GIS: Guilty of committing ecological fallacy?

David W. Wong

Professor

Geography and GeoInformation Science

George Mason University

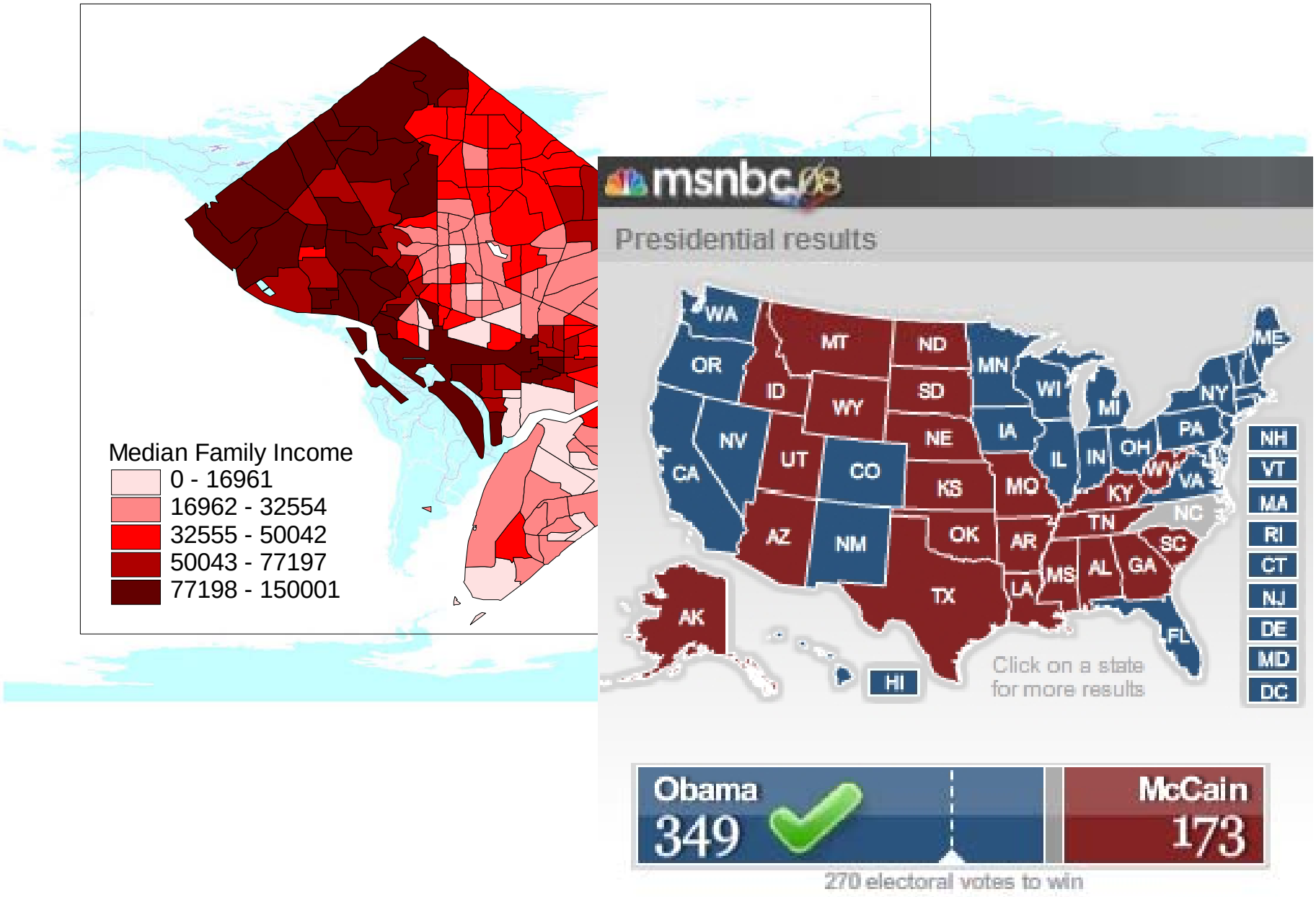
dwong2@gmu.edu

Ecological Fallacy (EF)

