CRP 4080: Introduction to Geographic Information Systems for planners

Lecture 7: Census Data Analysis

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

Announcement

- Final Project proposal due 10/11 – allow group work
- Thursday 10/10 – guest speaker: Keith Jenkins, GIS Librarian
- Lab 6 due 10/10 (Thursday) before the class
- Midterm exam 10:15am – 12:15pm on 10/17
- Lab 7 due 10/24 (Thursday) before the class

Final project

- Proposals (2 paragraphs)
  - Due: October 11 (submit via Canvas), non-graded.
  - Project idea (research question)
  - Possible data sources (with website links if possible)
  - If possible, list potential challenges of this project.
- Project idea must require use of GIS (ie have some sort of spatial component)
- Final Project examples have been posted on Canvas.
- If you have a dataset and unsure if it is good to use, email the dataset to me and I will have a look during the fall break.

Next week: Thursday (10/17)

CRP4080 Midterm Quiz

Cover material from labs 1 – 6

Midterm Quiz: You are allowed to refer to books as well as your labs and lecture notes for this exam. There are three types of questions:

Multiple choice and fill in the blank (15 pts): **Each question is worth 1 pt.**

True/False (10 pts.) **Each question is worth 1 pt.**

Short Answers (25 pts): **Please answer 5 of the 8 questions (5 pts each).**

Quiz Type	Graded Quiz
Points	50
Assignment Group	Assignments
Shuffle Answers	No
Time Limit	120 Minutes
Multiple Attempts	No
View Responses	Always
Show Correct Answers	No
One Question at a Time	No

Due	For	Available from	Until
Oct 17 at 12:15pm	Everyone	Oct 17 at 10:15am	Oct 17 at 12:15pm

Multiple choice (15 pts):

The term “false easting” refers to:

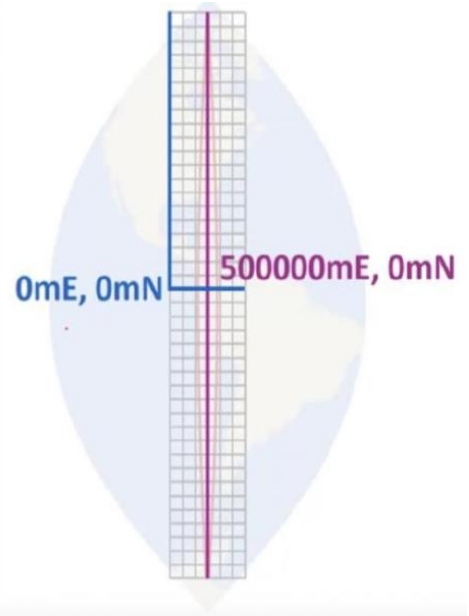
(a) an origin point south and west of a particular state plane zone

(b) intersection of equator and the zone's central meridian

(c) A value assigned to a central meridian to ensure that no coordinates within the zone have a negative value

(d) projected distance from the equator

In order to avoid negative value within the Cartesian Coordinate System, a false easting value of 500,000 meters is applied for all regions, and a false northing value of 10,000,000 meters is applied for only regions in the southern hemisphere.



Multiple choice (15 pts):

Which of the following is not a topological overlay?

(a) Union

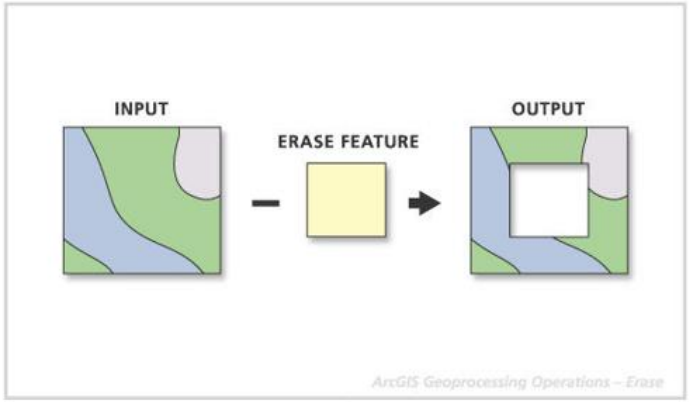
(b) intersect

(c) Erase

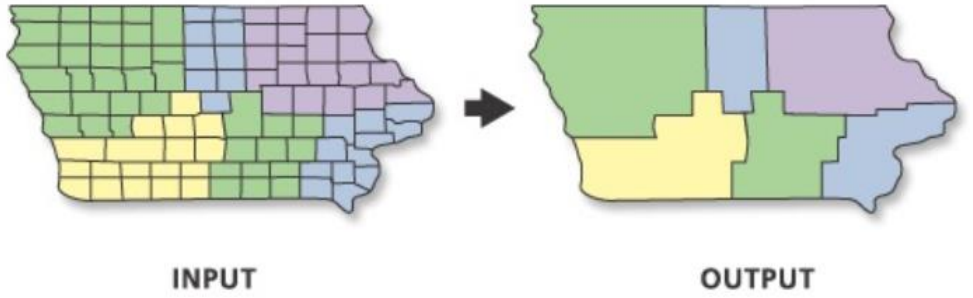
(d) Dissolve

Topological overlay operations:
Combining two layers to create a new output feature class containing information from both of the inputs.

Erase

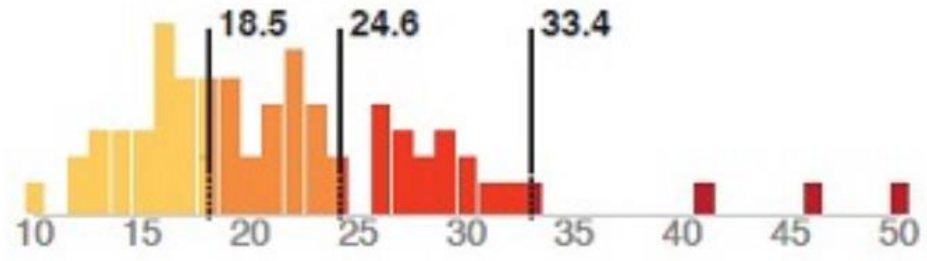


Dissolve

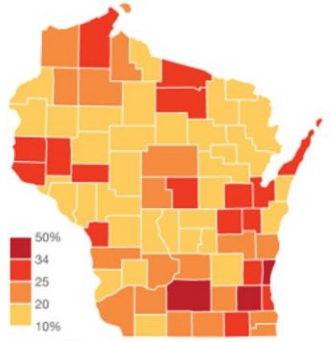


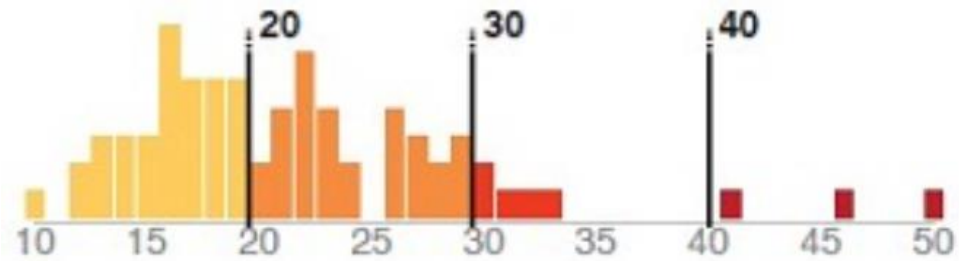
Exercise - Data Classification?

Quantile?
Natural Break?
Equal interval?
Standard Deviation?

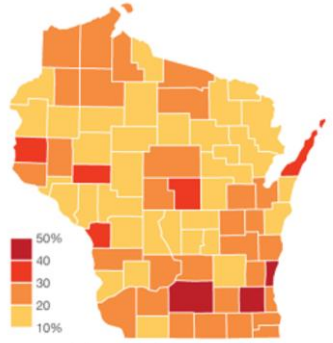


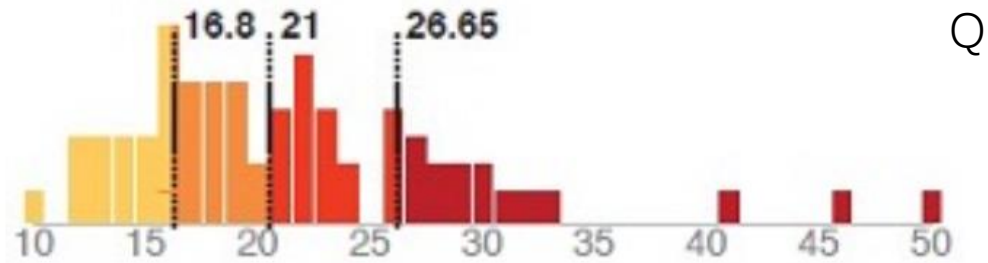
Natural Break



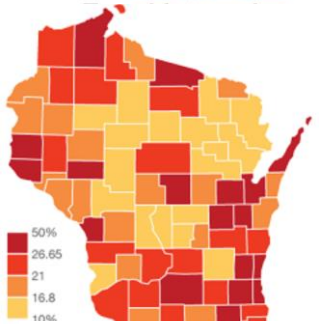


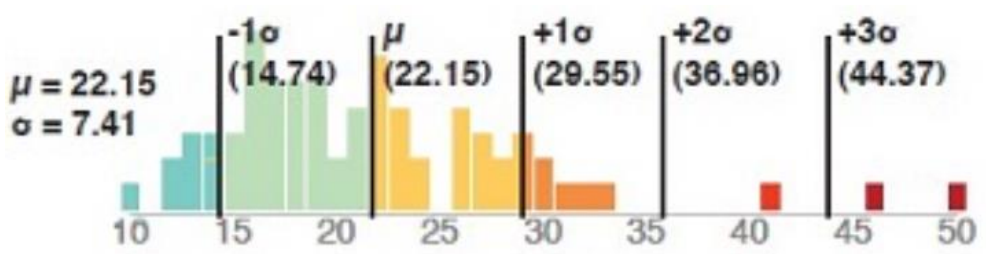
Equal Interval



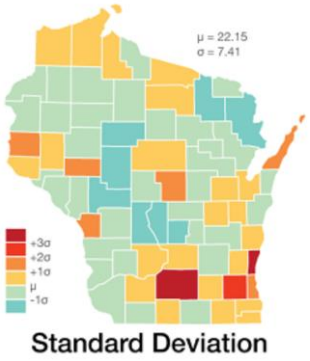


Quantile





Standard Deviation



Thematic Maps — Measurement scales

Qualitative maps make use of **nominal** (e.g. land cover class) or **ordinal** (e.g. a ranking) measurement scales,

quantitative maps make use of **interval** (e.g. Celsius temperature) or **ratio** measurement scales.

