An AMOEBA Procedure for Visualizing Clusters

Marta Jankowska[1], Jared Aldstadt[2],
Arthur Getis[1], John Weeks[1], Grant Fraley[1]

[1]San Diego State University/ Department of Geography
mjankows@rohan.sdsu.edu

[2]University at Buffalo/ Department of Geography
geojared@buffalo.edu

Outline

- Background
- Data
- What is AMOEBA?
- Visualization of clusters
- Controlling cluster size
- Results
- Conclusions

Neighborhoods in Accra

- Project: "Health, Poverty and Place: Modeling Inequalities in Accra Using RS and GIS"

  - Key hypothesis: neighborhood level effects are influencing morbidity and mortality patterns of individuals in Accra, Ghana

- Neighborhood definitions

  - Purpose (depends on context)

  - Variables (social, physical)

  - Size (aggregation)

AMOEBA
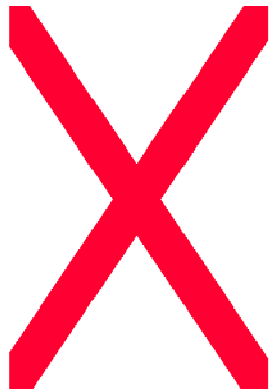
- AMOEBA: A Multidirectional Optimum Ecotope-Based Algorithm, as developed by Aldstadt and Getis (2006).

  - Hot and cold spot cluster identification

  - Spatial association of unit to surrounding units (neighborhood theory predicts like attracts like)

- Applicability to the project

  - Can AMOEBA create significant neighborhood boundaries?

  - Exhaustive feature allows all units to be included in clusters

  - Cluster size limitation feature

Data

Study Site and Census

- Accra, Ghana

- 2000 Ghana Census
  - 1,731 census tracts or enumeration areas (EAs)
  - 1,658,883 individuals
  - 366,853 households

Ghana

Accra

Gulf of Guinea

0   750 1,500   3,000
                    Meters

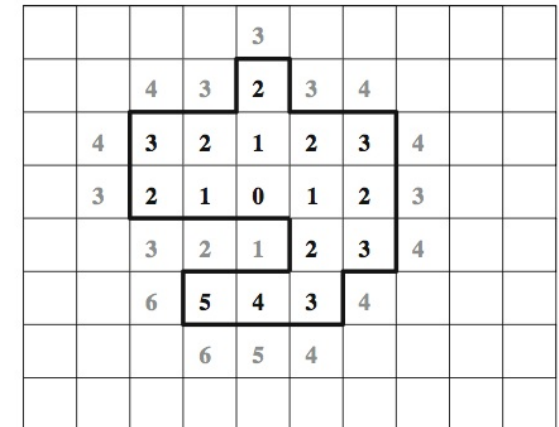census tract boundaries
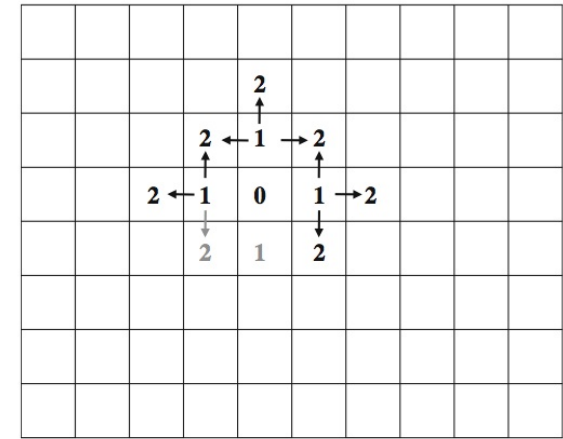
Data
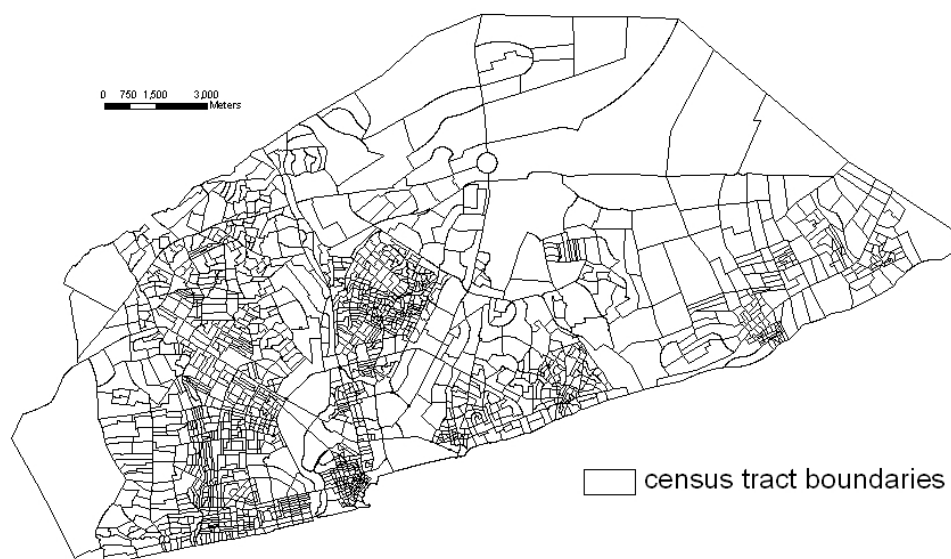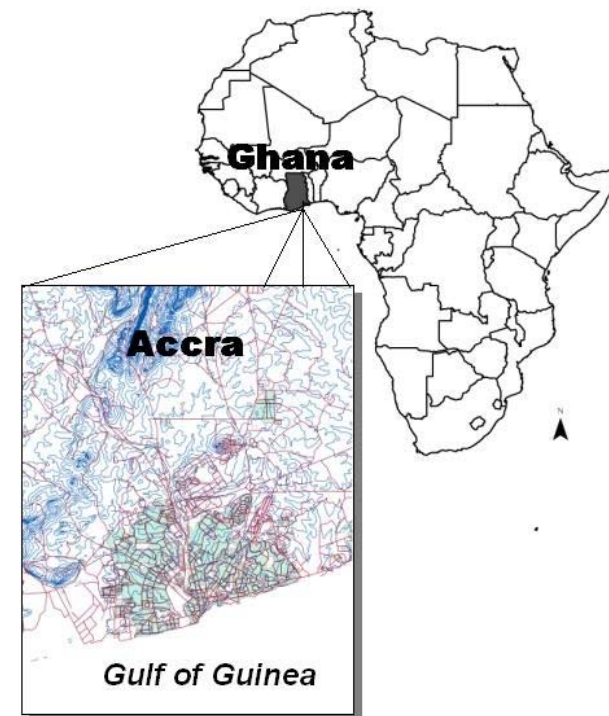
Variables