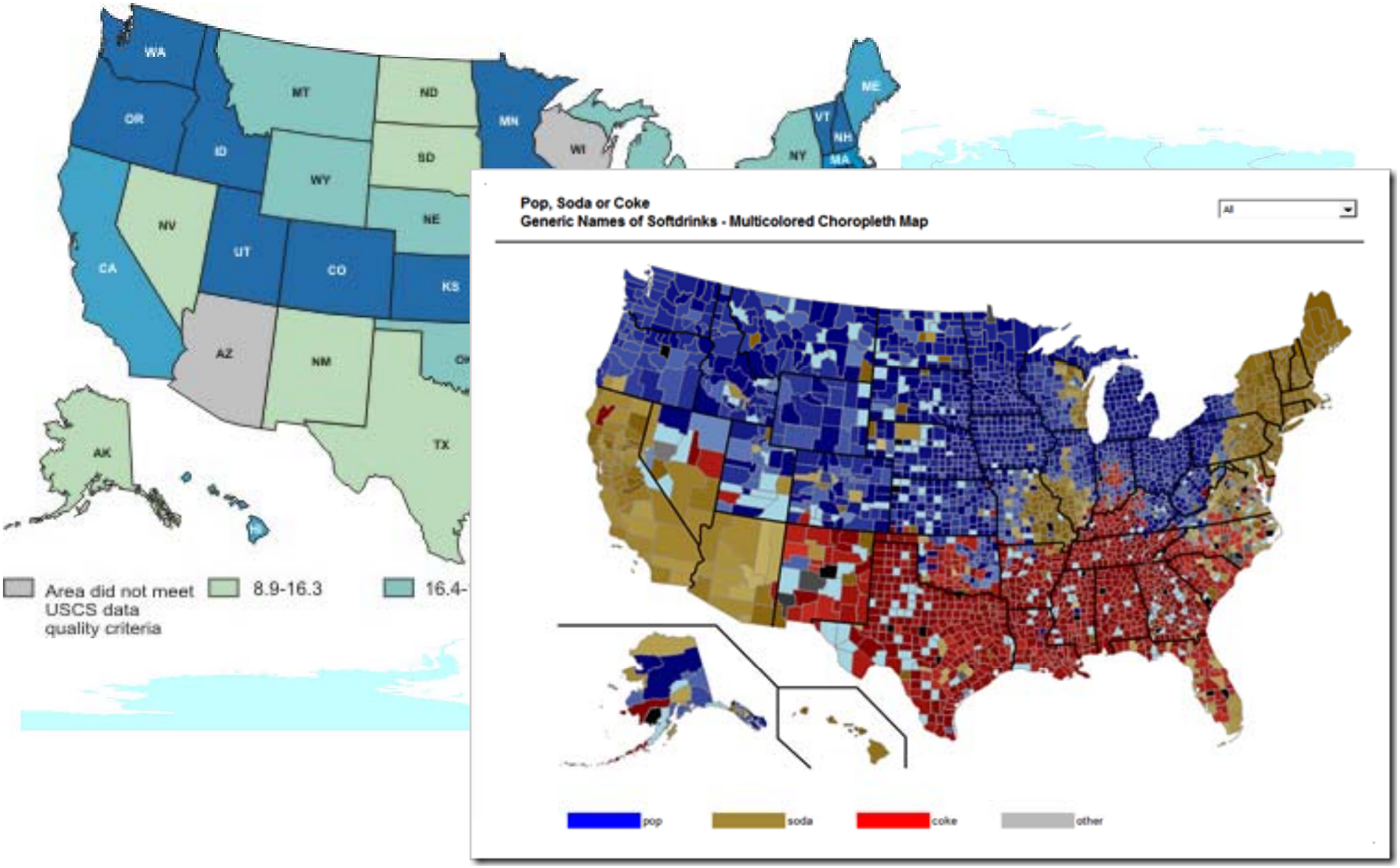
- Many slightly different definitions (trace back to ecological studies)
- Casual one: inferring individual behavior based upon the analysis of aggregated data
- Related to the modifiable areal unit problem (MAUP) – briefly discussed later

EF and GIS

- GIS deals with spatial data
- Most/traditional spatial data are spatially aggregated
- E.g.
  - Census data (by administrative or statistical units)
  - Business data (county business patterns)
- Uses of GIS: mostly for mapping
- Observations/units of mapping: areas, regions
  - Using spatially aggregated data
- Characteristics the areas or regions – o.k.
- Inferring behavior of individuals within the area???
- Analyses in GIS possibly committing EF


Median Family Income
0 - 16961
16962 - 32554
32555 - 50042
50043 - 77197
77198 - 150001
msnbc 08
Presidential results
WA
OR
CA
ID
NV
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MS
MI
IN
KY
TN
AL
OH
WV
GA
FL
SC
NC
VA
PA
NY
ME
AK
HI
NH
VT
MA
RI
CT
NJ
DE
MD
DC
Click on a state
for more results
Obama
349
McCain
173
270 electoral votes to win

