PATTERNS OF SPATIAL DISTRIBUTION FOR REPORTED GRAFFITI IN SAN FRANCISCO

SPATIAL ANALYSIS USING ARCGIS 10

GEOGRAPHY 4/592 (USP 592) – GIS & APPLICATIONS

PORTLAND STATE UNIVERSITY

VERONICA MEGLER

JOAN BURGGRAFF

DEBBIE BLACKMORE

7 MARCH 2011

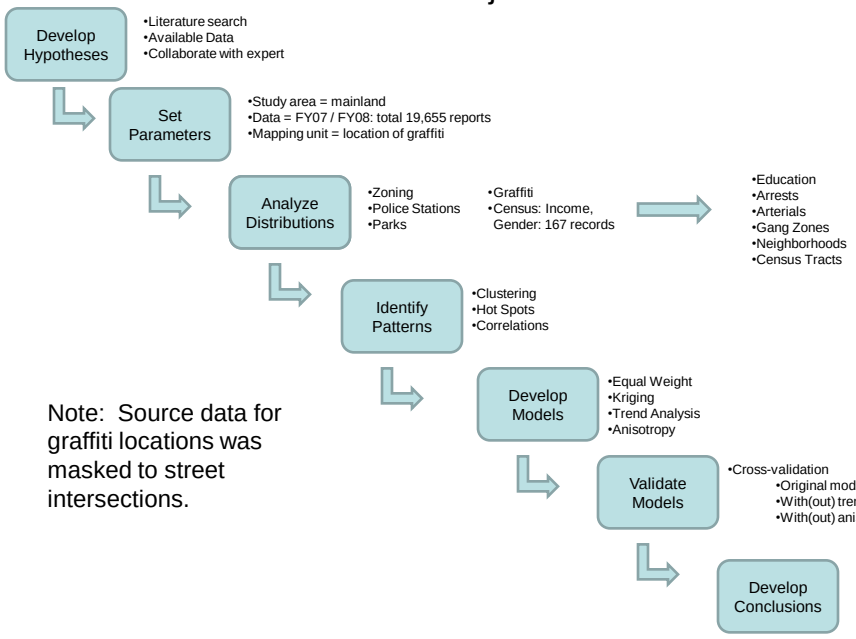
Graffiti in San Francisco - Problem

- Problem
  - The City of San Francisco spends more than $20M annually for graffiti removal. Their goal: Zero Graffiti.
  - Focus of study is *reported* graffiti (herein referred to as "GRAFFITI"), i.e., the subset of graffiti that exists AND is reported to the City for removal.
  - PSU Geography Professors David Banis and Hunter Shobe are exploring how GIS analysis can help the City target their graffiti abatement efforts.

- Project Hypothesis: The distribution of reported graffiti is spatially correlated with one or more of the following available datasets:
  - Arrests
  - Arterials
  - Education
  - Gender
  - Income
  - Parks
  - Police stations
  - Watched streets
  - Zoning
  - Other?

Norteño Tag.jpg
http://www.flickr.com/photos/xoconostle/1571162862/

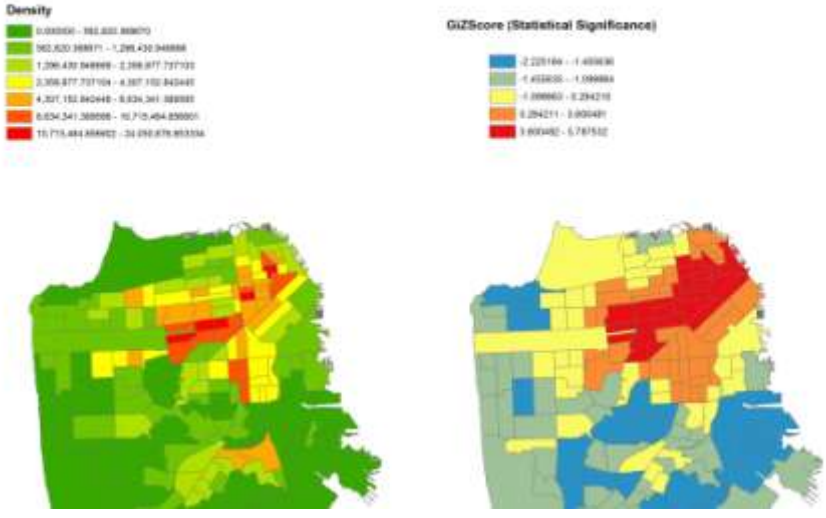
Graffiti in S.F. – Our Project Methods and Data



Note: Source data for graffiti locations was masked to street intersections.

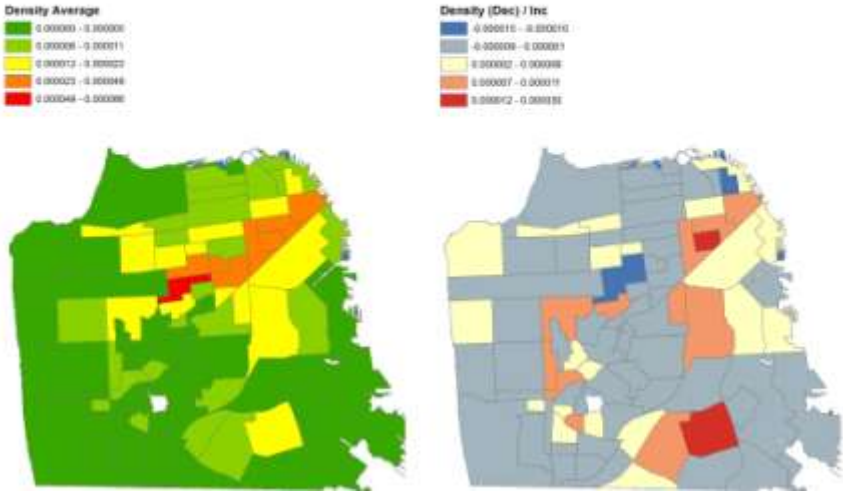
3

Graffiti: Density by Census Tract FY07-FY08



4

Graffiti: Density by Neighborhood FY07-FY08



5

Graffiti in San Francisco – Mapping Units

- Selected administrative boundaries were deemed 'arbitrary' relative to graffiti locations:
  - Supervisor District
  - Neighborhood
  - Census Tract
- Reported graffiti locations were the GIS analysis unit of choice.