Nominal, data is categorized without any order or ranking. Examples include categories like land-use types or types of vegetation

ordinal, not only categorizes data but also ranks it in order. However, the intervals between ranks are not necessarily equal.

interval, data is ordered and the intervals between values are equal, but there is no true zero point (e.g., 0°C doesn’t mean 'no temperature').

Ratio, equal intervals between values and a true zero point (e.g. density, income, etc..) .

Thematic Maps — Measurement scales

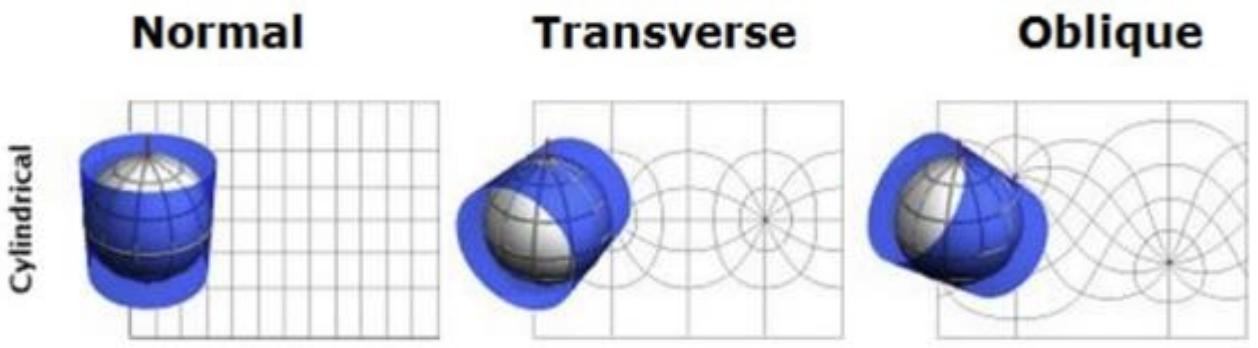
Each scale is represented once in the list below.

- Favorite candy bar (nominal)
- Weight of luggage (ratio)
- Egg size (small, medium, large, extra large, jumbo) ordinal
- Number of children in a family (ratio)
- Temperature in Celsius (interval)

True/False (10 pts):

A Transverse Mercator projection would be most appropriate for projecting the country of The Gambia in West Africa. (False/True)

Mercator projection is a kind of **Cylindrical Projection**. **Cylindrical Projection** wraps a cylinder around the globe



True/False (10 pts):

An Orthophoto is an aerial photo that has been corrected to account for camera tilt and lens distortion. (False/True)

Ordinary Orthophoto Mosaic: significant positioning errors are shown

True Orthophoto Mosaic: positioning errors are removed

Short answer

Map critique

- Basic elements: map title, legend, north arrow, scale bar, notes, etc…
- Does the map effectively convey the info?

Basic descriptive statistics

- Mean, Median, Max, Min, Skewness, Kurtosis; Histogram, QQ plot

Commonly used Surveys and Census

- the American Community Survey; Decennial census

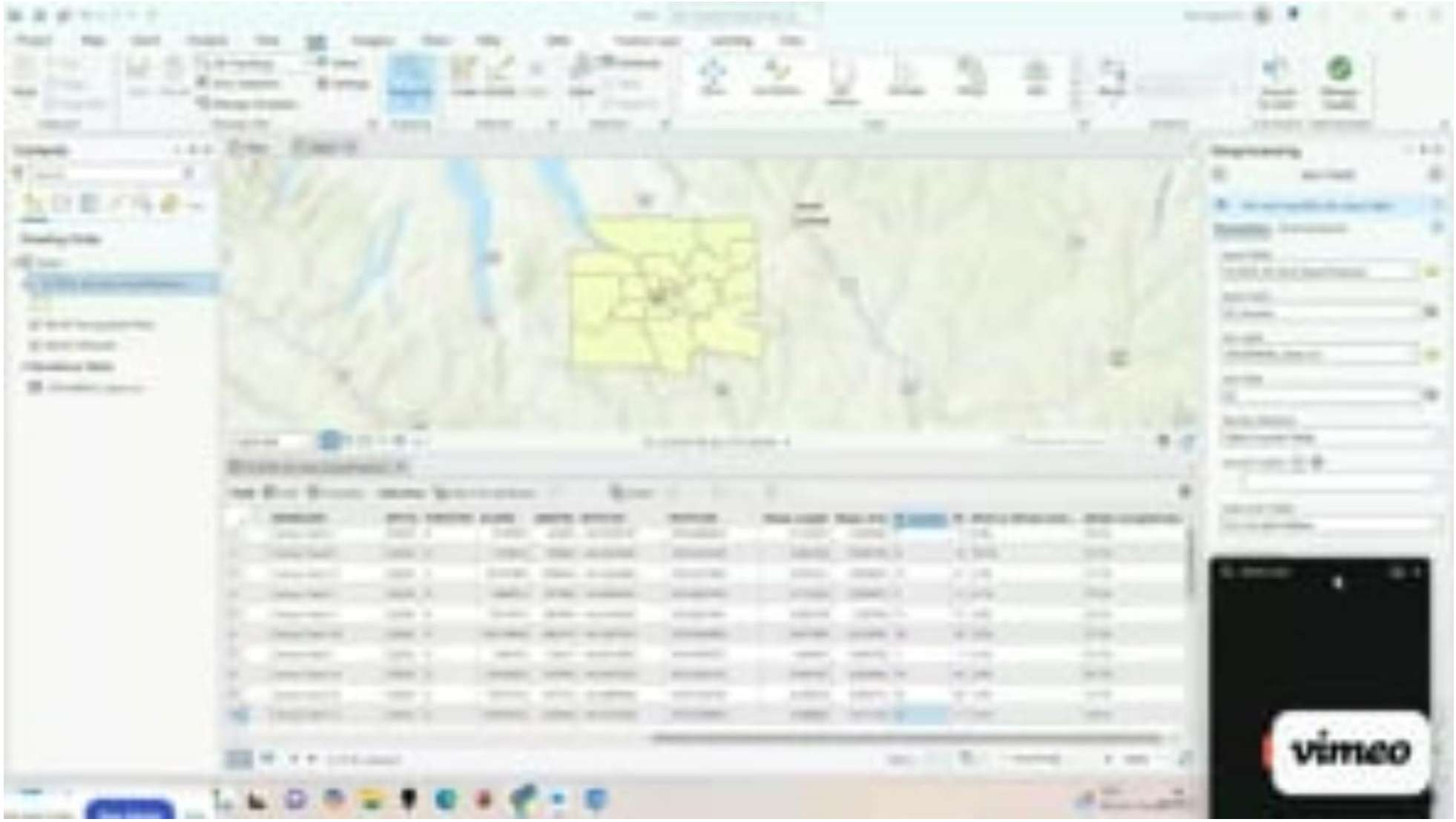
Projection

- What is projection? Which is unprojected (lng/lat) and which is projected?

Site Selection

- Geoprocessing tools; the steps of finishing the spatial analysis.

Join issue in the Lab 6



Research design in GIS: MAUP

- ***Modifiable Areal Unit Problem:*** different aggregation of the same data can result in very different sets of statistical relations being observed amongst a set of variables.

- Two observed effects of MAUP**, the scale problem** and **the aggregation or zonation issue.**

MAUP: Scale effect

Scale effect: tendency for different results to be obtained when aggregating information from a set of areal units into fewer and larger zones (i.e., a small-scale representation)

8	6	2
6	6	2
6	4	4

$n = 9,\ \bar{x} = 4.89$

20 $\bar{x} = 6.67$	18 $\bar{x} = 6$	8 $\bar{x} = 2.67$

$n = 9,\ \bar{x} = 5.13$

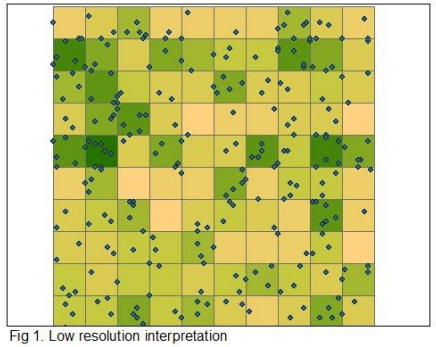


Fig 1. Low resolution interpretation

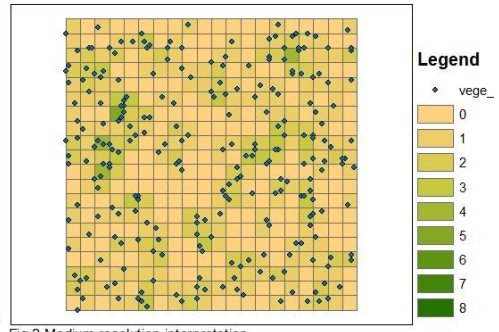


Fig 2.Medium resolution interpretation

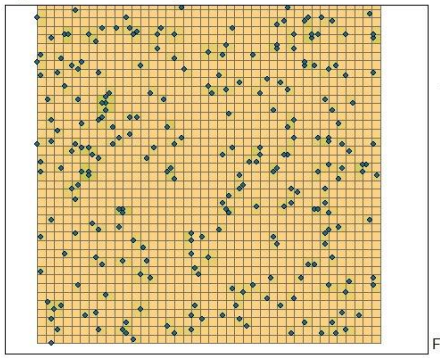
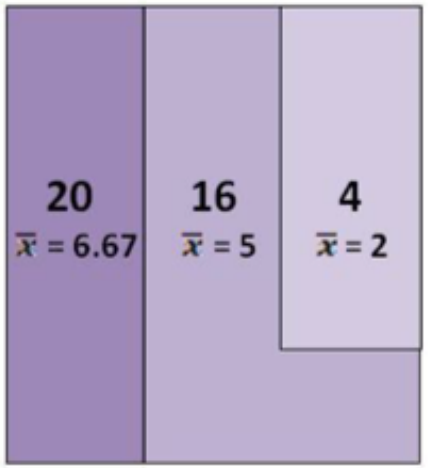


Fig 3.High resolution interpretation

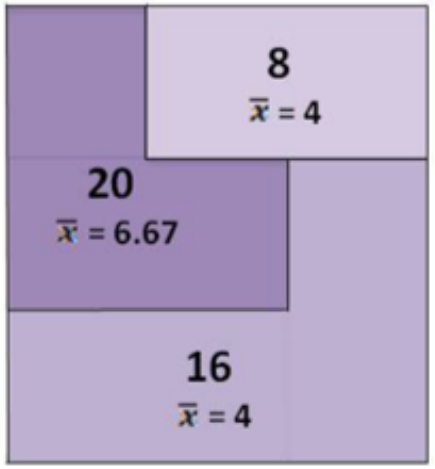
This data represent the distribution of plant species over a place. Analysing the distribution of these data at different resolution or scale yield in different results. The low resolution scale (Fig 1) is more general and while the high resolution scale (Fig 3) is more detailed.