Melanoma of the Skin Incidence Rates* by State, 2006



Non-GIS Analyses

Formulate from spatial exploratory analyses

Dep $Var_i = f (var_{1i}, var_{2i}, \ldots)$

=>

$Health\ Outcome_i$, $Political\ pref_i$

$= f (Income_i, \%Minority_i, \%Degree_i, \ldots)$

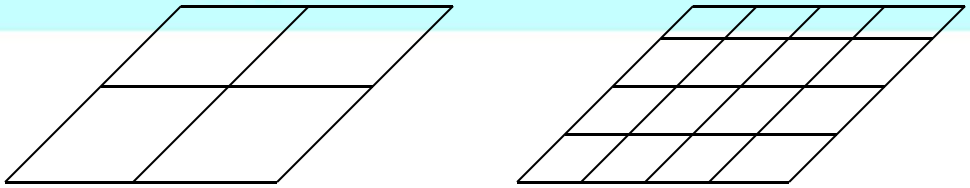
=> Individuals of lower income, in a minority group without a degree are more likely….

The MAUP

Two sub-problems: Scale and Zoning (aggregation)

Scale Effect:

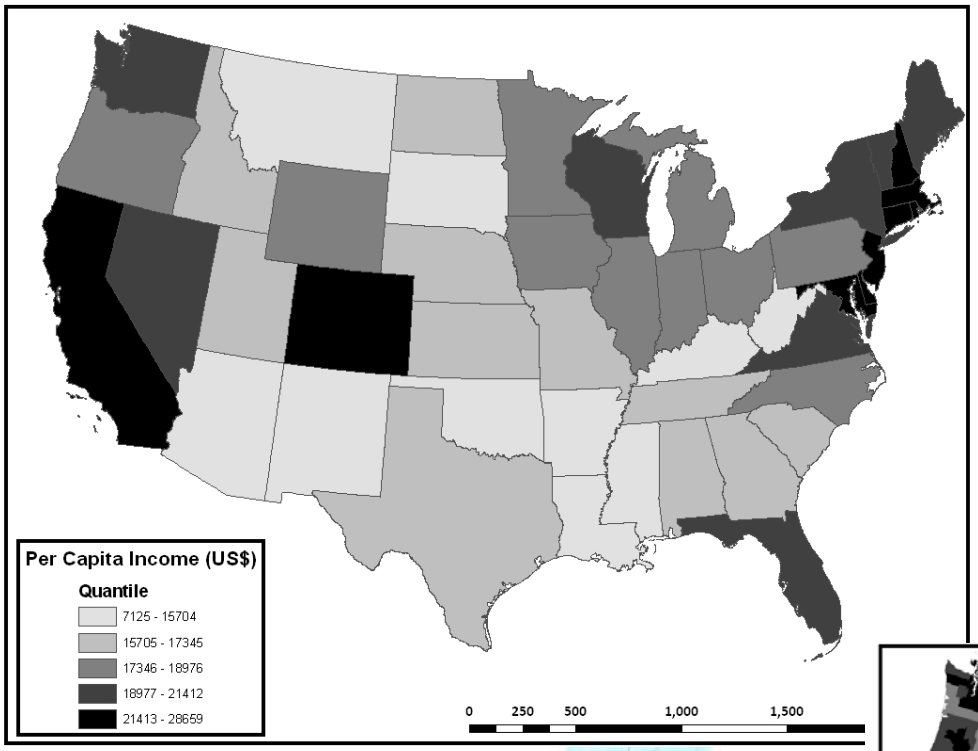
- Variability of analytical results when using data gathered for different sizes of areal units (resolution)
- E.g., counties vs. census tracts