- AMOEBA is a single variable algorithm, so neighborhood definition must be expressed as one variable.

- Individual (social)
  - Literacy
  - Professional vs. informal workforce

- Household (physical)
  - Liquid and solid waste disposal
  - Household building materials

- Combined in a principle components analysis, and first principle component, called the slum/SES component, is used as the input to AMOEBA
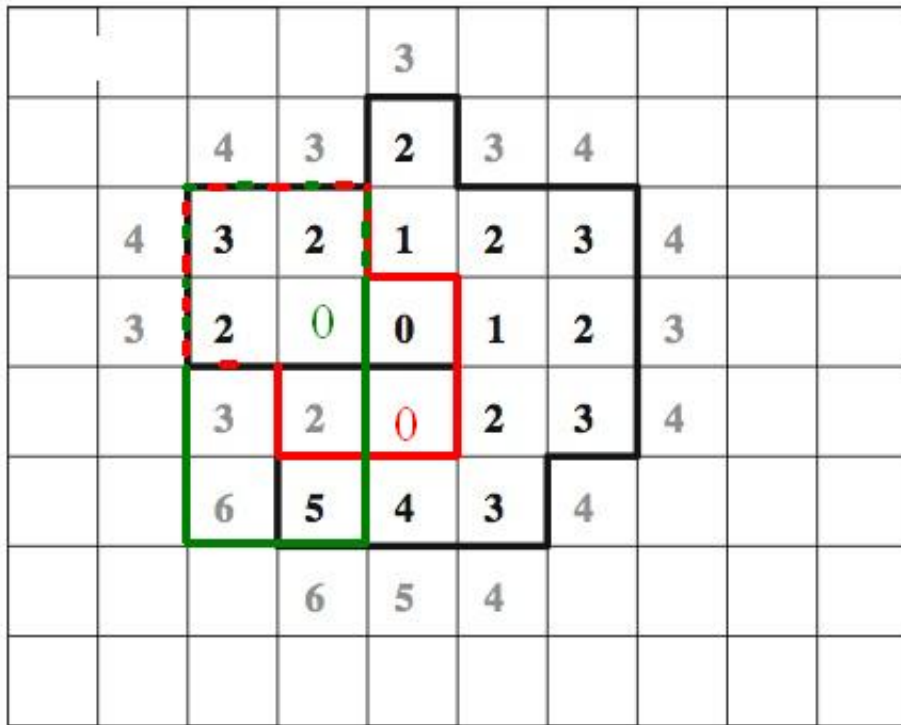
Building Clusters

