

SKETCHING MENTAL MAPS OF URBAN SPACES FOR THE VISUAL ANALYSIS OF SPATIAL DATA

A Dissertation
Presented to
The Academic Faculty

by

Alex Godwin

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Interactive Computing

Georgia Institute of Technology
August 2018

Copyright © 2018 by Alex Godwin

SKETCHING MENTAL MAPS OF URBAN SPACES FOR THE VISUAL ANALYSIS OF SPATIAL DATA

Approved by:

Professor John Stasko, Advisor
School of Interactive Computing
Georgia Institute of Technology

Professor Alex Endert
School of Interactive Computing
Georgia Institute of Technology

Professor Keith Edwards
School of Interactive Computing
Georgia Institute of Technology

Professor Christopher LeDantec
Digital Media
Georgia Institute of Technology

Professor Jason Dykes
Department of Computer Science
City, University of London

Date Approved: July 26, 2018

For my wife, Emily.

PREFACE

This dissertation is original work by the author, Alex Godwin. Portions of this dissertation have been published, or have been submitted for review prior to publication. There have been significant contributions by co-authors for portions of this work, which have been noted in the relevant sections.

TABLE OF CONTENTS

DEDICATION	iii
PREFACE	iv
LIST OF TABLES	viii
LIST OF FIGURES	ix
SUMMARY	xviii
I INTRODUCTION	1
1.1 Motivation	6
1.1.1 Capturing Mental Maps	8
1.1.2 Exploration of Qualitative Data from Non-experts	9
1.1.3 Utilizing Values for Policy	9
1.2 Barriers to Digital Mental Maps	10
1.2.1 Research Questions	12
II RELATED WORK	14
2.1 Image of Urban Spaces	14
2.2 Mapping Point-based Data	15
2.3 Participatory GIS	19
2.4 Interaction Techniques	20
III INTERACTING WITH CIVIC DATA	23
3.1 Sketching Spatial Data on Maps	23
3.2 Sketched Roads and Police Patrols	26
3.2.1 HotSketch	28
3.2.2 Usage Scenario	34
IV REACHABILITY MAPS OF URBAN SPACES	37
4.1 Districts and Paths	37
4.1.1 Reachability	38

4.1.2	Reachability Weighted Statistics	39
4.1.3	Road Sampling and Region Score	40
4.1.4	Region Weighting	42
4.1.5	System of Coordinated Views	51
4.1.6	Evaluation	53
4.1.7	Data	54
4.1.8	Comparison of Techniques	56
4.1.9	Discussion	60
V	ANALYZING CIVIC DATA AND URBAN SPACES	65
5.0.1	Design Goals	67
5.0.2	Approach	68
5.0.3	Community Asset and Crime Mapping Tool	69
5.1	Nodes, Paths, and Edges	71
5.1.1	Paths Revisited	73
5.1.2	Nodes	76
5.1.3	Edges	77
5.1.4	Crowd-Sourced Mental Maps of Property Crime	78
5.1.5	Discussion	80
5.2	Online Collection and Interpretation of Atlanta Mental Maps	84
5.2.1	Motivation	86
5.2.2	Method	88
5.2.3	Questionnaire	98
5.2.4	Analysis	100
5.3	Interpretation of Mental Maps	108
5.3.1	Method	108
5.3.2	Questionnaire	110
5.3.3	Analysis	111

VI CONCLUSION	112
6.1 Future Work	113
6.2 Discussion and Reflections	117
REFERENCES	119

LIST OF TABLES

1	Incidents of crime in Atlanta for one year	55
2	Participant Demographics	101

LIST OF FIGURES

1	Images of the city of Boston compiled from verbal interviews in which participants described the city, field observations conducted by the researchers, and sketched maps rapidly drawn by participants [75]. . .	2
2	Students asked “Where would you like to live?” provide responses that can be aggregated into an overall desirability map for a wide area [52].	3
3	Depictions of Crime in the South Allison Hill community of Harrisburg, Pennsylvania [27]. In Figure 3a, a group of youth from the neighborhood have created a map that indicates “bad” street corners and the alley ways that they use to navigate around the intersection. In Figure 3b, another view rate of crime is constructed to show the rate of police calls near each intersection, but the representation lacks the additional context of how this data affects the individual experience of the residents near those intersections.	15
4	Representations of the event distributions in an urban area. The heat map in Figure 4b reveals distribution more faithfully, but is disconnected from the political boundaries used to interpret the data. The choropleth map in Figure 4c creates a false impression of the relative distribution of events.	16
5	Racial Dotmap of Atlanta [21]. Each dot represents a single person, and is given a representative color to indicate that person’s race. The spacing and use of color are employed effectively to reveal areas of demographic homogeneity, more diverse areas, and the edges between them.	17
6	The Modifiable Areal Unit Problem (MAUP), in which the use of different binning strategies produces markedly different results for analyzing the distribution of events within the same geographic region. Each choropleth map depicts the distribution of one year of crime in Atlanta, but the use of different areal units produces different insights.	18
7	Jer Thorp’s Map Room in St. Louis, in which participants project images of urban data onto a large canvas and collaboratively trace, sketch, draw, and paint their own image of the city onto it [99]. . . .	20
8	Preliminary map visualization incorporating geospatial features and freehand annotations. Sketched paths in blue have been fused to the closest match on the road network from an origin to multiple destinations. Freehand closed shapes have been used to update and <i>stain</i> the paths that pass through the shapes in red.	24

9	The HotSketch dashboard. A route has been drawn through a commercial shopping area on a Friday night at 11:30pm. The type of crimes along this route and at this time are presented in the category panel in the upper right. Below that, panels show the distribution for crimes by day of week, hour of day, and season (week of year) that are contributing to the count in the category panel. At the bottom of the dashboard, a histogram shows the distribution of the crimes along the planned route	28
10	Detailed view of crimes in the dashboard after zooming into the shopping center at the end of the path in Figure 9. Violent crimes are shown as red circles; non-violent crimes are rendered as smaller black circles. The heatmap is displayed behind the event circles to retain context. By tapping or clicking on the location of a crime, a detailed panel on the left reveals additional information. Below that, a street view is shown for the crime location.	29
11	Panels provide a detailed look at the distribution of relevant events through histograms. For each panel, the current time is filled in black; other bars are colored by the score for individual kernel windows (darker bars score higher than light bars).	33
12	An officer has drawn an initial route through her beat based on the driving directions given by a wayfinding app. There is a history of activity at the start and end of her route, but few incidents along the route. The distribution of violent crimes in this area are high, especially compared to the commercial shopping area in Figure 10. To provide police presence in high activity areas, she extends part of her trip through a busy area to the south.	35
13	While it does increase the length of her drive to her destination, the addition of a detour through a busy street has significantly increased her dwell time in areas of high activity. The trip histogram at the bottom now includes a new collection of activity for the detour. The distribution of violent crimes along this route are also high, particularly for aggravated assault and pedestrian robbery. While this distribution is similar to that of her original route, the volume of historical crimes along the new route is nearly double that of the former.	36
14	The officer switches views to get a detailed look at a neighborhood with a history of frequent criminal activity. By clicking on the dots representing the relevant crimes, the officer becomes familiar with the details of past incidents and the context of the locations in which they occurred.	36

