L8

Design for Maps II
Data & map types, classifying/representing data, color optimization

GEP 299/680
Adam Jessup
10.30.2018

Data types

When we think about 'data' we can think of it in the following ways…

- By **dimension**–point, line, area, volume (text)
- By *continuity*–discrete versus continuous
- By <u>level</u>–nominal, ordinal, interval, ratio (i.e. the Stevens Scale)

Data types

When we think about 'data' we can think of it in the following ways...

- By **dimension**–point, line, area, volume (text)
- By *continuity*–discrete versus continuous
- By <u>level</u>–nominal, ordinal, interval, ratio (i.e. the Stevens Scale)

Most all data will require <u>classification</u> and <u>normalization</u> prior to representation (i.e. mapping)

➔ E.g. Collect data at points, count the # of points in polygon, normalize by area, then classify for representation in a choropleth map

Data types/map types matrix

Data Types				
	Point	**Line**	**Area**	**Volume**
Nominal	City	Road	Name of unit	Precipitation or soil type
Ordinal	Large City	Major Road	Rich county	Heavy precipitation, good soil
Interval	Total Population	Traffic flow	Per capita income	Precipitation in mm, cation exchange
Ratio				

Map Types				
	Point	**Line**	**Area**	**Volume**
Nominal	Dot map	Network map	Colored area map	Freely colored map
Ordinal	Symbol map	Ordered network map	Ordered colored map	Ordered chromatic map
Interval	Graduated symbol map	Flow map	Choropleth map	Contour map
Ratio				

Discrete vs Continuous

Discrete

- Much geographic data relates to specific points, lines, and/or polygons (i.e. areas)

- Discrete data assumes a uniform distribution of data values within and different among the aggregation units

- Units, especially in the case of areas, are often merely the way the data is aggregated (e.g. counties, census tracts, etc.)

Discrete vs Continuous

Discrete

- Much geographic data relates to specific points, lines, and/or polygons (i.e. areas)

- Discrete data assumes a uniform distribution of data values within and different among the aggregation units

- Units, especially in the case of areas, are often merely the way the data is aggregated (e.g. counties, census tracts, etc.)

Q: What's the most common map type for visualizing discrete data?

Discrete vs Continuous

Discrete

- Much geographic data relates to specific points, lines, and/or polygons (i.e. areas)

- Discrete data assumes a uniform distribution of data values within and different among the aggregation units

- Units, especially in the case of areas, are often merely the way the data is aggregated (e.g. counties, census tracts, etc.)

Q: What's the most common map type for visualizing discrete data?

Choropleths!

Discrete vs Continuous

Choropleth >> from the Greek

Chóros (space/area/region) + Plíthos (multitude/many) = 'many regions'

Discrete vs Continuous

Choropleth >> from the Greek

Chóros (space/area/region) + Plíthos (multitude/many) = 'many regions'

...a thematic map in which areas are shaded or patterned in <u>proportion</u> to the measurement of the statistical variable being displayed on the map, such as <u>population</u> <u>density</u> or *per-capita income*

Discrete vs Continuous

Choropleth >> from the Greek

Chóros (space/area/region) + Plíthos (multitude/many) = 'many regions'

...a thematic map in which areas are shaded or patterned in <u>proportion</u> to the measurement of the statistical variable being displayed on the map, such as <u>population</u> <u>density</u> or *per-capita income*

...provide an easy way to visualize how a measurement varies across a geographic area or show the level of variability within a region

Discrete vs Continuous

2014-2015 Immunization Percentages
in California Child Care Facilities

% of children from reporting facilities
that are up to date on their vaccinations

< = 50%
51 - 60%
61 - 70%
71 - 80%
81 - 90%
91 - 100%
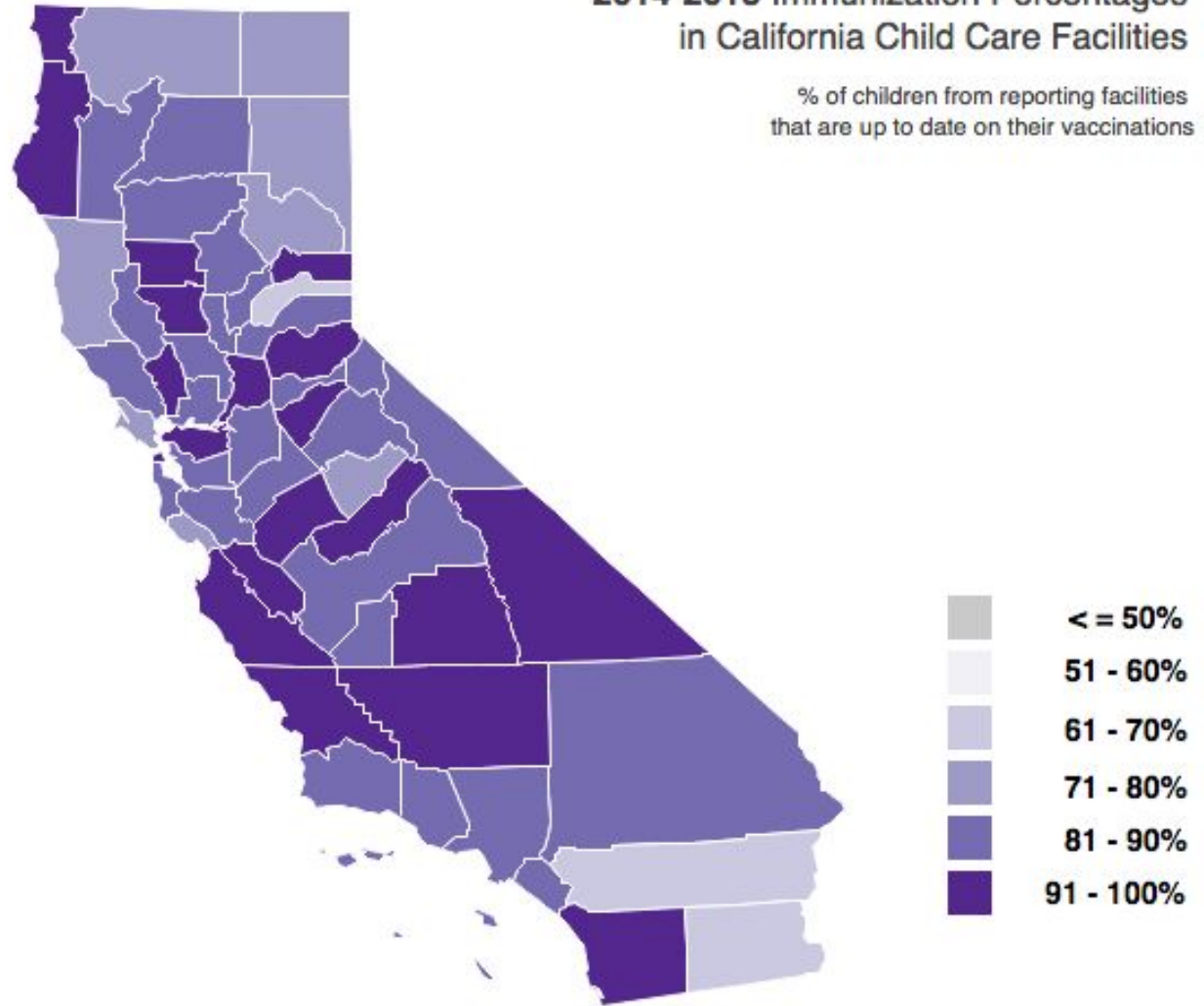
Source: CA Dept of Public Health

Powered by ZingChart

Select Demographic Variables for New Haven, CT

2 miles

Population Density per mi^2 (thousands)
- 25 - 50
- 14 - 25
- 7 - 14
- ≤ 7
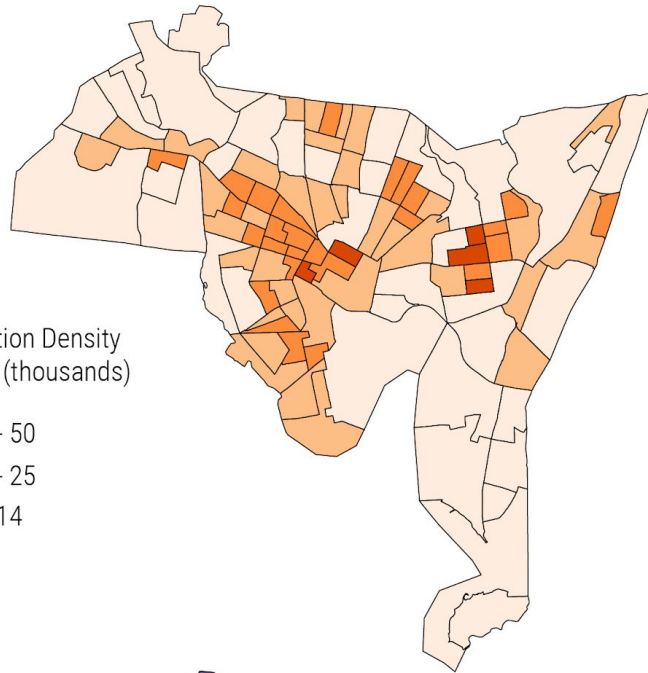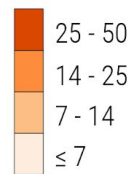
Minority (%)
- 81 - 100
- 61 - 80
- 31 - 60
- ≤ 30
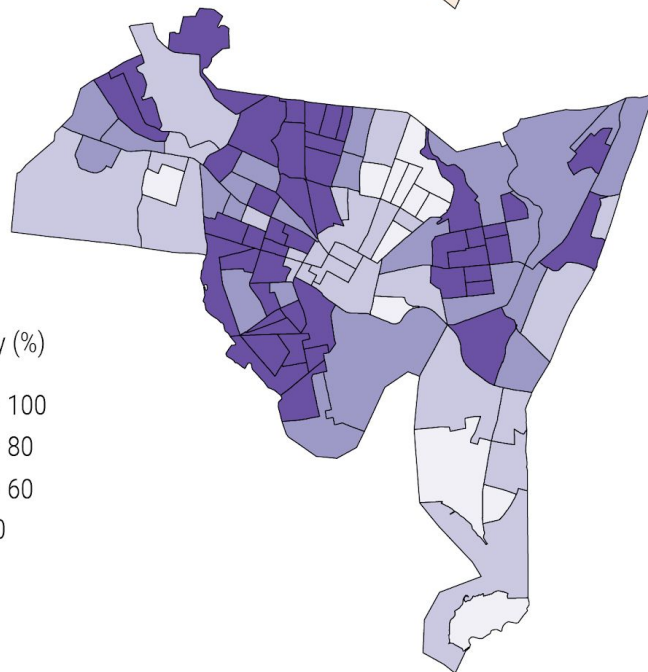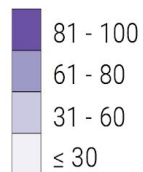
Female Householders with related children (%)
- 81 - 100
- 61 - 80
- 41 - 60
- ≤ 40
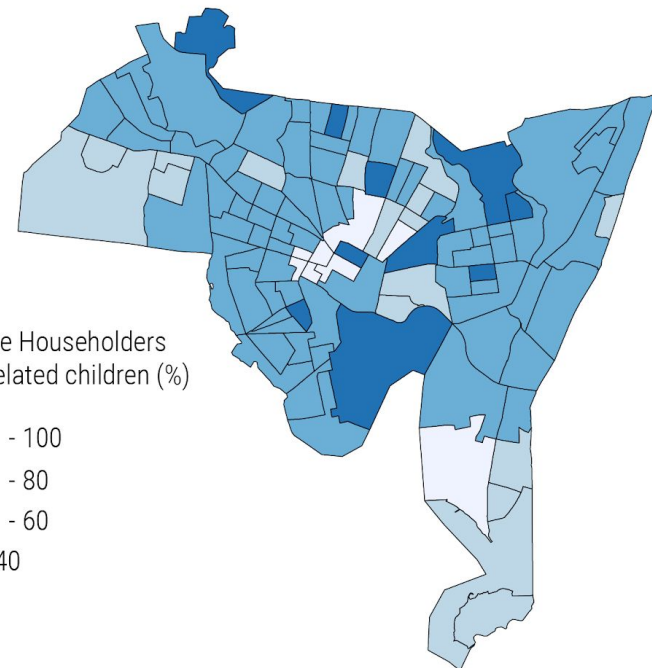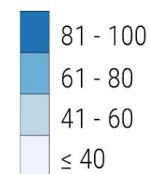
Discrete vs Continuous

Continuous

- Some geographic variables are measurable anywhere in space (e.g. air temperature/pressure)

- Creates a 'surface' or 'field'

- Can be visualized like topography (i.e. isolines and other terrain representation methods)

- Isolines often favored for sake of simplicity

Biomass in tons/hectare (2000)

