



Some Spatial Statistics

Overview

- The initial dataset
- Adding Census data
- Adding distance data
- Combining into block groups
- Clustering
- OLS
- GWR



The Initial Dataset

- A set of maps for Pilsen, Humboldt Park, and Garfield Park color-coded for 9 variables:
 - Abandoned, significantly improved (rehabbed), business, vacant business, new construction, racialized white space, racialized non-white space, civic/community space, and affordable housing
 - The four we're interested in:
 - Investment: improved & new construction
 - Disinvestment: abandoned & vacant

Digitizing



- Paper maps scanned
- Georeferenced
- Digitized by overlaying parcel file and coding dummy variables

Digitizing



- Paper maps scanned
- Georeferenced
- Digitized by overlaying parcel file and coding dummy variables

Digitizing



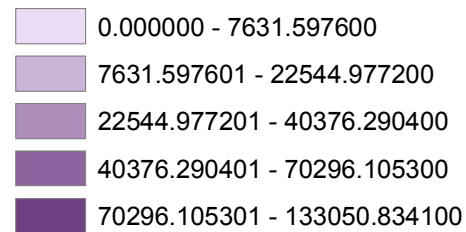
- Paper maps scanned
- Georeferenced
- Digitized by overlaying parcel file and coding dummy variables



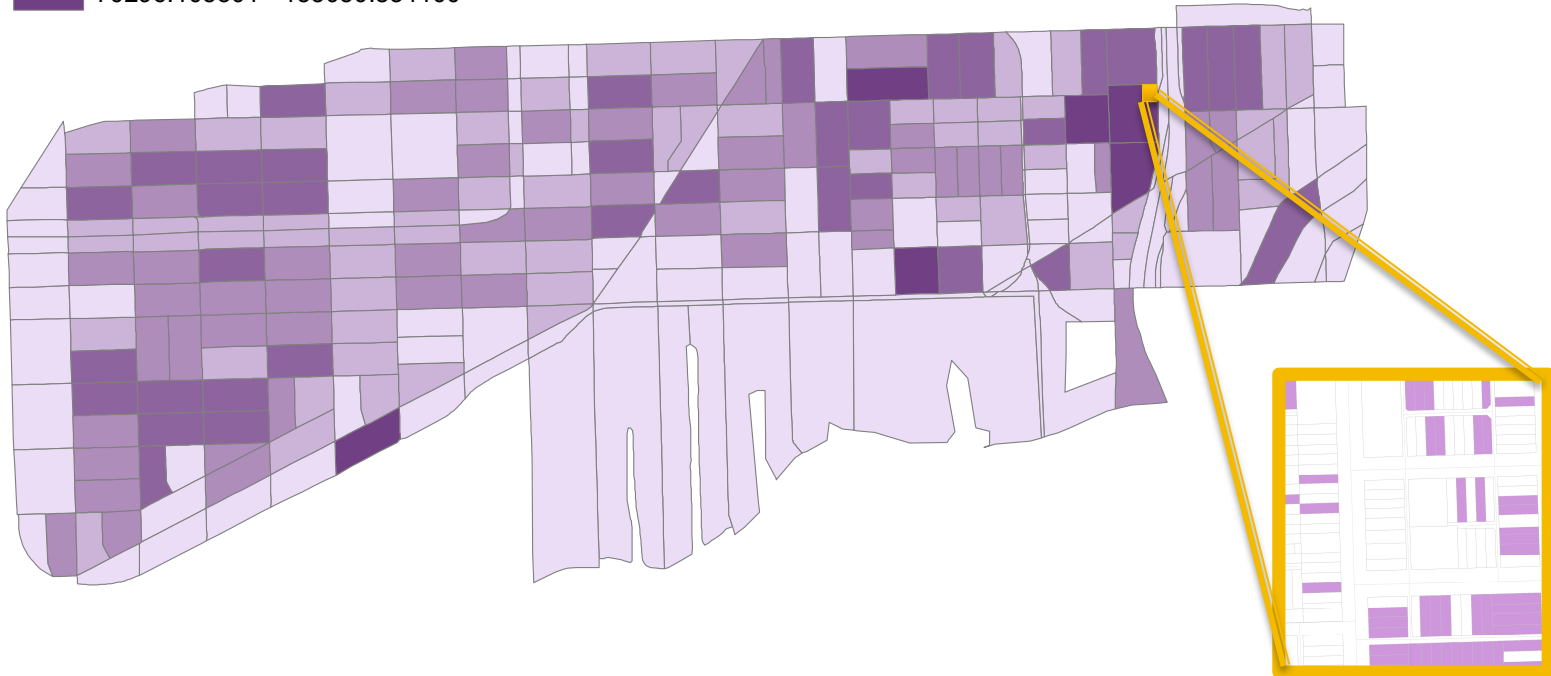
Aggregating

Pilsen Blocks

Improvement Area

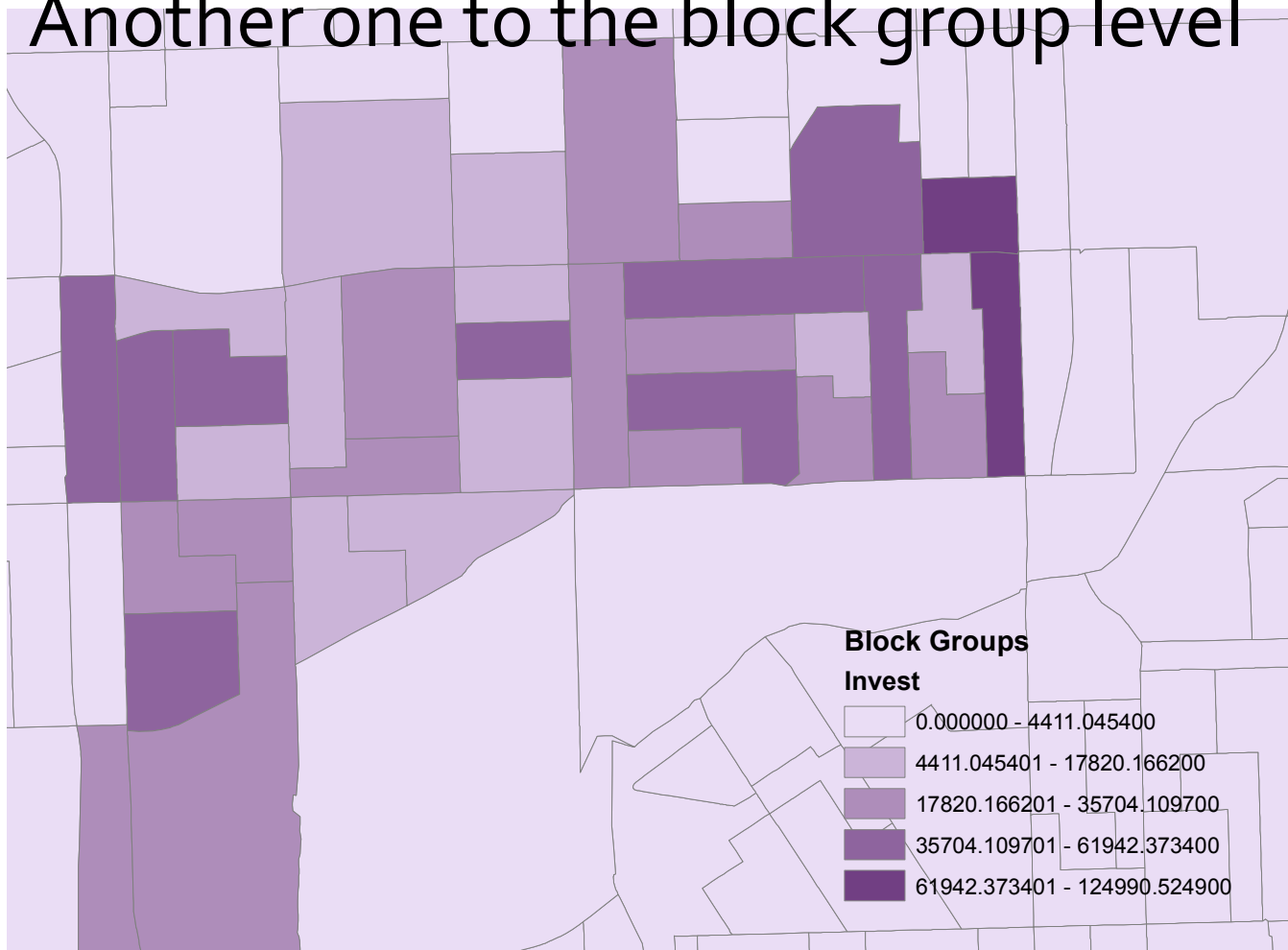


- Created one file aggregated to block level



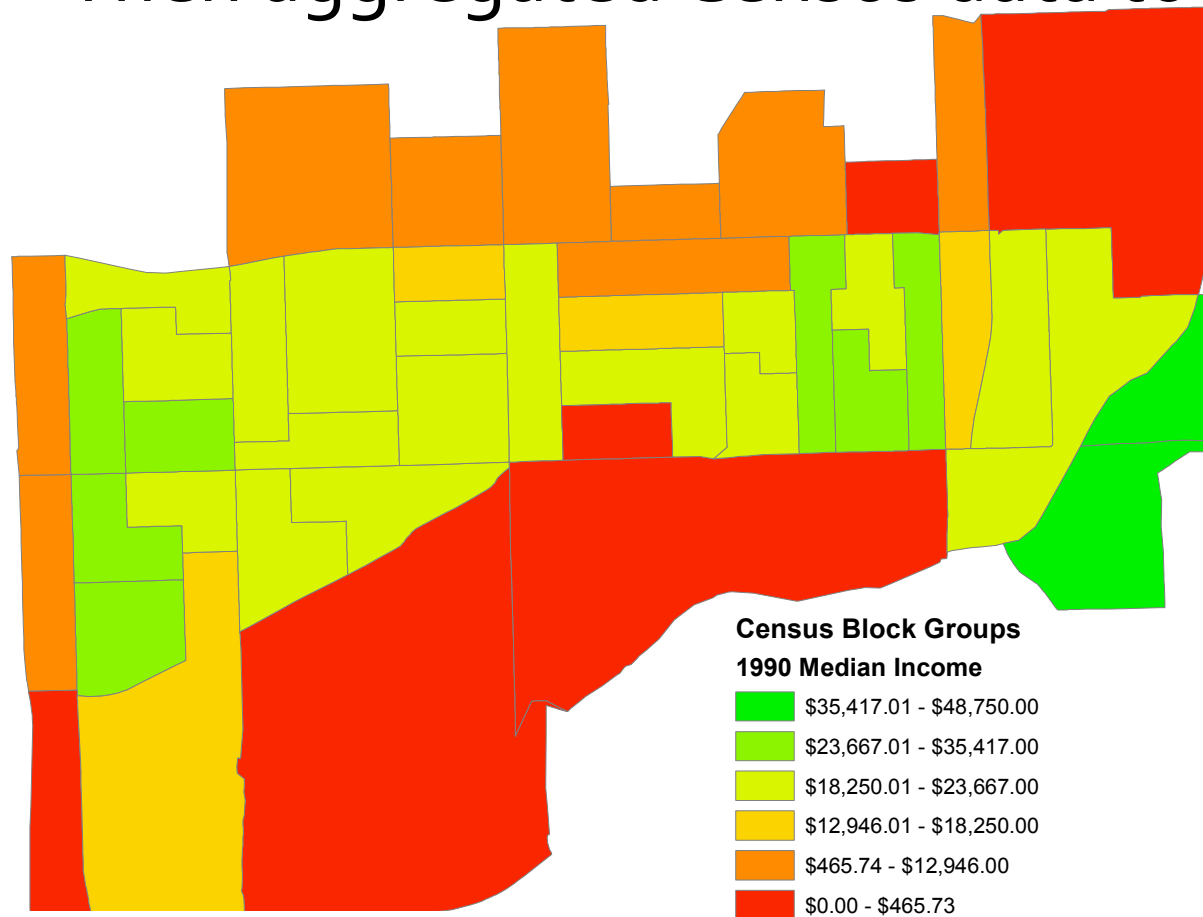
Aggregating

- Another one to the block group level