15	Representations of events in proximity to the road topology of an urban area. In Figure 15a, road scores have been joined to the underlying road network, revealing the infrastructure that occurs in close proximity and frequency to the event data. In Figure 15b, the regions have been colored based on the mean road score for the segments contained in that region. In Figure 15c, the two representations have been overlaid to explore the correlation between road sampling and region score. Region colors have been selected using Color Brewer [55].	40
16	Histogram of the road-event scores in a city. Scores are represented in log-scale due to the inverse power distribution. The scores for a region selected in the map are shown in red.	41
17	Geographically weighted mean applied at a range of scales. In Figure 17a, events are counted within regions. In subsequent figures, h_g is increased in increments of 5km.	44
18	Scalogram of the relative change in region score as the bandwidth h_g is gradually increased from 0 to 15km in Figure 17; lines are colored by their score at 15km. The scores exhibit substantial volatility from 0-5km that stabilizes from 5-10km. By 10km, most regions maintain their ordering and tend towards the mean.	44
19	Reachable locations in the periphery of a bounded region are included in its score. From left to right, the maximum distance is gradually increased to illustrate the non-uniform reachability of the region to nearby locations. While there are many reachable locations to the north, southwest, and south, other areas around the periphery of the region remain more sparse.	45
20	Reachability weighted (RW) mean applied to an event data set at a range of scales. In Figure 20a, no smoothing is applied between regions and the mean event proximity to roads is displayed. In subsequent figures, h_g is increased in increments of 5km. A lightly-shaded outlier is present in the center of the map, warranting further scrutiny. . . .	48
21	Exploration of bandwidths through coordinated views. An overview of all regions can be used to compare reachability weighted mean scores overlaid with road scores. Linked views provide (from top to bottom): details for road score distribution; a scalogram of variation in region score as bandwidth increases; and a detail isochrone map for a selected region.	51
22	A detailed look at a region of the map that has a lower score than other regions. Investigation of the overlay in Figure 22a reveals an area where no events have been sampled through roads, even though roads are visible in the isochrone of Figure 22b. This indicates that event data may be missing for the area.	54

23	Comparison of weighting methods. In the first row, events are counted without smoothing by three different modifiable areal units: police beat, neighborhoods, and 2010 census districts. In the middle row, geographically weighted (GW) mean has been applied to these regions with a bandwidth of $h_g = 5km$, however, the choice of boundaries yields significantly different results. In the bottom row, reachability weighted (RW) mean has been applied to the regions with the same bandwidth. Unlike the other methods, the color encoding in RW is relatively stable across the mapping schemes despite the changes in region size, shape, and count.	57
24	Random samples used to compare the variation in assigned score across areal unit schemes. Sample point locations are generated using Bridson's method [15] with a minimum spacing of 0.5km.	60
25	Comparison of sample point assignment as a consequence of modifiable areal units. In this figure, percentile rankings have been assigned based upon the number of events within the region that the sample point is located within. Correlation between areal unit schemes indicates that there is significant variability and little consistency, leading to very different outcomes in choropleth map interpretation and insights. . .	61
26	Comparison of sample point assignment as a consequence of modifiable areal units. In this figure, percentile rankings have been assigned based upon the geographically weighted mean (using Euclidean distance) of the region that the sample point is located within. Correlation between areal unit schemes indicates that there is significant variability but some consistency between beats and neighborhoods, leading to different outcomes in choropleth map interpretation and insights depending on the binning strategy used.	62
27	Comparison of sample point assignment as a consequence of modifiable areal units. In this figure, percentile rankings have been assigned based upon the reachability weighted mean (using path network distance) of the region that the sample point is located within. Correlation between areal unit schemes indicates that there is significant consistency and little variability, leading to more consistent outcomes in choropleth map interpretation and insights.	63
28	Map Module - The map displays drug and alcohol related crime (colored hexes) and vacant houses selected (blue circular pins).	70
29	Events are assigned to the closest lixel in the network before diffusing to nearby lixels. The local bandwidth h_l represents the maximum distance that a crime should considered to be associated with a representative street location. The global bandwidth h_g represents the maximum distance that connected and nearby streets are affected by nearby crimes.	73

30	2015 Property Crimes in Baltimore. In Figure 30b, the path KDE is rendered as line thickness and saturation, with thicker and darker lines indicating a higher network score. In Figure 30c, the path network is shown in yellow, with green highlights indicating the lixels that have been amplified due to node proximity (church and school nodes are black circles). In Figure 30d, the red highlights indicate lixels that have been dampened due to edge proximity (highway, water, and train edges are black lines mostly occluded by the path highlights).	75
31	The paths of an urban area are used to represent spatial data through network KDE with varying lixel size. Paths with a nonzero score are drawn in blue; path thickness is mapped to each lixel's net KDE percentile score for the map. Bandwidth h_g sizes for all images are 50m. Reducing lixel length increases the fidelity of the model but increases the computation cost for deriving the network KDE.	76
32	The paths of an urban area are used to represent spatial data through network KDE with varying bandwidth size. Lixel size for all images is 10m. Paths with a nonzero score are highlighted in blue; highlight thickness is mapped to each lixel's net KDE percentile score for the map; yellow paths for each line are thicker for lixels in the top 2%. In Figure 32a, the bandwidth h_g is set at 10m, which reveals only the most localized trends in path pattern variation. In Figure 32b, h_g is increased to 100m, which allows for identification of larger patterns throughout the network. Finally, in Figure 32c, h_g is set at 1km, which depicts the overall trend of the map but obscures fine-grained patterns in this data.	77
33	The disparities between the crime map and the church and school mental map. Areas where church and school lixel values are higher are in thick red lines, while areas where the original map is higher are in thick blue lines. The greatest disparity between the mental map and the raw data occurs around West Midtown and the intersection between the major highways in the center of the map near downtown. .	81
34	The disparities between the crime map and the church and school mental map. Areas where church and school lixel values are higher are in thick red lines, while areas where the original map is higher are in thick blue lines. the largest disparity occurs with high mental map distributions in East Baltimore in the neighborhoods of Oliver, Dunbar-Broadway, and Middle East. Many other neighborhoods, notably downtown and Fells Point, are lower, though Harlem Park to the west and the stadium area to the south also have higher property crime distributions.	82

35	The disparities between the crime map and the church and school mental map. Areas where church and school lixel values are higher are in thick red lines, while areas where the original map is higher are in thick blue lines. For the city of Chicago, most of the disparities in the mental map exist on the edge between the Loop and South Loop neighborhoods. To the south, neighborhoods along the lake and to the southwest along the highway also have an increase, as well as the northern neighborhoods around Goose Island.	83
36	Random samples used to group participants to the closest “home” neighborhood and to elicit rankings of the areas in the city of Atlanta. After seeding each NPU with a single sample point at its centroid, sample point locations were generated using Bridson’s method [15] with a minimum spacing of 1km. In the home selection interface, a textured circle, rather than a point marker, is used to convey that the area nearest to the center of the circle is also included in that region. . . .	88
37	All participants were prompted: <i>Of all of the neighborhoods in Atlanta, which five do you feel the most fear from crime? Which five do you feel the least fear from crime?</i> By clicking on neighborhoods in the map, locations were added to reorderable lists. Red markers are ranked high, blue markers are ranked low. A similar interface was used to elicit responses to the question: <i>Of all of the neighborhoods in Atlanta, which five do you feel is the most expensive to live in? Which five do you feel is the least expensive to live in?</i>	89
38	Participants in the Simple Group were prompted: <i>Suppose you were suddenly given the chance to choose where you would like to live- an entirely free choice that you could make quite independently of the usual constraints of income or job availability. Where would you choose to go?</i> By clicking on neighborhoods in the map, locations were added to reorderable lists. Blue markers are ranked highly desirable, red markers are ranked less desirable. The task is a digital version of the map ranking task used by Gould and White [52].	90
39	Complex group participants were prompted: <i>Suppose you were suddenly given the chance to choose where you would like to live- an entirely free choice that you could make quite independently of the usual constraints of income or job availability. Where would you choose to go?</i> The task is a dynamic version of the blank sketching process used by Lynch and Appleyard to elicit imageability from participants on paper maps [75, 8], and uses the digital interface to dynamically adjust the color in response to the user’s sketched lines.	91