Aboveground Woody Biomass (tons per hectare)

0   25   100   225   400
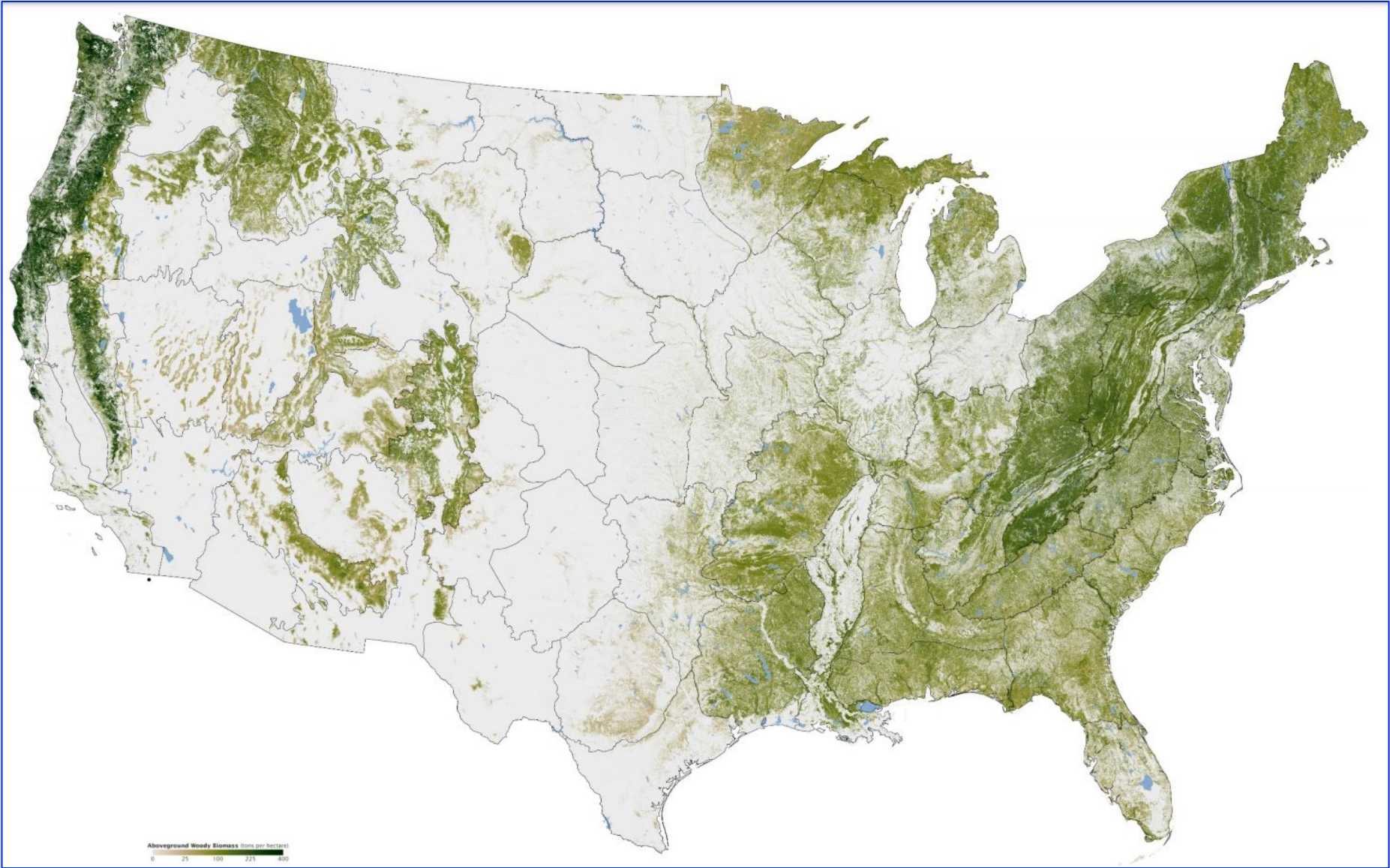
NZ Population Distribution

AUCKLAND

WELLINGTON

CHRISTCHURCH

DUNEDIN

2013 Census Resident Population

- < 500
- 500 - 10,000
- 10,000 - 50,000
- 50,000 - 100.000
- > 100,000
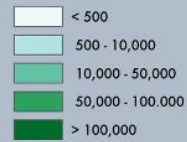
Data levels or levels of measurement

Q: Does quantitative data have 'levels' of measurement?

Data levels or levels of measurement

Q: Does quantitative data have 'levels' of measurement?
Indeed! (see slide #2)

From Stanley Smith Stevens (1946)

- **Nominal**
  - Has <u>name</u> or <u>class</u> only
- **Ordinal**
  - Has <u>rank</u> only
- **Ratio**
  - Has value on scale w/ <u>absolute</u> <u>zero</u> (e.g. Kelvin)
- **Interval**
  - Has value on <u>arbitrary</u> scale (e.g. Fahrenheit)

Nominal data and map types

- Relates to or existence of a class
- Importance of place names and legends
- Cartographic methods
  - **Point**–labels at at locations
  - **Lines**–network shown w/ symbols
  - **Areas** (polygon)–classes shown by color and pattern
- At the simplest data level, no real quantitative analysis possible

Line nominal map

Nominal area map

ETHNIC GROUPS

- Bantu
- Nilo-Hamitic
- Nilotic
- Hamitic
- Area of extremely complex tribal and ethnic mixture

KISII Tribal name

Ordinal data and map types

- Ordinal involves some kind of ranking
- One class or feature is 'above' or 'below' another
- Cartographic methods
  - **Point**–use symbol size, shape, and color
  - **Line**–different symbols, line weights, color
  - **Area**–color and/or pattern (legends often high, medium, low, or similar)

Average Gas Prices at
Selected Montana Towns

March 24-26, 2003

Average Price

- 1.599 - 1.68
- 1.68 - 1.71
- 1.71 - 1.74
- 1.74 - 1.79
- 1.79 - 1.85
- 1.85 - 1.969

Ordinal point map
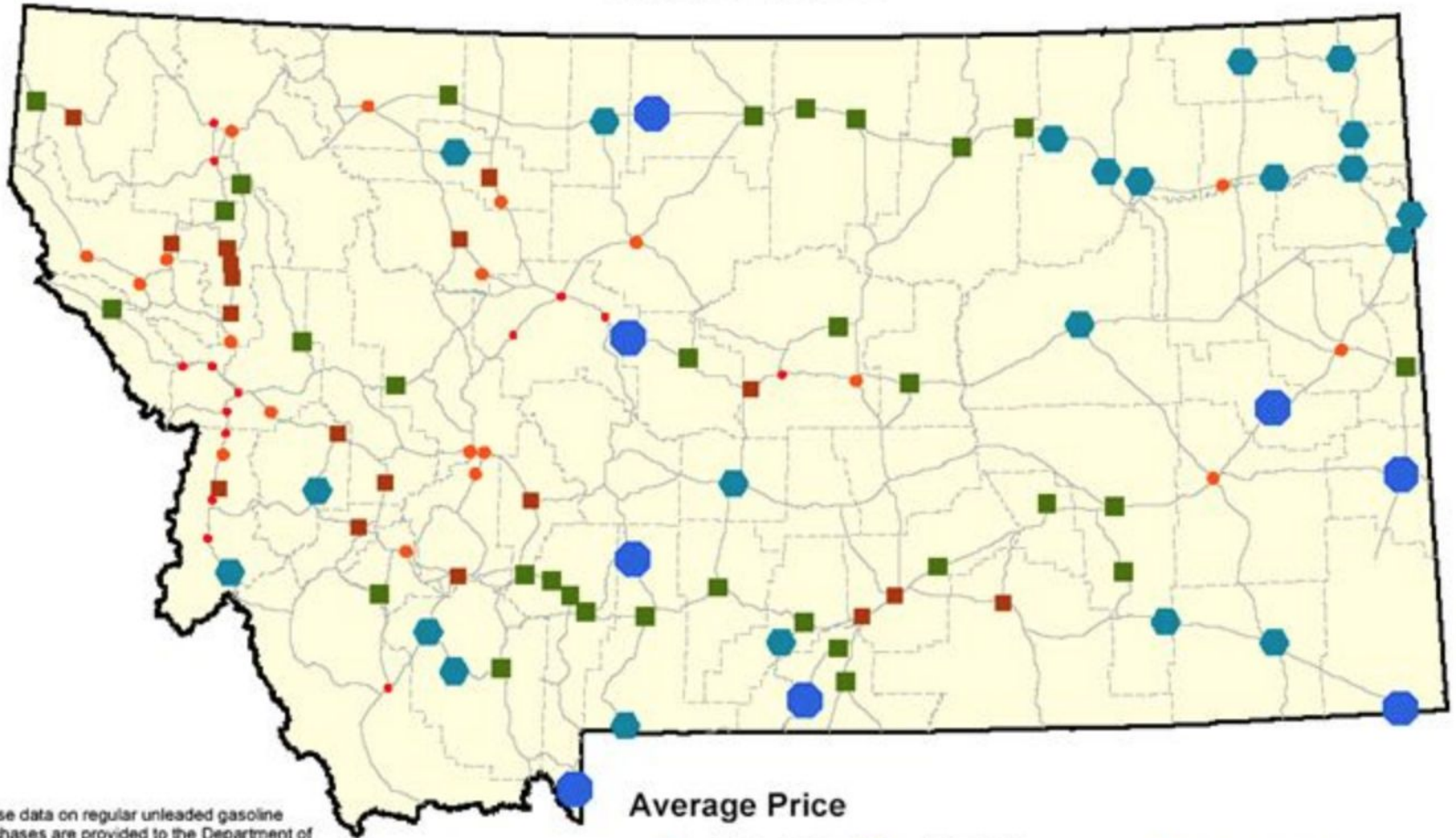
These data on regular unleaded gasoline
purchases are provided to the Department of
Environmental Quality by Oil Price Information
Servic...
card tr...
and F...
card s...
valid s...
generally track closely with actual average purchase price, especially in
the larger towns where the number of observations is greater.

Montana State Library
Natural Resource
Information System

Map Number 03deq0195 -- 3/27/2003

Ratio data and map types

- Numerical data value on a scale with an absolute zero

- Can be physical absolute (e.g. wind speed) or ratio of two numbers (people / mi^2)

- Cartographic methods similar to <u>interval</u>

  - **Point**–compound point symbol with encoded data

  - **Line**–vectors, isolines

  - **Area**–choropleth and other methods (e.g. dasymetric)

3D isoline population density map

Line ratio map: wind flow vectors

300 mb; July 2011

→ 10 m/s

GPH at 300 mb (m)