6

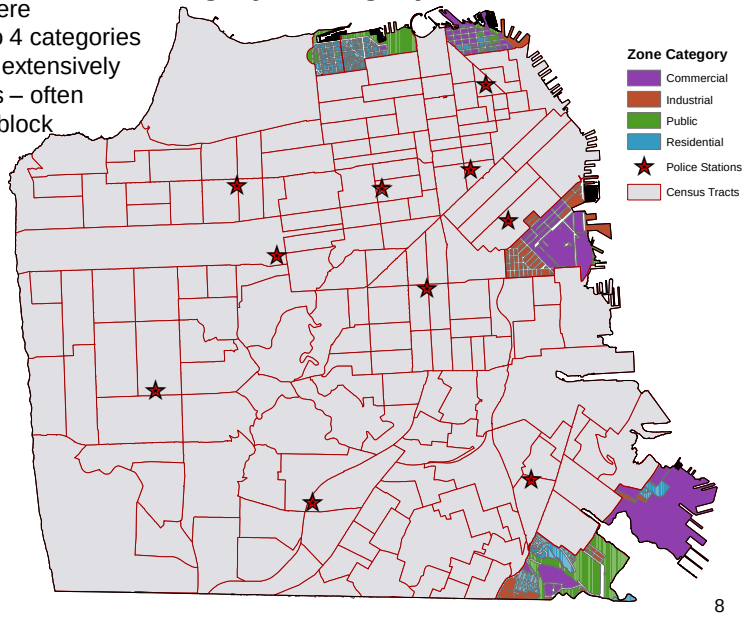
Graffiti in San Francisco – Analyze Distributions

- Zoning
- Police Stations: compare distance to graffiti and arrests
- Gender
- Income
- Parks
- Arrests: compare distribution of graffiti vs. that of arrests
- Arterials
- Gang Injunction Zones
- Education
- Watched streets

7

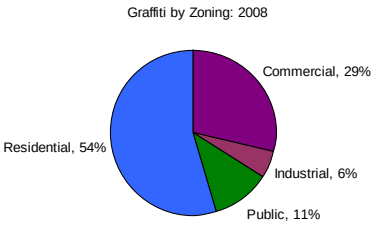
Zoning by Category

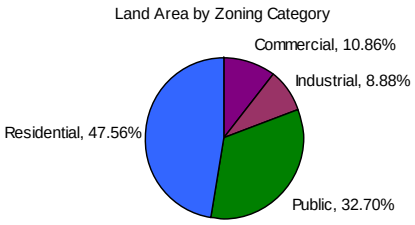
41 land zones were consolidated into 4 categories
Note: SF zoning extensively mixes all classes – often within the same block

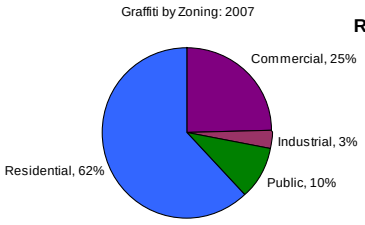


8

Relationship of Graffiti Reports and Zone Category







Relative Likelihood of Graffiti + Report per Zoning Category

	2007	2008
Commercial	3.05	3.06
Industrial	0.24	0.38
Public	0.30	0.32
Residential	1.15	1.11

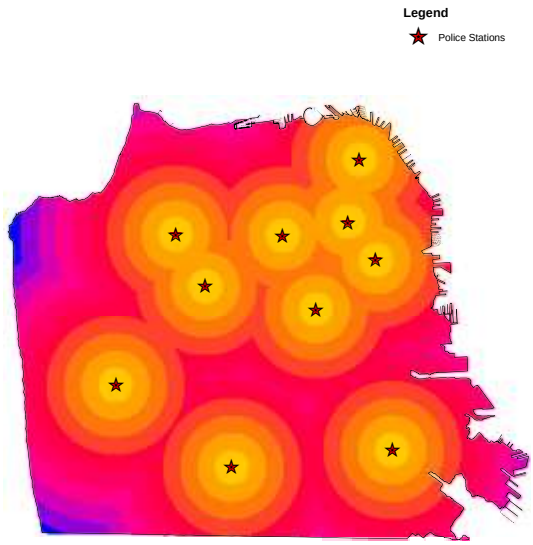
Conclusion: Strong correlation between zone type and reports

9

SF Police Station Locations

Zone	Zoning Near P.S.	Zone % of Total Land	% of reports / Zoning
Commercial	20%	11%	33%
Industrial	8%	9%	3%
Public	39%	33%	11%
Residential	33%	48%	53%
	100%	100%	100%

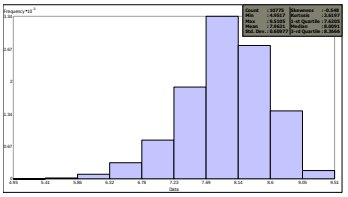
Conclusion: Police station locations are largely independent of zoning



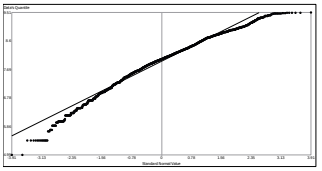
10

Graffiti: Distance from Police Stations

Histogram
Transformation: Log

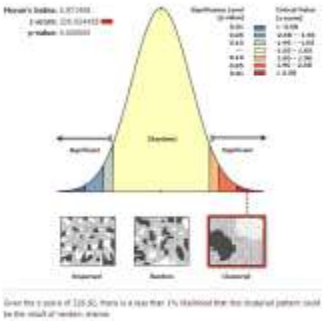


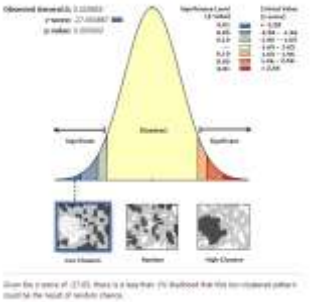
Normal QQPlot
Transformation: Log



Data Source: FiscalSRPublic08 selection Attribute: distance_f

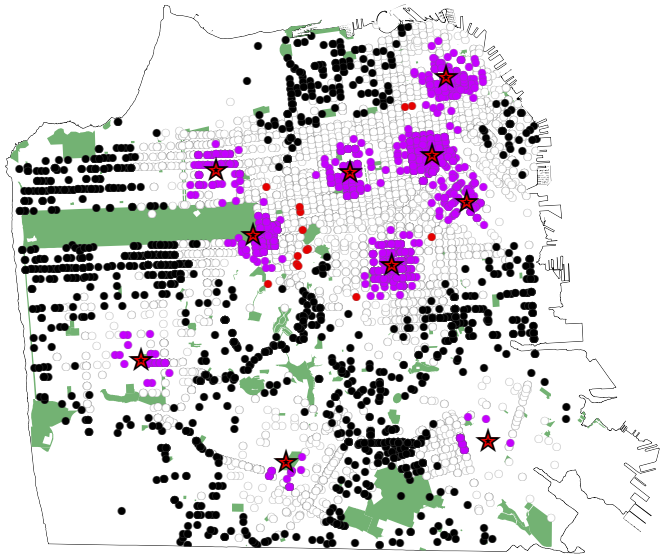
Data Source: FiscalSRPublic08 selection Attribute: distance f