40	Participants in the simple group were shown this style of map moments after expressing the desirability of the neighborhoods in the city, which has been constructed using a regularly spaced grid of sample locations that are assigned a hue based on proximity to the nearest neighborhood ranked by the participant. For consistency with the Complex group, isolines have been drawn to separate major groups by their desirability.	92
41	Participants in the Simple Group sketch the important features of their neighborhood. The task is a digital version of the blank sketching process used by Lynch and Appleyard to elicit imageability from participants on paper maps [75, 8].	94
42	Participants in the Simple Group were shown this style of map after sketching the important features of their neighborhood. It depicts the features that they drew in the task but reveals the underlying map tiles for context.	95
43	Participants in the Complex Group sketch the important features of their neighborhood. The task is a digital version of the blank sketching process used by Lynch and Appleyard to elicit imageability from participants on paper maps [75, 8], enhanced with dynamic feedback from interacting with the reachability model of connected streets. . .	96
44	Participants in the Complex Group were shown this style of map after sketching the important features of their local neighborhood. The sketched features have been applied to a network reachability model in which edges dampen the reachability of nearby lixels and nodes amplify nearby lixels. Lixel that are close to any feature sketched by a user are considered <i>source lixels</i> , and are assigned a score of 1. In the generated representation, the network KDE score is encoded in the hue and thickness of the lines overlaid on the map.	97
45	Neighborhood planning units E (midtown) and M (downtown) were the most common home NPUs for participants in the study (NPU-E : 56; NPU-M : 18).	100
46	Perception of Crime (dots) vs Crime Locations (background). On the left are perceptions obtained from participants in all neighborhoods, on the right are perceptions obtained only from participants in NPU-E. The foreground dot hue pattern depicting a separating line of fear is in contrast to the background hue pattern depicting crime central to midtown and west of downtown.	103

47	Perception of Crime (dots) vs Crime Locations (background) obtained only from participants in NPU-M. The foreground dot hue pattern is unlike the dot pattern for NPU-E, and is only focused on a small portion of the area around the west of downtown. Notably, a small portion of that area on the west of the NPU is seen as ranked quite low for fear from crime.	104
48	Perception of Crime (dots) vs Desirability (background). On the left are perceptions obtained from participants in all neighborhoods, on the right are perceptions obtained only from participants in NPU-E. Notably, while NPU-E participants do not perceive most of NPU-E as fearful, they also do not desire strongly to live there, instead showing a preference for the eastern edge of the region and beyond.	105
49	Perception of Crime (dots) vs Desirability(background) obtained only from participants in NPU-M. Unlike the residents of NPU-E, the perceived fear from crime at neighborhoods within the NPU does not dissuade residents from desiring to live there.	105
50	Perception of Price (dots) vs Median House Listing Price (background). On the left are perceptions obtained from participants in all neighborhoods, on the right are perceptions obtained only from participants in NPU-E. Notably, NPU-E participants are fairly accurate at predicting home prices, save for the Home Park area north of campus where perception of crime is similarly different from the available data and the northeast portion of the map which is much more expensive than participants of all groups perceived.	106
51	Aggregated mental map features of NPU-E obtained from residents of that neighborhood during a Complex sketching task.	106
52	Aggregated mental map features of NPU-E obtained from residents of that neighborhood during a simple sketching task.	107
53	Interface for analyzing the interpretation of mental maps. In this image, the image on the left is aggregated from user reported perceptions of crime, housing price, and Complex desirability as a background layer. The representation on the right depicts the aggregated network features from the Complex group for each of the individual neighborhoods that make up make up the NPU. Controls in the center can be used to select other hues for the foreground or background layers on the left.	108

- 54 Interface for analyzing the interpretation of mental maps. In this image, the image on the left is aggregated from user reported perceptions of crime, housing price, and simple desirability as a foreground hue for the dots. The representation on the right depicts the aggregated network features from the simple group for each of the individual neighborhoods that make up the NPU. Controls in the center can be used to select other hues for the foreground or background layers on the left. 109

SUMMARY

This document details research conducted at Georgia Tech to determine technology-based methods for improving the dialogue between civic leaders and ordinary citizens. I discuss software designs for helping people to capture their perception of the important parts of a city and the connections between these parts as a *mental map* that may be different from “official” maps constructed by land surveyors or political agencies. I also discuss ways that these mental maps can be put to use to understand how groups of citizens see the city, and I show the differences that can exist between different groups and between the overall perception of the city and official maps.

CHAPTER I

INTRODUCTION

Managing a city is a difficult problem. They are, by definition, made up of a huge number of people that each have different beliefs, values, and identities. A city reflects these values, and for cities like Atlanta, even people living in close proximity can have wildly different perceptions of what is valuable and what should be changed. In midtown, for example, the skyline is shifting rapidly as new apartment complexes and condominium high rises are constructed to serve the booming housing demand in the city's core. This is fantastic news if you can afford the luxurious options on display, but not everyone can. Many other neighborhoods are struggling to balance the pace of development with preserving an existing cultural identity. The beltline is rapidly connecting neighborhoods and reshaping the economy in the core of the city, but also bringing property taxes to a point that threatens long-term residents [87]. Like many cities, the topic of gentrification is raised frequently, with contested definitions and no clear solution. Throughout the city, twenty-five Neighborhood Planning Units (NPU) are charged with providing a forum for community members to work with city administrators and urban planners to make decisions on zoning and land use, but major differences in beliefs, demographics, and income create substantial divides in the opinions for how this outcome can best be achieved.

This is not, by any means, a new problem. Kevin Lynch was one of the first to study the problem of disparate mental maps held by members of the community (Figure 1), and worked to capture these maps as physical sketches that could be discussed in group settings to establish the navigability of areas in a city [75]. Donald Appleyard built from this work by documenting the conflicting beliefs and values of urban

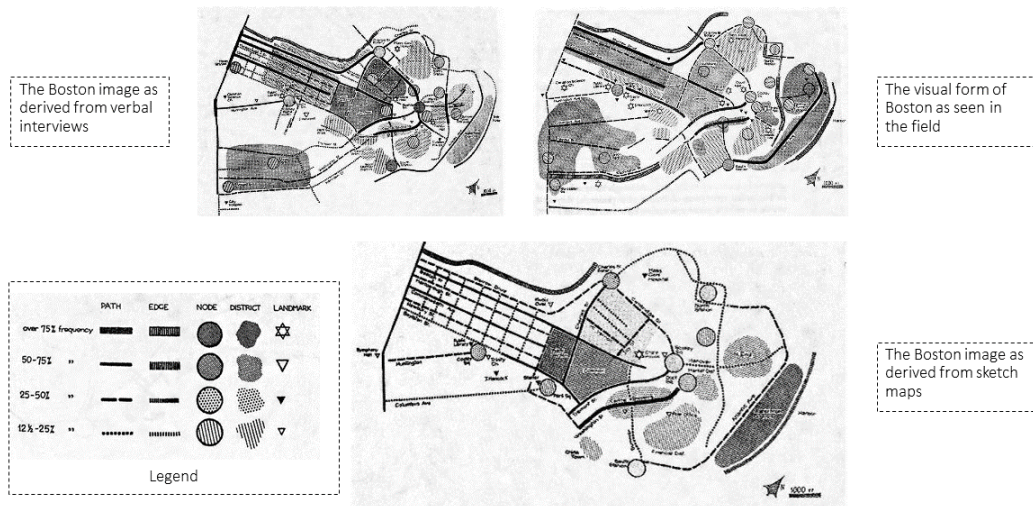


Figure 1: Images of the city of Boston compiled from verbal interviews in which participants described the city, field observations conducted by the researchers, and sketched maps rapidly drawn by participants [75].