Census Data & Other Variables

- Then aggregated Census data to block groups



Census Data & Other Variables

- Census data for 1990 & 2000 collected to calculate the change in percents from 1990 to 2000:
 - Total Population
 - Non-Latino/Hispanic, white
 - Latino/Hispanic change
 - African-American
 - Tenure of occupied housing units & vacancy
 - Median owner housing value change
 - Median Year Structure Built
 - Median household income
 - Number in all service occupations
 - Number in professional occupations



Census Data & Other Variables

- Performed Census calculations by
 - 1. Downloading 1990 and 2000 Census Block Groups
 - 2. Joining 2000 Census data to 2000 block groups
 - 3. Joining 1990 Census data to 1990 block groups
 - 4. Joining
 - 5. Calculating

Census Data & Other Variables

- Then used Network Analyst to calculate the distance from each block group centroid to points representing
 - The central business district (point somewhere in the Loop)
 - Highway ramps (points for centroid)
 - CTA stations
 - Metra stations (or places with stops)

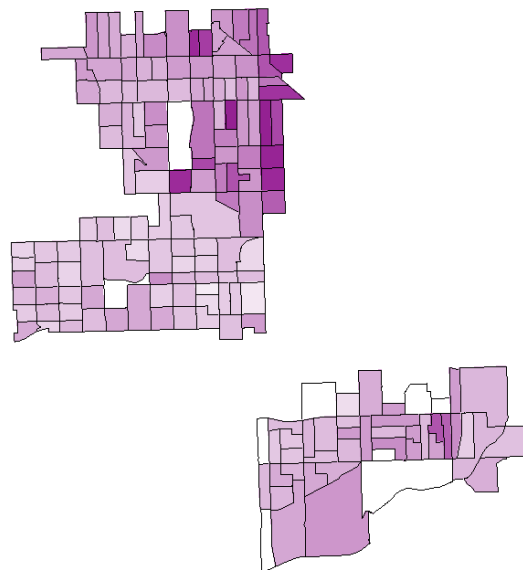
Put it All Together

Then joined all that Census and distance stuff together with the investment stuff at the block group level!