9100 9200 9300 9400 9500 9600 9700 9800
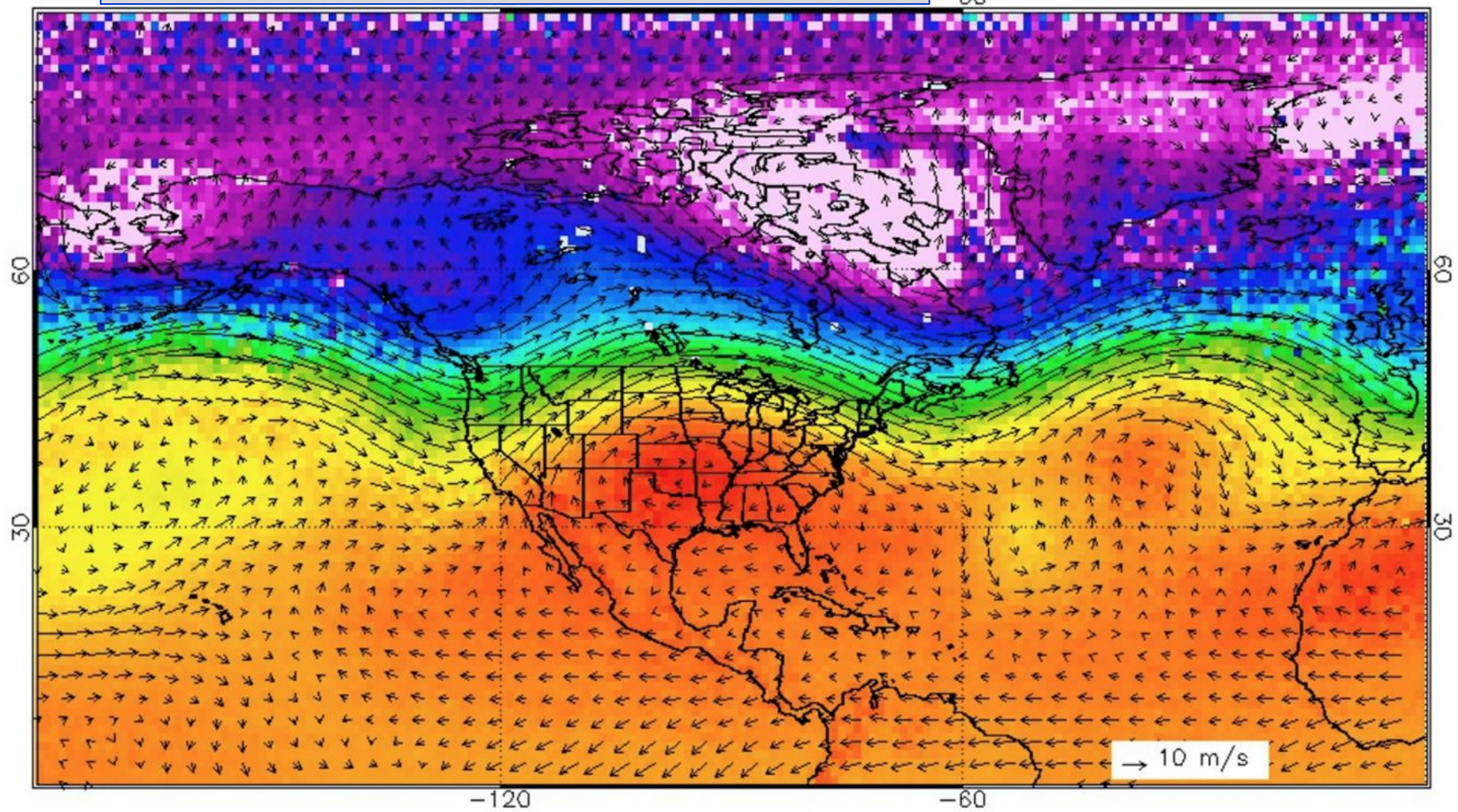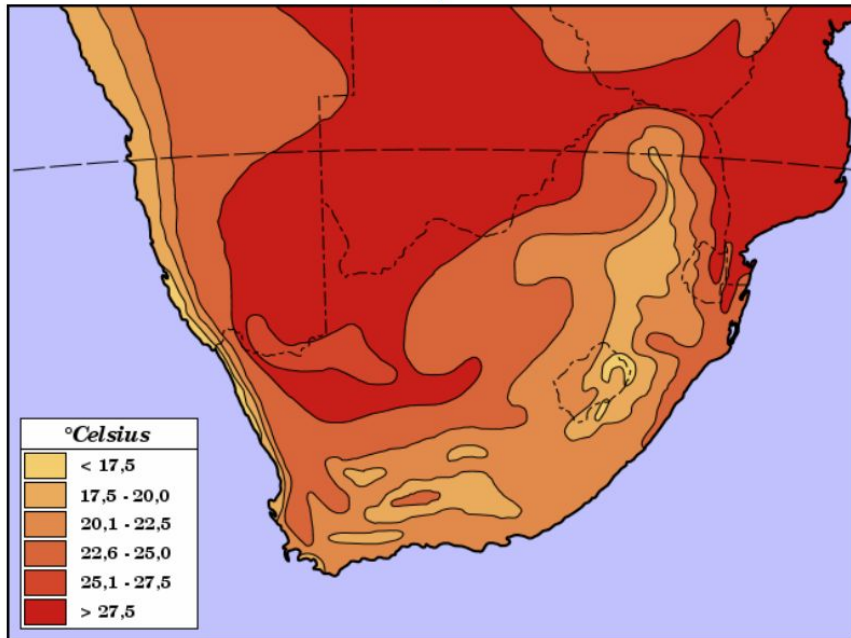
Population Density by County 2010
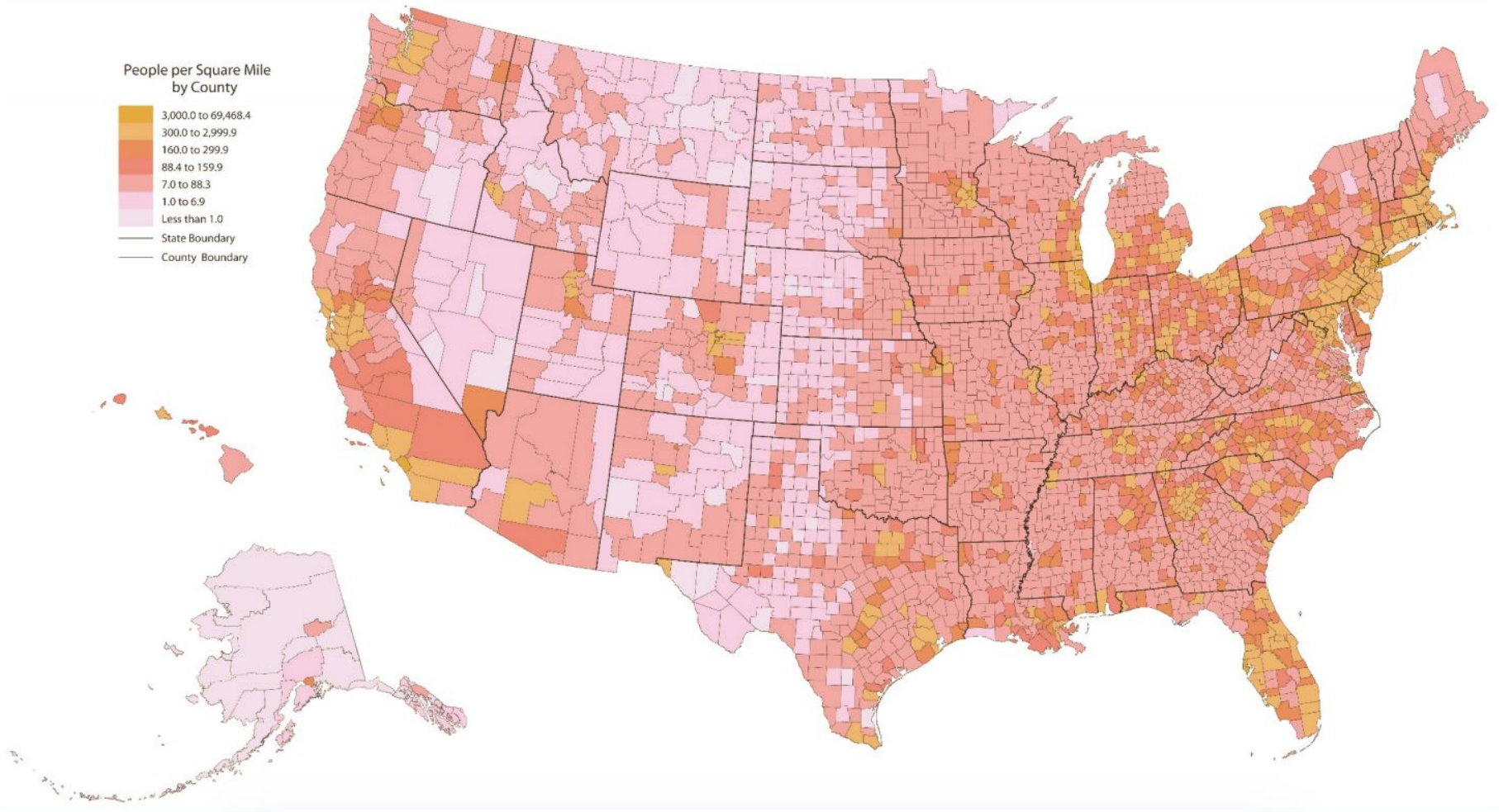
Area ratio map: choropleth
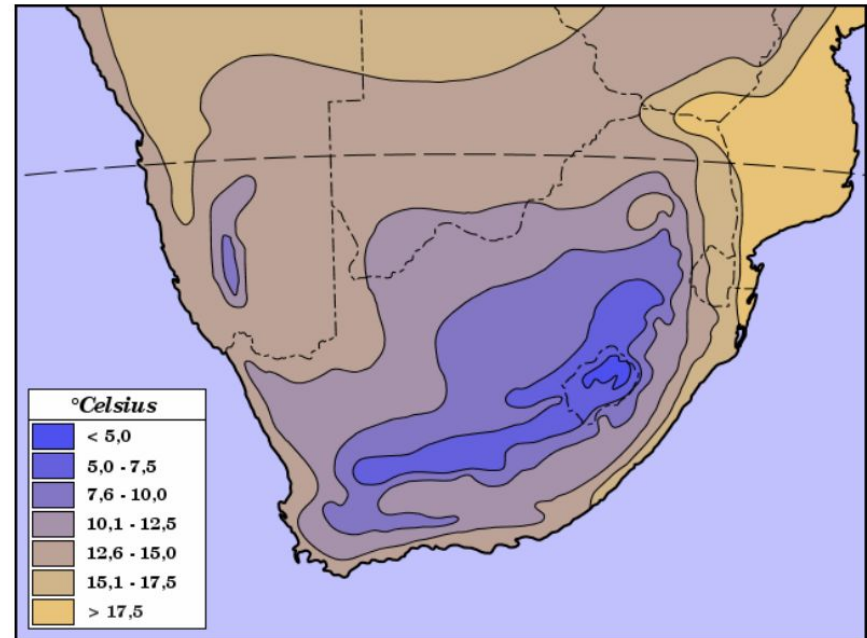
Interval data and map types