planners and inhabitants during the design and development of Ciudad Guyana [8]. Cities also grow and change as a consequence of new technology, environmental factors, and the evolving requirements of their citizenry. Major metropolitan areas now have entire departments dedicated to analyzing and mapping important data through geographic information systems (GIS). As documented extensively by Appleyard, the problem with keeping city design and mapping solely within city administration is that it tends to reflect the beliefs and values of the urban planners and those in power. Inhabitants of a city can have very different conceptual understandings for what features of a city are important, and ascribe meaning and value to different recognizable elements.

At a much larger scale, Peter Gould and Rodney White studied the use of transcribed mental maps to explore factors of desirability in the United States, the United Kingdom, Sweden, and Canada [52]. In the U.S., they surveyed university students

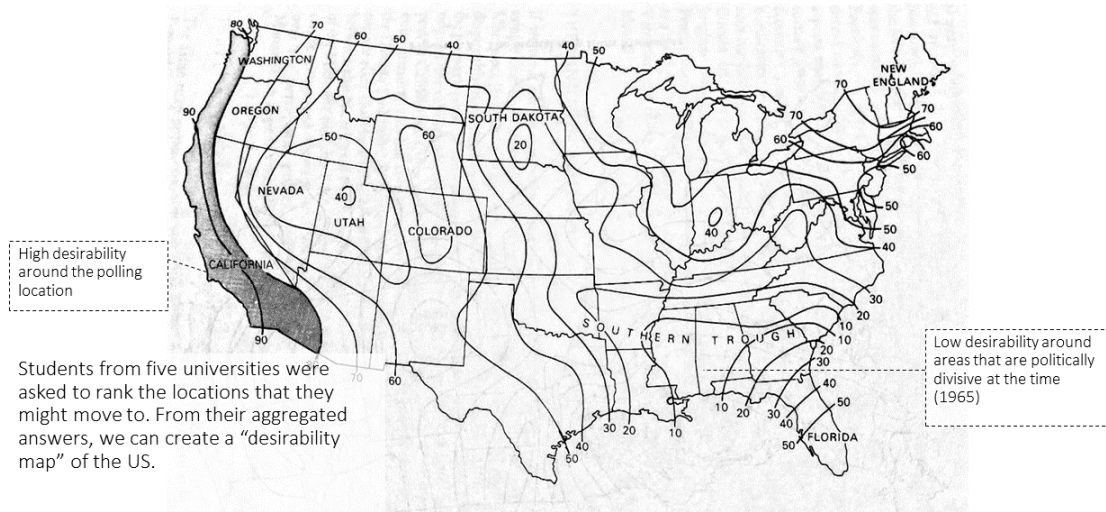


Figure 2: Students asked “Where would you like to live?” provide responses that can be aggregated into an overall desirability map for a wide area [52].

in five different regions of the country to determine their ranked preference for moving to each of the fifty states after graduation [51]. By simply answering the question, “Where would you like to live?” students from each region exhibited markedly different preferred locations, revealing varying importance placed on perceived environmental quality, political quality, economic opportunities, and social and cultural aspects (Figure 2). These answers were obtained through simple ranking tasks in which participants indicated on a paper map their preferred order of moving to each of the possible locations. When aggregated, these responses produce a rich contour map of a large area in which lines separate areas that are perceived as more or less desirable.

Geographic information systems (GIS) have advanced substantially in the years since this research was introduced, and modern systems provide a wealth of capabilities for the analysis of complex spatial data. Unlike the paper maps used by Gould, White, Appleyard, and Lynch, the typical interface of a GIS is complex, feature-rich, and powerful. Many cities, such as Atlanta and Boston, have departments dedicated

to supporting the analysis of spatial data in city planning through GIS systems. Commercial systems, such as ArcGIS, make it possible to create complex maps and share them embedded directly into websites. Through open-source libraries like Leaflet, Mapbox, and Turf, developers can create powerful mapping sites for more customized analysis of specific types of data.

Unfortunately, use of a GIS system or developing a site can prove very intimidating for inexperienced users, who are more at ease using natural speech and sketching to access map-based data than through a GIS [91]. Hand-drawn materials (e.g., engineering sketches) can often improve collaboration by serving as boundary objects between different stakeholders [56, 95], which allows different users to read domain-specific information from the same object and improve distributed cognition. Plans can take many different forms, but are created by urban planners and civic engineers to specify the changes that alter a physical space from its current form into another. For plans that involve changes to a physical location, such as a proposed housing development or sports arena, maps provide a layer upon which notes, changes, and annotations can be drawn to respond to details of the plan or suggest alterations. Previous research has indicated that rapid, sketch-based prototyping of user interfaces can facilitate a more free-form exploration of alternatives [68]. These aspects can substantially improve the value of an illustration as an artifact that promotes distributed cognition among stakeholders, allowing them to rapidly externalize and share mental models within a team. Finally, a “sketchy” rendering style can be used in information visualization to encourage annotation and to convey the unfinished quality of uncertain data [111].

Geospatial data within urban spaces is inextricably tied to the architectural forms and topological features of the city, which can have varying levels of importance for different inhabitants. Lynch classified these features into five primary elements: nodes (areas of heightened activity and interaction), paths (channels that people take to

move around the city), edges (barriers that divide one area from another), districts (well-defined regions that a person can be within), and landmarks (simple and highly visible features useful for navigation) [75]. These elements have been used extensively to analyze the structure of urban spaces, identify locations for improved design, and to compare cities to each other in content and aesthetic appeal.

In a detailed analysis of the US cities of Boston, Jersey City, and Los Angeles, Lynch detailed how these elements can be used to ascertain a city's imageability: "that quality in a physical object which gives it a high probability of evoking a strong image in any given observer." The image of the city can contain many features that are shared by its inhabitants, but large groups of people may have different opinions on the essential elements. These differences can be used to inform the analysis of behavioral data, such as crime patterns, that are governed by human movement through urban spaces [13, 9]. Criminals and non-criminal alike may overlap significantly in the images that they maintain of the city, but acute differences in these images and the built environment can reveal themselves to a criminal as an opportunity for mischief [79]. Brantingham and Brantingham have provided an extensive analysis of the ways in which Lynch's elements, notably nodes, paths, and edges, contribute to trends in behavior [14].