- Stand alone, single variable algorithm built in Python 2.5.2
- Creation of weights matrix using $Gi*$ or any other local statistics
- Seed location
  - Neighbors that increase $Gi*$ are included in the cluster
- Results in indication of
  - Hot and cold spots
  - How spatially autocorrelated a cluster is

Aldstadt and Getis, 2006

Exhaustive AMOEBA

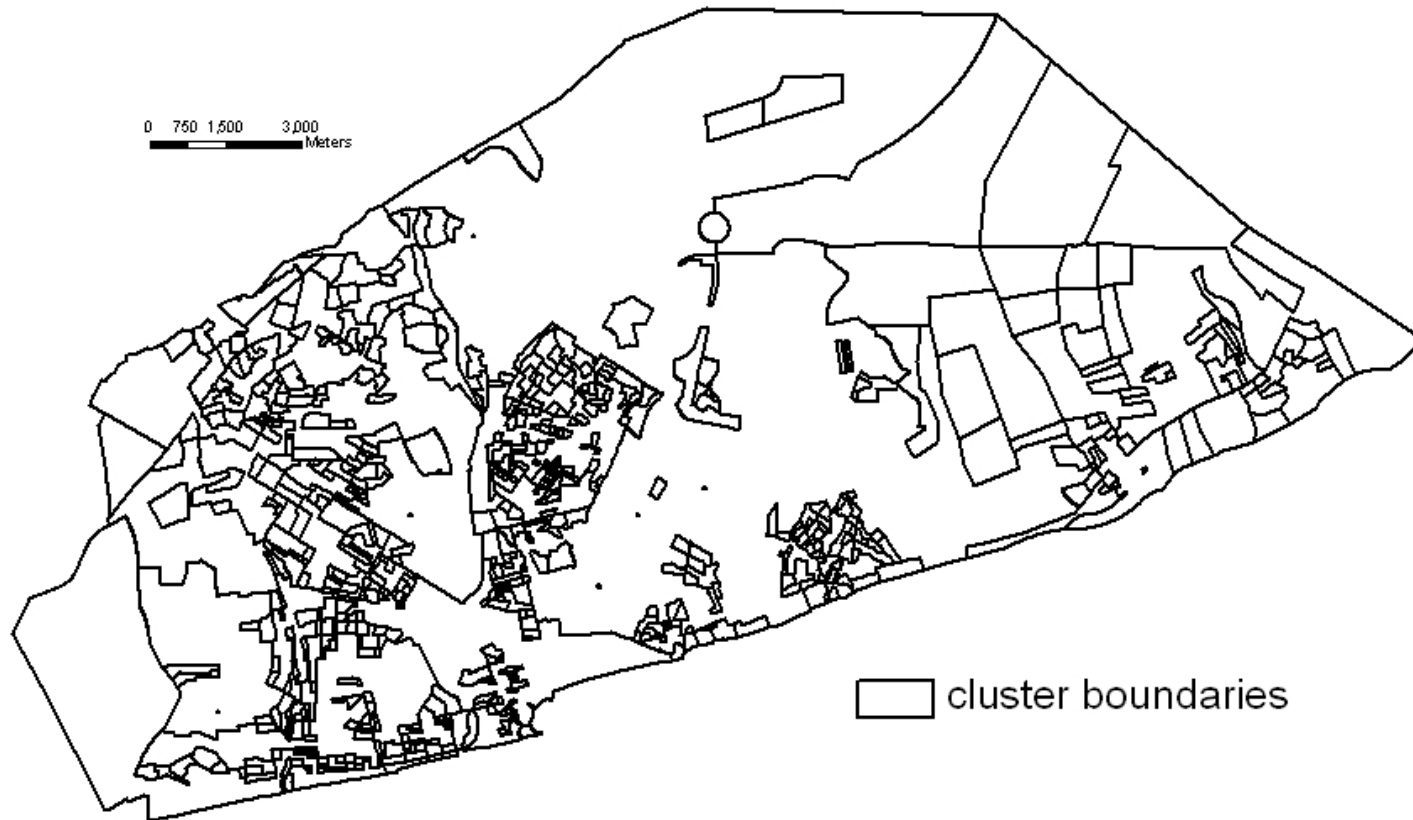- Project utilizing newer version of AMOEBA - Exhaustive
  - All units are included in clusters
- Seed is generated in every unit
  - All units belong to multiple clusters
  - Considerable visualization challenges