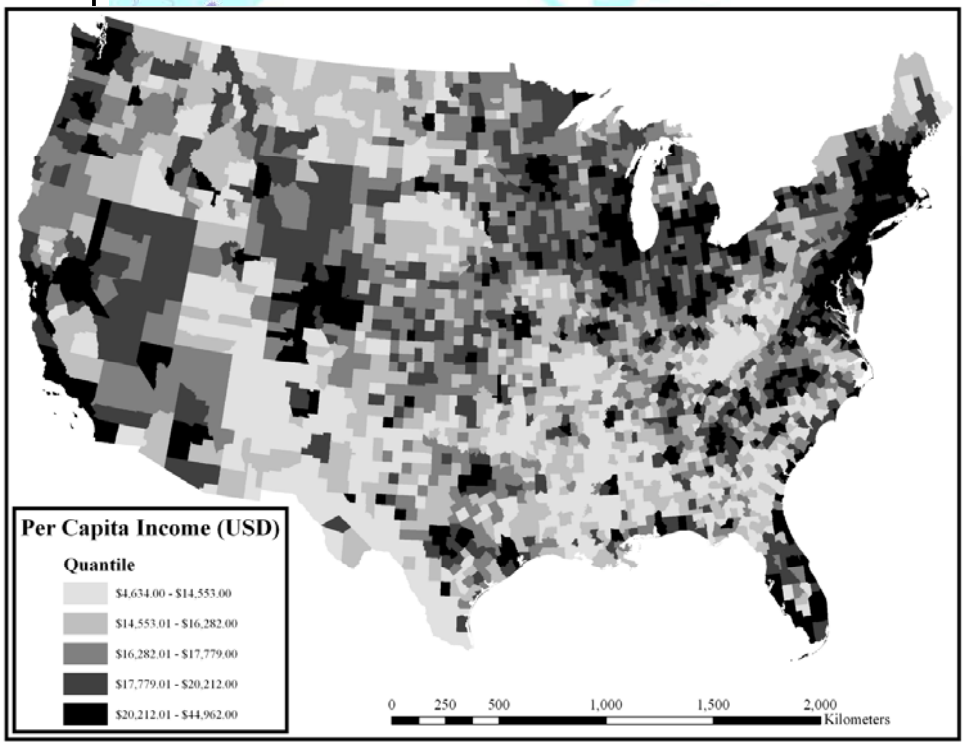


The MAUP

Zoning or Aggregation Effect:

- Variability of analytical results when using data gathered for different spatial partitioning systems.
- E.g., census tracts vs. zip code units
- Solutions: G. King – error bound?





Per-capita Income by State, 1999

Can we avoid committing EF when using GIS?

- Subjects of analysis: areas or individuals
- Do chosen variables for mapping and analyze describe the subjects?
- Explain and interpret the variables explicitly (do they make sense in the context?)
- Methods of analysis/modeling
- Use individual level data
  - Availability?
  - Present the results geographically?
- Use segregation study as an example

Measuring Segregation: Standard Approach

$$D = \frac{1}{2}\sum_i \left| \frac{wi}{W} - \frac{bi}{B} \right|$$

Dissimilarity Index:

- b_i and w_i : black and white population counts in areal unit *i*
- *B* and *W*: total black and white population counts of the entire study region
- $0 \leq D \leq 1$

Sortable List of Dissimilarity Scores

[1990 rank in red,1980 rank in green]

Rank	Area Name	white/black	white/hispanic	white/asian	black/hispanic	black/asian	hispanic/asi
1	Detroit, MI PMSA	85(2)(4)	46(113)(94)	46(99)(99)	78(3)(5)	81(3)(11)	59(44)(59)
2	Gary, IN PMSA	84(1)(1)	48(41)(23)	32(216)(212)	71(6)(26)	82(1)(4)	56(14)(33)
3	Milwaukee-Waukesha, WI PMSA	82(4)(12)	60(22)(27)	41(100)(258)	78(5)(7)	64(34)(22)	52(130)(41)
4	New York, NY PMSA	82(8)(17)	67(6)(7)	51(41)(10)	56(78)(100)	78(11)(25)	57(39)(30)
5	Chicago, IL PMSA	81(3)(3)	62(9)(8)	44(63)(24)	77(2)(1)	82(2)(1)	63(4)(7)
6	Newark, NJ PMSA	80(5)(16)	65(4)(5)	35(261)(248)	61(27)(43)	75(14)(46)	60(7)(10)
7	Cleveland-Lorain-Elyria, OH PMSA	77(6)(5)	58(18)(19)	38(152)(189)	75(1)(3)	75(7)(13)	61(11)(29)
8	Buffalo-Niagara Falls, NY MSA	77(10)(21)	56(30)(54)	47(28)(8)	63(12)(25)	68(13)(18)	55(6)(15)
9	Flint, MI PMSA	77(9)(7)	27(182)(145)	35(164)(139)	64(19)(51)	77(5)(7)	42(103)(91)
10	Cincinnati, OH-KY-IN PMSA	75(17)(25)	30(230)(192)	42(95)(112)	62(23)(80)	68(37)(54)	39(209)(167)
11	Saginaw-Bay City-Midland, MI MSA	74(7)(10)	44(65)(53)	42(92)(40)	48(138)(184)	76(4)(8)	56(28)(18)
12	Nassau-Suffolk, NY PMSA	74(13)(31)	47(92)(131)	36(242)(259)	46(103)(72)	68(30)(48)	48(109)(146)
13	St. Louis, MO-IL MSA	74(12)(15)	29(262)(213)	42(117)(83)	63(9)(24)	70(23)(20)	38(198)(178)
14	Benton Harbor, MI MSA	74(26)(60)	40(143)(125)	44(24)(39)	66(15)(31)	75(6)(26)	45(152)(217)
15	Bridgeport, CT PMSA	74(32)(55)	67(3)(2)	35(199)(297)	23(260)(203)	49(194)(100)	43(162)(21)
16	Miami, FL PMSA	74(35)(19)	44(44)(37)	31(295)(291)	74(4)(2)	66(63)(32)	45(88)(100)
17	Kankakee, IL PMSA	73(21)(48)	44(136)(115)	35(109)(60)	42(63)(58)	75(24)(37)	57(78)(55)
18	Bergen-Passaic, NJ PMSA	73(16)(20)	58(16)(12)	36(223)(202)	44(178)(170)	70(27)(38)	56(22)(25)
19	Birmingham, AL MSA	73(28)(47)	48(179)(146)	47(38)(42)	65(14)(177)	75(8)(62)	47(164)(87)
20	Youngstown-Warren, OH MSA	73(15)(29)	49(53)(39)	38(80)(74)	52(68)(108)	67(19)(33)	53(19)(27)
21	Philadelphia, PA-NJ PMSA	72(14)(28)	60(10)(10)	44(82)(100)	59(24)(35)	66(39)(65)	56(20)(24)
22	Gadsden, AL MSA	71(44)(83)	41(167)(200)	47(5)(36)	44(46)(124)	57(77)(141)	46(41)(125)
23	Peoria-Pekin, IL MSA	71(47)(84)	34(162)(141)	45(54)(101)	46(127)(130)	63(132)(167)	46(85)(116)
24	Fort Wayne, IN MSA	71(19)(36)	46(115)(98)	42(127)(81)	45(82)(95)	63(95)(84)	44(159)(76)
25	Monroe, LA MSA	71(40)(65)	27(259)(162)	47(8)(4)	55(30)(135)	66(21)(88)	45(27)(69)
26	Beaumont-Port Arthur, TX MSA	71(48)(45)	50(131)(153)	54(4)(1)	50(88)(61)	58(99)(66)	37(148)(129)