MAUP: Aggregation problem

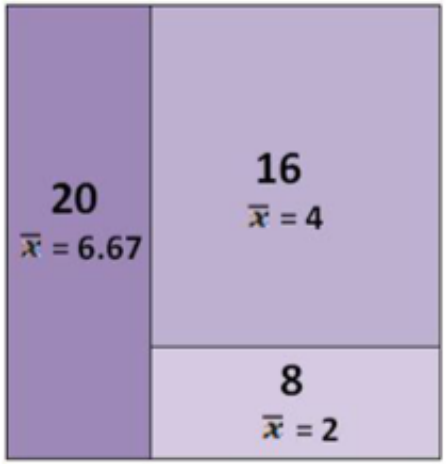
Aggregation problem: variability in results obtained through variations in the shape of areas.



$n = 3,\ \overline{x} = 4.57$

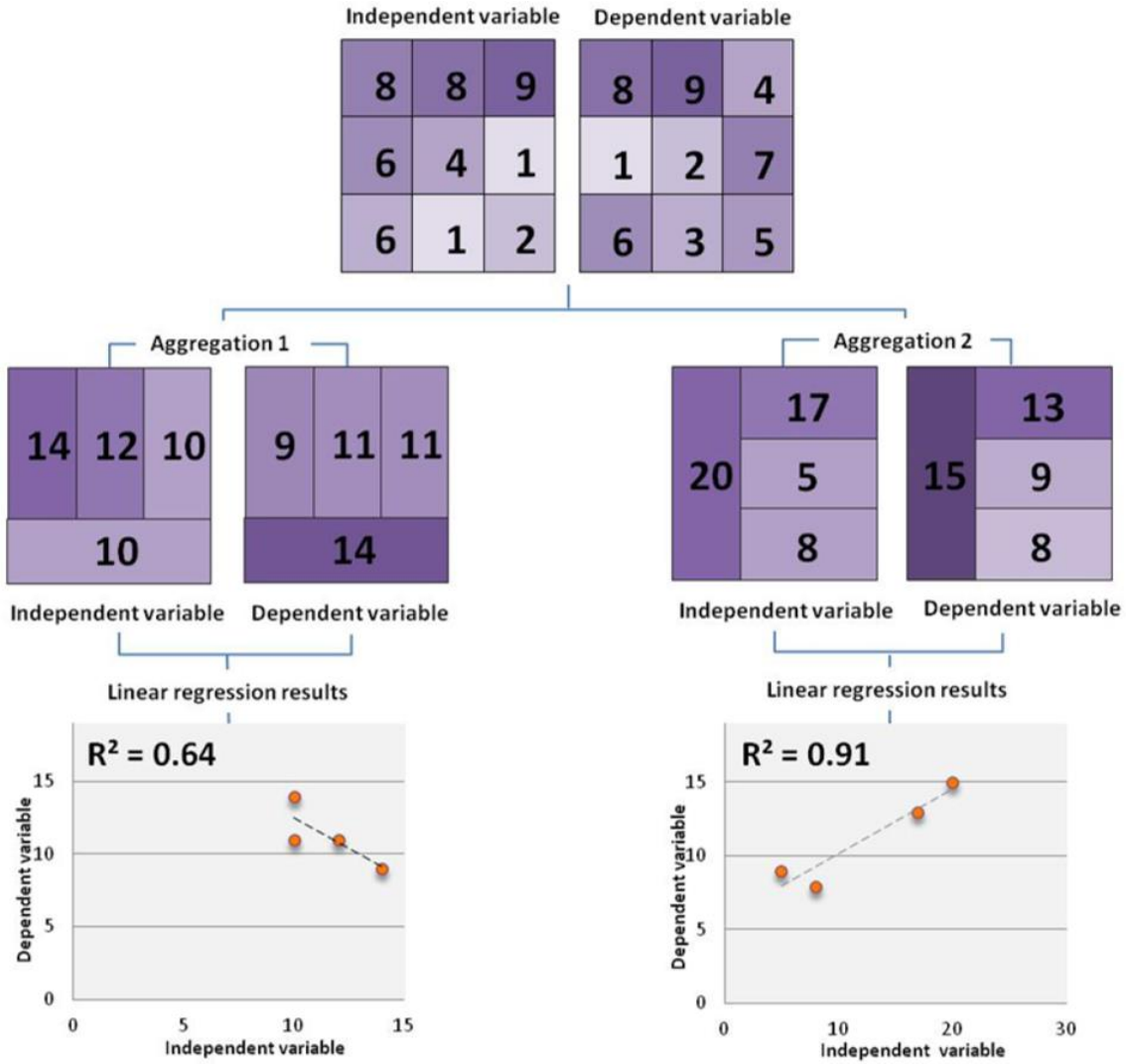


$n = 3,\ \overline{x} = 4.23$



$n = 3,\ \overline{x} = 4.89$

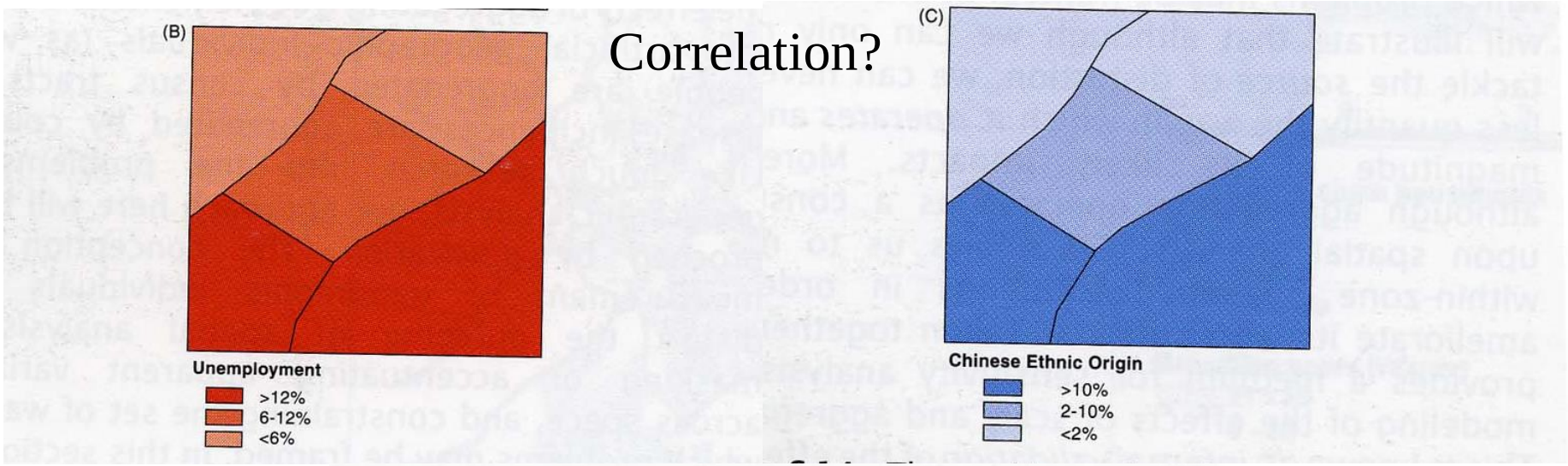
MAUP: Aggregation problem

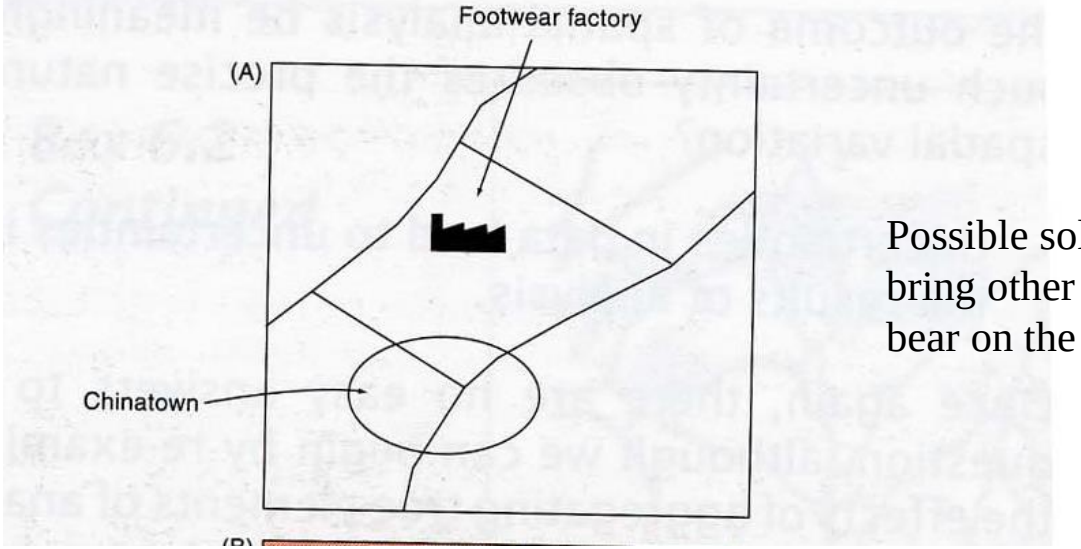


Research design in GIS: Ecological Fallacy

- ***Ecological fallacy:*** attempts to extend the conclusions of one level of spatial aggregation to another (usually finer).
  - happens when you incorrectly apply group-level findings to individuals.
- We cannot ascribe the characteristics of *areas* to smaller units within those areas
- If social policies are based on such conclusions, there could be unfortunate consequences in terms of wasted resources and/or money.