Visualizing Overlapping Clusters

- AMOEBA outputs a number of statistics for each observation including:
  - Gi*max - the largest Gi* value of all the clusters the observation can be found in
  - Cluster max - an id number of the largest Gi* value cluster the observation belongs to
  - Pmax - the p value for the observations' Gi*max value

Visualization of clusters

Cluster max

0   750 1,500   3,000
■■■■■□□□□□■■■■■ Meters

☐ cluster boundaries

Visualization of clusters

Gi*max

Exhaustive AMOEBA
Gi*max values of ses/slum component

Gi*max

■ (dark green)	< -1.2 Std. Dev.
■ (green)	-1.2 - -0.75 Std. Dev.
■ (light green)	-0.75 - -0.25 Std. Dev.
■ (cream)	-0.25 - 0.25 Std. Dev.
■ (yellow)	0.25 - 0.75 Std. Dev.
■ (orange)	0.75 - 1.3 Std. Dev.
■ (red)	1.3 - 1.4 Std. Dev.

0   1,500   3,000   6,000 Meters
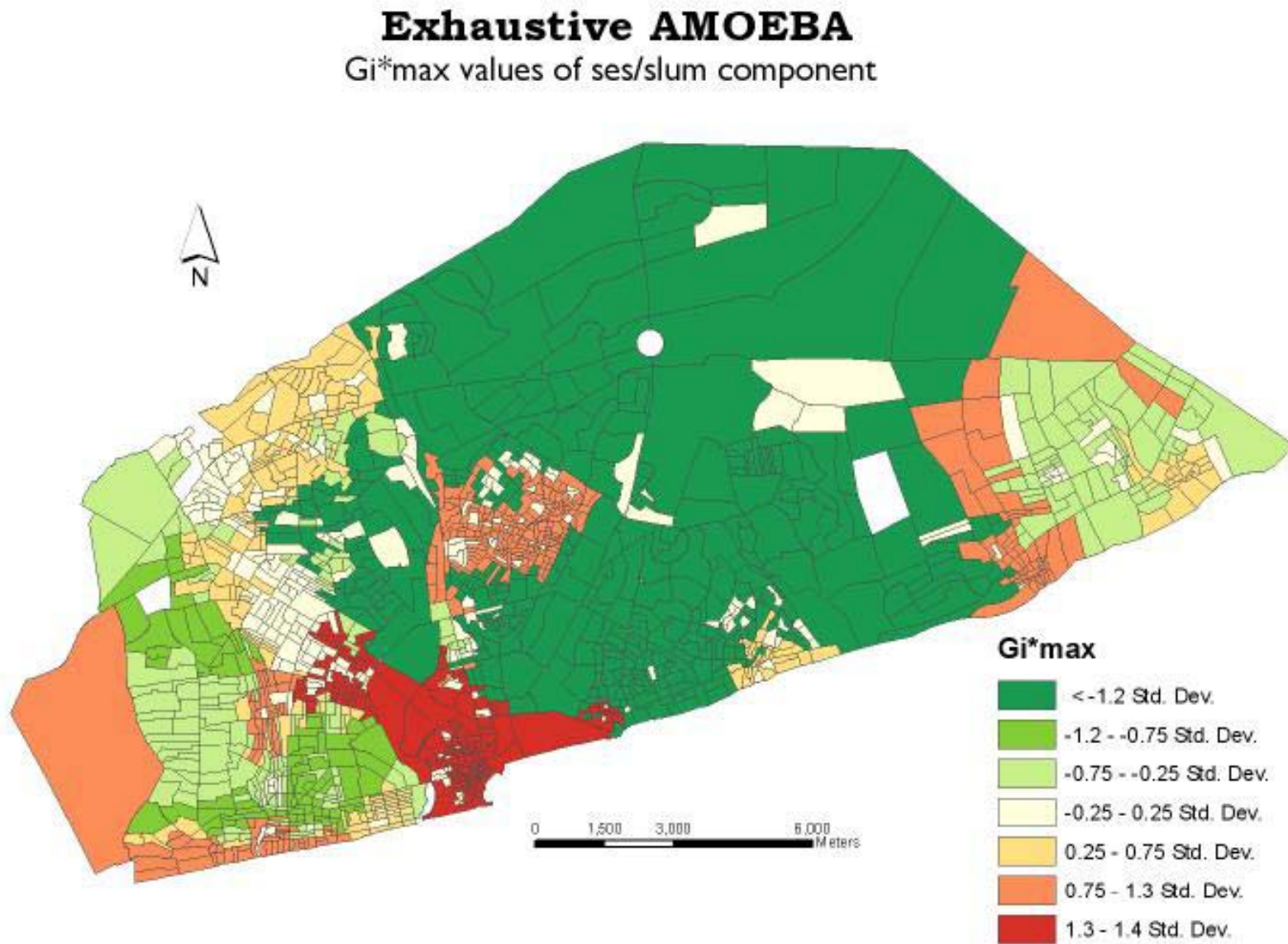
Visualization of clusters