How Segregation Values are used?

- Traditionally, cross-region comparison at the Metro or city level using census tract level population data
- A summary value for the entire region – global measures
- More recently, “zooming” into the neighborhood level to conduct local analysis (local measures)

How Segregation Values are used?

Reasons for Local Analysis:

1. Recognizing the (spatial) variation of segregation within a region – part of the trend in geography to perform local analysis and the development of local statistics
2. The development of multi-level framework in social science in general and in social epidemiology in specific

- Individual-level outcome = f (individual characteristics, neighborhood characteristics, regional/metro characteristics)

Development of Local Segregation Measures

- Decomposing global measures to smaller census units (tracts, or block groups)
- Each unit has a segregation value
- Some local measures are spatial modifications of traditional measures (e.g., exposure index)
- Some are local spatial autocorrelation measures from Geography

How local is local?

- From metro to census tract, block group or block
- How far can / should we go?
- Smaller units – more homogeneous, more segregated according to D (part of the MAUP)
- Individual level – perfect segregation using D
- A Different conceptualization of segregation based on interaction

Typology of Segregation Measurement (I)

Socio-geographical Spaces

1. Residential Space
   - Most emphasized, high significance on **impacts**
     - overburdening of undesirable conditions
     - difficulties to access facilities or opportunities
2. School Space (for school-age children)
   - Closely tied to residential space in N. Am
   - Similar impacts as in residential space
   - Additional: socio-psychological isolation

Typology of Segregation Measurement (II)

Socio-geographical Spaces

3. Work Space (for working adults)
   - 8 hrs or more (> 50% of non-sleeping hrs)
4. Others: cultural, religious, entertainment

- Using “Activity Space” concept to evaluate segregation in multiple socio-geographical spaces

Conceptual Framework

- Activity Space (Abler et al., 1971)
  - Geographical space or area defined by locations of various types of activity that an individual participates
- Identify individual's activity spaces
- Overlap of activity spaces of individuals in different population groups – reduced segregation

Operational Method

1) Derive activity spaces of individuals based upon locations individuals have visited
2) Evaluate the population (racial-ethnic) characteristics of visited locations or neighborhoods
3) Compare the racial-ethnic characteristics of individuals/subjects with visited neighborhood characteristics
4) Differences: reduced segregation
5) Individual exposures to other groups can be aggregated to neighborhood levels by residential locations

Operational Method (ii)