11

Anselin Local Moran's Clusters: (log) Distance from Police Stations



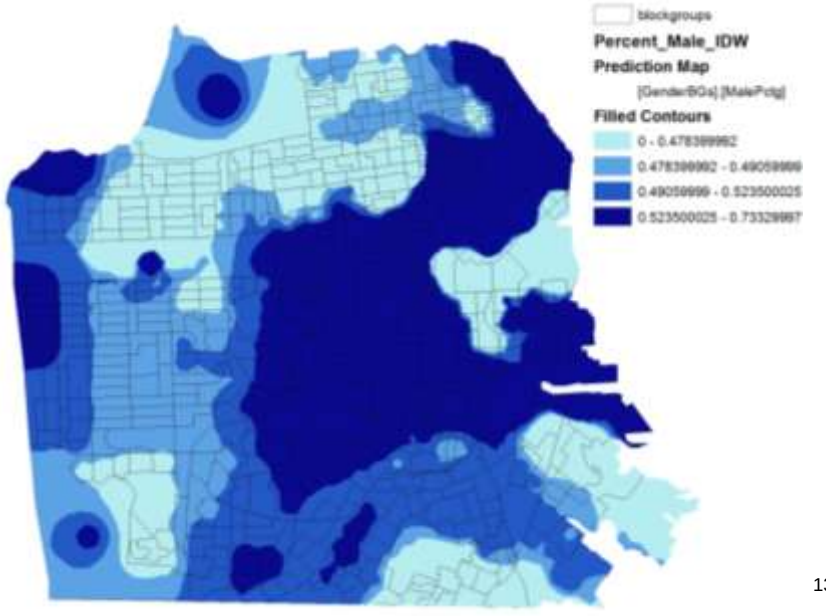
Clustering

- Not Significant
- High / High
- High / Low
- Low / High
- Low / Low
- Police Stations
- Parks

Cluster	# Reports
None	6,740
High/High	1,946
High/Low	14
Low/High	0
Low.Low	2,075