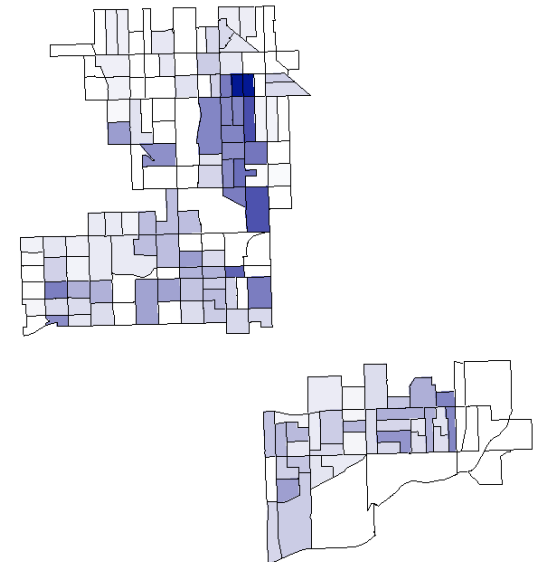
Distance to CTA Station



Median Home Value 2000

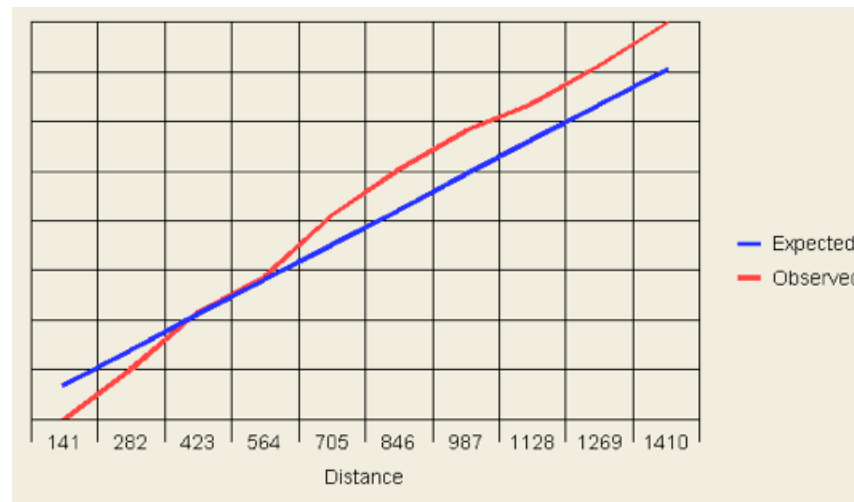


Observed Investment, 2009



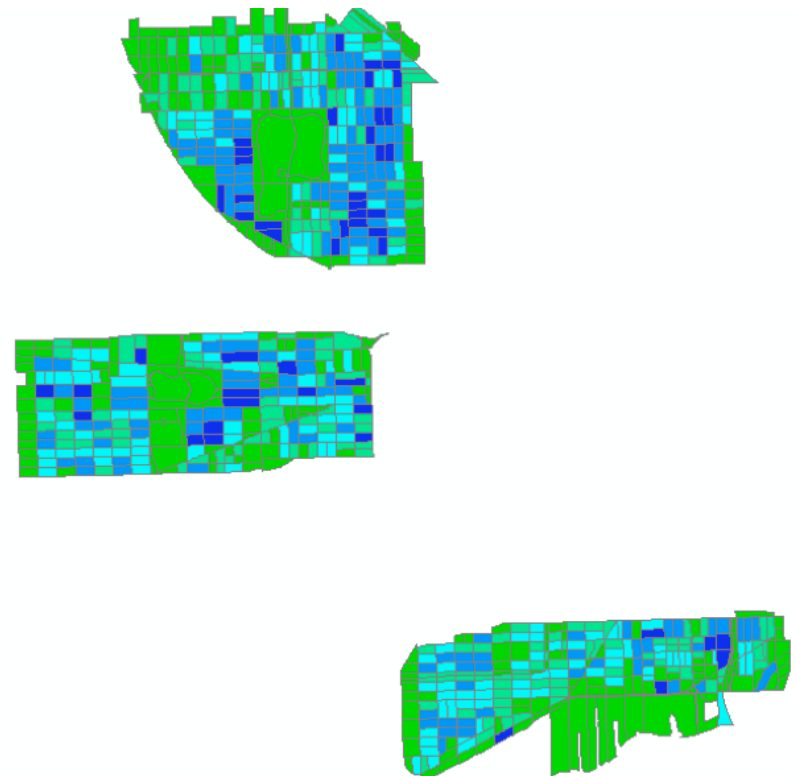
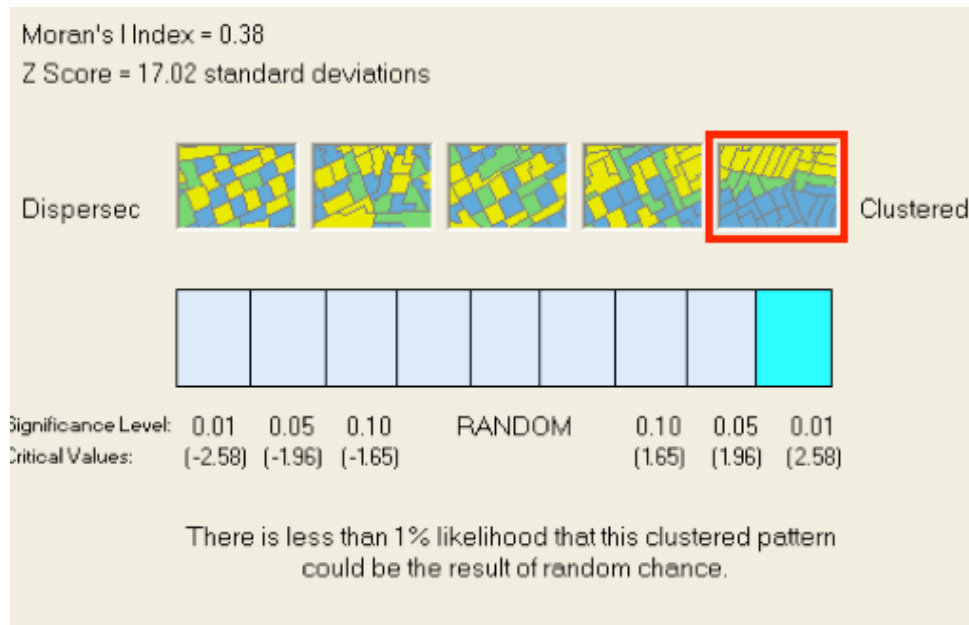
Spatial Statistics

- Next step was running Ripley's K a few times to find the neighborhood size at which clustering is maximized for later analysis



Spatial Statistics

- Then Moran's I to assess global clustering of the investment variable at the block level – it's clustered!