- Ω_{ij} :the set of locations representing the activity space of individual j in group a, residing in unit i
  - These locations: S_{ijk} where k = 1, 2, 3, …n_{ijk}, $S_{ijk} \in \Omega_{ij}$
- the total population in group b that individual j is exposed to = the total population of group b in the activity space Ω_{ij}

$$B_{ij} = \sum_{k=1}^{n_{ijk}} b_{S_{ijk}}$$

- total population that an individual j in i is exposed to within his or her activity space Ω_{ij} is

$$T_{ij} = \sum_{k=1}^{n_{ijk}} (a_{S_{ijk}} + b_{S_{ijk}})$$

- the exposure of an individual j in group a in unit i to group b within the activity space of j will be

$$E_{ij,a*b} = \frac{1}{A} * \frac{B_{ij}}{T_{ij}}$$

Operational Method (iii)

$$E_{ij,a*b} = \frac{1}{A} * \frac{B_{ij}}{T_{ij}}$$

- individual version of Lieberson's exposure measure
- For the exposure of all individuals in group a in unit i to group b, we need to sum the individual exposure level for all individuals a_i

$$E_{i,a*b} = \sum_{j \in i}^{a_i} \frac{1}{A} * \frac{B_{ij}}{T_{ij}}$$

- the exposure of population group b in the same unit i to group a in the study region by just switching the related terms

An Illustrative Example

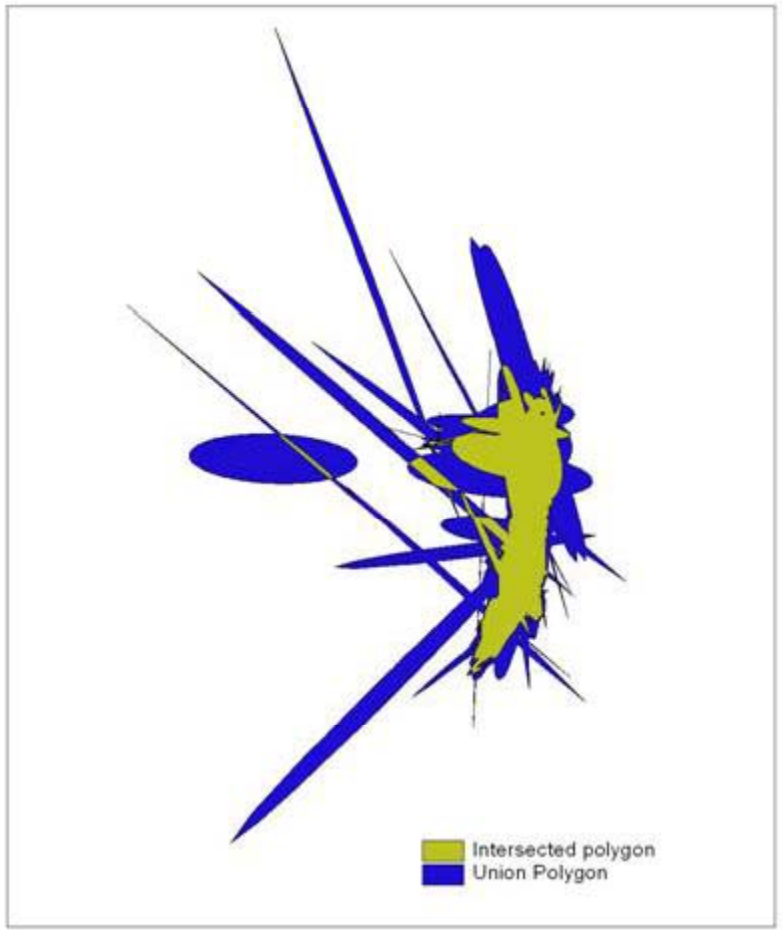
- travel or commuting data at the individual level – travel diary
- report where they have visited in given days
- a tri-county area of southeast Florida (Miami-Dade County, Broward County, and Palm Beach County)
- each participating household reported all travel-related activities for every resident of the household (including infants) for an entire 24-hour survey day
- **race-ethnicity variable not included

Data Processing

- dataset were evaluated, examined, reformatted and imported to a GIS
- assigned a race, either white or black, to the sampled individuals probabilistically within a census tract based upon the white-black proportion of the tract provided by the 2000 Census data
- the results: do not reflect the true racial-ethnic experience of the population in the tri-county area, but just to illustrate how the framework can be implemented