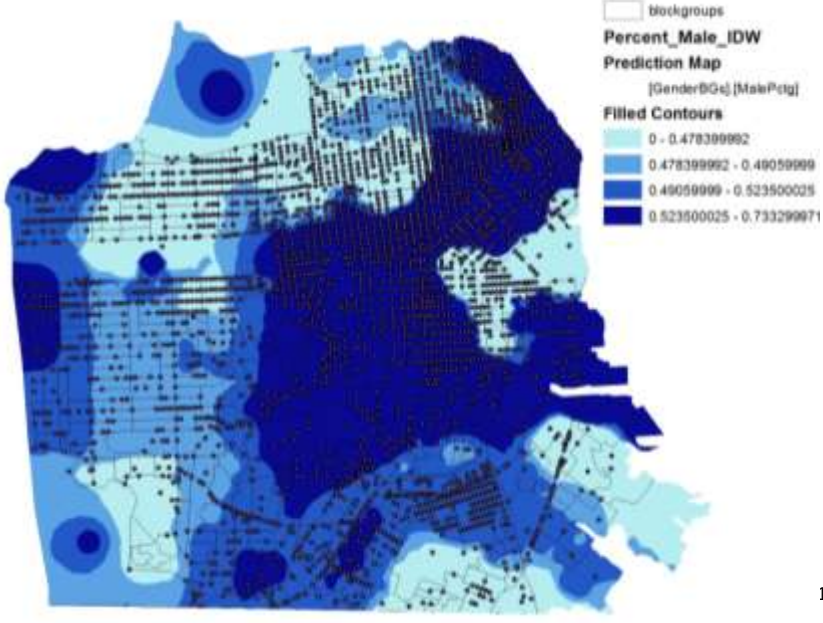
12

Percentage of Males



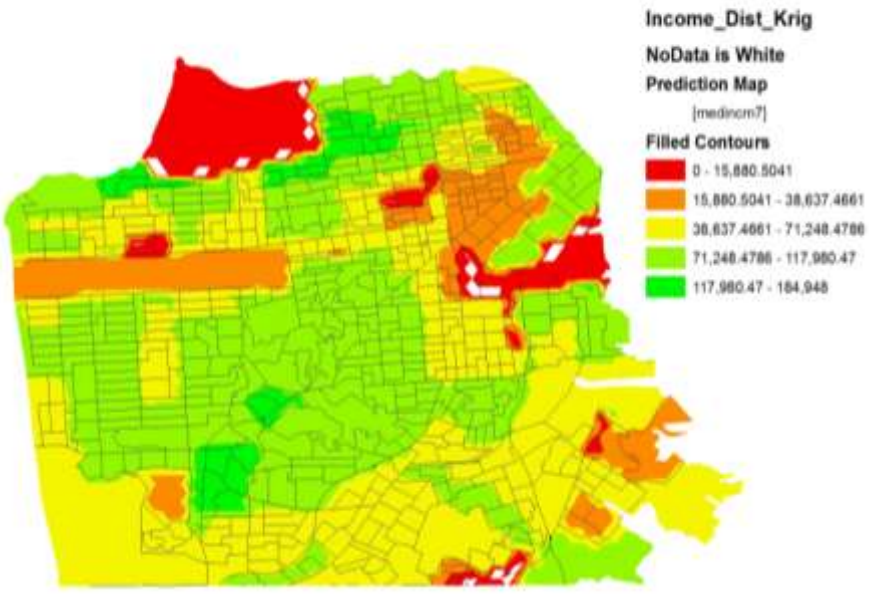
13

Percentage of Males and Graffiti Points

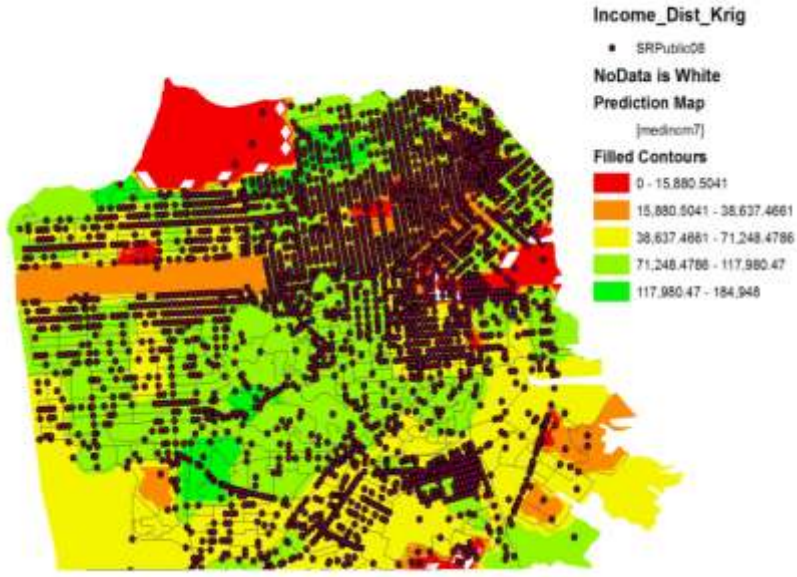


14

Income Distribution



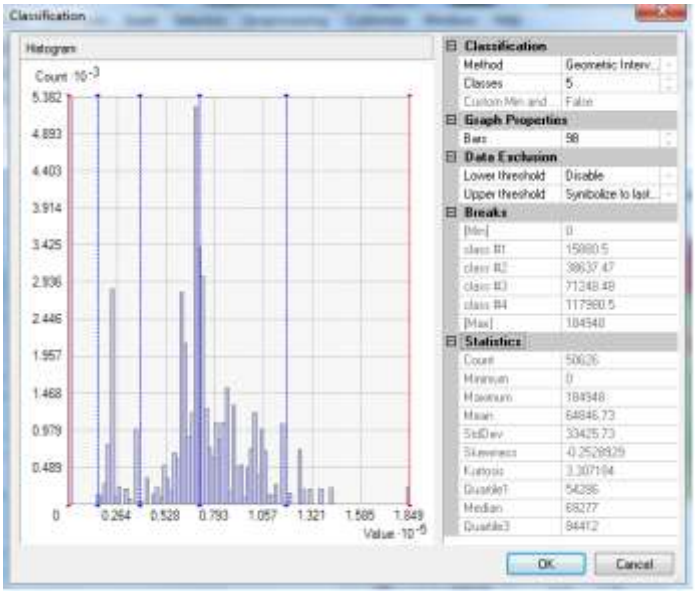
Income and Graffiti Points



16

Income Statistics

Income statistics appear to show a relatively normal distribution

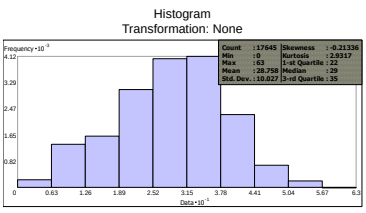


17

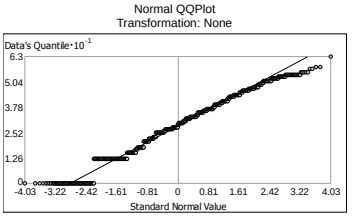
Graffiti: Distance to Parks

18

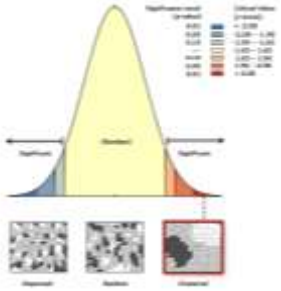
Graffiti: Distance to Parks

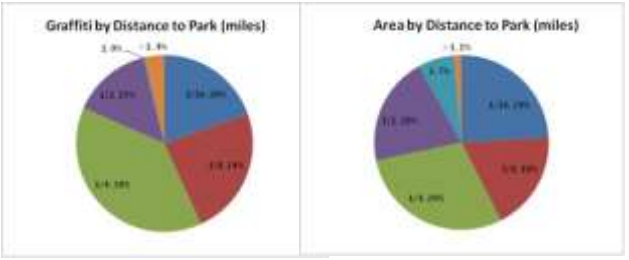


Data Source: fiscalSRpublic0708 nonull clip Attribute: sqrt dist



Data Source: fiscalSRpublic0708 nonull clip Attribute: sqrt dist

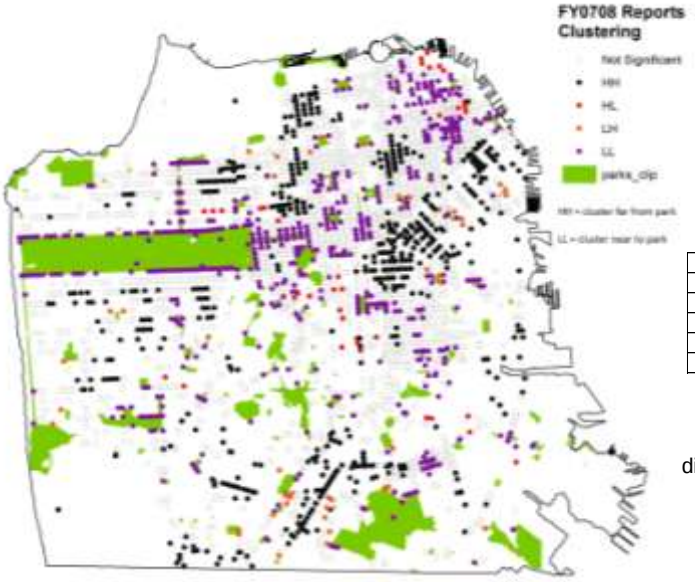




Given the z-score of 319.49, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

19

Clusters: (sqrt) Distance from Parks



Cluster	#
None	9,919
HH	3,848
HL	37
LH	43
LL	3,798

Conclusion: Bimodal distribution; close to parks and far from parks.

20

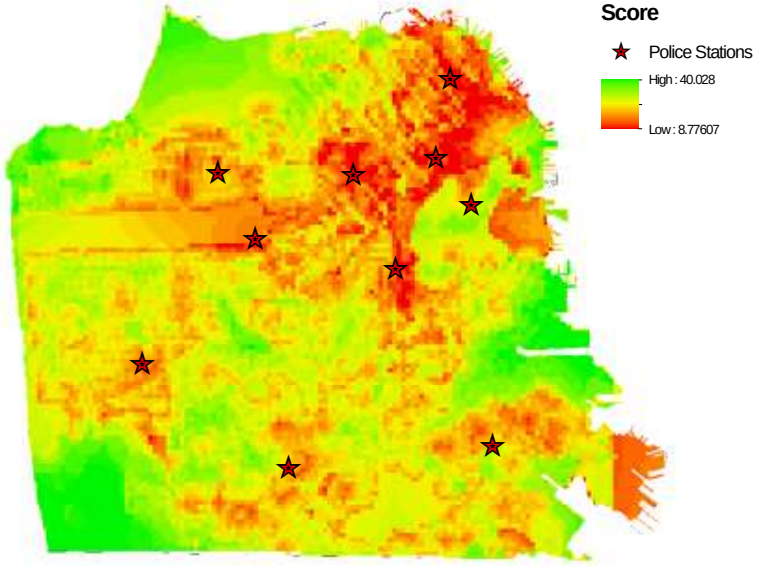
Graffiti in San Francisco – Model

GRAFFITI: Which factors were analyzed?			
feature	source	model	correlation?
Graffiti Reports	city	yes	control
Zoning	city	yes	residential; negative to industrial and public
Police Stations	city	yes	negative, distance to police station
Parks	city	yes	positive, closer to parks
Income	census	yes	negative to income
Gender	census	yes	positive to % male population
Education	census	no	no obvious relationship
Arrests	city	no	slightly positive; not statistically significant
Arterials	city	no	positive; consider transit routes
Gang Zones	city attorney	no	no obvious relationship

- For each selected factor, data was normalized and classified into 10 levels
- Selected factors were combined into a single model
  - No weighting was indicated from literature search or discussion with expert
  - Initial model was built using equal weights