Ecological Fallacy

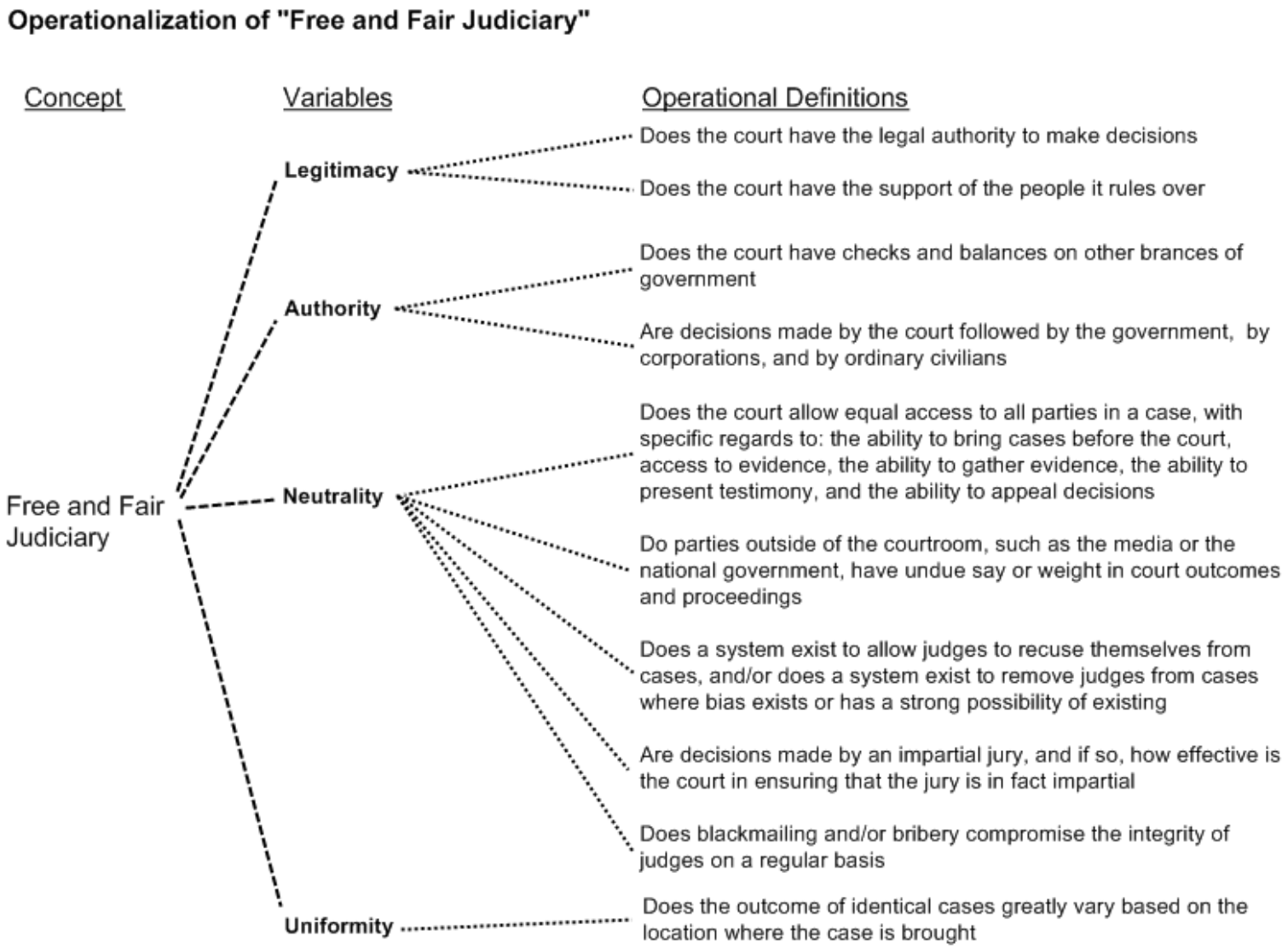




Possible solution: bring other data to bear on the analysis

Research Design: operationalization

Operationalization is the process of transforming abstract concepts into measurable variables that we can analyze.



Research Design in GIS: operationalization

Concept: Sprawl

Variables:

- Street Connectivity/Accessibility
- Low Density
- Segregation of Uses
- Lack of strong center

Operational definitions

Average block length

Number of street links divided by the number of nodes (intersections)

Urban sprawl

Urban development

Concept: Gentrification

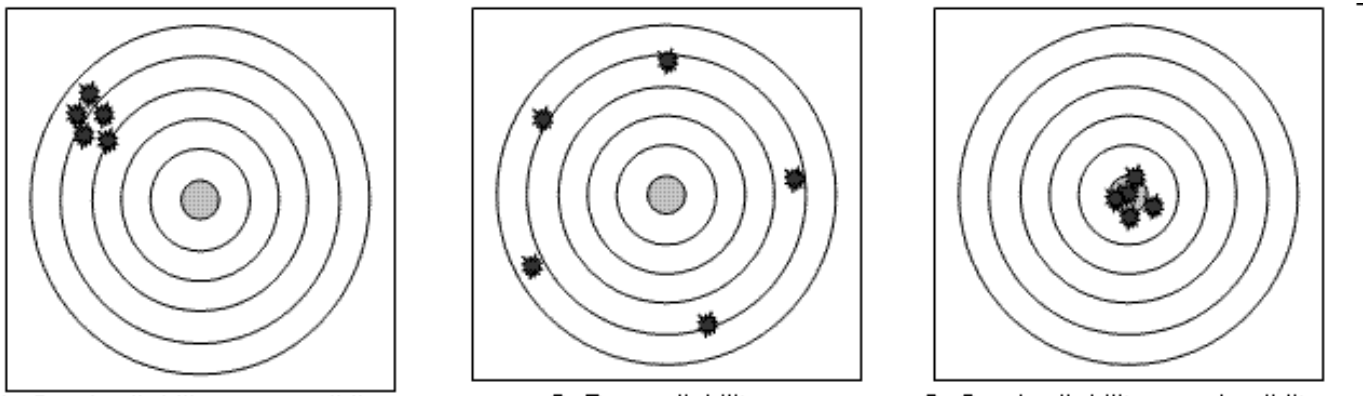
Variables:

- Changing income patterns → *Operational definitions* Average Household income by census tract
- Race → Percent non-white by census tract

Can we use the average block length to measure street connectivity?
Theory: shorter blocks = more intersections, more connectivity, easier to walk

Methodological challenges:

- **Construct validity:** extent to which constructs are successfully represented by operational variables: Does average block length actually measure street connectivity?
- **Internal validity:** extent to which cause-effect relationships can be established: If average block length changes, does street connectivity change?
- **External validity:** extent to which results can be generalized to other settings and populations: Can we use average black length as a measure of connectivity elsewhere?
- **Reliability:** consistency with which a given variable is measured or given result is obtained: if we measure block length will it always give the same result?