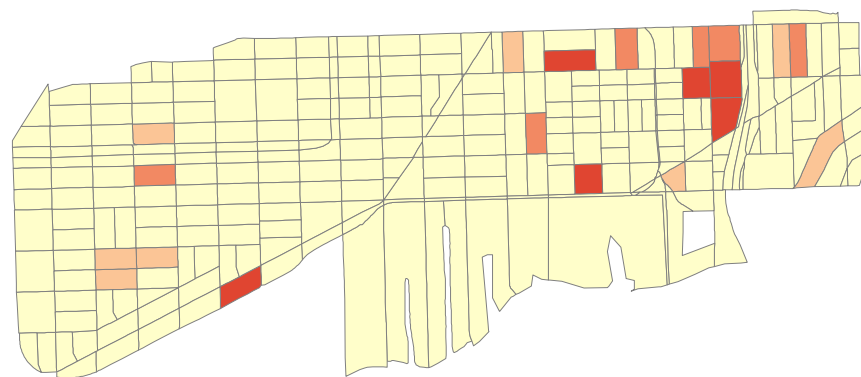


Spatial Statistics

- Ran Getis-ord G_i^* to look at hot spots at block level for individual neighborhoods (Pilsen, Humboldt Park, and Garfield Park)
 - With Euclidean Inverse Distance with distance of 987 feet
 - Represented by Z scores to indicate the intensity of clustering

Spatial Statistics

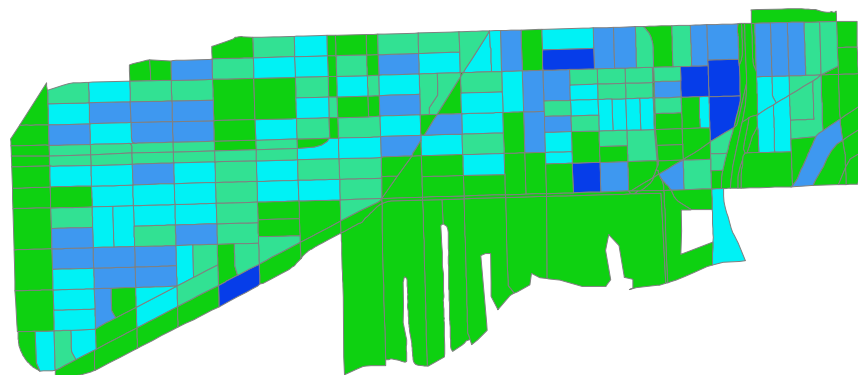
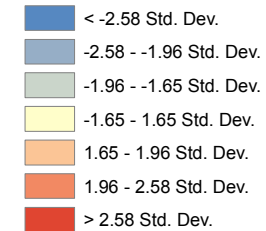
■ Getis-ord G_i^* for Pilsen



Legend

PilsenHotSpot

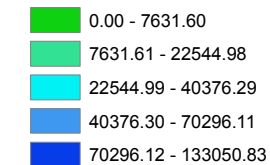
GiZScore



Legend

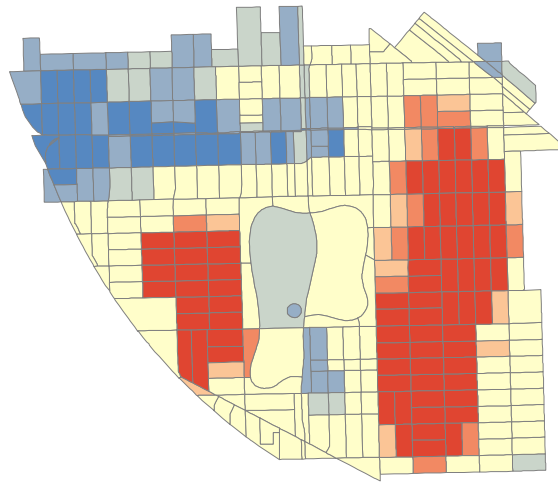
PilsenJoinBlks

NCandSI_Area



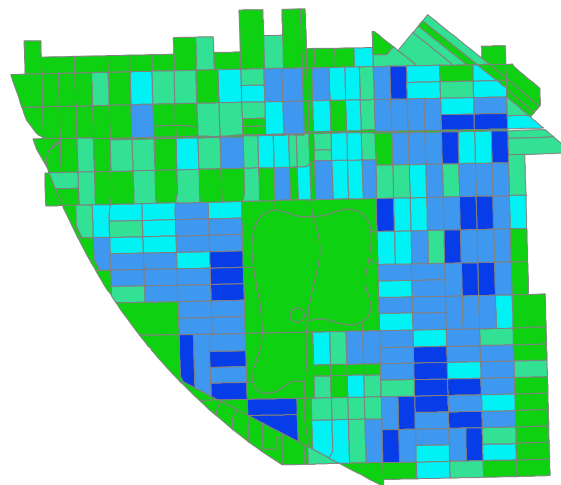
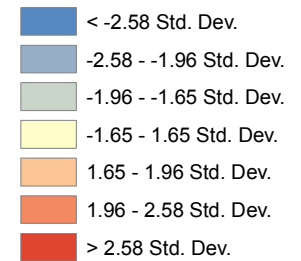
Spatial Statistics