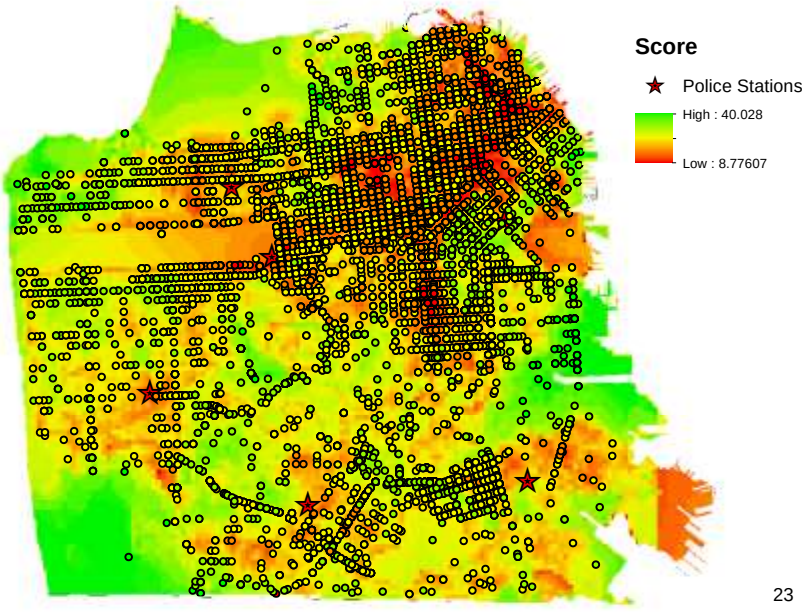
21

Explanatory Model: Equal Weights



22

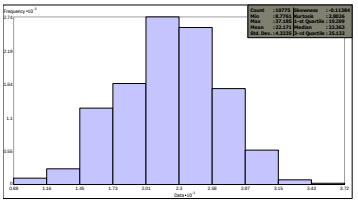
Explanatory Model with Graffiti Reports



23

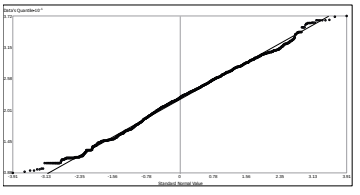
Explanatory Model Statistics

Histogram
Transformation: None

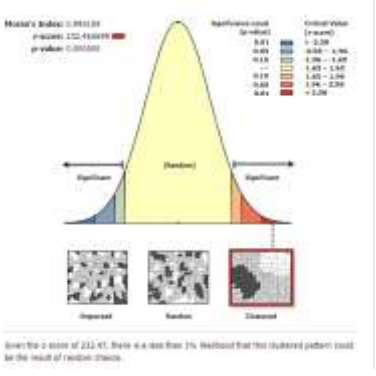


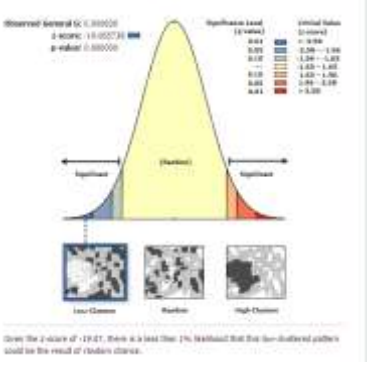
Data Source: Fiscal8SlSm5a Attribute: RASTERVALU

Normal QQPlot
Transformation: None



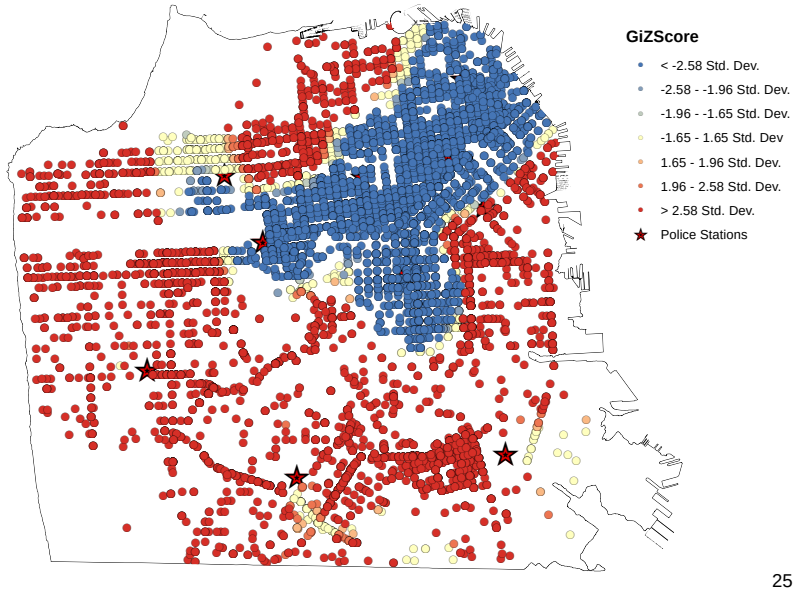
Data Source: Fiscal8SlSm5a Attribute: RASTERVALU





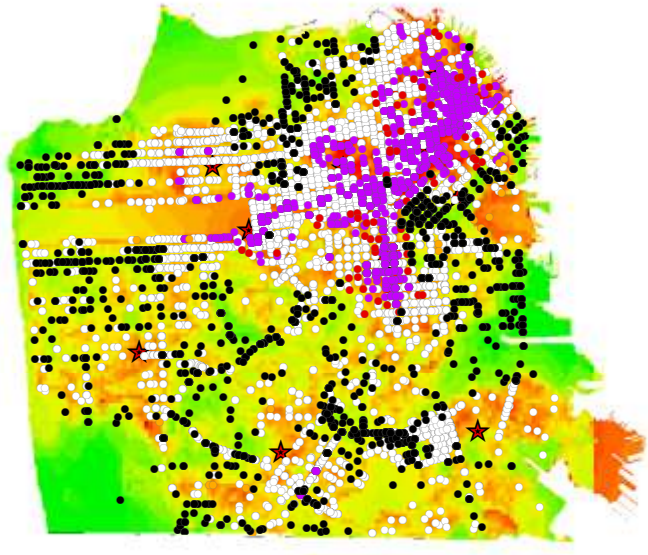
24

Model Cluster Significance: Getis-Ord Gi



25

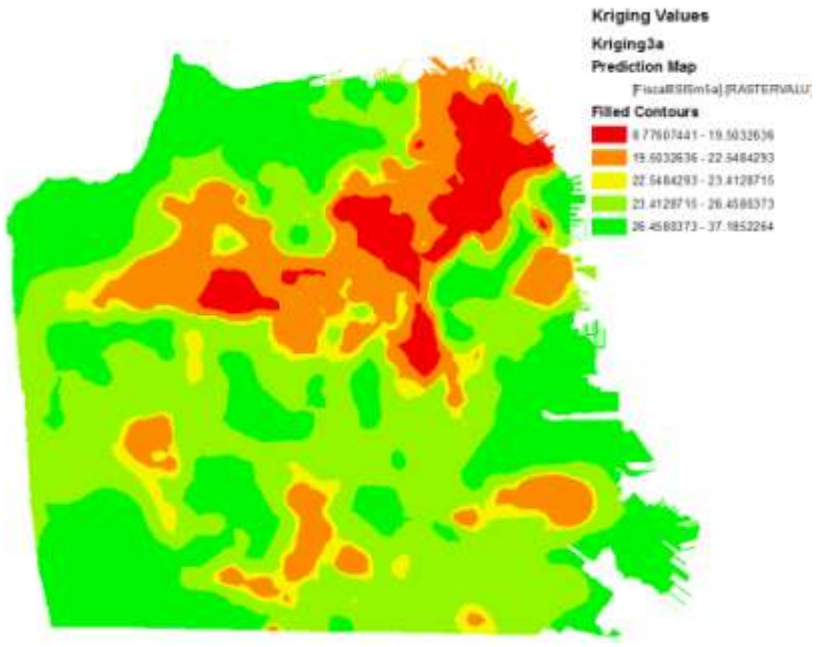
Clustering: Anselin Local Moran's



Cluster	# Reports
None	6,385
High/High	1,999
High/Low	62
Low/High	20
Low/Low	2,309

26

Kriged Map of Model



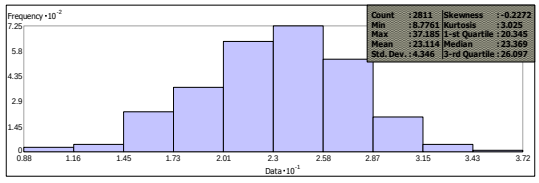
Kriging

Creating an ordinary predictor kriged map, it was found that, given equal weights, our model appeared strongly predictive.