- Numerical data value but on an arbitrary scale
- Often reflective of 'counts' (e.g. total population)
- Cartographic methods
  - **Point**–proportional symbol, usually geometric object, varies in size, sometimes classed
  - **Line**–flow map, line width proportional to value
  - **Area**–prism map, shaded map, choropleth
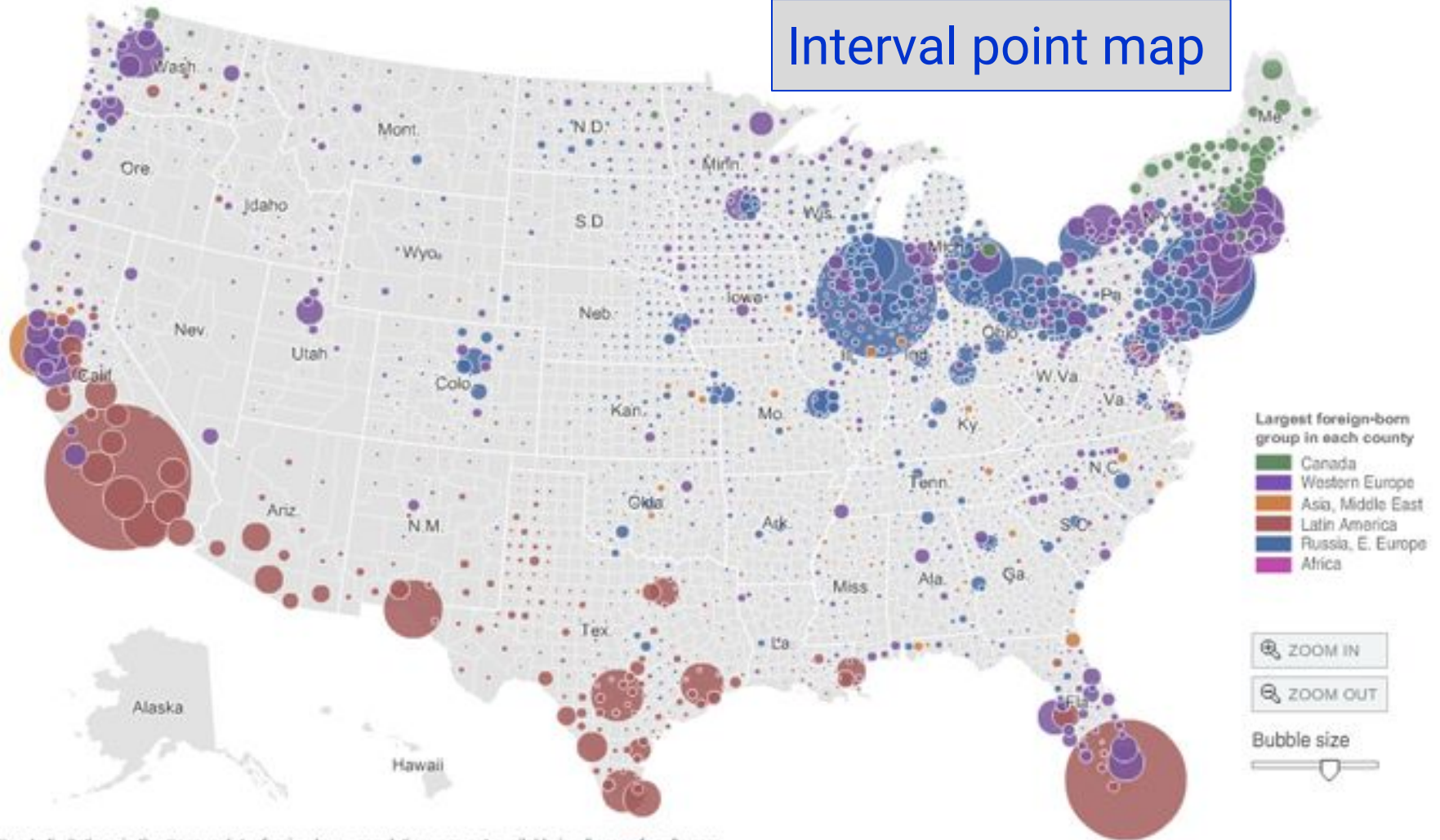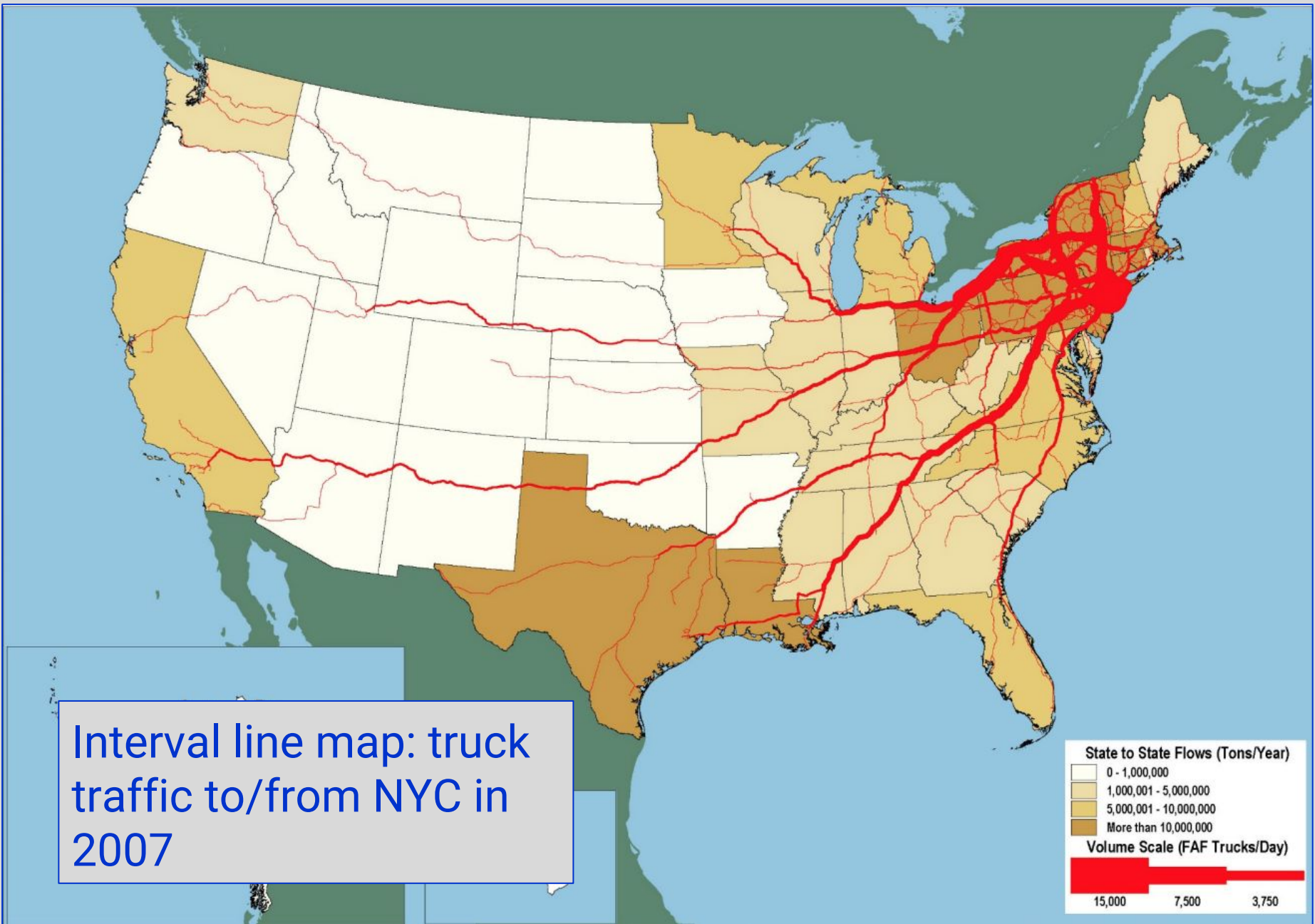
Interval data and map types

January	June
°Celsius	°Celsius
< 17,5	< 5,0
17,5 - 20,0	5,0 - 7,5
20,1 - 22,5	7,6 - 10,0
22,6 - 25,0	10,1 - 12,5
25,1 - 27,5	12,6 - 15,0
> 27,5	15,1 - 17,5
	> 17,5

Interval point map

Largest foreign-born group in each county
- Canada
- Western Europe
- Asia, Middle East
- Latin America
- Russia, E. Europe
- Africa

ZOOM IN
ZOOM OUT

Bubble size

Note: Due to limitations in the Census data, foreign-born populations are not available in all areas for all years.

Interval line map: truck traffic to/from NYC in 2007

State to State Flows (Tons/Year)

- 0 - 1,000,000
- 1,000,001 - 5,000,000
- 5,000,001 - 10,000,000
- More than 10,000,000

Volume Scale (FAF Trucks/Day)

15,000   7,500   3,750

Visualizing continuous and discrete data

Continuous/Sequential

- Symbology blurs boundaries, use for hotspot effect

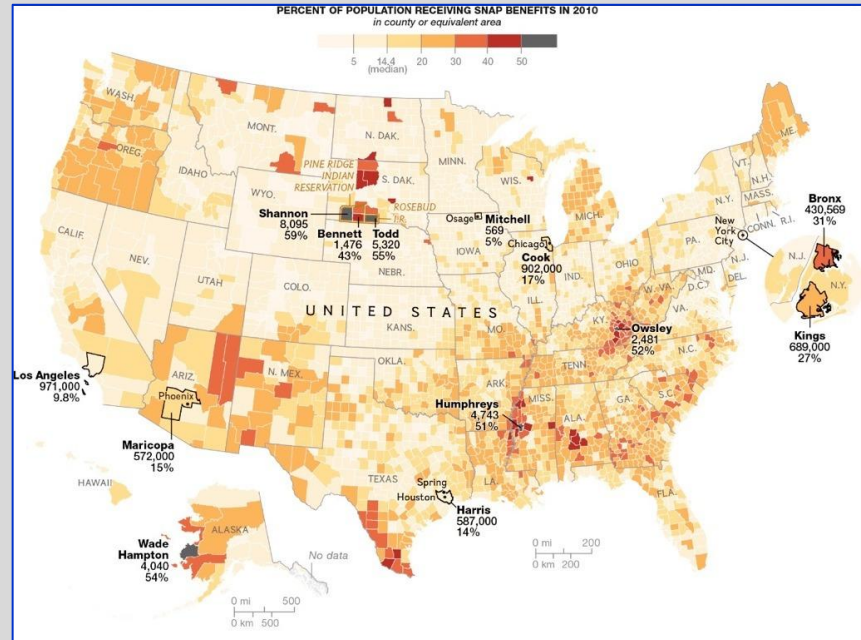A Three-Color Visualization of the Climate of the Coterminous United States for Continental and Multi-Scale Conservation Efforts
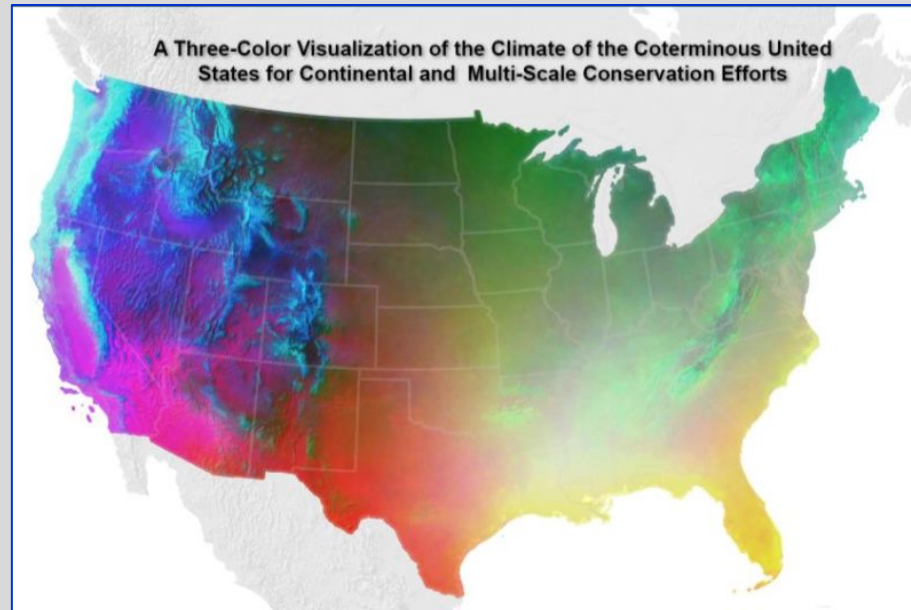
Discrete/Qualitative

- Symbology emphasizes boundaries, use for classifying data
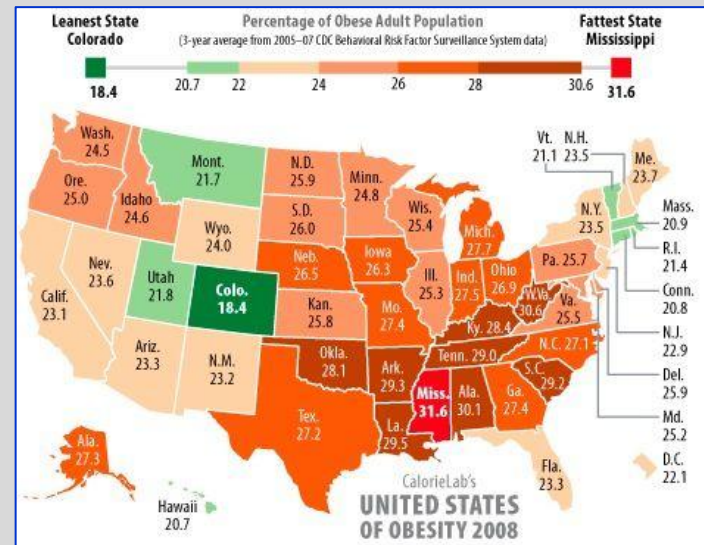

Climate Zones
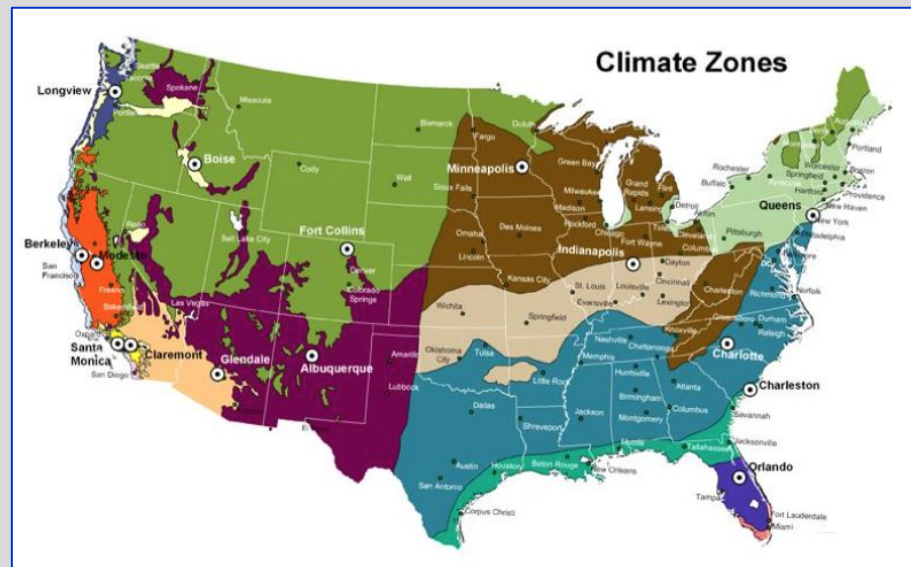
Visualizing in one or more colors