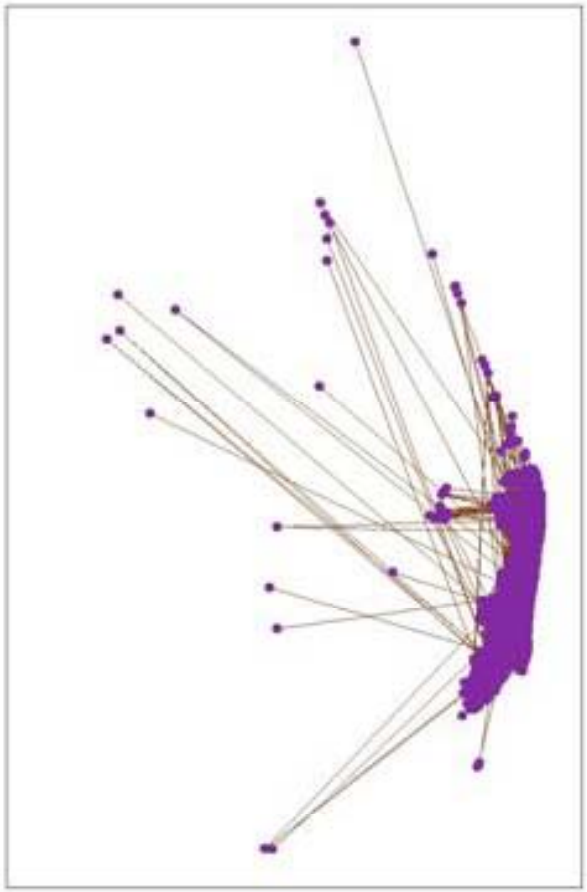
Regional Level

- Compare two population groups (White-Black) at regional scale
- Given visited locations by individuals, derive ellipses (act space) by individuals
- Merged individual activity spaces by two groups – polygons of group activity spaces
- Overlay the polygons to derive the S index
- S = 0.3034 (hi-low – relative)

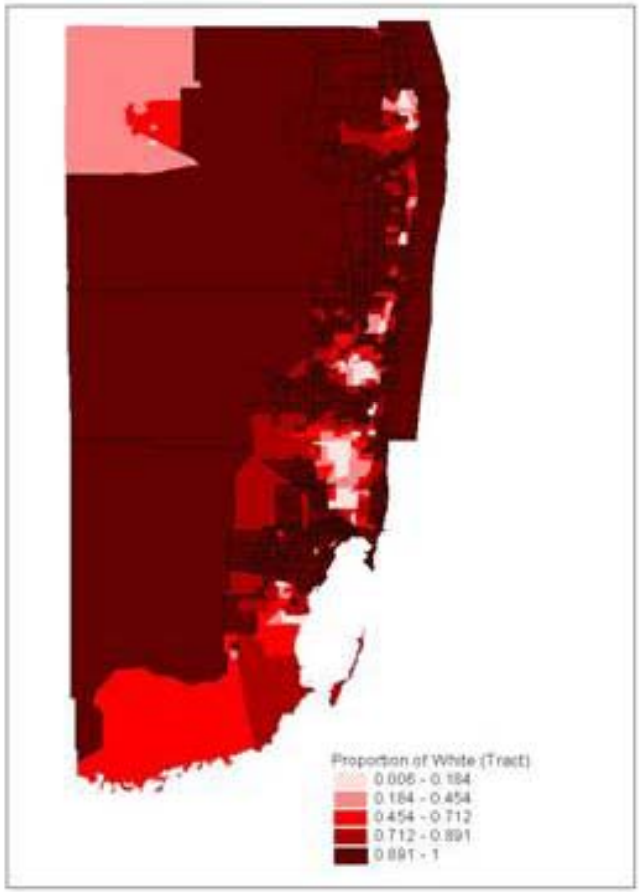


Local Level

- segregation at the local level
- reflect how the neighborhood residents were exposed to or isolated from the other group based upon their activity spaces
- For each census tract: compute the exposure measures of all individuals as defined by their activity spaces
- Two act-space implementation methods:
  1. Including all tracts intersected by the lines connecting all visited locations (line-based)
  2. including those tracts with the visited locations, but not the tracts in-between the locations (point-based)