P max

P-value of AMOEBA

p value
- 0.00 - 0.01
- 0.01 - 0.05
- 0.05 - 0.10
- 0.10 - 0.60

Gi*max
- < -1.2 Std. Dev.
- -1.2 - -0.75 Std. Dev.
- -0.75 - -0.25 Std. Dev.
- -0.25 - 0.25 Std. Dev.
- 0.25 - 0.75 Std. Dev.
- 0.75 - 1.3 Std. Dev.
- 1.3 - 1.4 Std. Dev.

0   1,500   3,000   6,000
                          Meters
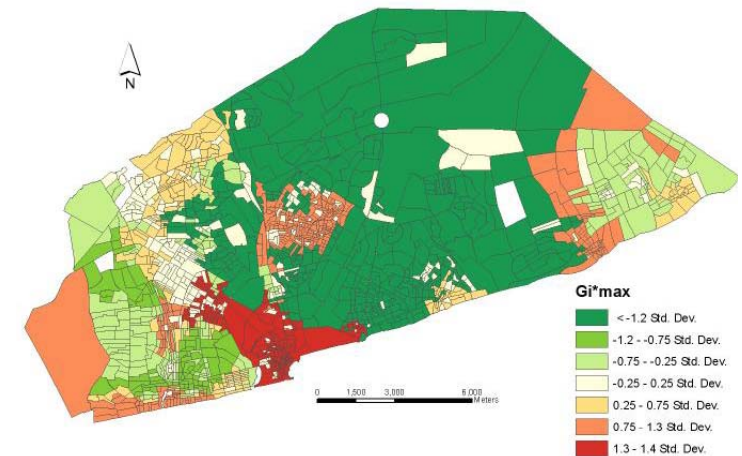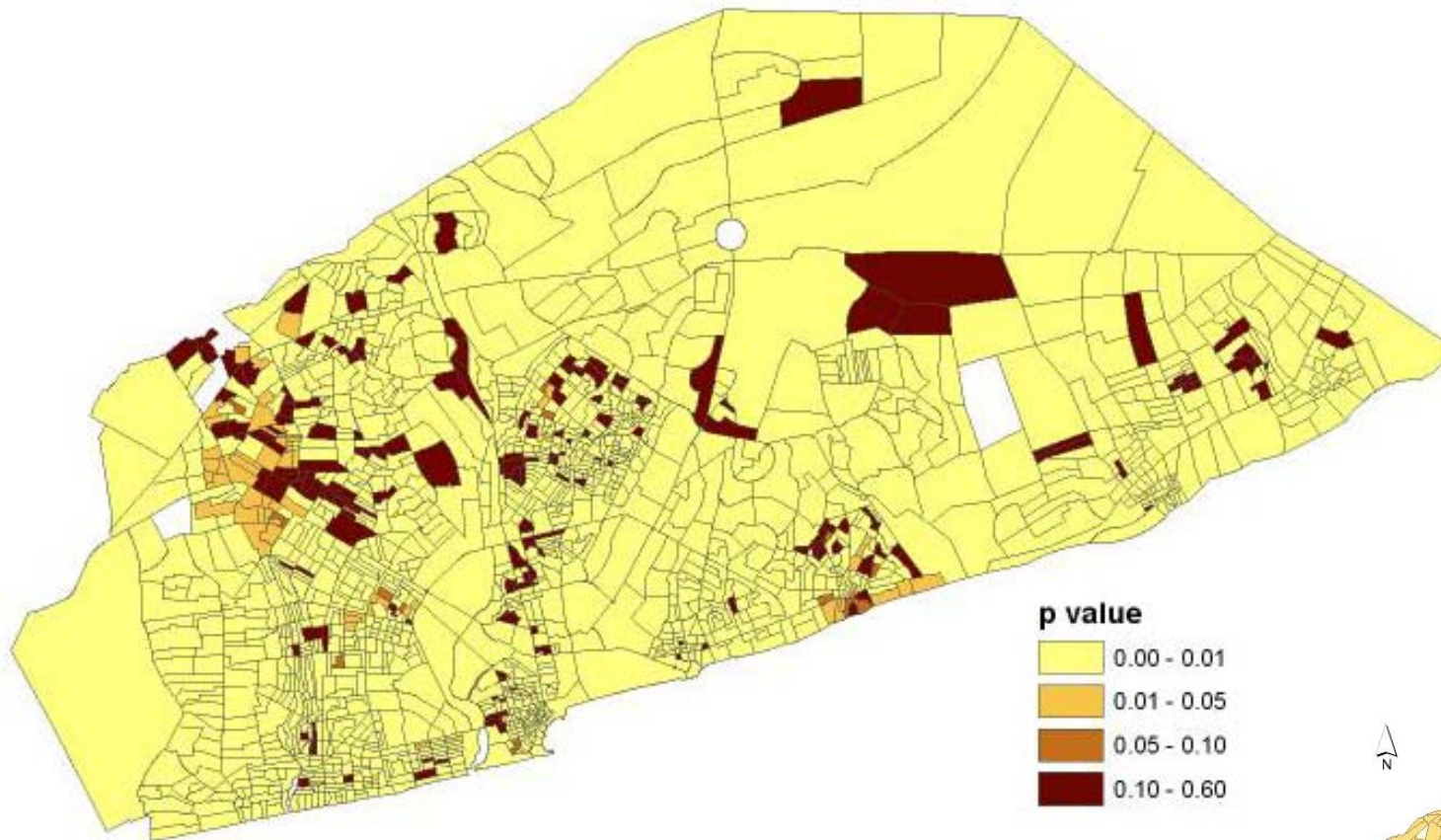
Limitations on Cluster Size