Monochromatic/

Sequential

- Single color/hue, sequential values (can be continuous or discrete)
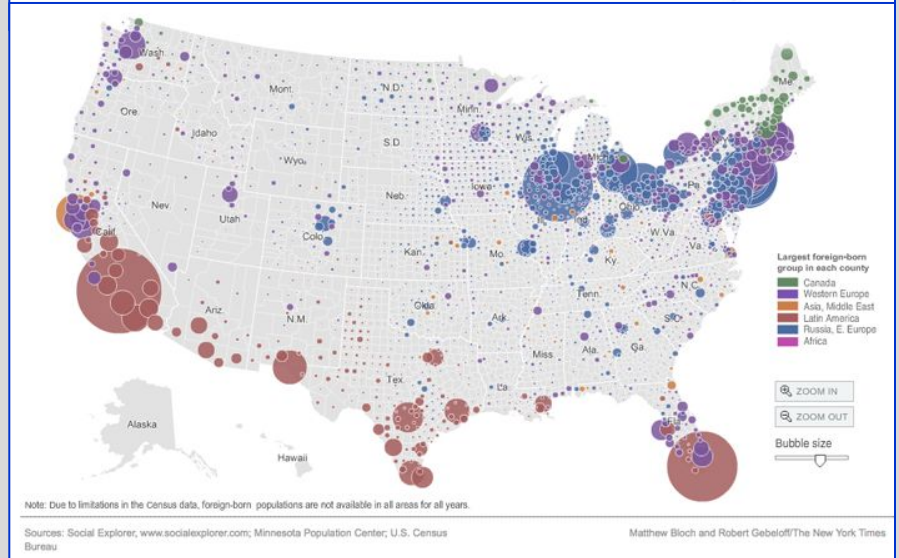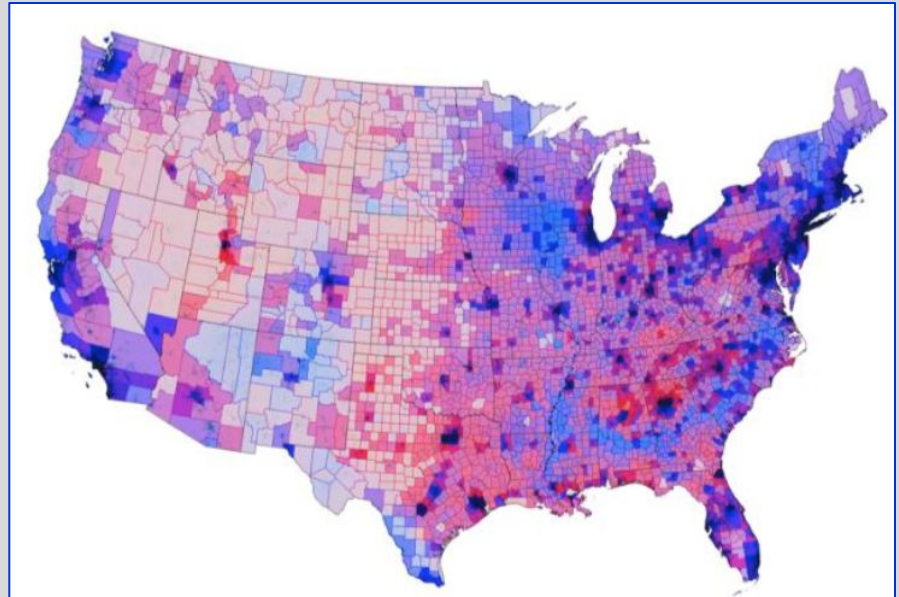

Diverging

- Spans 2 or more colors, implies average, above and below average (can be continuous or discrete)
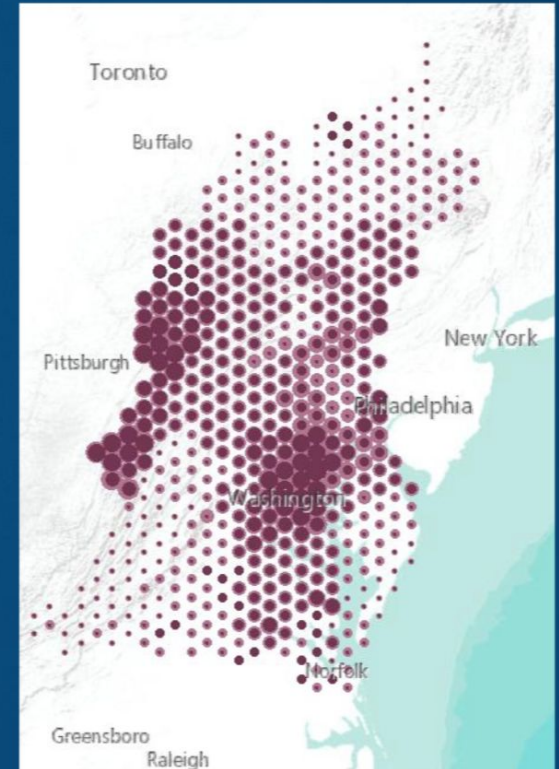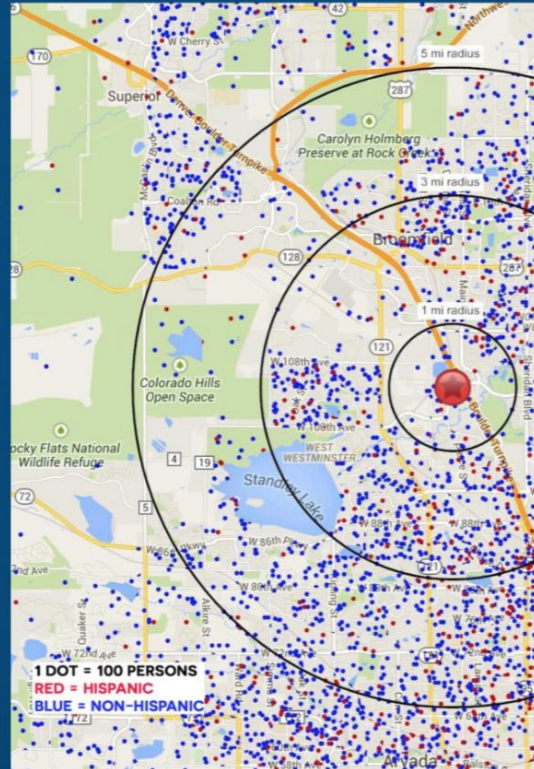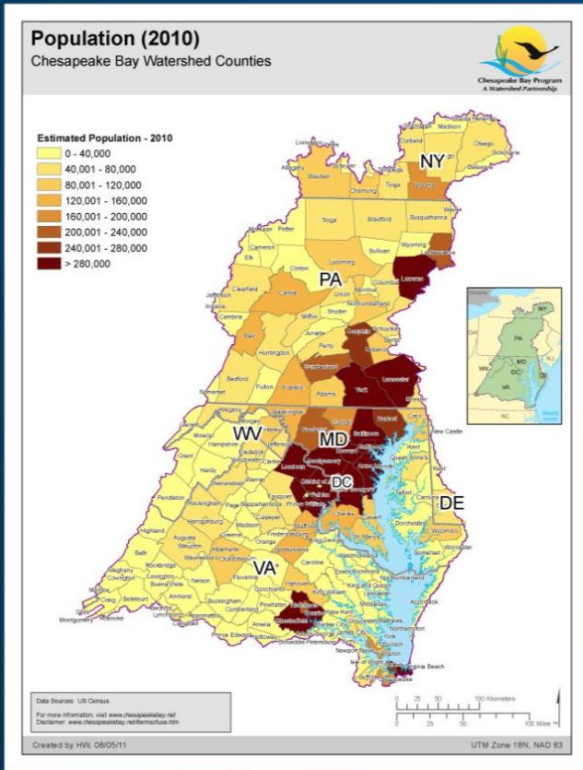
Visualizing two or more variables

- Using color alone (red and blue work well)

- Using color and symbology together

Surface/polygon mapping (area)

Choropleth   Dot density   Binning (proportional symbols)

New Haven Smokers

*due to confidentiality requirements, dots have been distributed randomly within each administrative unit and do not depict actual residential locations of smokers in the study area
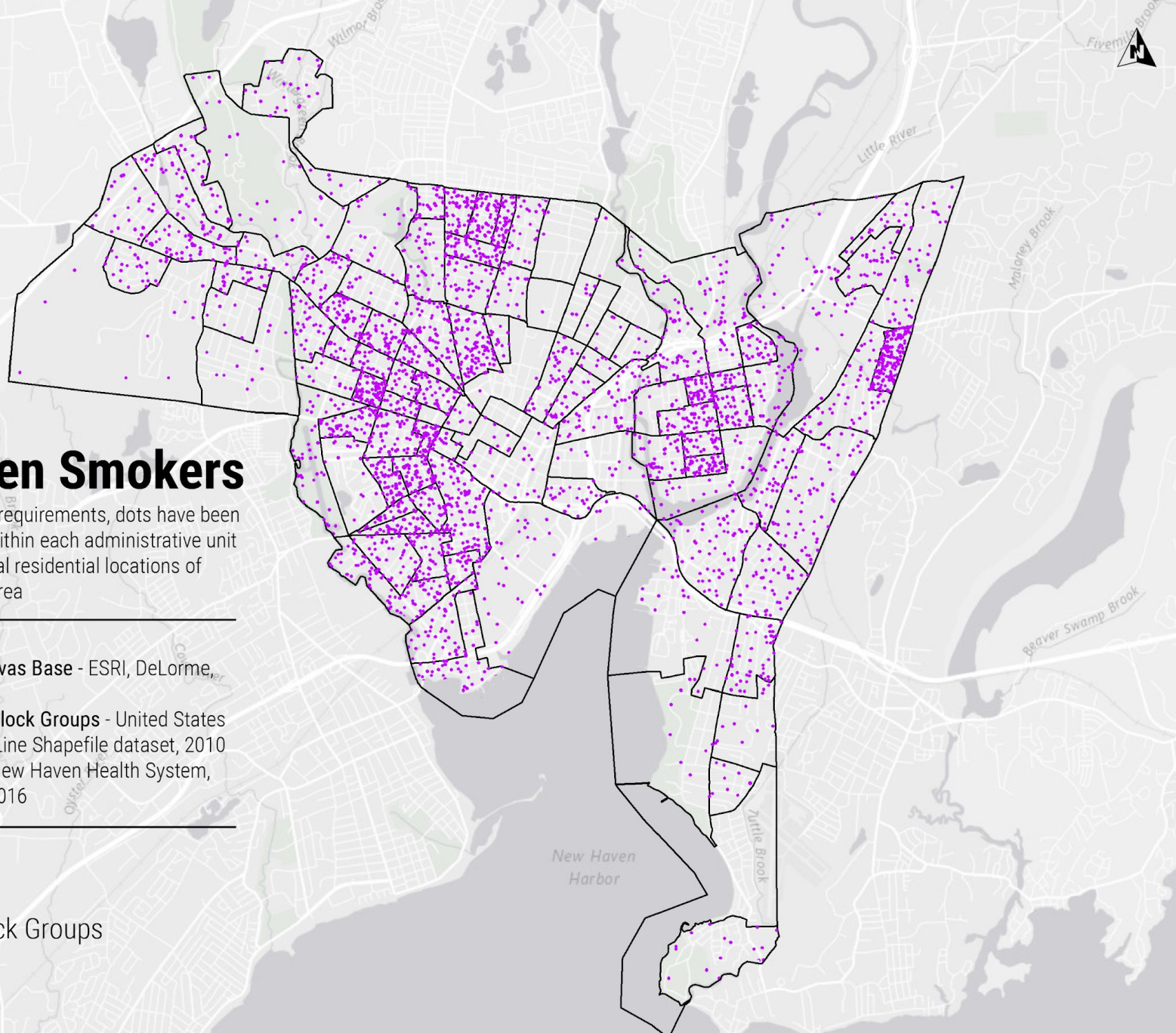
Data Sources
• **World Light Grey Canvas Base** - ESRI, DeLorme, HERE, MapmyIndia
• **New Haven Census Block Groups** - United States Census Bureau, Tiger/Line Shapefile dataset, 2010
• **Smoker Data** - Yale, New Haven Health System, Registry of Smokers, 2016