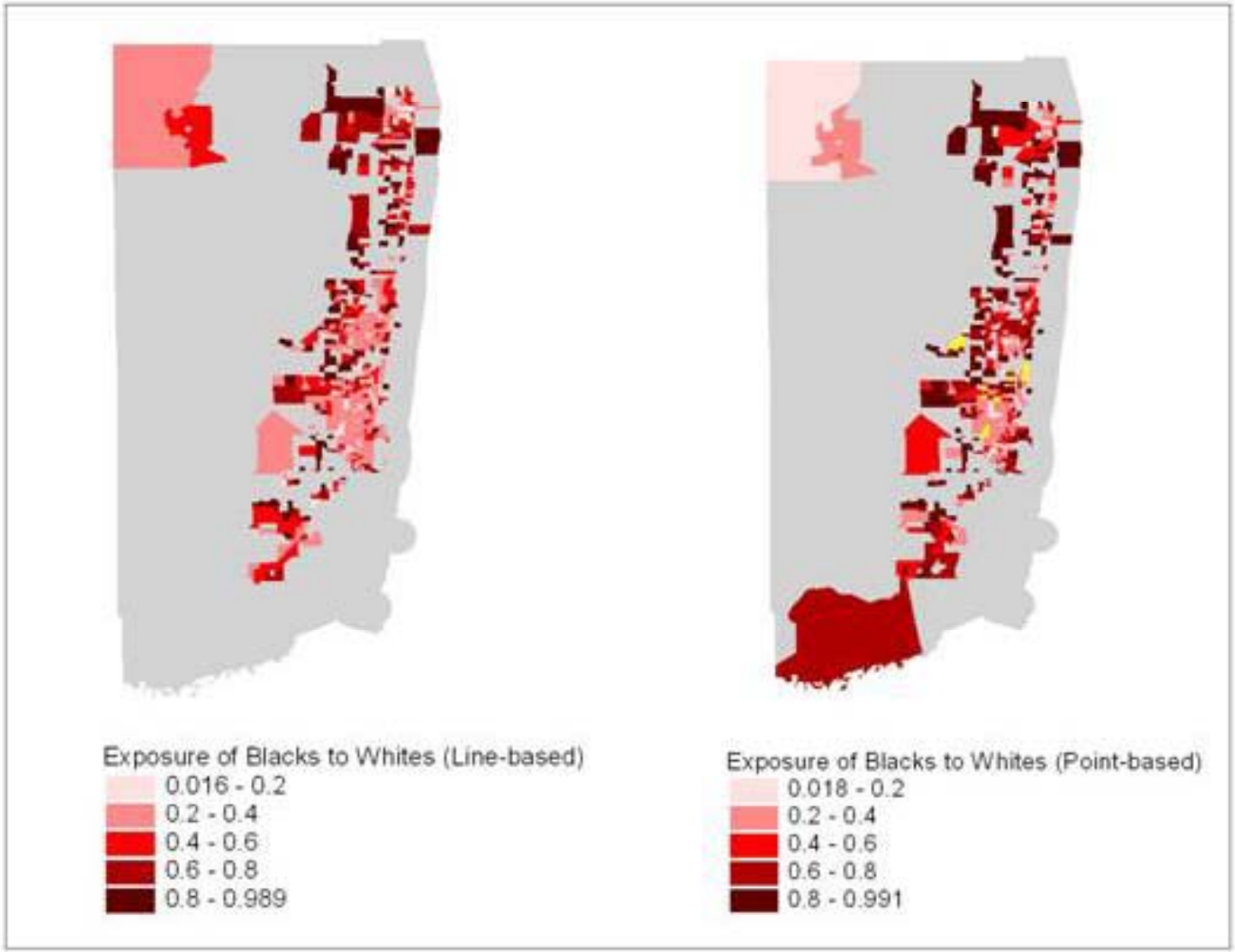
(a)


Proportion of White (Tract)
0.006 - 0.184
0.184 - 0.454
0.454 - 0.712
0.712 - 0.891
0.891 - 1

(b)

Exposure of Whites to Blacks

Exposure of Blacks to Whites



Summary & Conclusions

- Evaluate segregation at both the individual and neighborhood levels
- Data-individual level, results – aggregated to the spatial levels of interest
- Results shown in GIS, but have an individual-behavioral foundation
- Need individual-level data
- Aggregated data: used support the analysis
- Be careful to draw inference on individuals when using aggregated data in GIS
- ***Maps lie***

J Geogr Syst
DOI 10.1007/s10109-010-0112-x

ORIGINAL ARTICLE

Measuring segregation: an activity space approach

David W. S. Wong · Shih-Lung Shaw

Thank You Very Much!!

Questions?

Summary & Conclusions -2

- validity of the exposure concept: contact theory in social psychology
- positive intergroup contact can reduce prejudice
- environment or geographical pattern: facilitates intergroup contact
- exposure values: potentials for groups to interact
- Many existing technical issues (e.g., how long do they visited the location? Data Quality)
- But demonstrated the feasibility of the framework that higher quality data can support

Implications

- Residential space – important, but comprehensive evaluation of segregation should include other spaces
- Implementation does not identify those spaces explicitly, but can be done
  - Exposure weighted by duration in each space
- Civil rights – focused on residential/school space mostly
- If other spaces are important in evaluating segregation, importance of residential/space policies should be re-evaluated

What do we measure? - 1

Massey & Denton: five dimensions:

- evenness, exposure-isolation, concentration, centralization and clustering

Reardon and O'Sullivan (2004):

- two composition dimensions of evenness-clustering and isolation-exposure
- centralization is out

Brown and Chung (2006):

- evenness is similar to concentration
- clustering represents low exposure
- two: evenness-concentration and clustering-exposure.

$$S = 1 - \frac{Area(E_1 \cap E_2 \cap E_3 \ldots E_n)}{Area(E_1 \cup E_2 \cup E_3 \ldots E_n)}$$