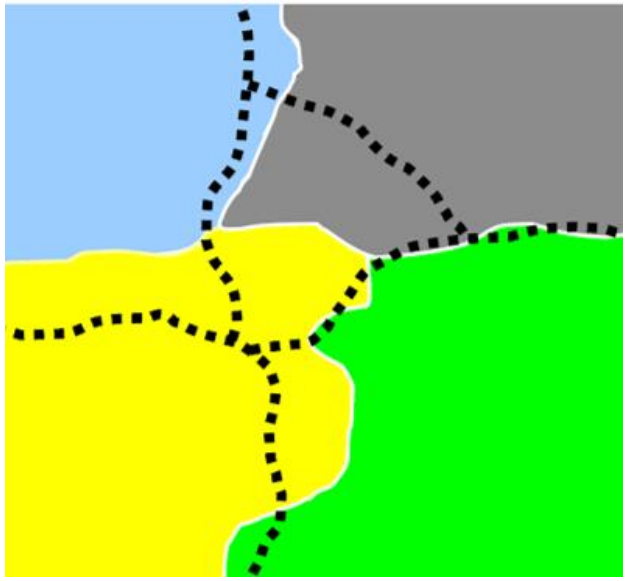


1. Good reliability, poor validity.
2. Poor reliability, good validity (on average).
3. Good reliability, good validity.

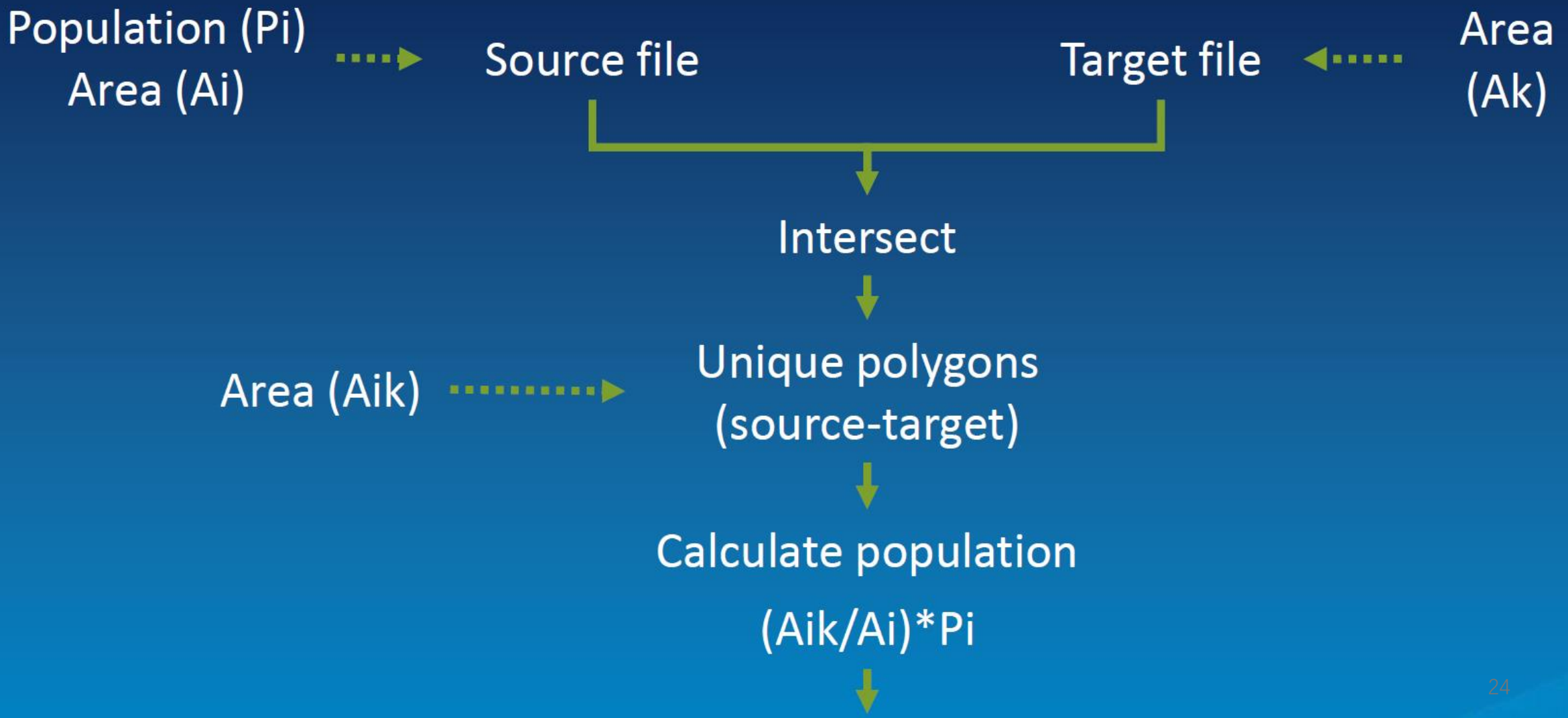
Research Design in GIS: Apportion

Data apportionment allows you to use attributes available within census geographies, such as total population, to calculate information for your custom geographies

- **If we are aggregating data to a lower resolution (i.e. larger areas)**
  - **Geographic boundaries nest within each other**
  - **We can use Dissolve or Summary Statistics**
  - **Possibly use spatial join?**
- **But what if it's a completely different geography?**



Simple area weighting

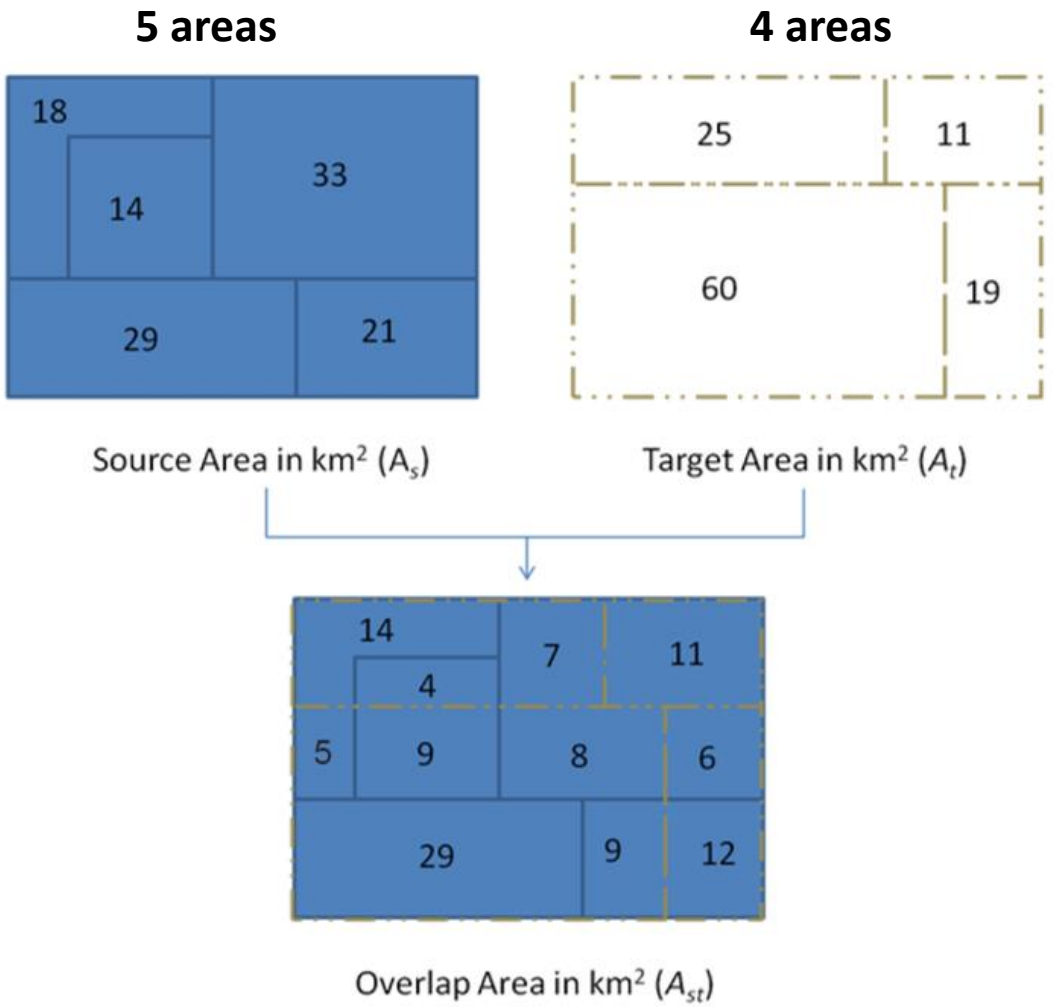


Research Design in GIS: Apportion

Simple Area Weighting

$$V_t = \sum_{s=1}^{s} \frac{V_s \ . A_{st}}{A_s}$$

where: V_t is the value in target zone t;
V_s is the population in source zone s;
As is the area of source zone s; and
A_{ts} is the area of target zone t overlapping source zones.



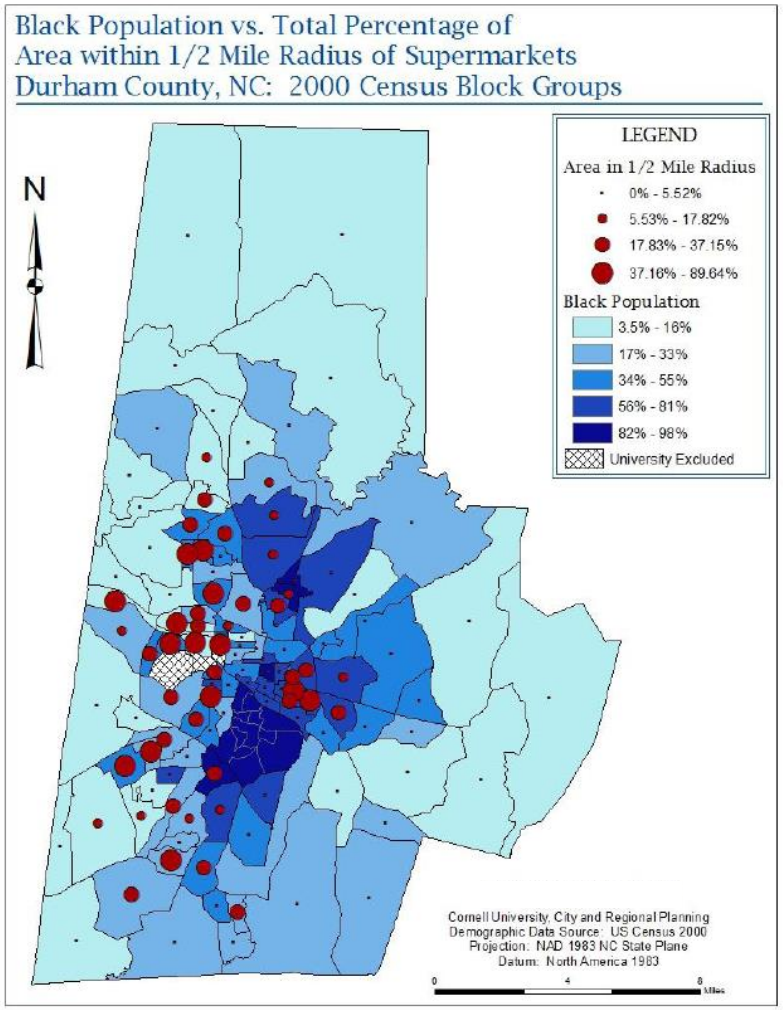
Assumptions

- Only suitable for population counts
  - not suitable for variables that are related to the areal unit in which they are located
- Assumes that the data is homogenous i.e. evenly distributed across an area
- Other approaches should be used where other informative data is available