■ Getis-ord Gi*



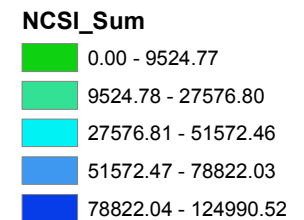
Legend

HumboldtHotSpot GiZScore



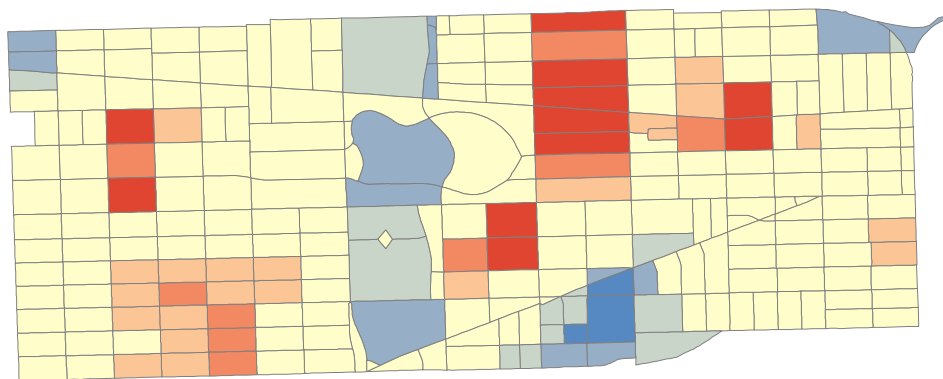
Legend

HumboldtJoinBlks NCSI_Sum



Spatial Statistics

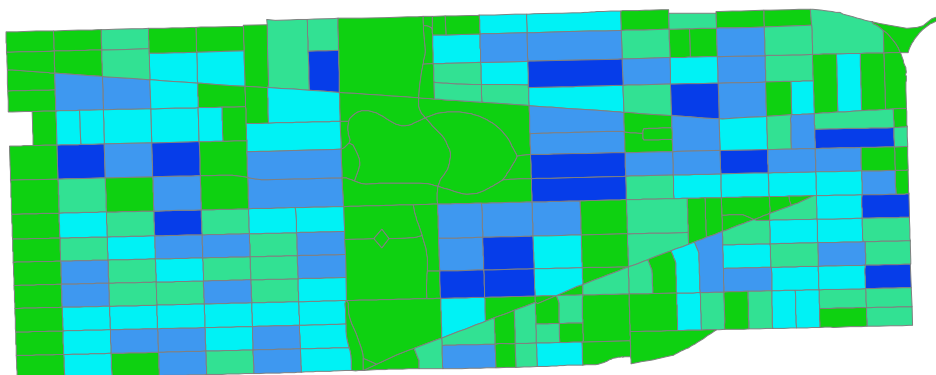
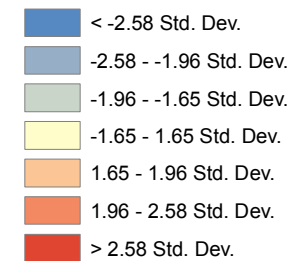
■ Getis-ord G_i^*



Legend

GarfieldHotSpot

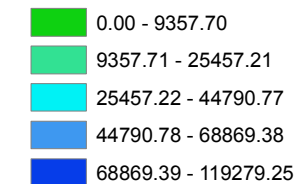
GiZScore



Legend

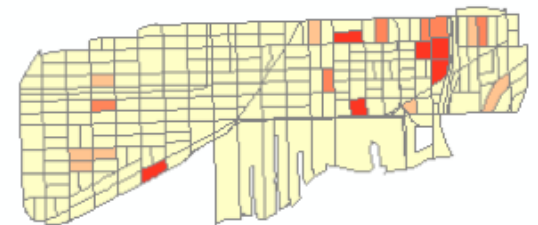
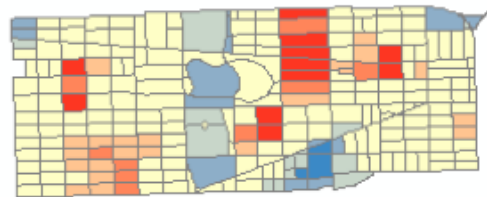
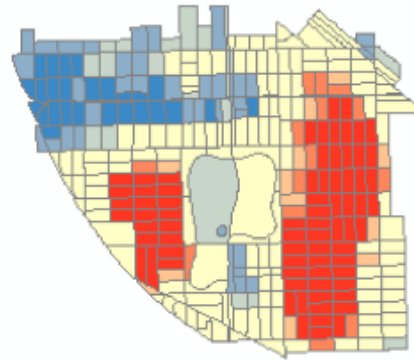
GarfieldJoinBlks

NCSI_Sum



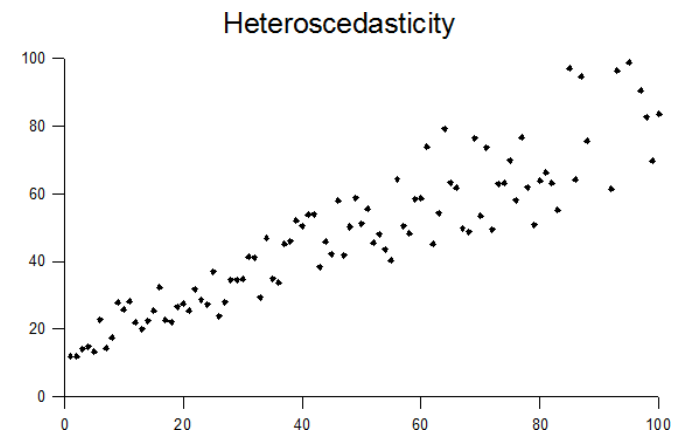
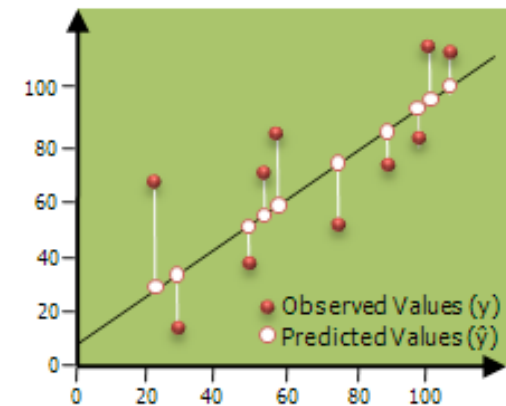
Spatial Statistics

- Getis-ord G_i^*
 - High Z scores with small p values indicate clustering (can reject null hypothesis). Z indicates how much a score varies from null hypothesis.



Spatial Statistics

- Ordinary Least Squares
 - A global linear regression analysis to predict or model a dependent variable in terms of its relationship to the explanatory variables
 - Goal is calibrate a model that does not show signs of heteroscedasticity



Spatial Statistics

Variable	Coefficient	StdError	t-Statistic	Probability	Robust_SE	Robust_t	Robust_Pr	VIF [1]
Intercept	0.000010	0.00014	0.068573	0.945408	0.000285	0.034501	0.972515	-----
CTA_DIST	-0.000000	0.00000	-0.245112	0.806698	0.000000	-0.296277	0.767424	1.799682
METRA_DIST	0.000000	0.00000	1.143687	0.254526	0.000000	1.165343	0.245680	1.517225
RAMP_DIST	0.000000	0.00000	2.761510	0.006452*	0.000000	2.954007	0.003633*	2.534616
CBDDIST	-0.000000	0.00000	-3.512766	0.000592*	0.000000	-4.127279	0.000065*	2.147854
POPCHG	0.000001	0.00000	0.337725	0.736039	0.000003	0.304581	0.761104	1.189181
WHITEMAG	0.000008	0.00001	0.445905	0.656302	0.000012	0.636889	0.525143	1.132263
LATINOMAG	-0.000012	0.00005	-0.206207	0.836900	0.000046	-0.252431	0.801047	1.150062
BLACKMAG	0.000030	0.00002	1.154184	0.250211	0.000025	1.193364	0.234562	1.159839
RENTMAG	0.000129	0.00008	1.559510	0.120939	0.000073	1.763419	0.079819	1.451517
OWNMAG	-0.000042	0.00003	-1.335626	0.183649	0.000017	-2.408825	0.017175*	1.173702
VACMAG	-0.000012	0.00003	-0.370367	0.711628	0.000026	-0.463343	0.643782	1.138165
VALUEMAG	-0.000049	0.00002	-1.844624	0.067016	0.000028	-1.719827	0.087480	1.158197
MEDYRBLT	0.000000	0.00000	2.446954	0.015523*	0.000000	1.188956	0.236287	1.337742
MDINCMAG	-0.000001	0.00000	-0.199519	0.842118	0.000002	-0.299022	0.765333	1.247041
SERVEMAG	-0.000012	0.00002	-0.538204	0.591219	0.000012	-0.958491	0.339308	1.202168
PROFMAG	0.000001	0.00000	0.121348	0.903565	0.000007	0.155189	0.876869	1.155754