Polygon Limitation
Clusters limited to 10 polygons

Gi*max
- < -1.3 Std. Dev.
- -1.3 - -0.75 Std. Dev.
- -0.75 - -0.25 Std. Dev.
- -0.25 - 0.25 Std. Dev.
- 0.25 - 0.75 Std. Dev.
- 0.75 - 1.2 Std. Dev.
- 1.2 - 1.7 Std. Dev.

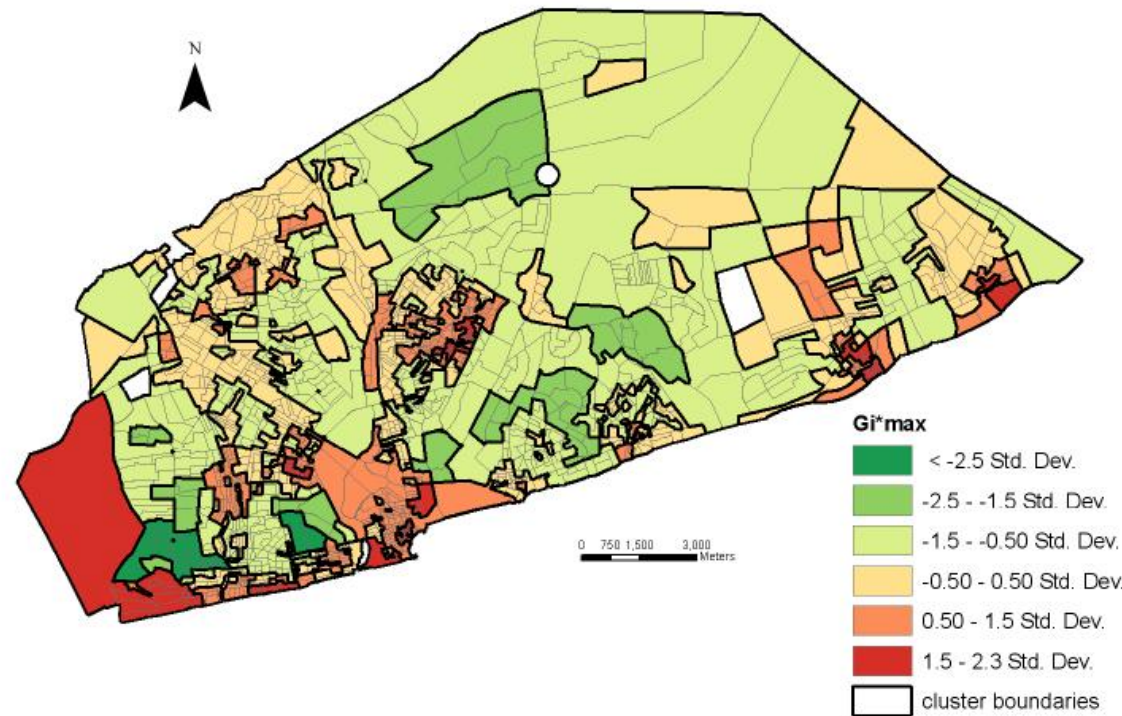0   1,200   2,400   4,800
Meters
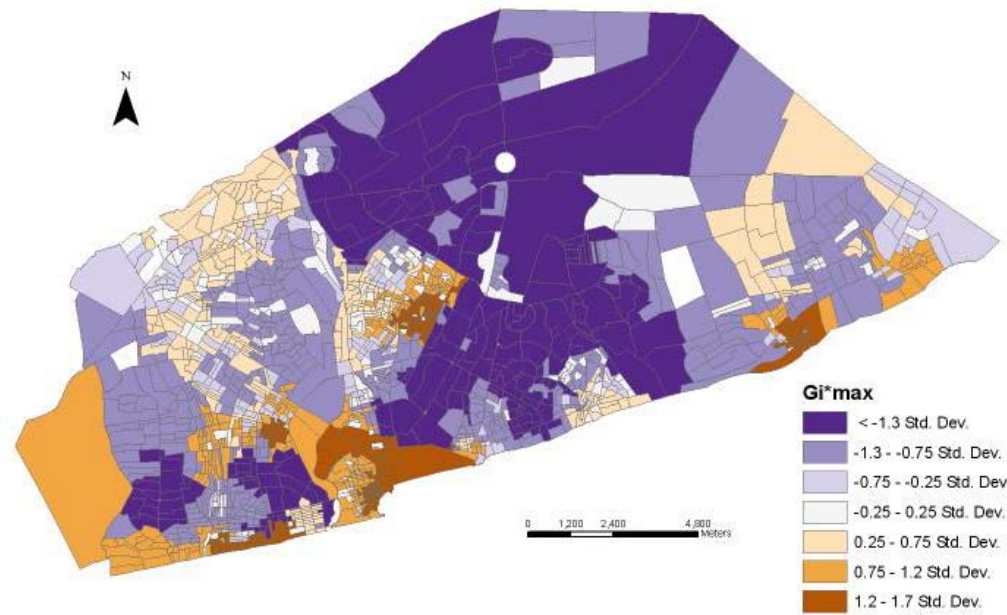