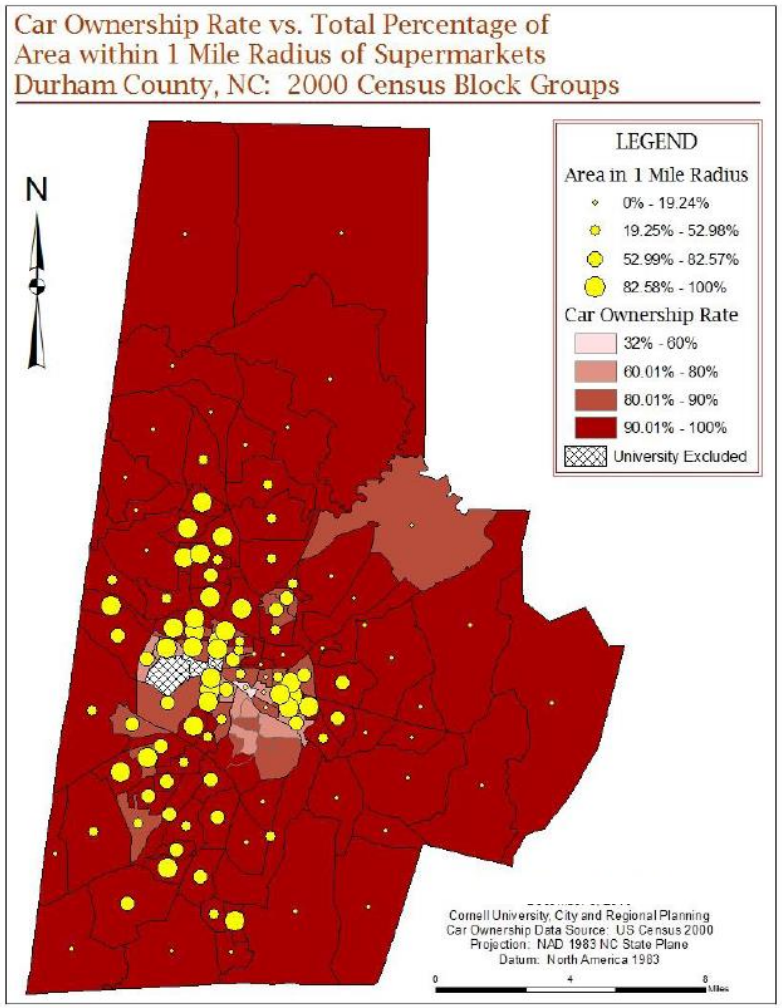
Research Design in GIS - *Triangulation*

- *Triangulation:* multiple variables/data sources all pointing in the same direction

MAP 15



MAP 21



Research Design in GIS - *Triangulation*

- Comparing results across different scales (neighborhood, city, regional, state, etc.)
- Comparison of study area (food desert) with control (entire county) across a number of relevant variables

Table 3: Aggregated Statistics for Urban Food Desert vs. Durham County

	Block Groups	Total Pop	Black	Density (sq.mi)	MHI	MHI % of county	Pov Rate	Car Own	1/2 Mile Radius	1 Mile Radius
Food Desert	11	16701	92%	3520	24487	57%	30%	82%	0%	0%
County	126	217396	40%	734	42741	100%	13%	92%	15%	44%

Environmental Justice case study

Evidence of environmental inequality: literature

- Are the benefits and costs of environmental protection programs distributed unfairly?

- What is the relationship between race and proximity to locally unwanted land uses? Do minority, marginalized, and low-income communities shoulder a disproportionately higher share of environmental risks?

Proving Inequity Exists

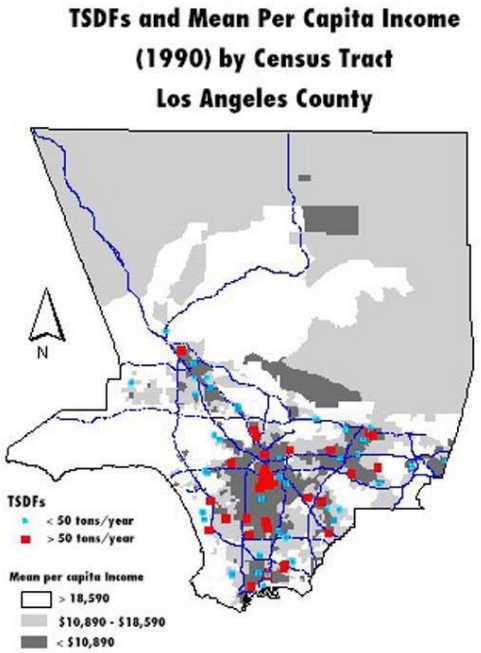
Is race a determining factor in assessing exposure to environmental risk, and is this exposure deliberate?

- Inequitable outcomes easy to prove, because we can observe distributions of things
- Inequitable processes harder to prove, because procedures have formal and informal aspects
- For ex., racism is hard to prove in court: need to show intent, overcome statistical issues
  - Multi-collinearity (income, education)
  - synergies (racism as cumulative, complex phenomenon)

Occidental College Study

The study examined 82 hazardous waste treatment, storage or disposal facilities in Los Angeles County. Using GIS mapping showed what many had long argued: Minorities and poor people are far more likely to live near potential environmental hazards.

- Minorities in Los Angeles County are three times more likely than whites to live within half a mile of hazardous waste treatment or dumping centers
- Race was even more important than income in determining whether a neighborhood had a toxic waste dump
- Some populations are in a one-mile radius of six or seven hazardous waste sites



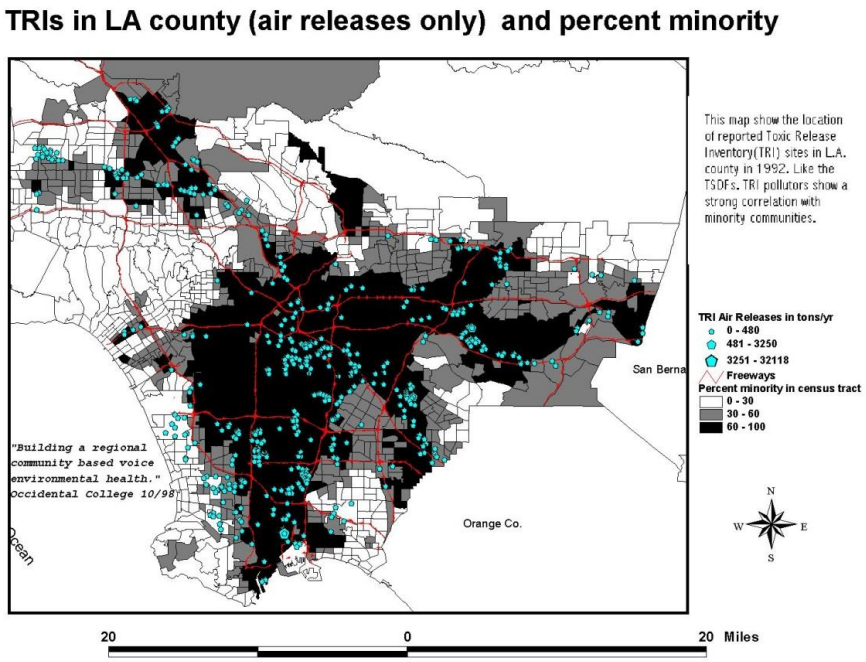
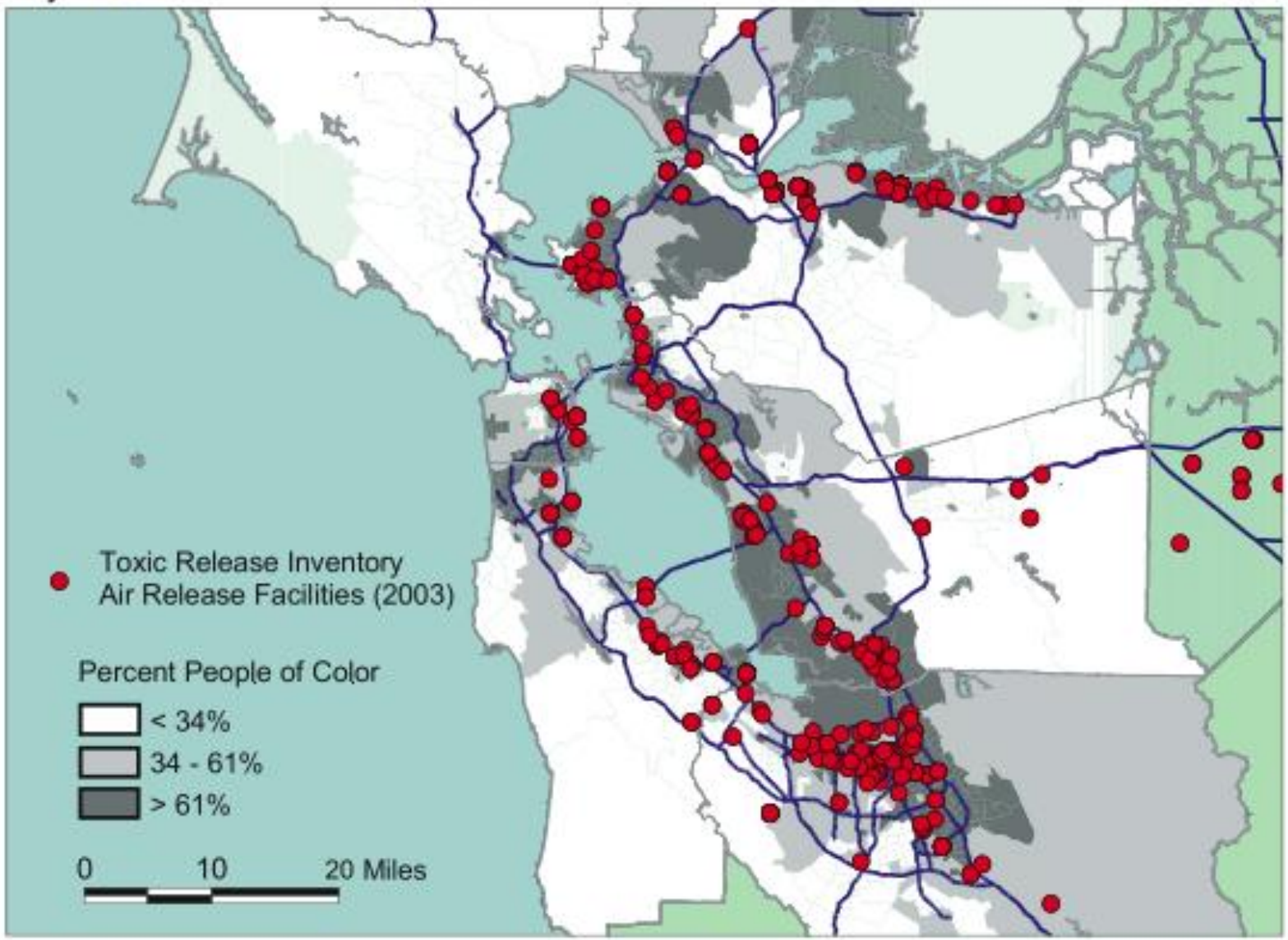
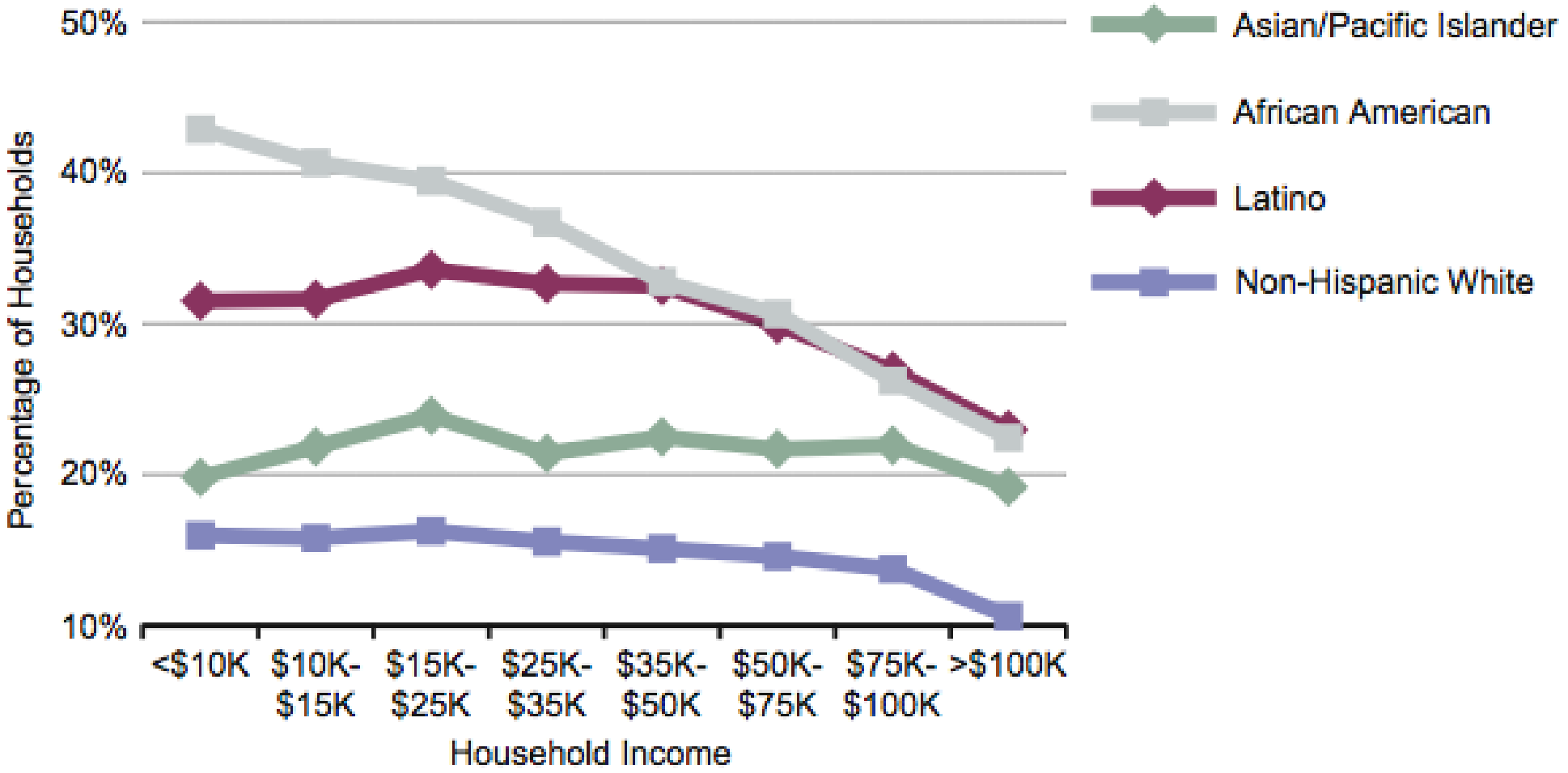