While it may seem unusual to incorporate differing and subjective perspectives into the visual analysis of behavioral data, it should be noted that such data is often fraught with ambiguity. For example, in an analysis of crime data during a series of violence prevention programs in Cardiff, researchers determined that fewer than one-third of violent incidents in the UK and Scandinavia requiring emergency treatment appear in police records [41]. The absence of these incidents from the official record points to a striking difference of perspective between the police and the inhabitants of the city, notably, victims of violence. Clearly, alternative perspectives are necessary to provide insight into the fuller nature of urban spatial data. This

local knowledge can be incorporated through a variety of mechanisms, notably Public Participation in Geographic Information Systems (PPGIS) [93, 107] and Bottom-Up GIS (BUGIS) [97]. Lynch’s imageability elements serve as an ideal basis for capturing mental maps of urban spaces as digital representations that facilitate the incorporation of differing perspectives in the analysis of spatial data [65].

1.1 Motivation

It is true that many domains and forms of analysis may not require an understanding of city data along navigable elements; for example, it is of little value to estimate the movement of seasonal weather patterns between states, as temperature fronts do not follow interstate highways and they do not respect perceived neighborhood boundaries. People do follow roads, however, and when trying to understand the likelihood of travel time delays and accidents due to inclement weather it can be useful to model its effects within the context of the underlying road network topology. From this perspective, point-based events (e.g., accidents, flood locations) can be modeled as adverse elements that close individual links within the road network, forcing travelers to take other routes [63]. This approach would also not work for certain types of discrete spatial events, such as bird strikes, that do not occur within the city or are affected by human behavior or mobility. For domains that are strongly tied to human mobility, the importance of road network topology is difficult to understate: it is the principal medium through which we move between locations whether on foot, on a bicycle, or in a car. For example, the elements of an area are strongly tied to both the amount and types of crime that occur there [9]. Areas with high levels of thoroughfare are at increased risk of crime because of their path reachability, and closed-off streets and culs-de-sac are at a reduced risk. Connectivity is often preferable for increased walkability and cycling, but urban planners must also take into account the effects that the design of an area have on negative aspects of living there.

An area of recent concern is the path reachability of an area and its effect on access to substantive and affordable food [20]. Based on the method of travel, it can take hours to move from a “food desert” to a store providing healthy groceries and back. Because of this limited access, the residents of a city have an increased prevalence of illnesses related to heart disease and diabetes. Understanding the relationships between nodes with available food and paths connecting them is the first step in providing additional access and designing new programs that help. There is a commercial interest as well, in that other types of retail stand to benefit from understanding the prevalence of other similar businesses in proposed locations and the subsequent competition that might be faced. By using the techniques described in this research, legislators would be more capable of comparing the different districts of a city and the differing access to healthy food options that are perceived as available to inhabitants depending on what elements of the city they view as most important. This could be an important first step for identifying districts that need additional support and the edges that separate them, either through tax incentives that increase the number of healthy food nodes or through improved public transportation that improves the connectivity to neighboring areas with better options.

The primary benefit of incorporating citizen-sourced mental maps in the analysis process is to identify disparities between perception of a quality within the urban landscape and the quantitative data about that quality. For example, members of a community might be fearful because it feels to them that crime is increasing, when the authoritative data of crime distributions indicates no significant change. Recognizing these differences in mental maps is just the first step to understanding how they affect perceptions of quantitative data. The goal of my work is to design and evaluate interaction techniques that can be used to capture mental maps of urban spaces, allowing powerful capabilities of visualization and analysis while retaining the simplicity and ease of understanding provided by hand-drawn maps. There are three

overarching problems that visual analysis of spatial data with mental maps can help solve:

- GIS systems are complex and expensive, so public participation in GIS is largely through paper-based capture methods based on sketching exercises that are not easily shared or stored and that are time consuming to aggregate.
- Current approaches for GIS rarely incorporate qualitative data obtained from citizens within approaches for analyzing quantitative data. If kept separate, qualitative data can be difficult to interpret and therefore easy to ignore.
- Qualitative data collected from people is separate from and relatively ignored in the face of compelling quantitative data, leaving citizen concerns unaddressed. New policy that affects citizens does not incorporate the values and beliefs of the people who live in an area.

1.1.1 Capturing Mental Maps

Paper-based sketches and surveys provide an intuitive route for obtaining a wide variety of citizen-sourced images of the city, but this approach typically relies on participants attending workshops at a central location to transcribed maps by hand. This can create complications in obtaining a large number of participants who are unable to attend the meetings in person due to transportation and scheduling challenges. After these are completed, the physical depictions of mental maps can serve as a powerful and compelling overview of the way in which participants perceive the built environment around them, and provide context for their individual or group experiences in the city. As static images, these sketches can provide insight into the value that citizens place on different features of the city and their hopes for the city's plans to preserve those features, but the sketches themselves do not promote further interaction or investigation beyond the short window of time that researchers have

to ask the participant questions about the image. Finally, collating and aggregating these artifacts into a single data representation is time-consuming and subjective. Through online digital sketching techniques, people can transcribe and annotate their own

1.1.2 Exploration of Qualitative Data from Non-experts

In cities like Atlanta, there are areas that are changing rapidly through the demolition of existing structures and the construction of new buildings. While this is expected for the sake of progress, the current residents do not always have the ability to engage with local government in determining how these changes will take place. Residents of affluent neighborhoods often have access to professional services, such as GIS experts, architects, and urban planners that can help analyze impending changes to the neighborhood and interpret the wishes of the residents into a modified version of these plans. Such advocacy can be an important, and costly, requirement for maintaining a voice during negotiations for use of public space. Without such resources, however, residents of lower-income neighborhoods find it difficult to maintain equal footing with the professionally developed plans proposed by the city. Improved interaction techniques could help create a more even footing for residents of these neighborhoods by providing improved access to methods of analysis for local, map-based data. These interaction techniques could be utilized on smart phones and tablet devices, which are often still present in homes without more costly desktop computers.

1.1.3 Utilizing Values for Policy

Map visualizations can often incorporate landscape topography, transportation topology, spatiotemporal data, and annotations of important elements provided by designers. What is frequently missing is the incorporation of city elements considered vital to novice GIS users who do not have the ability to download, manipulate, and host complex maps. This makes it impossible to identify differences in beliefs and values

about the city and spatial data under scrutiny. Without a nuanced conversation of these differences in perspective, it can be challenging to move forward with policy that is equitable to experts and novice users alike. Incorporating individual mental maps in the analysis of spatial data relevant to city policy questions can bring these differing perspectives to the forefront and allow for a meaningful discussion of the elements that make up these differences in values. This will allow for more rapid discussion of these ideas, while also allowing novice users to quickly contribute to discussions of map-based data.

1.2 Barriers to Digital Mental Maps

There are many software systems available for constructing and analyzing spatial data, from feature-rich commercial software like ArcGIS to free and open-source solutions like QGIS. Other tools for data visualization, such as Tableau and d3.js, provide some capabilities for mapping spatial data though that is not their primary purpose and the supported features are more limited. Many other solutions are available, but there is currently no software solution available to the public designed solely for capturing digital representations of mental maps.

While these systems, particularly ArcGIS, can be used for comparing or aggregating maps, the steep learning curve is prohibitive for those without a background in GIS. Once collected in paper form, each individual hand-drawn mental map needs to be transcribed by a member of the interview team before they can be analyzed in aggregate. This is a time-consuming and labor-intensive process. Matei et al., for example, asked participants in Los Angeles to use crayons to draw mental maps in order to study the perception of the environment and its correlation to feelings of fear [80]. Each of the 215 maps were translated by research personnel to numeric matrices in ArcView (an entry-level form of ArcGIS available at the time) before

analysis could take place. While free online tools do exist that can be used to collect digital sketches of the environment, from JavaScript libraries like Leaflet.js that support mapping to more complete applications like GeoJson.io, there is still no tool to facilitate aggregated analysis of collections of mental maps and few methods for joining mental maps to official data directly.