- Neighborhoods cannot be too large
  - Need method of breaking up large clusters
- Introduce limitations
  - Unit limitation
  - Area limitation
  - Variable of choice limitation (eg. population)
- Visualization the same
  - Clusters are smaller creating visually smaller neighborhoods
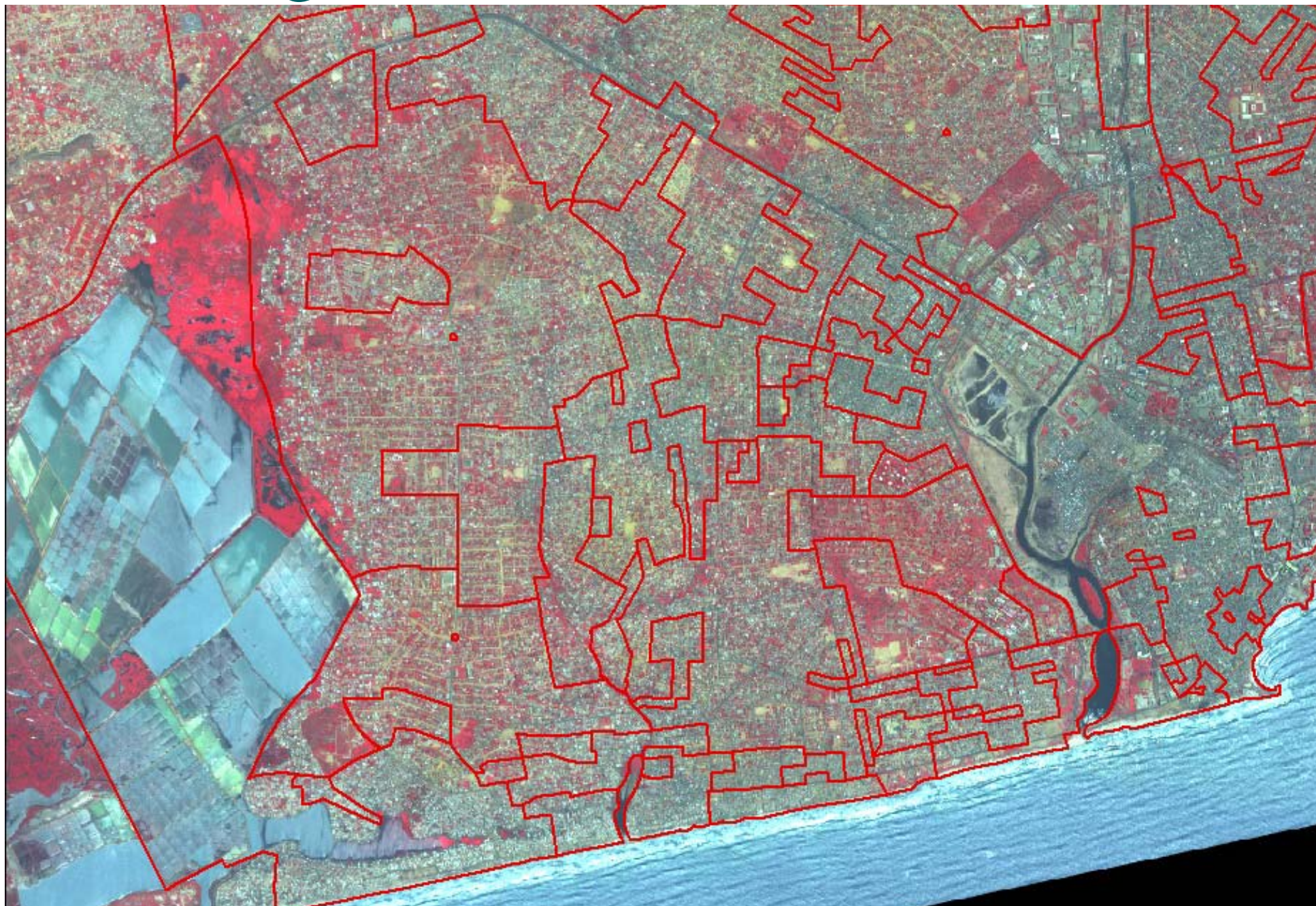
Confirming Neighborhoods