Figure 1: Locations of Facilities with Air Releases (as Recorded in the Toxic Release Inventory or TRI) Relative to Neighborhood Demographics in the 9-County Bay Area



Center for Justice, Tolerance and Community at UCSC

Figure 3: Percentage Households within One Mile of an Active TRI (2003) by Income and Race/Ethnicity in the 9-County Bay Area



Lab #7: Environmental Justice case study

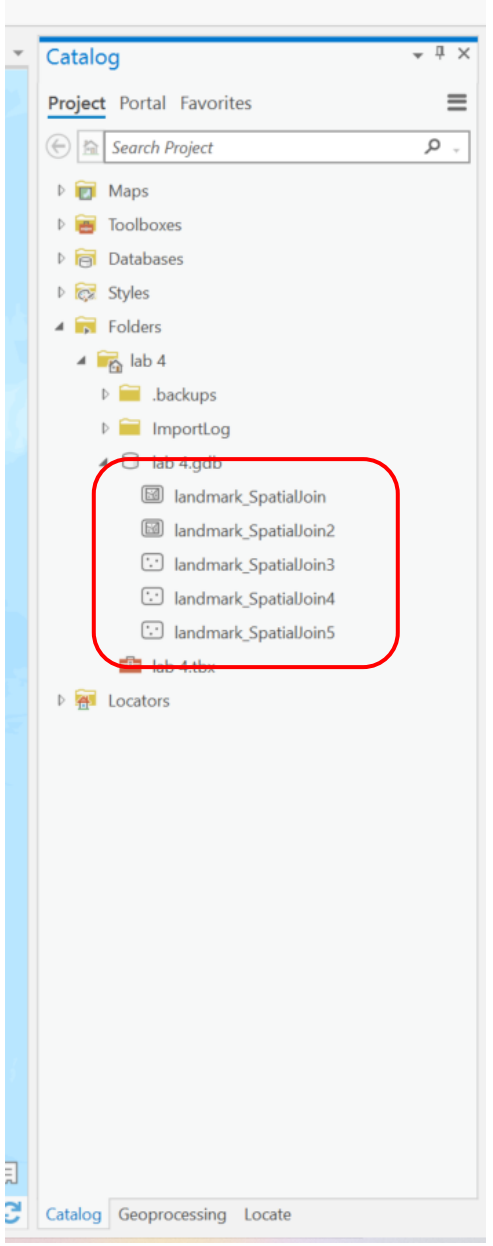
- In this exercise, we will utilize GIS to explore and document issues of environmental justice in Massachusetts.
  - What is the relationship between race and proximity to locally unwanted land uses?
  - Are certain segments of the population disproportionately impacted?

- We will utilize geoprocessing functions to determine the percentage of Blacks and Hispanics living within 1 mile of land fills (apportion!)

- We will compare our results to statewide statistics to come to conclusions about the distribution of Blacks and Hispanics across the state and within metropolitan Boston

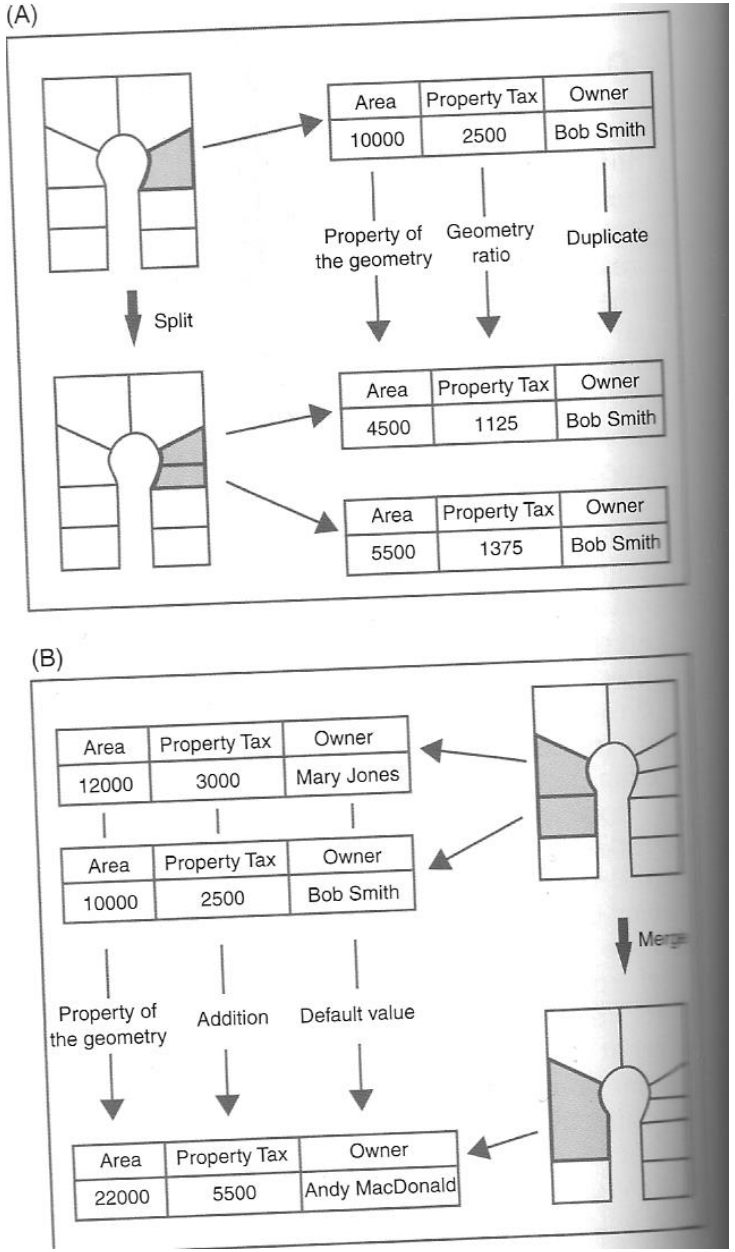
Geodatabase file format

- Creating a new project automatically create a geodatabase to store the data files…
- an object-oriented framework for geographic information
- Users can add behavior, properties, rules and relationships to data to better represent real-world objects. You can use this behavior to support sophisticated modeling of networks, data entry error prevention…
- Supports topology: allows you to represent shared geometry between features
- Support's multi-user editing.