-The histogram showed a neat normal distribution, with a difference in median and mean values of only .255

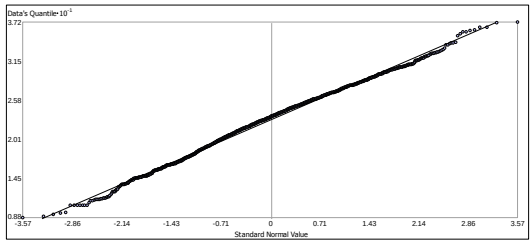
-The QQ Plot also showed a very strong connection between the data and a standard normal set of values

Histogram
Transformation: None



Data Source: Fiscal8SlSm5a Attribute: RASTERVALU

Normal QQPlot
Transformation: None



Data Source: Fiscal8SlSm5a Attribute: RASTERVALU

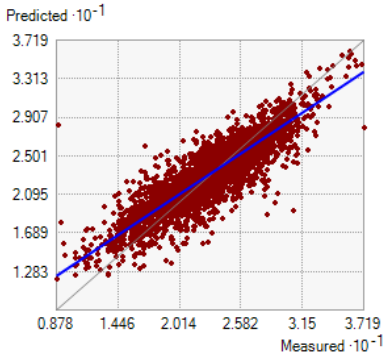
Cross-validation

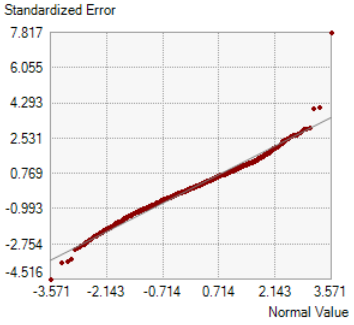
When contrasting measured versus predicted values, the model appeared quite accurate, as shown by the yellow line in the picture.

With 2811 samples examined, a mean error close to 0, a root-mean-square error and average standard error relatively low, and a root-mean-square standardized error nearing 1, this model would appear to be strongly predictive.

Mean error:	-0.01106
Root-Mean-Square:	2.16624
Average Standard Error	2.39799
Root-Mean-Square Standardized	0.90815

The QQPlot shows prediction errors and their closeness to a normal distribution.