- As of yet the best neighborhoods boundaries result out of a combination of limitations and the exhaustive technique
  - Area and polygon limitation resulting in 154 neighborhoods
- Resulting neighborhood boundaries can be tested against on-the-ground observation (social), as well as satellite imagery (physical)

Limitation Combination and Resulting Cluster Boundaries
Polygon and area limitations resulting in 154 clusters

Gi*max

■	< -2.5 Std. Dev.
■	-2.5 - -1.5 Std. Dev.
■	-1.5 - -0.50 Std. Dev.
■	-0.50 - 0.50 Std. Dev.
■	0.50 - 1.5 Std. Dev.
■	1.5 - 2.3 Std. Dev.
□	cluster boundaries

0   750 1,500   3,000
                   Meters

Results

Quickbird vs. AMOEBA

Conclusions and Where Next?

- Visualization
  - Can get it as close as possible to actual clusters, but Gi*max classification schemes are still relatively arbitrary

- Limiting cluster size
  - Smaller cluster sizes result in less statistical significance - a balance that must be constantly considered

- Neighborhood confirmation
  - AMOEBA is drawing useful and significant neighborhood boundaries - more testing for visualization needed

- Single seed exhaustive AMOEBA?

Acknowledgements

References

Aldstadt, J. and A. Getis (2006). "Using AMOEBA to Create a Spatial Weights Matrix and Identify Spatial Clusters." Geographical Analysis 38: 327-343.