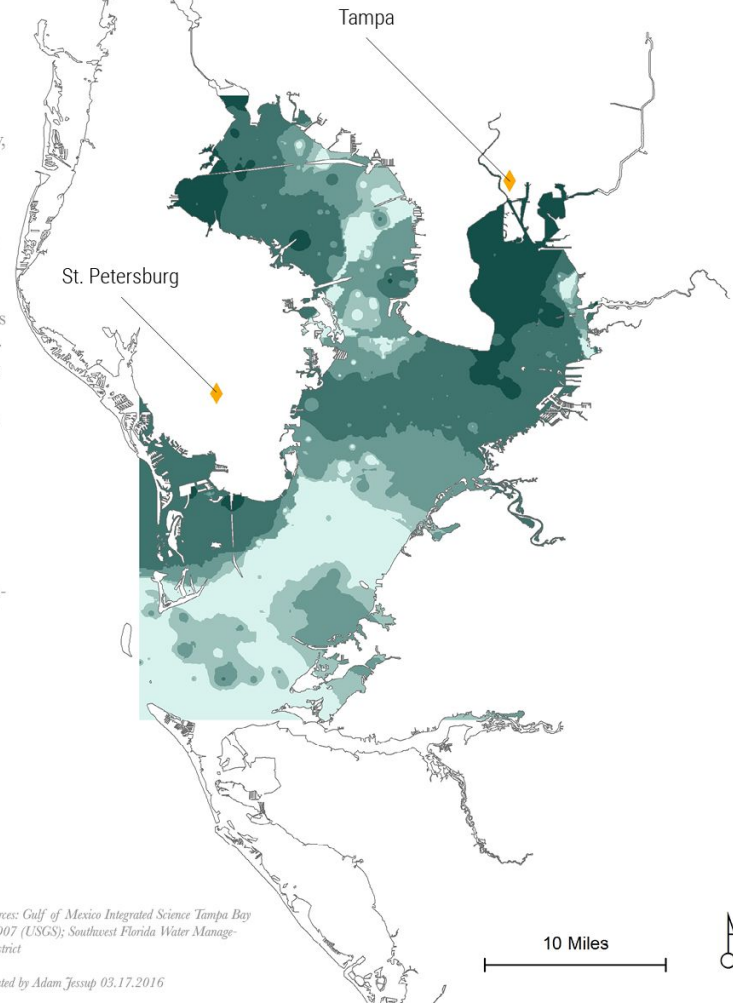1 dot = 1 person

☐ Census Block Groups

2 miles

IDW-Derived Percentage of Total Organic Compounds in Tampa Bay Sediment
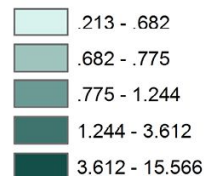
Study Area

Tampa

St. Petersburg

As part of the Gulf of Mexico Integrated Science Tampa Bay Study, many sediment cores have been collected throughout the estuary. Sites were chosen for the purposes of downcore stratigraphic and paleoecological analyses and are not uniformly distributed throughout the bay, but data from the top cores surface sediments can be interpreted to understand current estuarine bottom characteristics. A description of the bottom surface, such as sediment grain size, percent carbonates and organics, are important when evaluating hydrologic and sediment models of estuaries.

– E. Terence Edgar, U.S. Geological Survey Center for Coastal and Watershed Studies; Thomas M. Cronin, U.S. Geological Survey, National Center; Gregg Brooks, Eckerd College 2007
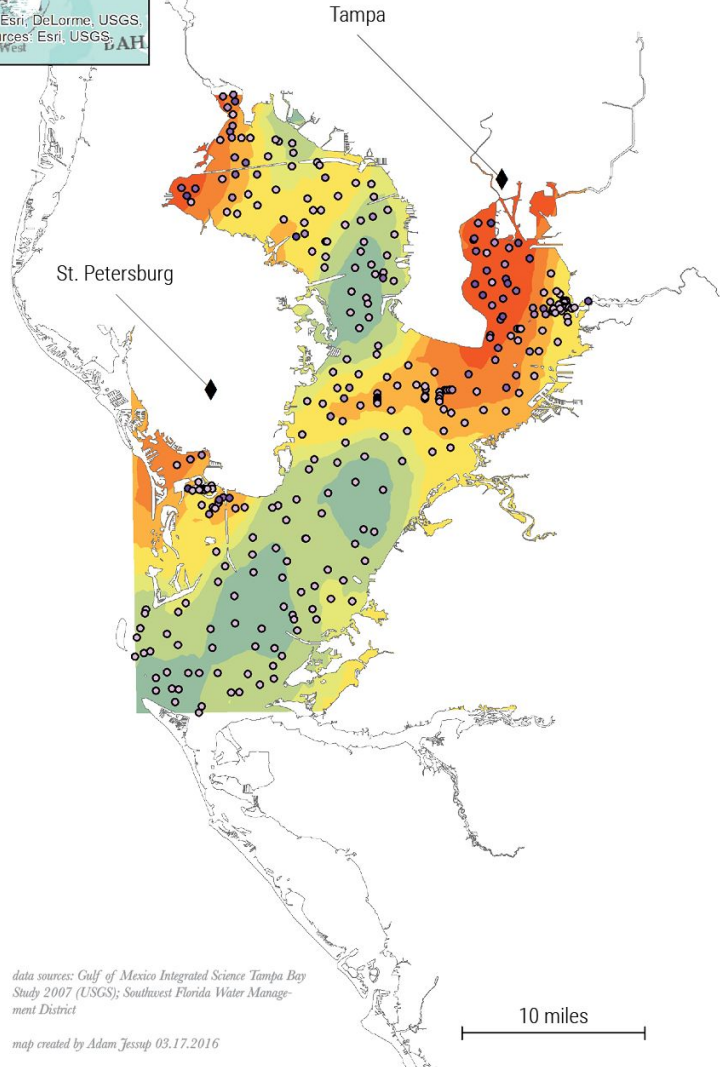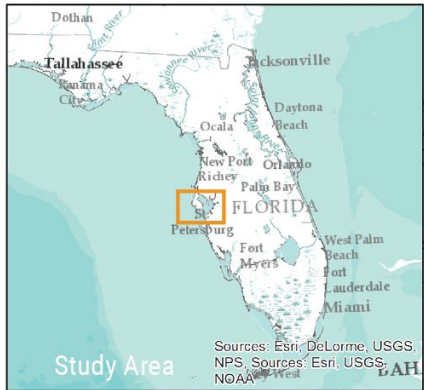
% Total Organic Carbons

- .213 - .682
- .682 - .775
- .775 - 1.244
- 1.244 - 3.612
- 3.612 - 15.566

10 Miles

Tampa Bay Sediment
Anisotropic CoKriging with logTOC, logMUD, logSILT

Study Area

Sources: Esri, DeLorme, USGS,
NPS, Sources: Esri, USGS,
NOAA

Tampa

St. Petersburg

Sample Points
Total Organic Carbons

○ 0 - 1.8
○ 1.8 - 5
○ 5 - 9.5
○ 9.5 - 13.6
○ 13.6 - 19.2

CoKriging Variables (logTOC, logMUD, logSILT

- -2.3 – -1.49
- -1.49 – -.951
- -.951 – -.599
- -.599 – -.368
- -.368 – -.216
- -.216 – .016
- .016 – .368
- .368 – .904
- .904 – 1.72
- 1.72 – 2.96

data sources: Gulf of Mexico Integrated Science Tampa Bay
Study 2007 (USGS); Southwest Florida Water Manage-
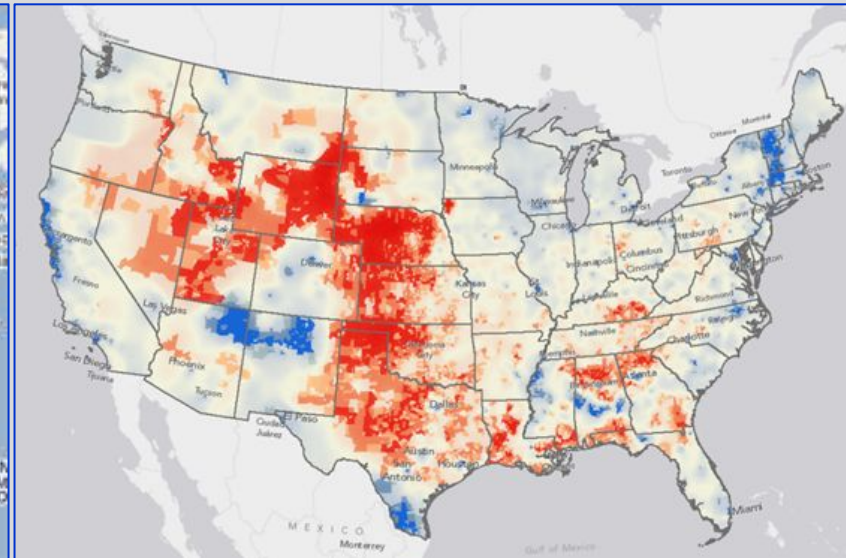ment District

map created by Adam Jessup 03.17.2016

10 miles

Heat map vs hotspot map

Heat map

Hotspot map

Q: What are the differences?

Classification schemes

Common methods

- Equal interval
- Quantile
- Mean (standard deviation)
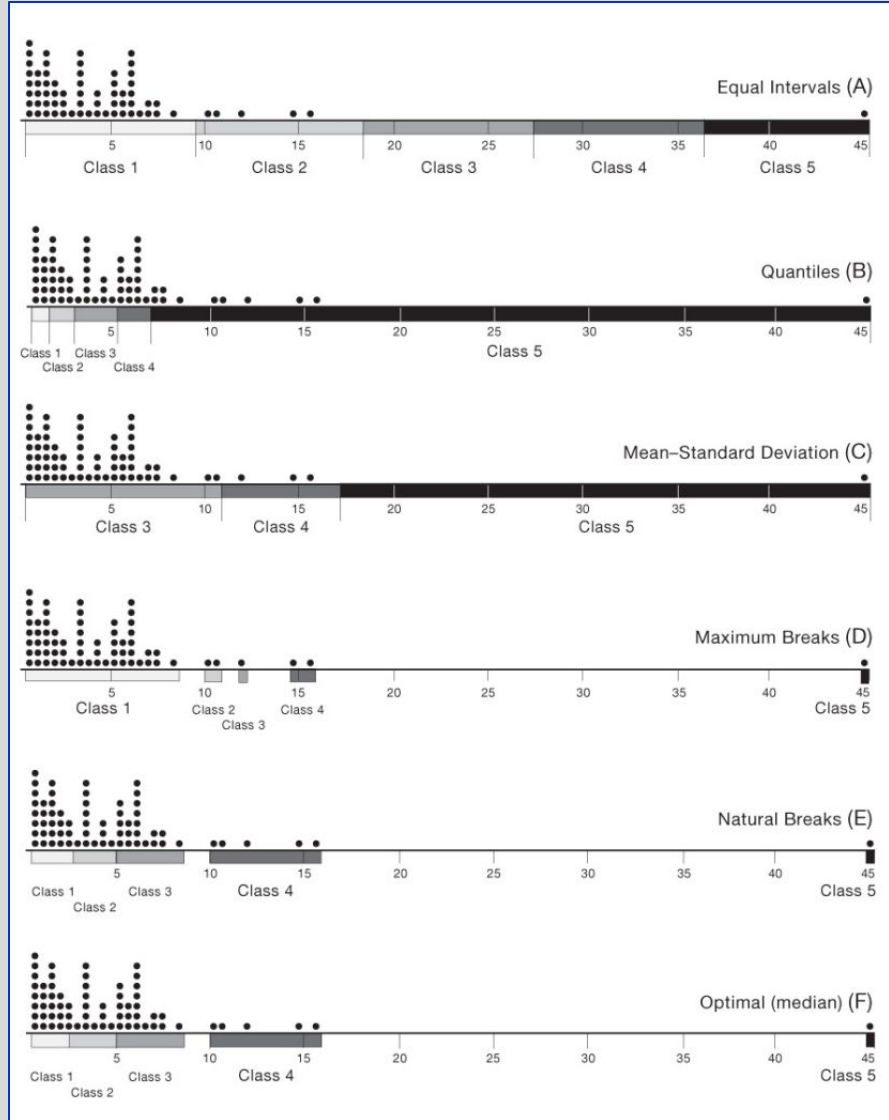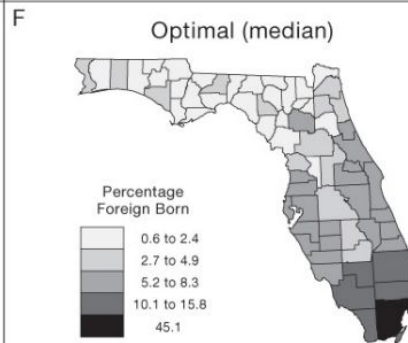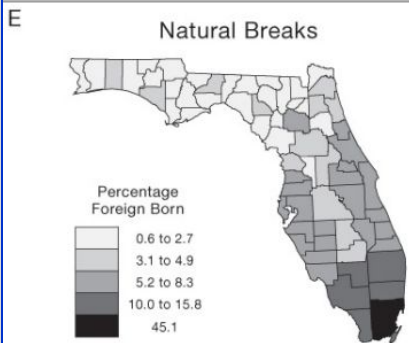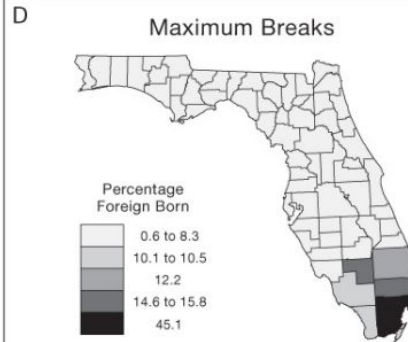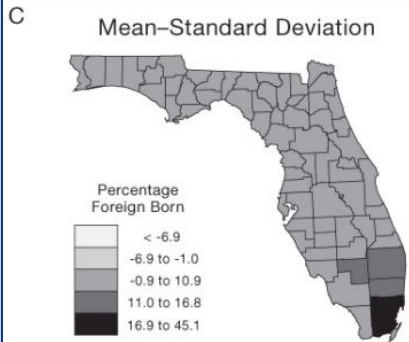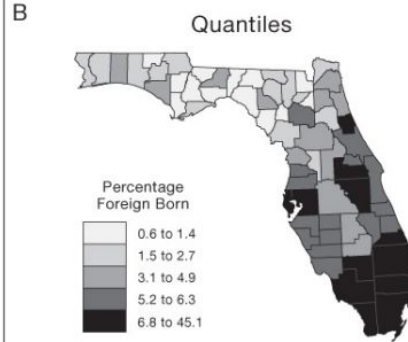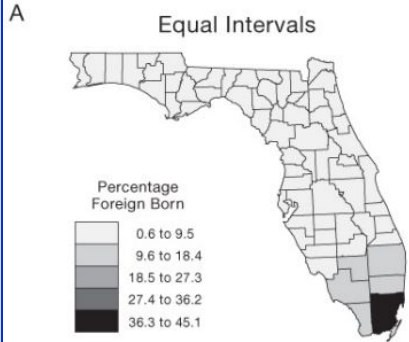- Natural breaks (Jenks)