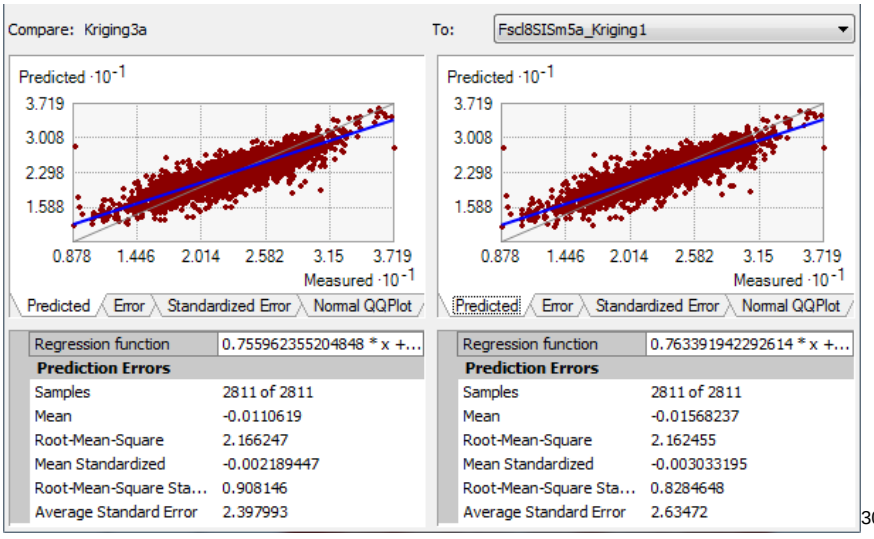




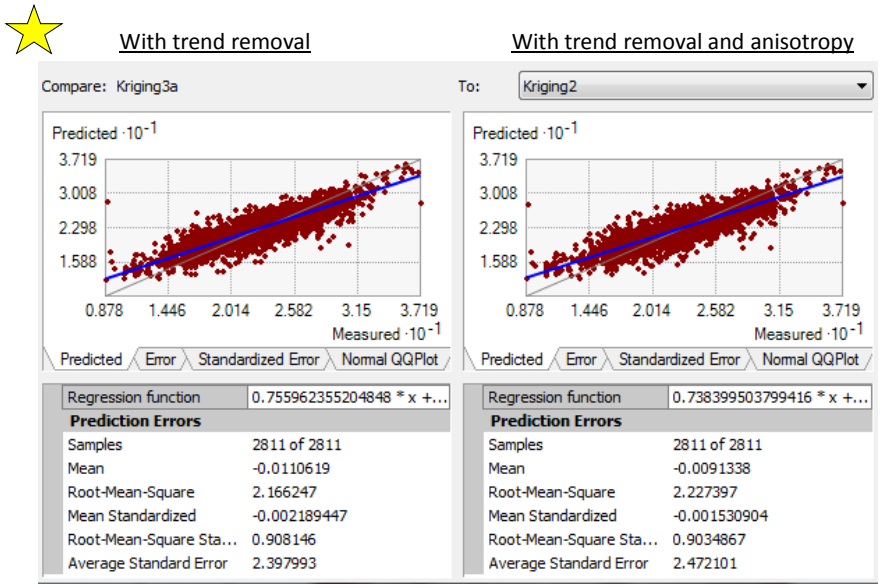
Comparing Models

With trend removal

Without trend removal

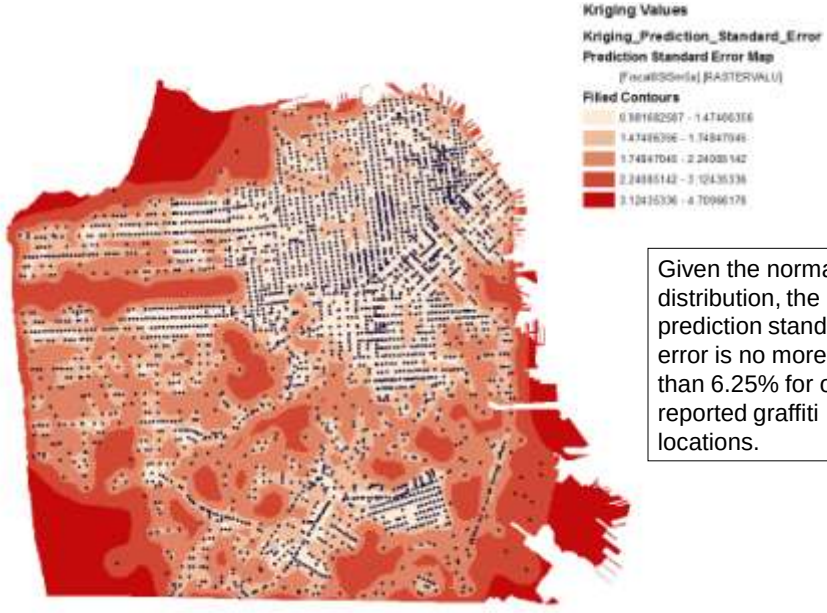


Comparing Models



31

Prediction Standard Errors



Given the normal distribution, the prediction standard error is no more than 6.25% for our reported graffiti locations.

32

Graffiti Reports in San Francisco – Considerations

External factors:

- Sampling/reporting bias
  - Relationship between graffiti and graffiti reports is unexplored
  - How do physical and environmental factors affect occurrences of graffiti?
  - How do social and administrative factors affect the reporting of graffiti incidents?
- Limited prior research available on factors correlated with occurrences of graffiti
- No information on long-term recorded trends, or on comparative weightings of factors

Project factors:

- Limited (potentially correlated) data available to analyze
  - E.g.: Transit lines, locations of bars, police beats, location of donut shops
- Masking of report locations affected analysis
- Insufficient data for interpretive predictive analysis

- Additional analyses should be performed
  - Factors should be analyzed for independence
  - Regression analyses should be performed to improve factor weightings

33

Reported Graffiti in San Francisco – Conclusions

- We developed a strong predictive model of graffiti reports in San Francisco

- Our model concludes the following factors strongly predict higher numbers of reported graffiti:
  - Land zoning (highest for commercial property, lowest for industrial)
  - Increasing distance from police stations
  - Increasing as the percentage of resident males increases
  - Increasing as income decreases
  - Decreasing distance to parks

- Our model includes conventional assumptions of graffiti causes (low income neighborhoods, distance from police stations) with unconventional results discovered through spatial analysis (percentage males)

- Only one of these factors is amenable to modification by City Council: distance to police stations
  - Additional research should be performed to validate the model and study the impact of increasing the number of police stations

34

Graffiti in San Francisco – References

[1] J. Zou, “Graffiti in New York City: Part I « Statistical Fantastic,” 23 February 2010. [Online]. Available: http://statzou.wordpress.com/2010/02/23/graffiti-nyc/. [Accessed: 05-Mar-2011].

[2] J. Zou, “Graffiti in New York City: Part II « Statistical Fantastic,” 23 February 2010. [Online]. Available: http://statzou.wordpress.com/2010/02/23/graffiti-in-new-york-city-part-ii/. [Accessed: 05-Mar-2011].

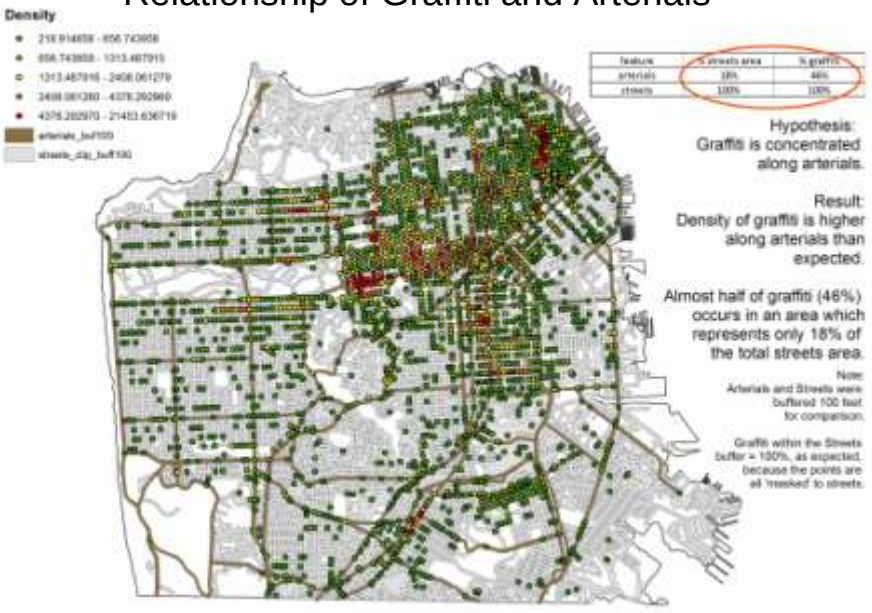
[3] Jennifer Hung, Kevin Ly, and Victor Ngo, “Graffiti In The Urban Everyday: Comparing Graffiti Occurrence With Crime Rates, Land Use, And Socioeconomic Indicators In Vancouver, BC”, April 19th, 2010. University of British Columbia, Geography 270 (201): Introduction to Geographic Information Science

“City agencies as a whole spend $20 million…” Katy Gathright, “City’s Struggle Against Graffiti Tries Rewards, Murals and Profiling”, 3 August 2010. [Online]. http://sfpublicpress.org

Graffiti diagram on slide 2: Norteño Tag.jpg, http://www.flickr.com/photos/xoconostle/1571162862/

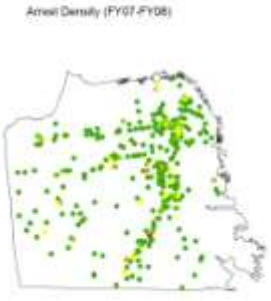
Other Analyses Performed
(but not used in our model)

Relationship of Graffiti and Arterials



feature	% streets area	% graffiti
arterials	18%	46%
streets	100%	100%

Relationship of Graffiti and Location of Arrests



Hypothesis: Arrests are made where Graffiti is reported.

Steps:
Calculate density of arrests.
Calculate density of SRs.
Compute correlation between layers.

Result:

Arrests are somewhat positively correlated to graffiti locations.