This is unfortunate, especially considering that much previous work has indicated not only refinements to the paper-based techniques of Lynch and Gould & White but also the benefits that mental mapping can bring to civic discussion. Public perception of urban data often differs from official data sources, and these differences can give context to the data or highlight deficiencies. Expressive tools for digital mental mapping can support public participation in GIS (PPGIS), providing an opportunity to bring these differences to light. These barriers for digital mental mapping indicate a need for additional research in this area to build and explore these capabilities in web-based tools. In this thesis, I will discuss my research to explore techniques for capturing and analyzing digital representation of mental maps.

Thesis Statement Quantitative spatial data, such as crime locations and housing prices, can be used to substantially augment our understanding of topics of interest to civic planning, including human behavior. However, it does not necessarily capture the qualitative values or understanding of an environment from the people that live within it, which is important for promoting public participation in GIS. Interactive mental maps may provide an external cognition aid that allows people to confront and interpret their qualitative understanding of an urban environment during quantitative analysis of spatial data. A mental map encodes the aspects of an area that a person considers core to the nature of that place, and the creation of a digital mental map can both incorporate the significant elements of the urban environment and reflect the beliefs of the person who created the map. Creating digital expressions of these

elements and beliefs within software systems may facilitate cooperative discussion of spatial data between everyday people and experts.

1.2.1 Research Questions

How do we analyze civic data? There are two substantial challenges to urban data analysis that this thesis seeks to explore. These challenges lead us to these primary research questions:

- There are many different ways to represent the same underlying data set, and this can lead to misinterpretation and miscommunication. How can interactive visualizations improve interpretability of representations of mental maps?
- People have different values and opinions on what is important about their city: what should be changed, and what should be preserved. How can interactive visualizations improve current methods for capturing and depicting mental maps at the neighborhood and city levels?

In Chapter 2, I discuss related work and describe its relation to the hardware and software I have developed during my own research in this area. In Chapter 3, I describe software interaction techniques using a combination of sketching, multi-touch, and mouse commands that I developed to explore spatial data. Once collected from many different participants, there are many differences that can emerge between the features present maps and the shapes of neighborhood boundaries. **There are many different ways to represent the same underlying data set, and this can lead to misinterpretation and miscommunication.** To explore these differences, I describe a technique in Chapter 4 for analyzing the same data set from multiple perspectives by using road networks to aggregate and smooth choropleth maps. In that chapter, I use the technique to analyze crime in the city of Atlanta across multiple areal unit schemes for neighborhoods, police beats, and census districts. Even with such techniques, **people have different values and opinions on**

what is important about their city, which can affect whether they think events in an area of the city are more or less relevant to them. In Chapter 5, I discuss a method for flagging important elements in a digital mental map and incorporating them into the analysis of distributions of spatial data. I describe preliminary results using crowd-sourced data from OpenStreetMap, then detail an experiment I conducted using participants from the city of Atlanta. Finally, in Chapter 6, I provide some reflections on this research and discuss some potential directions for future work.

CHAPTER II

RELATED WORK

2.1 Image of Urban Spaces

As a support mechanism for visualization, imageability is frequently used to determine the characteristics of a scene that will allow a user to navigate through a 2D or 3D environment and better understand the data. An early attempt by Ingram and Benford set the stage for many later efforts in using imageability elements (e.g., nodes, paths, edges) to improve the legibility of a data visualization. Many of these efforts seek to automatically detect notable elements, though, rather than allowing the user to explicitly record a mental map of the data [59, 60]. For example, Chang et al. combined a building aggregation algorithm and a demographic data exploration panel to analyze the differences in census data between neighborhoods [23]. Similarly, van Wijk et al. support the creation of wayfinding maps through the simplified representations of urban networks given a focal origin node [104]. Glander and Döllner also use focal points and building aggregation, but incorporate a balanced tree of landmark elements to help navigate a 3D representation of the city at varying levels of abstraction [45].

An explicitly-defined mental map of the city provides intrinsic value, as it allows a community member to share their perspective of the city. Most representations of an urban area are “owned” by those in local governments, and reflect a top-down perspective of how elements of the city are used. Unfortunately, this perspective is often exclusive to many of the inhabitants of the community, particularly those who are not in power [67, 30]. This knowledge capture facilitates the expression of the interests and concerns of individual community members as well as groups of

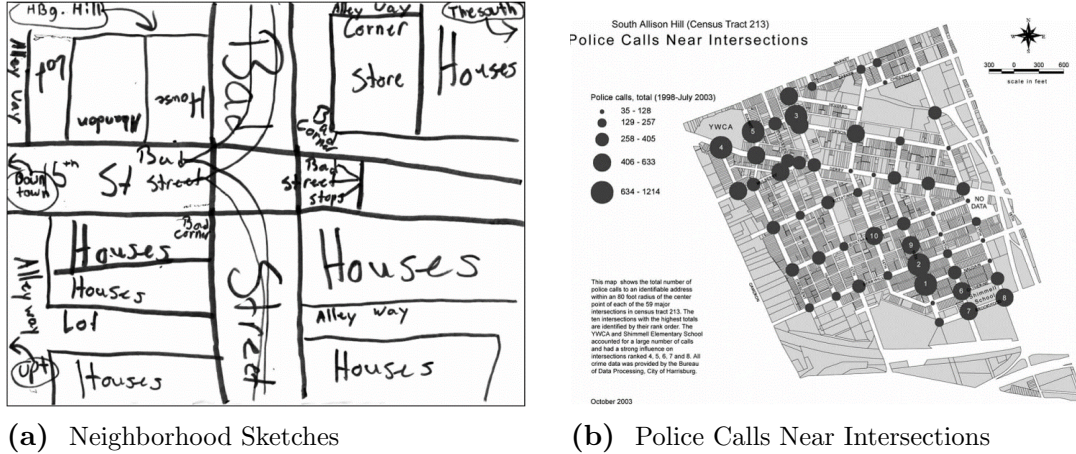


Figure 3: Depictions of Crime in the South Allison Hill community of Harrisburg, Pennsylvania [27]. In Figure 3a, a group of youth from the neighborhood have created a map that indicates “bad” street corners and the alley ways that they use to navigate around the intersection. In Figure 3b, another view rate of crime is constructed to show the rate of police calls near each intersection, but the representation lacks the additional context of how this data affects the individual experience of the residents near those intersections.

citizens [97]. It should be noted, however, that our technique is not intended to directly capture qualitative aspects of the community such as the “bad intersections” sketched by the youth Dennis studied during the South Allison Hills Youth Planning Project [27]. Instead, our technique uses the explicitly-defined mental map as a framework for interpreting quantitative data and comparing those interpretations.