Classification schemes

Common methods

- **Equal interval**
  - divides the range of attribute values into equal-sized subranges

- **Quantile**
  - assigns the same number of data values to each class, thus no empty classes or classes with too few/too many values, good for linearly distributed data

- **Mean (standard deviation)**
  - class breaks are created with equal value ranges that are a proportion of the standard deviation—usually at intervals of 1, 1/2, 1/3, or 1/4 standard deviations using mean values and the standard deviations from the mean

- **Natural breaks (Jenks)**
  - classes are based on natural groupings inherent in the data. Class breaks are identified that best group similar values and that maximize the differences between classes.
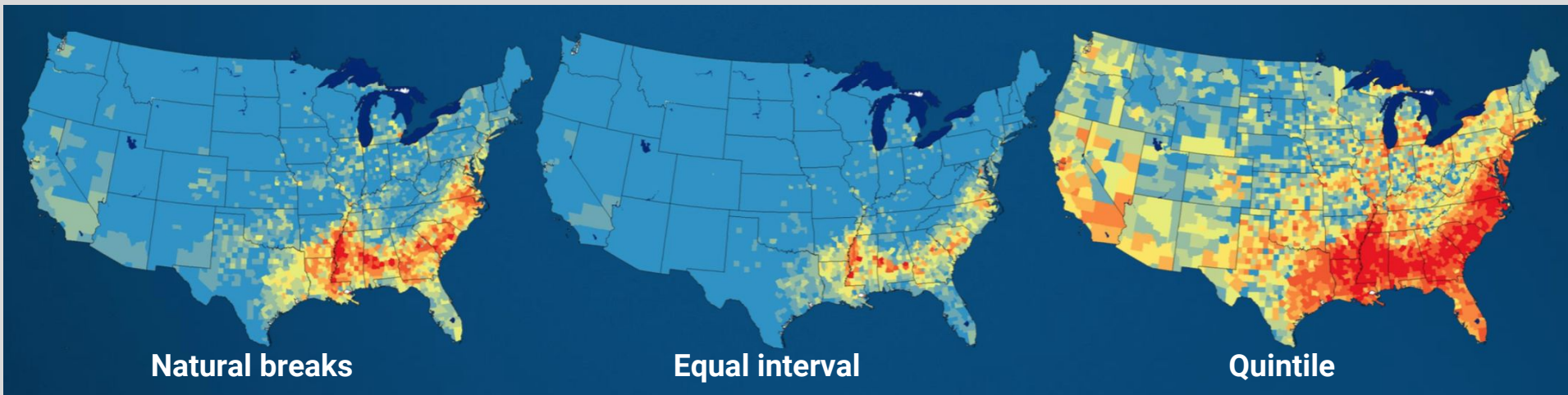
Same data, different distribution

Which method?

- Equal interval
- Quantile
- Mean (standard deviation)
- Natural breaks (Jenks)

If you're unsure, start with quintile

Natural breaks   Equal interval   Quintile

Which method?

- Equal interval
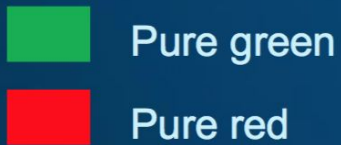- Quantile
- Mean (standard deviation)
- Natural breaks (Jenks)

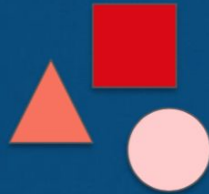If you're unsure, start with <u>quintile</u>...why?

Natural breaks   Equal interval   Quintile
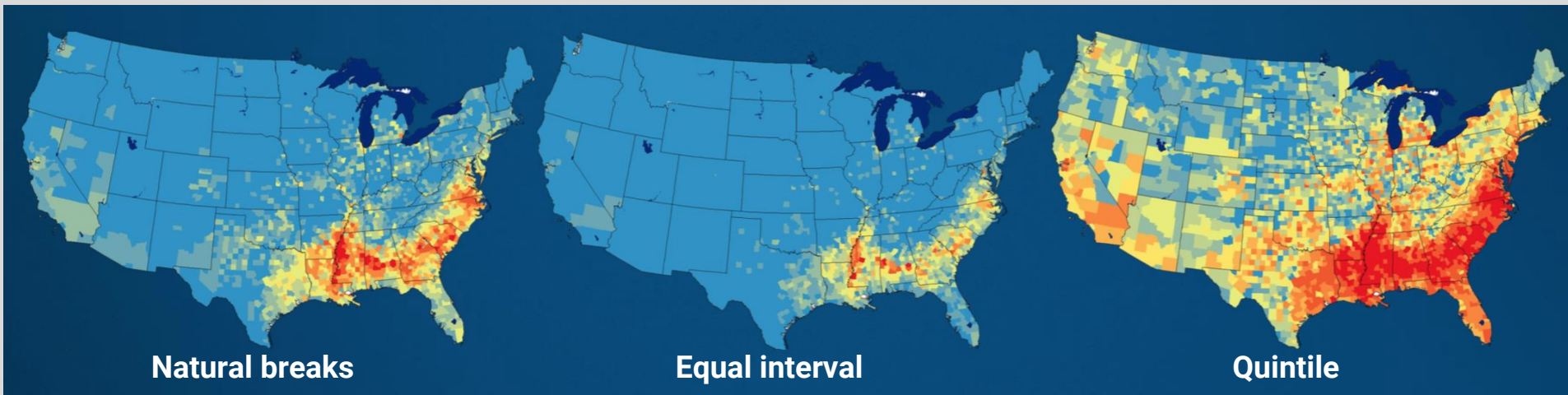
Designing for color-impaired audience
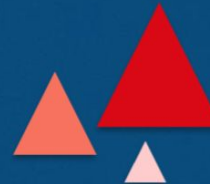
Graphic/visual variables

Position

Color

Size

Orientation

Shape
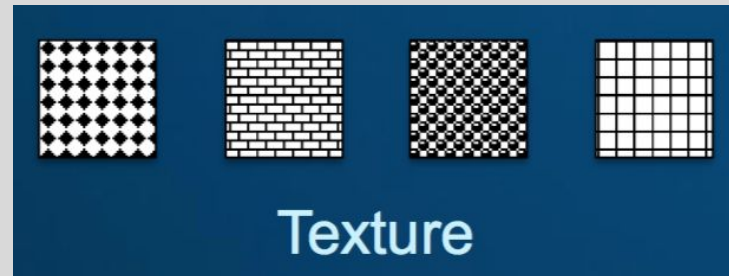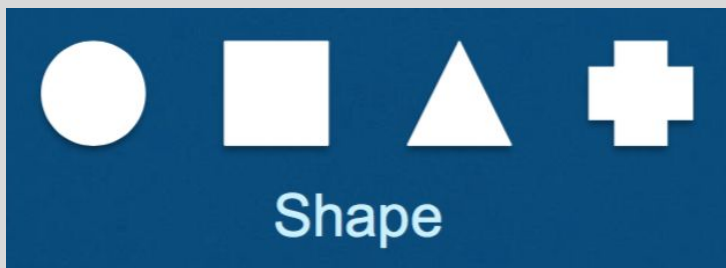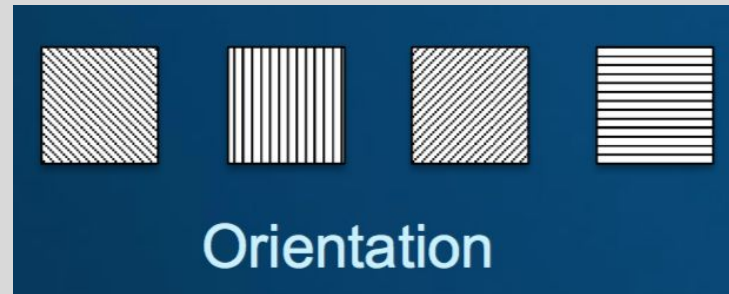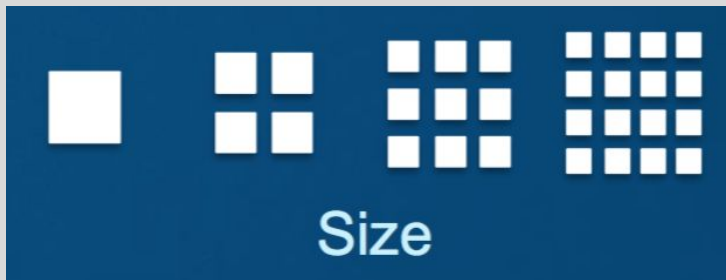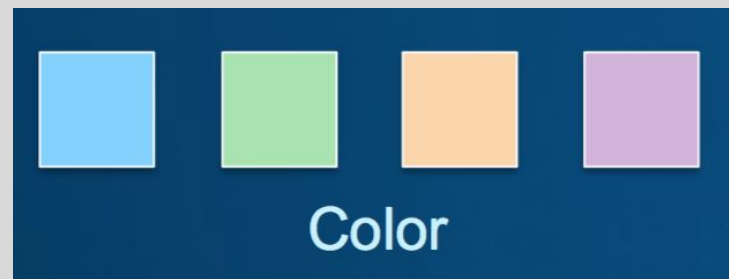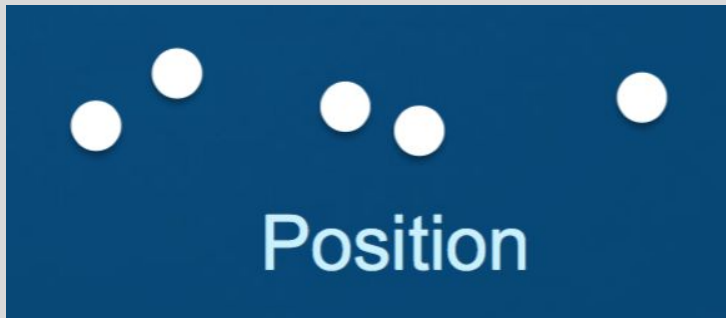
Texture

Value

Graphic/visual variables

Position–changes in the x, y location

Size change in length, area, repetition

Shape–changes in shape

Value–change from light to dark

Color–changes in hue

Orientation–changes in alignment

Texture–variation in "grain"

Constraints can make your maps better!

