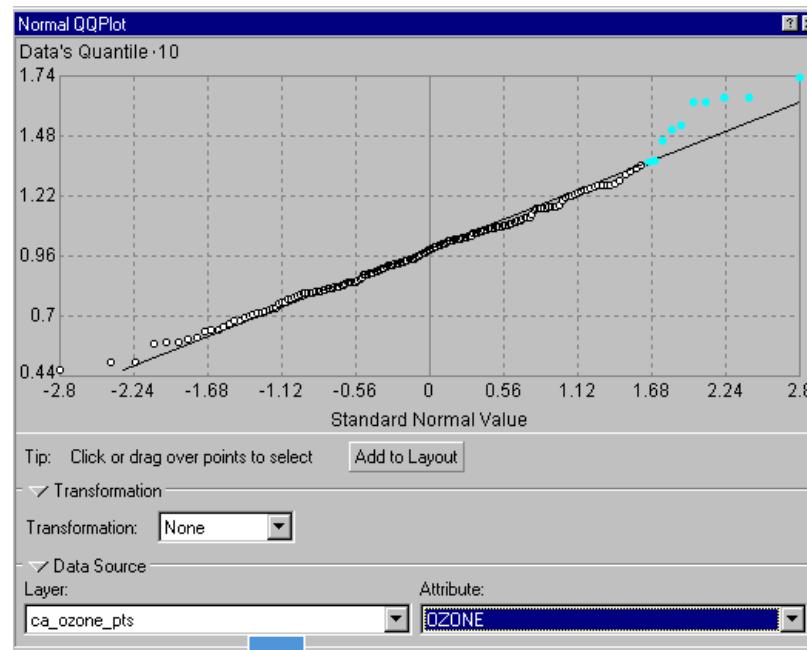
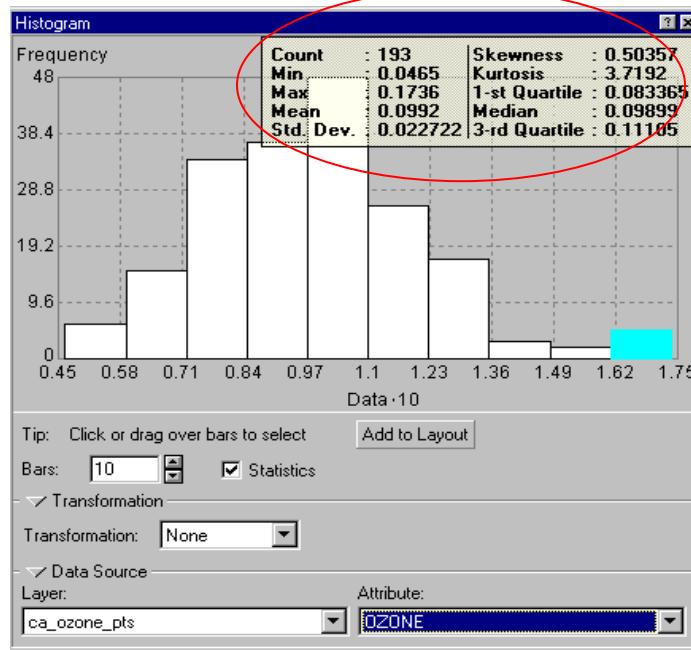
Exploratory spatial data analysis: Normal QQ plot

- A Q–Q plot is used to compare the shapes of distributions by plotting their quantiles against each other.
- 2 distributions are plotted relative to each other. If the two distributions being compared are similar, the points in the Q–Q plot will approximately lie on the line $y = x$.
 - Normal QQ Plot is created by plotting data values with the value of a standard normal where their cumulative distributions are equal
 - General QQ Plot is used to assess the similarity of the distribution of two datasets.



Normal QQ Plot

Normal QQ Plot



- A standard normal distribution are plotted on the x-axis, and the distribution of the dataset are plotted on the y-axis. You can see that the plot is close to a straight line. The main departure from this line occurs at high values of ozone concentration. These have been selected and are then highlighted in the data view.

References

- University Consortium for GIS
- Maantay, J., & Ziegler, J. (2006). GIS for the Urban Environment. Redlands, CA: Esri Press.