OLS Diagnostics

Number of Observations:	171	Number of Variables:	17
Degrees of Freedom:	154	Adjusted R-Squared [2]:	0.103321
Multiple R-Squared [2]:	0.187714	Adjusted R-Squared [2]:	0.103321
Joint Wald Statistic [4]:	44.021756	Prob(>chi-squared), (16) degrees of freedom:	0.000196*
Koenker (BP) Statistic [5]:	28.817894	Prob(>chi-squared), (16) degrees of freedom:	0.025193*
Jarque-Bera Statistic [6]:	85.683101	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Spatial Statistics

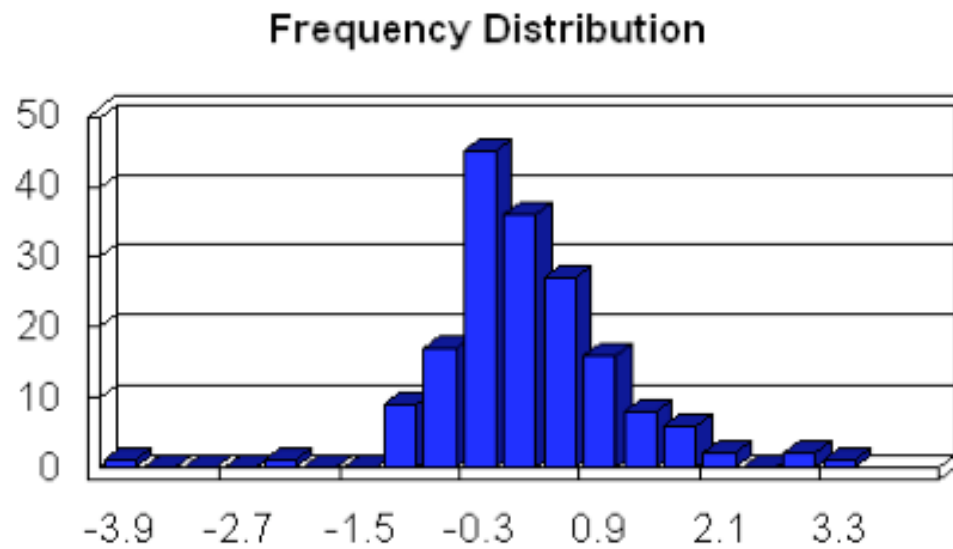
Variable	Coefficient	StdError	t-Statistic	Probability	Robust_SE	Robust_t	Robust_Pr	VIF [1]
Intercept	0.000010	0.000143	0.068573	0.945408	0.000285	0.034501	0.972515	-----
CTA_DIST	-0.000000	0.000000	-0.245112	0.806698	0.000000	-0.296277	0.767424	1.799682
METRA_DIST	0.000000	0.000000	1.143687	0.254526	0.000000	1.165343	0.245680	1.517225
RAMP_DIST	0.000000	0.000000	2.761510	0.006452*	0.000000	2.954007	0.003633*	2.534616
CBDDIST	-0.000000	0.000000	-3.512766	0.000592*	0.000000	-4.127279	0.000065*	2.147854
POPCHG	0.000001	0.000002	0.337725	0.736039	0.000003	0.304581	0.761104	1.189181
WHITEMAG	0.000008	0.000017	0.445905	0.656302	0.000012	0.636889	0.525143	1.132263
LATINOMAG	-0.000012	0.000056	-0.206207	0.836900	0.000046	-0.252431	0.801047	1.150062
BLACKMAG	0.000030	0.000026	1.154184	0.250211	0.000025	1.193364	0.234562	1.159839
RENTMAG	0.000129	0.000083	1.559510	0.120939	0.000073	1.763419	0.079819	1.451517
OWNMAG	-0.000042	0.000031	-1.335626	0.183649	0.000017	-2.408829	0.017175*	1.173702
VACMAG	-0.000012	0.000033	-0.370367	0.711628	0.000026	-0.463348	0.643782	1.138165
VALUEMAG	-0.000049	0.000027	-1.844624	0.067016	0.000028	-1.719827	0.087480	1.158197
MEDYRBLT	0.000000	0.000000	2.446954	0.015523*	0.000000	1.188956	0.236287	1.337742
MDINCMAG	-0.000001	0.000003	-0.199519	0.842118	0.000002	-0.299022	0.765333	1.247041
SERVEMAG	-0.000012	0.000022	-0.538204	0.591219	0.000012	-0.958491	0.339308	1.202168
PROFMAG	0.000001	0.000009	0.121348	0.903565	0.000007	0.155189	0.876869	1.155754

OLS Diagnostics

Number of Observations:	171	Number of Variables:	17
Degrees of Freedom:	154	Akaike's Information Criterion (AIC) [2]:	-2316.105292
Multiple R-Squared [2]:	0.187714	Adjusted R-Squared [2]:	0.103321
Joint F-Statistic [3]:	2.224280	Prob(>F), (16,154) degrees of freedom:	0.006471*
Joint Wald Statistic [4]:	11.881756	Prob(>chi-squared), (16) degrees of freedom:	0.000186*
Koenker (BP) Statistic [5]:	28.817894	Prob(>chi-squared), (16) degrees of freedom:	0.025193*
Jarque-Bera Statistic [6]:	85.683101	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

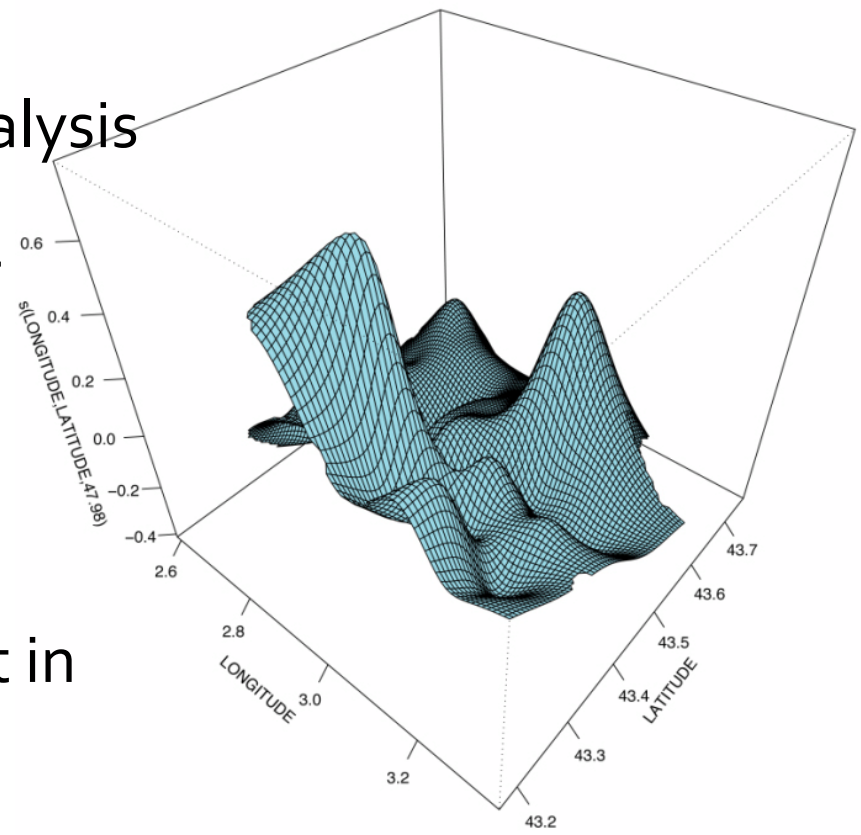
Spatial Statistics

- Frequency Distribution
- Moran's I on residuals – is clustered showing incomplete model...