2.2 Mapping Point-based Data

Three common methods for representing point-based events in a spatial context are the dotmap, heatmap, and choropleth map. These methods typically provide context for the distribution of the data at larger scales. The dotmap presents a relatively straightforward option, but comparisons of local areas by relative density can be nearly impossible when events are significantly overplotted (Figure 4a). Still, when scattered appropriately and with a limited color palette, dotmaps can be effective in revealing the relative distribution of data in an area. One recent examples of an effective dotmap is Dustin Cable’s racial dotmap (Figure 5, in which each dot represents a single person [21]. Through effective color selection and dot separation,

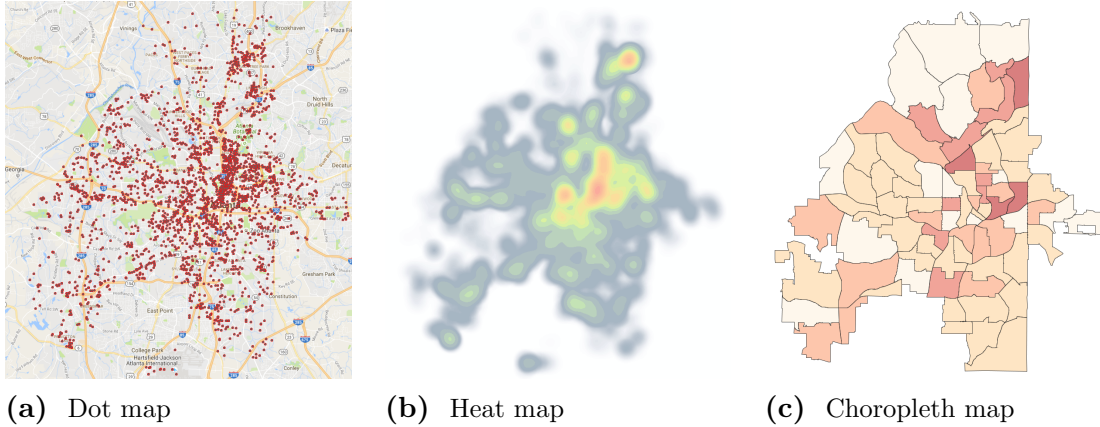


Figure 4: Representations of the event distributions in an urban area. The heat map in Figure 4b reveals distribution more faithfully, but is disconnected from the political boundaries used to interpret the data. The choropleth map in Figure 4c creates a false impression of the relative distribution of events.

Cable’s maps convey overall population distribution, demographic heterogeneity, and boundaries that separate racially homogeneous areas.

Heatmapping avoids the overplotting concern by mapping the discrete event space to a continuous domain [33] (Figure 4b). The relative density of areas on the map are more easily comparable, and heat maps can be extended to include temporal features in addition to spatial variation [77]. Of the methods for performing this mapping, kernel density estimation is acknowledged to outperform other approaches [22]. Kernels use a single bandwidth that must be carefully selected, which can highlight local or global features but not both. In recent years, significant attention has been paid to analysis of variation in scale and spatial extent. For example, Goodwin et al. have explored how correlation coefficients between variables alter dramatically across locations and analysis scale [49, 50]. Similarly, Turkey et al. have explored representation techniques for depicting spatial statistical summaries for multiple attributes across variation in bandwidth and location, but use abstract paths and rectangular windows that are not constrained to the underlying topology of movement used by people [102].

Heat maps also exclude the possibility of quantitatively comparing the values of

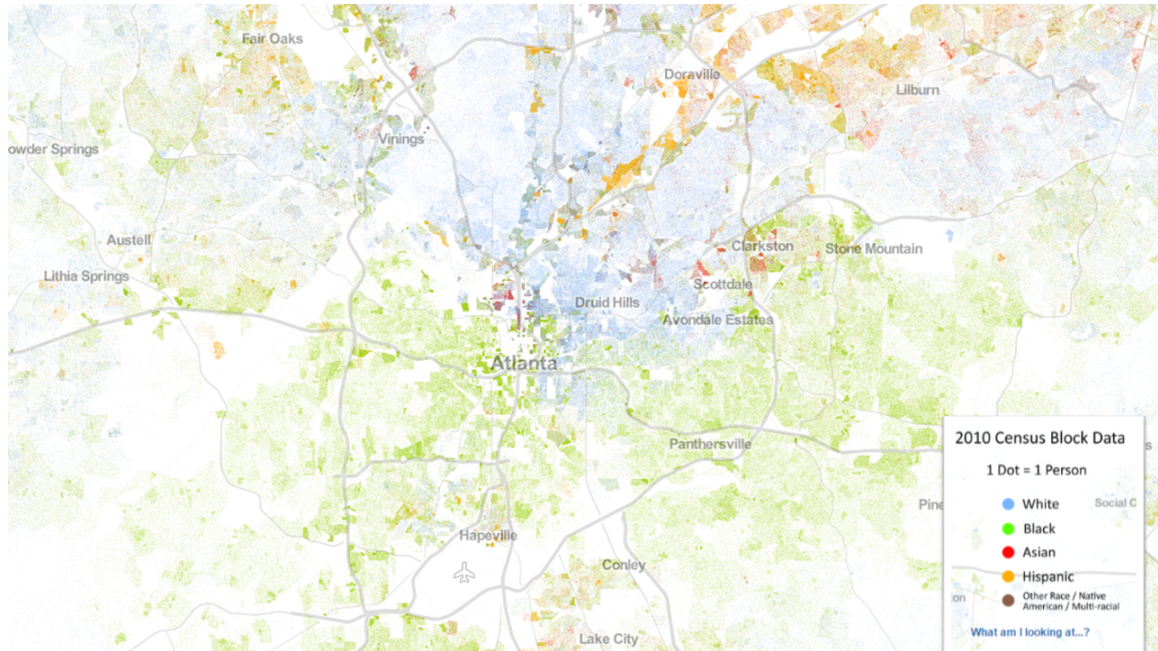


Figure 5: Racial Dotmap of Atlanta [21]. Each dot represents a single person, and is given a representative color to indicate that person’s race. The spacing and use of color are employed effectively to reveal areas of demographic homogeneity, more diverse areas, and the edges between them.

bounded regions on a map, which is the primary task supported by the choropleth map. It is not enough, however, to simply aggregate the number of points contained within each region due to unaccounted for differences in population density and other factors [69] (Figure 4c). The problem of area aggregation extends to the choice of bounding contours, and the substitution of alternative boundaries (e.g., census, public safety districts, neighborhood) can lead to significant differences in the resultant choropleth map. This issue, known as the Modifiable Areal Unit Problem (MAUP), weighs against every evaluation of the benefits of choropleth maps [84]. The use of bounded regions, while convenient for political comparisons at the global scale, can be misleading and inaccurate for comparison of trends at the local scale depending entirely on how the lines are drawn (Figure 6). Areal units create an artificial barrier between neighboring regions that ignores the significant effects they can have on each other [44, 103], a problem which is compounded in multivariate analysis

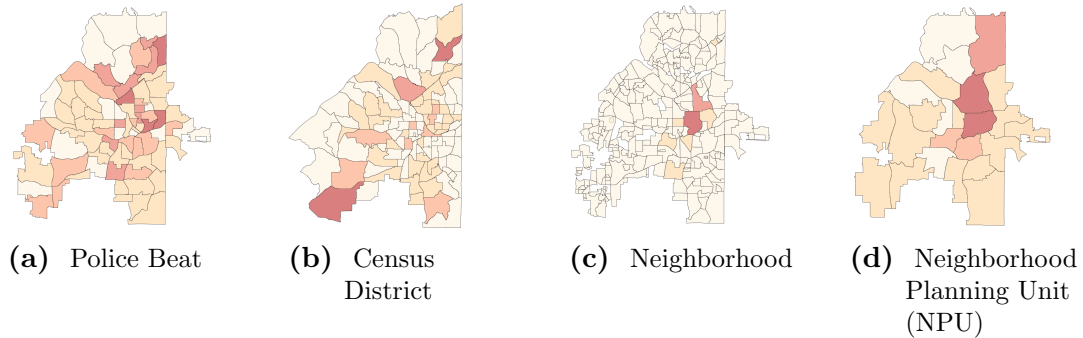


Figure 6: The Modifiable Areal Unit Problem (MAUP), in which the use of different binning strategies produces markedly different results for analyzing the distribution of events within the same geographic region. Each choropleth map depicts the distribution of one year of crime in Atlanta, but the use of different areal units produces different insights.