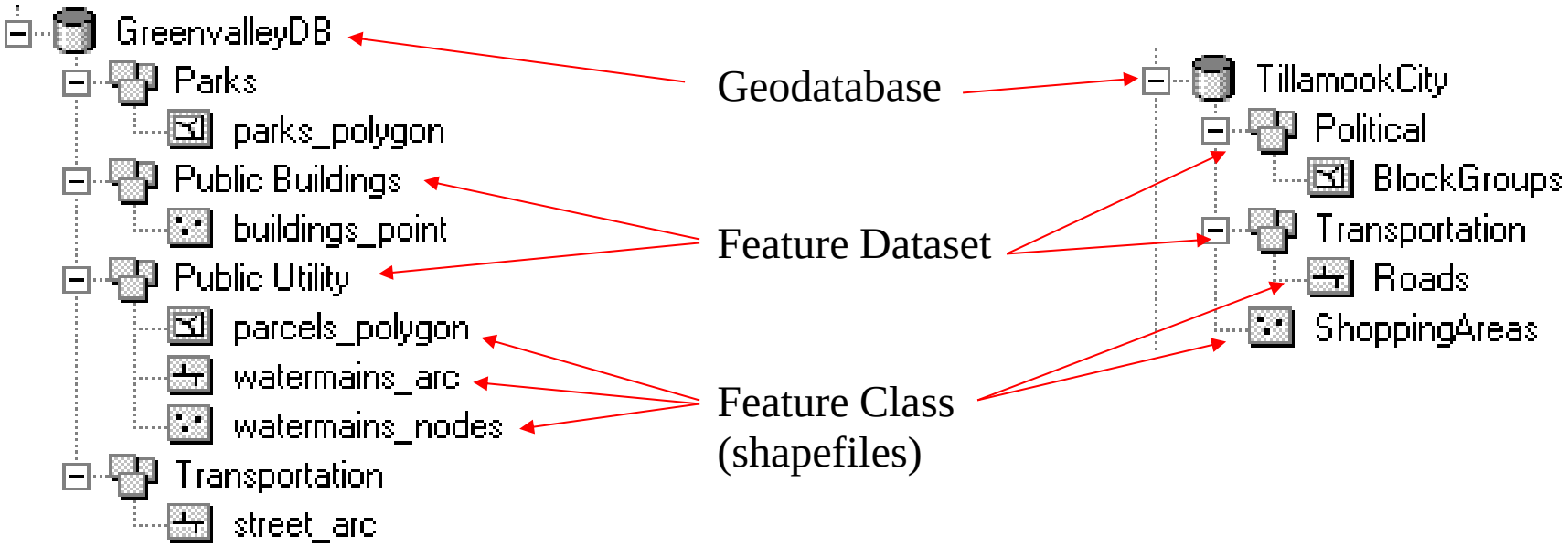
No need to recalculate areas: when a parcel is split, its land use code is transferred to both parcels, but its original land area value is divided in proportion to the areas of the split polgyons

Geodatabase files are transferrable: no need to worry about path files, etc.



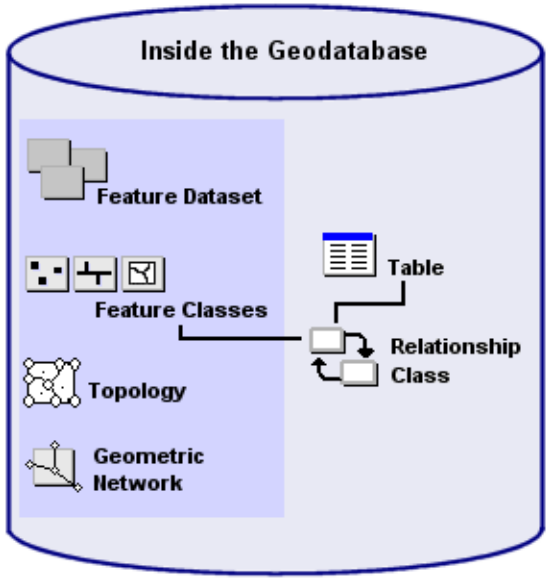
Geodatabase file format

At a minimum, consider it similar to a subdirectory with shapefiles
Unlike shapefiles, you can enforce extents, storage types, projections, topology rules, connectivity rules, network-specific rules, and so on



Geodatabase Data Model

- *Topology rules* – control geographic relationships between feature classes
  - objects in a building class are always inside the objects in a parcel class.
- *Network (connectivity) rules*: A feature can connected to another feature only if certain rules are met
  - Freeway on-ramps must be connected to freeways. A local road can only be connected to a freeway through an on-ramp.
  - Stream lines cannot overlap and must connect to one another at their endpoints
- *Relationship rules* - constrains cardinality (one-to-one, one-to-many, many-to-one, many-to-many) between feature classes and tables.
  - An owner can own multiple parcels, a parcel can have only one owner.



Top 10 States by African American Population Living within a Half Mile Radius of Oil and Gas Facilities (2010 Census)

State	African American Population within a Half Mile Radius	Percent of African American Population in State within a Half Mile Radius
Texas	337,011	10%
Ohio	291,733	19%
California	103,713	4%
Louisiana	79,810	5%
Pennsylvania	79,352	5%
Oklahoma	73,303	22%
West Virginia	13,453	17%
Arkansas	10,477	2%
Mississippi	10,448	1%
Illinois	10,227	1%
TOTAL	**1,052,680**	**2%**

Source: http://oilandgasthreatmap.com

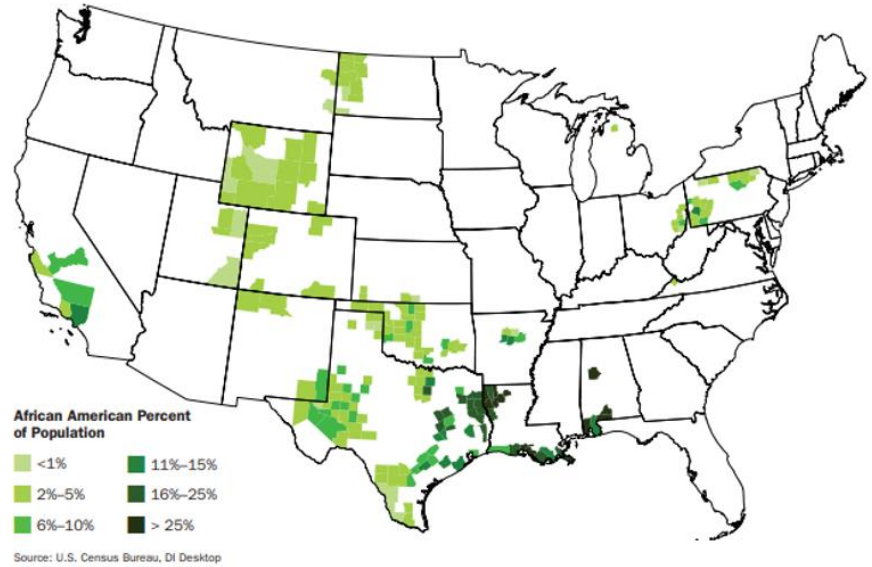
Top 10 States with African American Population Living in Counties Above EPA's Level of Concern for Cancer Risk (2015 Population Data)

State	Number of Counties Above EPA's Level of Concern for Cancer Risk	Total Population in High Risk Counties	Total African American Population in High Risk Counties	Percent of Population in High Risk Counties that is African American
Texas	82	4,189,179	528,357	13%
Louisiana	19	1,027,556	354,952	35%
Oklahoma	40	796,695	37,130	5%
West Virginia	28	804,850	30,589	4%
Pennsylvania	8	624,764	25,071	4%
North Carolina	1	169,866	22,682	13%
Mississippi	2	37,135	17,039	46%
Colorado	6	419,023	7,458	2%
Illinois	13	205,829	7,417	4%
New Mexico	3	247,495	7,093	3%
Total	**238**	**9,086,228**	**1,050,372**	**12%**

Source: "Fossil Fumes," US Census Bureau

FIGURE 2

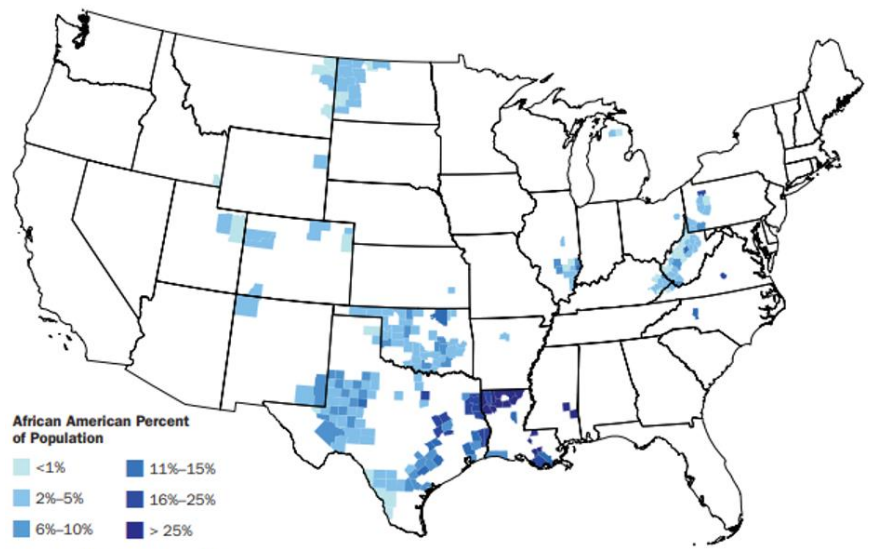
African American Percent of Population in 200 Counties with Highest Oil and Gas Production (2015)



Source: U.S. Census Bureau, DI Desktop

FIGURE 5

African American Percent of Population in Counties above EPA's Level of Concern for Cancer Risk from Oil and Gas Emissions



Source: "Fossil Fumes," U.S. Census Bureau