Spatial Statistics

- Geographically Weighted Regression
 - A local linear regression analysis that takes into account relationships that vary over space (heteroscedasticity)
 - The equation measures the relationship between independent and dependent variables around each point in space



Spatial Statistics

- Geographically Weighted Regression
 - A local linear regression analysis that takes into account relationships that vary over space (heteroscedastic)
 - The equation measures the relationship between independent and dependent variables around each point in space

$$y(g) = b_0(g) + b_1(g)x_1 + b_2(g)x_2 + e - \text{equation}$$

$$b(g) = (X^T W(g) X)^{-1} X^T W(g) Y - \text{weighting scheme}$$

Spatial Statistics

- Geographically Weighted Regression
 - Ran the model several times
 - Because you have a weighting scheme you can choose a kernel type: fixed (enter set distance/neighbors) or adaptive (if density varies) – adaptive is preferred method

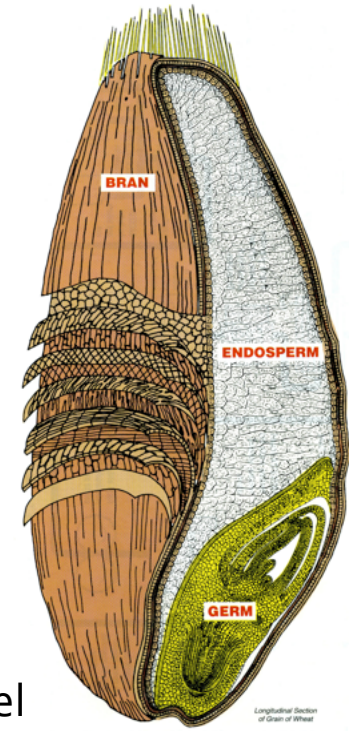


Spatial Statistics

- Geographically Weighted Regression
 - Next option is Bandwidth. There are three: AICc, CV, and Bandwidth parameter
 - Akaike (aKAeKe) - automatic method that finds the value that minimizes the AICc value (a good thing) – preferred method because it takes into account complexity of model
 - Cross Validation – finds the lowest CV score (?)
 - Specified bandwidth – if chose Fixed kernel type, then enter feet (if in feet) or adaptive, choose neighbors

Spatial Statistics

- Geographically Weighted Regression
 - Bottom line seems to be:
 - Use Adaptive kernel with AICc



Wheat Kernel:
Wrong Type of Kernel

Spatial Statistics

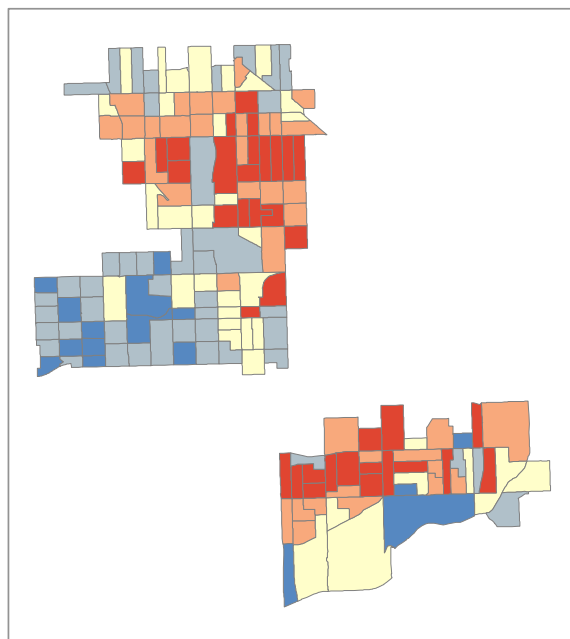
- Ran GWR several times – goal is to get lowest AICc
 - Dependent variable - % Net investment (investment – disinvestment / total area)
 - Method: Fixed with AICc at 987 feet
 - AICc of -2309. Had values ranging up to 3000. Lowest is best.



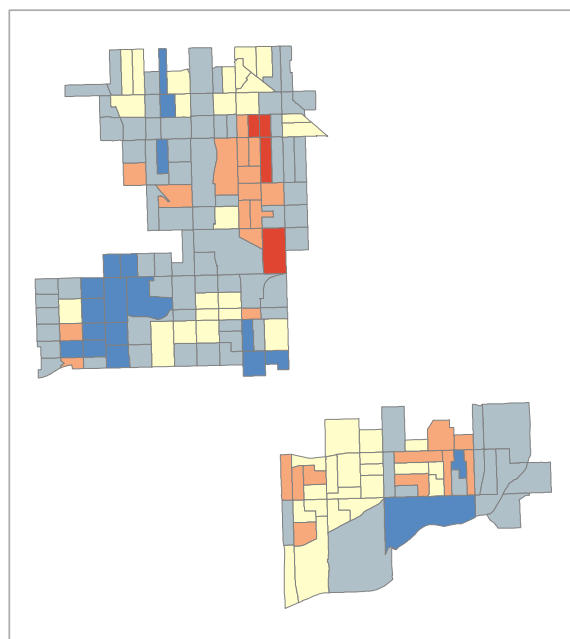
Spatial Statistics

■ GWR Output

Bandwidth	: 43655.95683645904
ResidualSquares	: 1.0507676616160562e-005
EffectiveNumber	: 19.910713601360758
Sigma	: 0.00026371602090289394
AICc	: -2309.7958875232102
R2	: 0.20688604825671042
R2Adjusted	: 0.10761791911161245

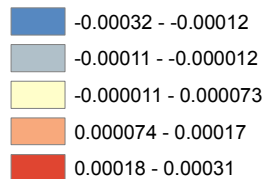


0 1 2 4 Miles



GWR

Predicted % Net Investment



Predicted & Existing Investment* for Pilsen, Humboldt Park, and Garfield Park:

A Geographically Weighted Regression Analysis with Observed 2009 Data and 1990 and 2000 Census Data

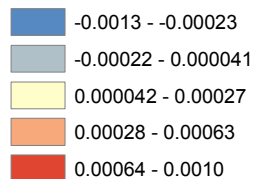
Map prepared by: A. Ryding
for UPP 462 - Intermed. GIS
04/20/2010

Projection: State Plane Illinois East FIPS 1201 (feet)
Data Source: 1990 & 2000 Census, SF3 block group data; Ramps/streets, CTA stations, Metra stations data downloaded from City of Chicago GIS Dept. website 04/20/10; existing investment values observed by J. Mumm, 2009