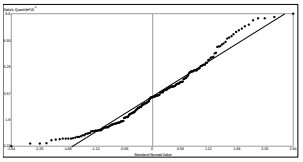
Correlation Coefficient = 0.24

CORRELATION MATRIX

# Layer	1	2
1	1.00000	0.23734
2	0.23734	1.00000

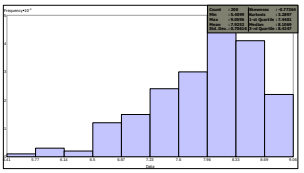
Arrests 2008 / Distance to Police Stations

Normal QQPlot
Transformation: None

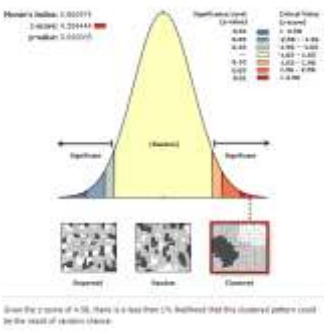


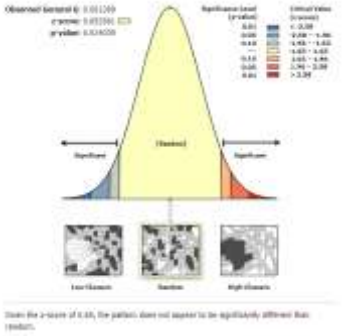
Data Source: Arrests2008Locdp Attribute: RASTERVALU

Histogram
Transformation: Log



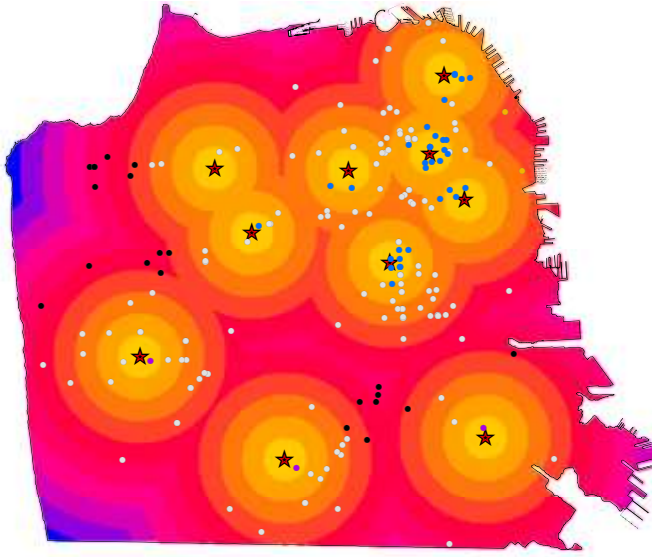
Data Source: Arrests2008Locdp Attribute: RASTERVALU





39

Arrest Clustering: (log) Distance to Police Stations



Clustering

- Not Significant
- HH
- HL
- LH
- LL
- Police Stations

Cluster	#
None	141
HH	21
HL	2
LH	3
LL	33

40

BACKUP

41

ZONE	SHORTNAME	DISTNAME
PUBLIC LANDS:		
P	PUBLIC	PUBLIC
P-W	PUBLIC-WTR	PUBLIC-WATER BODY
RESIDENTIAL:		
CRNC	CTN-RES/NEIGH COM	CHINATOWN- RESIDENTIAL/NEIGHBORHOOD COMMERCIAL
DTR	DTN- RES	DOWNTOWN- RESIDENTIAL
RC-3	MED DEN RES-COMM	RESIDENTIAL- COMMERCIAL, MEDIUM DENSITY
RC-4	HI DEN RES-COMM	RESIDENTIAL- COMMERCIAL, HIGH DENSITY
RED	SOMA RES ENCLAVE	SOUTH OF MARKET RESIDENTIAL ENCLAVE
RH-1	1-FAMILY RESIDENTIAL	RESIDENTIAL- HOUSE, ONE FAMILY
RH-1(D)	1-FAM RES-DETACHED	RESIDENTIAL- HOUSE, ONE FAMILY- DETACHED
RH-1(S)	1-FAM RES-SEC UNIT	RESIDENTIAL- HOUSE, ONE FAMILY- SECONDARY UNIT
RH-2	2-FAM RESIDENTIAL	RESIDENTIAL- HOUSE, TWO FAMILY
RH-3	3-FAM RESIDENTIAL	RESIDENTIAL- HOUSE, THREE FAMILY
RM-1	LO DEN RESIDENTIAL	RESIDENTIAL- MIXED, LOW DENSITY
RM-2	MOD DEN RESIDENTIAL	RESIDENTIAL- MIXED, MODERATE DENSITY
RM-3	MED DEN RESIDENTIAL	RESIDENTIAL- MIXED, MEDIUM DENSITY
RM-4	HI DEN RESIDENTIAL	RESIDENTIAL- MIXED, HIGH DENSITY
RSD	SOM-RES SERVICES	SOMA RESIDENTIAL- SERVICE

42

ZONE	SHORTNAME	DISTNAME
COMMERCIAL:		
C-2	COMMUNITY BUSINESS	COMMUNITY BUSINESS
C-3-G	DTN-GENERAL	DOWNTOWN-GENERAL
C-3-O	DTN-OFFICE	DOWNTOWN-OFFICE
C-3-O(SD)	DTN-OFFICE(SPEC DEV)	DOWNTOWN- OFFICE (SPECIAL DEVELOPMENT)
C-3-R	DTN-RETAIL	DOWNTOWN RETAIL
C-3-S	DTN-SUPPORT	DOWNTOWN SUPPORT
C-M	HEAVY COMMERCIAL	HEAVY COMMERCIAL
CCB	CTN-COM BUSINESS	CHINATOWN- COMMUNITY BUSINESS
CVR	CTN-VISITOR RETAIL	CHINATOWN- VISITOR RETAIL
HP-RA	HP-RED PLAN	HUNTERS POINT- REDEVELOPMENT PLAN
MB-O	MBAY OFFICE	MISSION BAY OFFICE DISTRICT
MB-OS	MBAY OFFICE SP	MISSION BAY- OFFICE SPACE
MB-RA	MBAY RED PLAN	MISSION BAY- REDEVELOPMENT PLAN
NC-1	NGBD COMM CLUSTER	NEIGHBORHOOD COMMERCIAL, CLUSTER
NC-2	SMALL SCALE NCD	NEIGHBORHOOD COMMERCIAL, SMALL SCALE
NC-3	MODERATE SCALE NCD	NEIGHBORHOOD COMMERCIAL, MODERATE SCALE
NC-S	SHOPPING CENTER NCD	NEIGHBORHOOD COMMERCIAL, SHOPPING CENTER
NCD	NGBD COMM DIST	NEIGHBORHOOD COMMERCIAL DISTRICT
SPD	SOM-SOUTH PARK	SOMA SOUTH PARK
SSO	SOM-SERV-2ND OFFICE	SOMA SERVICE/ SECONDARY/ OFFICE
INDUSTRIAL:		
M-1	LIGHT INDUSTRIAL	LIGHT INDUSTRIAL
M-2	HEAVY INDUSTRIAL	HEAVY INDUSTRIAL
SLI	SOM-SERV-LT INDUS	SOMA SERVICE- LIGHT INDUSTRIAL
SLR	SOM-SERV-LT IND-RES	SOMA SERVICE- LIGHT INDUSTRIAL- RESIDENTIAL

43

Zoning: Supporting Statistics

	# Reports	Shape_Area	% of reports	% of total land	Expected Reports
Commercial	3,698	369,430,319	33%	10.86%	1,207.75
Industrial	376	71,555,923	3%	8.88%	987.55
Public	1,168	148,020,345	11%	32.70%	3,636.60
Residential	5,879	705,199,347	53%	47.56%	5,289.20
	11,121	1,294,205,934	100%	100%	11,121.10