Buffer D
1 = 0.25
2 = 0.5
3 = 1

Ce

No
Tre

3

2

1

1 mi

a

b

c

d

Buffers

Affected Areas

Census Block Groups

• Hazardous Facilities
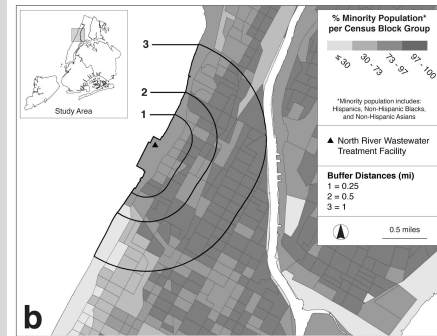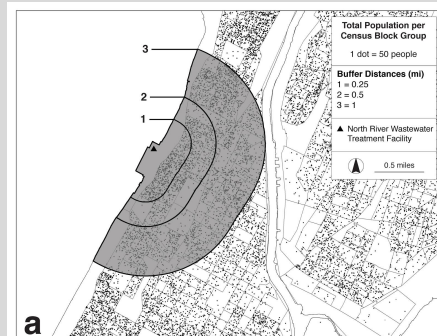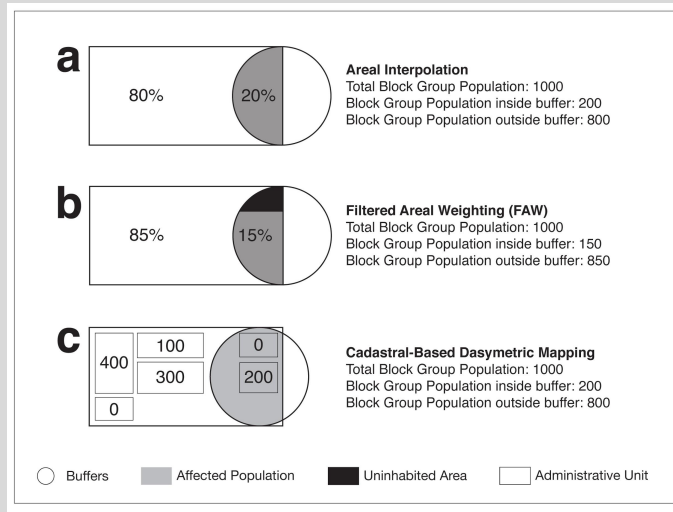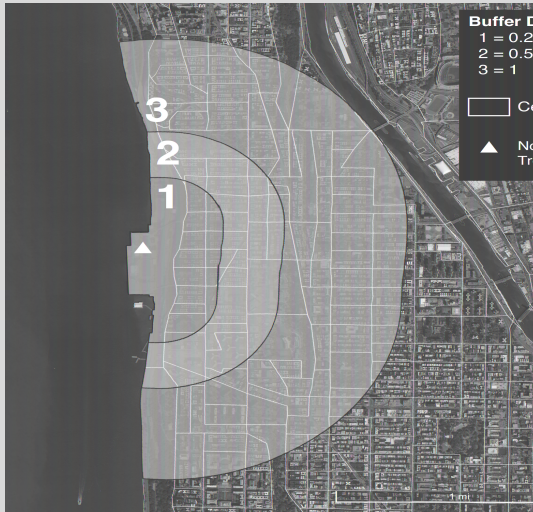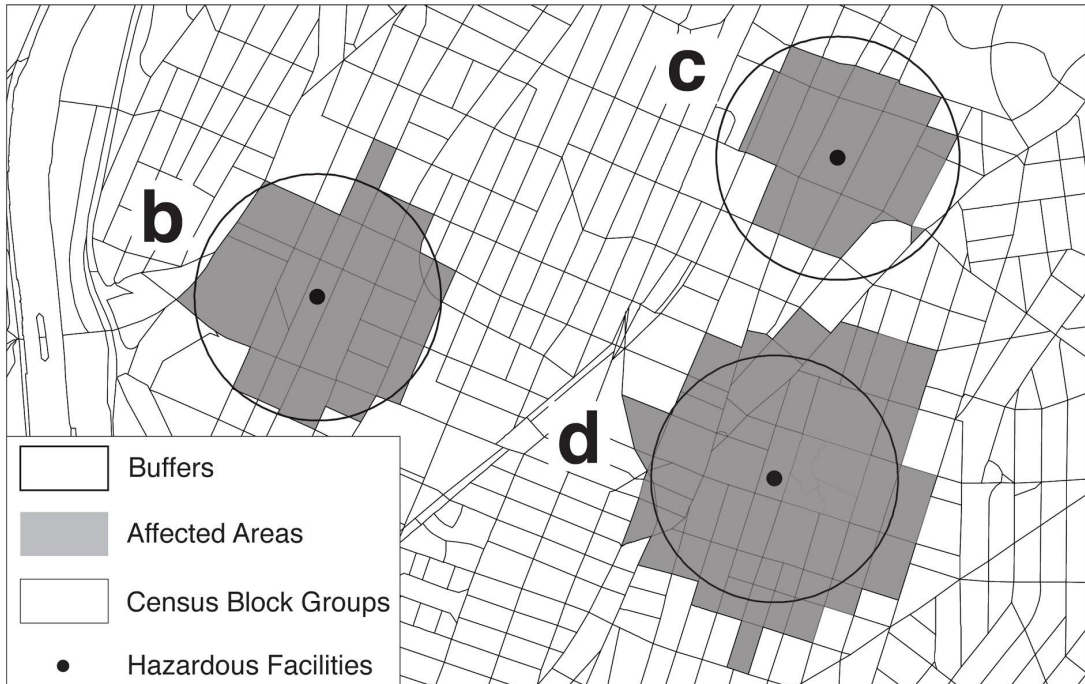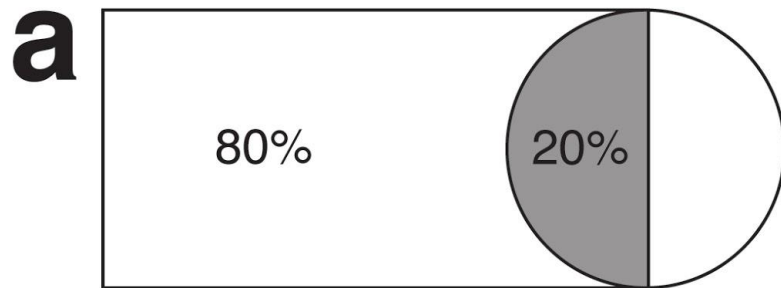
a Areal Interpolation
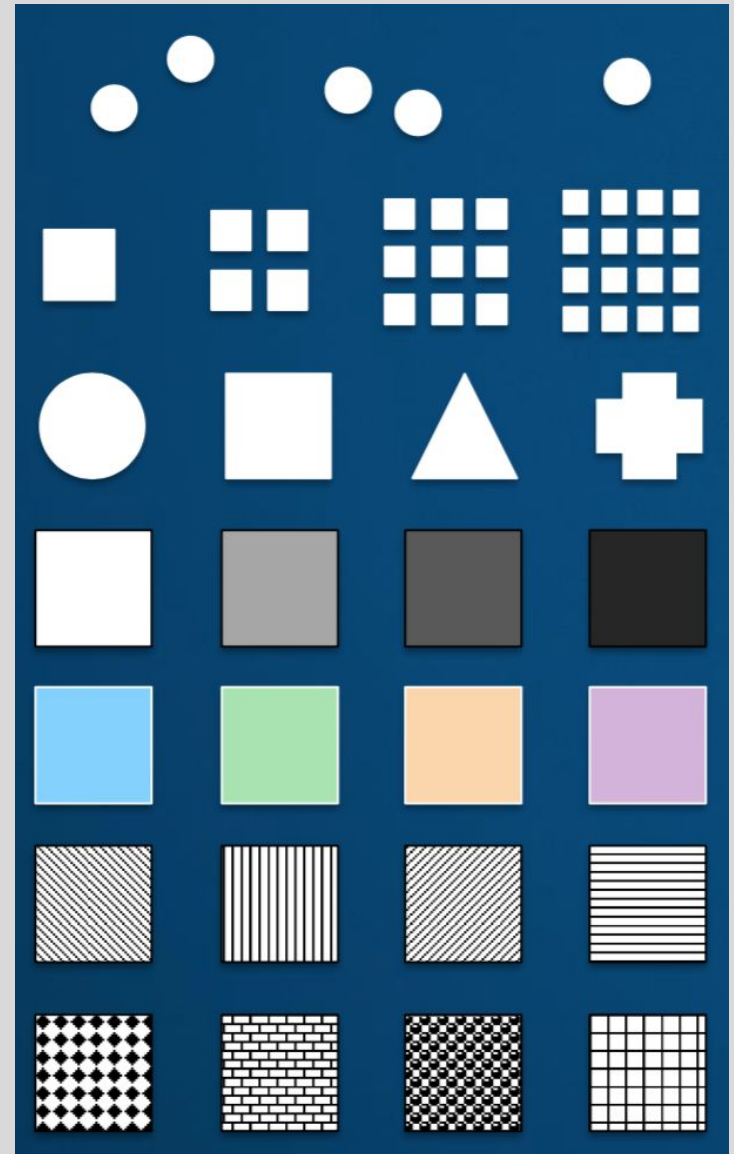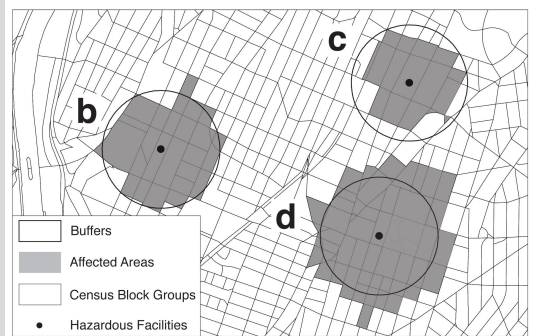Total Block Group Population: 1000
Block Group Population inside buffer: 200
Block Group Population outside buffer: 800
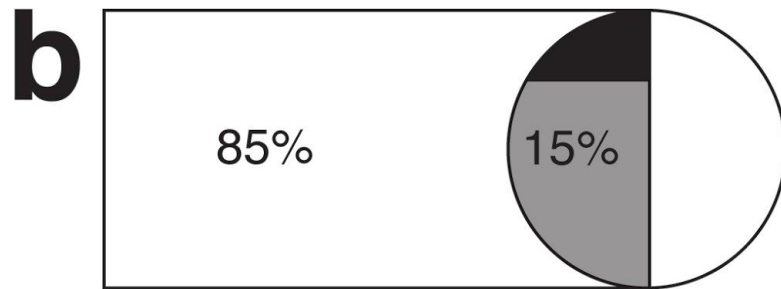
b Filtered Areal Weighting (FAW)
Total Block Group Population: 1000
Block Group Population inside buffer: 150
Block Group Population outside buffer: 850
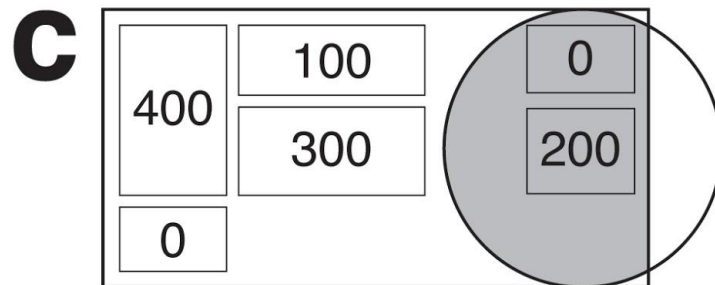
c Cadastral-Based Dasymetric Mapping
Total Block Group Population: 1000
Block Group Population inside buffer: 200
Block Group Population outside buffer: 800

⬤ Buffers   ▢ Affected Population   ▢ Uninhabited Area   ▢ Administrative Unit

a

Total Population per Census Block Group

1 dot = 50 people

Buffer Distances (mi)
1 = 0.25
2 = 0.5
3 = 1

▲ North River Wastewater Treatment Facility

0.5 miles

b

Study Area

% Minority Population* per Census Block Group

≤ 30   30 – 73   73 – 97   97 – 100

*Minority population includes: Hispanics, Non-Hispanic Blacks, and Non-Hispanic Asians

▲ North River Wastewater Treatment Facility

Buffer Distances (mi)
1 = 0.25
2 = 0.5
3 = 1

0.5 miles

Brief introduction to the Final Project

Intro to the Final Project [Grad Students]

Your final project is to create a **personal portfolio** website featuring your best cartographic work

Portfolio requirements
- Minimum of 10 projects
  - Projects can be maps from any class you've taken at Lehman, or any map you have created in general
  - 3 of these 10 should be "old" labs that you have revised/workshopped to reflect the design competencies you have gained throughout this course
  - At least half of these projects (minimum, 5) should include writing that contextualizes the map/visual product for the viewer
- Minimum of two pages
- bio/photo of yourself/personal logotype
- CV/resume/LinkedIn profile

Portfolios are due at 12pm on 12.04.18

Intro to the Final Project [Undergraduates]

Your final project is to create **graphic presentation** featuring your best cartographic work

Presentation requirements
- Minimum of 5 projects
  - Projects can be maps from any class you've taken at Lehman, or any map you have created in general
  - 2 of these 5 should be "old" labs that you have revised/workshopped to reflect the design competencies you have gained throughout this course
  - Choose one layout (your favorite/best) and step through the various design choices you made in the process of creating the map (e.g. *Why did you choose the colors/typefaces/layout scheme you did? etc.*)
- Presentation should be **>5** but **<10** minutes long

Presentations are due at 12pm on 12.04.18

For next week…

- **A7 is due at 12pm on 11.06!**

- **Readings**

    - **Bernhard,** *Color design for the color vision impaired*

    - **Aisch,** *Choropleths*

- Next week we begin thinking about 'portfolios on the web'

- FYI only **4 actual class meetings** after this one!