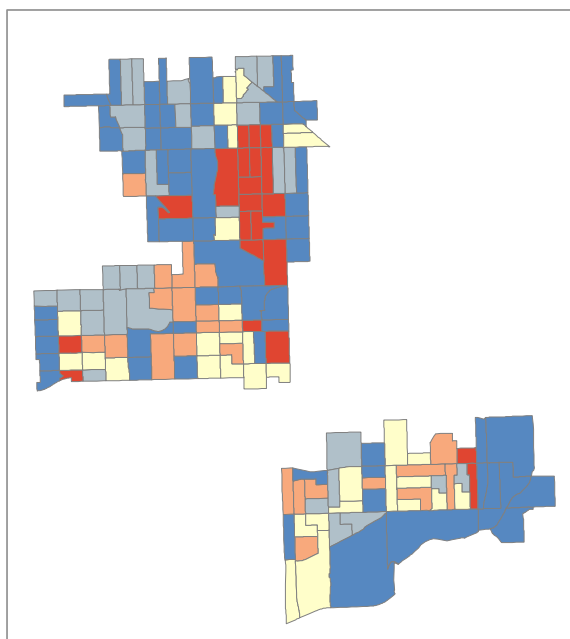
*Investment equals net investment (investment assigned parcels minus all disinvestment assigned parcels) and GWR data was calculated by change from 1990 percents to 2000 percents

Existing Values

% Net Investment

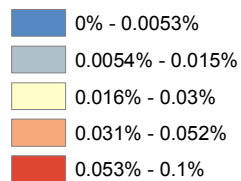


GWR & % Net Investment



Observed Variables

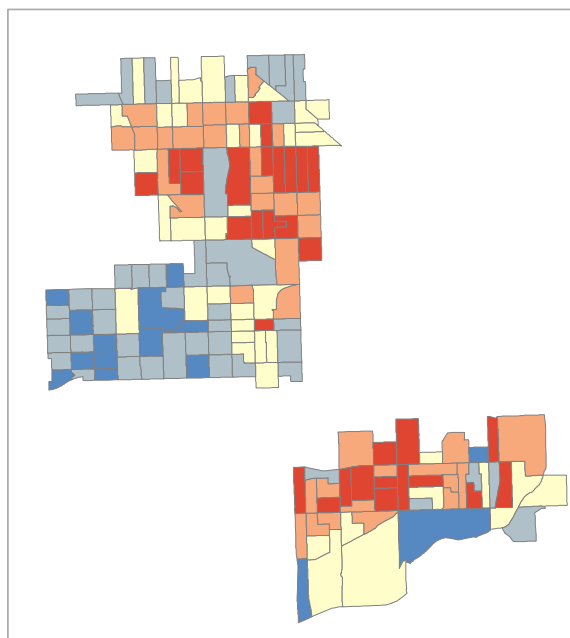
% Investment



Existing Percent Investment and Estimated Percent Investment* for Pilsen, Humboldt Park, and Garfield Park:

An Ordinary Least Squares Model
With 2009 Observed Data, and
1990 and 2000 Census Data

0 1 2 4 Miles



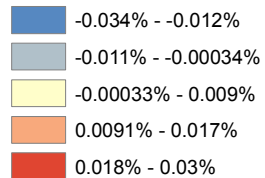
Map prepared by: A. Ryding
for UPP 462 - Intermed. GIS
04/20/2010

Projection: State Plane Illinois East FIPS 1201 (feet)
Data Source: Existing investment values observed by J.
Mumm, 2009; 1990 and 2000 SF3 Census data used for OLS

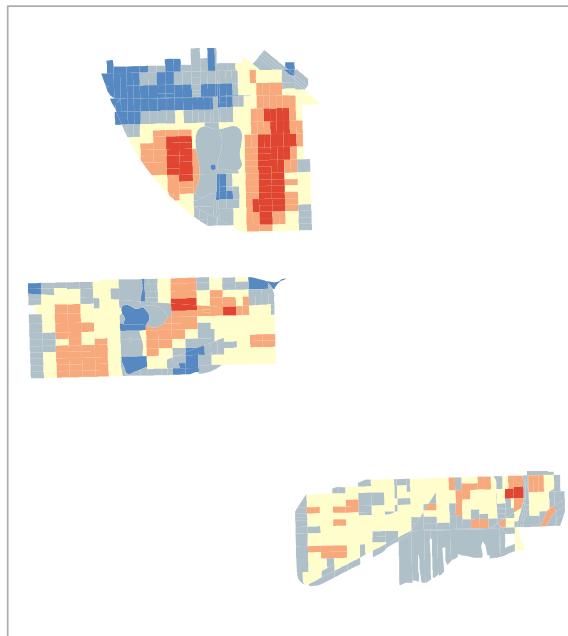
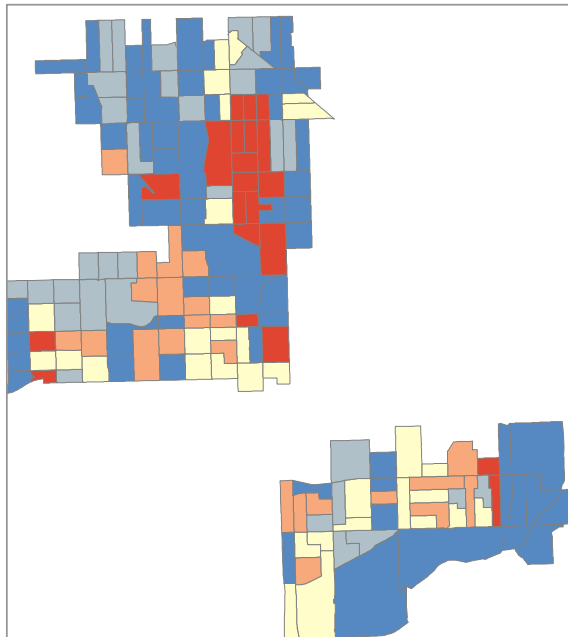
*Investment was calculated by totaling the square
footage for parcels marked as "significantly improved"
and/or "new construction."

Ordinary Least Squares

Estimated %

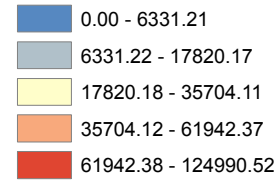


OLS & % Investment



Variables

Investment



Existing Investment Square Footage and Investment Hot Spots for Pilsen, Humboldt Park, and Garfield Park

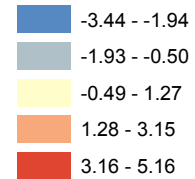
Map prepared by: A. Ryding
for UPP 462 - Intermed. GIS
04/20/2010

Projection: State Plane Illinois East FIPS 1201 (feet)
Data Source: Existing investment values observed by J. Mumm, 2009

*Investment was calculated by totalling the square footage for parcels marked as "significantly improved" and/or "new construction."

Hot Spots

Gi Z-Score



Hot Spots & Investment

The End