[42]. Geographical weighting (GW) can be used to partially circumvent this issue by smoothing the value of each region so that it incorporates the value of its neighbors for some statistical consideration [18], most commonly and intuitively the local mean [19]. A distance function (e.g., Gaussian, bisquare) is used to determine the relative influence that a nearby region will have on a local mean in relation to its distance to the region of interest. The choice of an allowable distance for affecting the mean, or window size, varies depending on the underlying data and domain, though automated methods do exist to find promising options [72]. Visual inspection of the effects of changing the window size can reveal regional characteristics that vary from the global to the local and indicate important correlations between data attributes as the scale is changed [31].

Regardless of the window size and distance function, the normative Euclidean distance typically utilized in GW and other multi-scale techniques does not account for human reachability by available paths, which can be critical for many of the urban domains discussed in the previous section. Alternative models for accounting for reachability include Bristlemaps, which aggregates discrete event locations to the roads of a city [64]. This is accomplished through a continuous mapping by kernel density estimation that highlights localized patterns in the data, but it does not allow

for direct global comparison of bounded regions across the map. The importance of visualizing topology in the context of spatial location has also been explored in GreenGrid, a system that blends cartograms and weighted edge representations for analyzing power grids [109]. Reachability has also been explored for local statistics in the context of travel time and road topology[113], but only for point locations and never for bounded regions [73, 74].

2.3 Participatory GIS

Research to improve public participation in GIS (PPGIS) has often found that creating rapid sketch-based representations best improve communication during participatory planning workshops, and that many current GIS tools fail to facilitate this capability [5]. Moreover, many GIS applications remain inaccessible to a wider audience [54, 93]. Al-Kodmany avoided this dilemma by pairing an expert-operated GIS with an artist that took requests from participants during planning sessions in the Pilsen community [4]. Many systems allow for sketching in a GIS context, though these are often more as a natural interaction technique for exploring data [102] or creating a query of existing spatial features [35]. Rarer is the capability to express elements within a spatial context that are of importance to the user. This type of expression could be acquired implicitly, by tracking the areas of the city and spatial data items that a user inspects. This approach has been used quite successfully in other contexts, notably text document analysis, by generating a semantic model from user interaction at varying levels of detail [37]. This type of interaction, however, often divides the user from understanding the internal mechanisms that are being used to generate a representation of the data. Instead, following Green’s et al.’s guidance on process initiators [53], our technique allows the user to explicitly capture her knowledge of the spatial environment as an explicit and direct interaction with the interface by sketching the elements of the city.



Figure 7: Jer Thorp’s Map Room in St. Louis, in which participants project images of urban data onto a large canvas and collaboratively trace, sketch, draw, and paint their own image of the city onto it [99].

2.4 Interaction Techniques

There are many techniques from human-computer interaction and visualization research that lend themselves to encoding a mental map, and they have different advantages and disadvantages based on the model of community engagement with GIS that they most directly support [71]. For example, the data artist Jer Thorp’s “St. Louis MapRoom” blends the idea of a *Map Room* provided by the city planning office to community organizations with the notion of a *Neighbourhood GIS Center* (Figure 7). Members of the community have access to official data that they would not otherwise be easily able to analyze, and the data is projected onto a collaborative canvas upon which people annotate, paint, and draw to blend the quantitative data with their own experiences of the city [99]. Consequently, this approach is very easy to use and promotes collaboration between community groups, but requires travel to the location of the map room in order to be utilized and requires significant external funding for the location, equipment, and maintenance.

Taking the idea of painting as annotation further, Huck proposed the use of a “Spraycan” tool to allow users to digitally record their perceptions of a data distribution on a map [58, 38]. The airbrush interface allows participants to create a fuzzy-point responses to prompts, such as the suitability of various sites for wind farms. This type of approach is suitable for use as browser-based GIS technique, promoting Leitner’s *Internet Map Server* model, which allows the community to have direct access to spatial data but comes with the disadvantage that the utilization is dependent on the hosting capabilities of the provider and makes it more difficult to have a conversation about the results or engage with experts.

Visualization researchers have developed many techniques for interacting with data beyond the mouse and keyboard in a style that is more similar to the rapid and easy sketching on a paper map. Notably, Lee et al. provide an in-depth treatment of “natural” interaction techniques that stray beyond the setup of WIMP (Windows, Icons, Menus, and Pointer) interfaces [70], many of which are highly suitable for collaborative work. Jansen and Dragicevic provide a model for characterizing interaction in both conventional and unconventional non-WIMP interfaces, from tangible actuator systems and legos to wall displays [62]. Isenberg and Isenberg focus more narrowly on the interaction techniques afforded by touch surfaces, primarily wall and table displays, but echo the input difficulties recognized by Lee et al. in arguing that the benefits of such systems have yet to be fully realized [61]. In the context of community participation in GIS, many of these techniques are promising, but in their current form rely on extremely costly hardware components that are out of reach for many neighborhoods. From a practical standpoint, the most likely realization of this arrangement could be through *GIS Facilities Hosted in Universities and Public Libraries*. This could allow communities to have access to more advanced equipment, staff, and volunteers, but is reliant on access to such a university and a number of other incidental factors. As such, the more an interaction technique can be adapted

to Leitner's *Internet Map Servers* model, the more widely it can be adapted to a wider variety of contexts and potentially support additional models. While this does not preclude the use of natural interaction techniques, it does require that they be accessible through WIMP controls in addition to touch and stylus interactions that are available to users with more advanced hardware.

CHAPTER III

INTERACTING WITH CIVIC DATA

In this chapter, I describe my work to date and preliminary findings for developing interactive techniques for sketching mental maps of urban spaces for visual analysis of spatial data. This chapter describes my work on the use of sketching and multi-touch interaction. I began this research with the intent to identify simple, direct methods for interacting with visual representations of spatial data. This work consisted of sketch and multi-touch methods that would be relatively easy for non-experts in GIS to grasp and included capabilities for path analysis in the context of spatial point data (Section 3.1). This foundation led to several connected efforts, most directly to the design of a route planning app that allows police officers to rapidly sketch patrol paths (Section 3.2). Though this is perhaps the most application-focused of my efforts to this point, it still relies heavily on the analysis of data in the context of an individual understanding of the city. In this case, the individual is a police officer, who applies her domain knowledge of the city to decide where her efforts will be most valuable.

3.1 Sketching Spatial Data on Maps

In a poster at IEEE VIS in 2015, I presented the details for an initial prototype of sketch-based interaction techniques in a system called SpaceSketch [46]. SpaceSketch is designed for use on a multi-touch enabled display with a stylus (Figure 8). The stylus is used to draw data sketches on the map, while touch controls are used for navigation within the display. There are two underlying data sets in the SpaceSketch prototype: spatial events and reachable paths. The spatial events have a fixed location and point in time, such as crimes or traffic collisions. A reachable paths dataset,

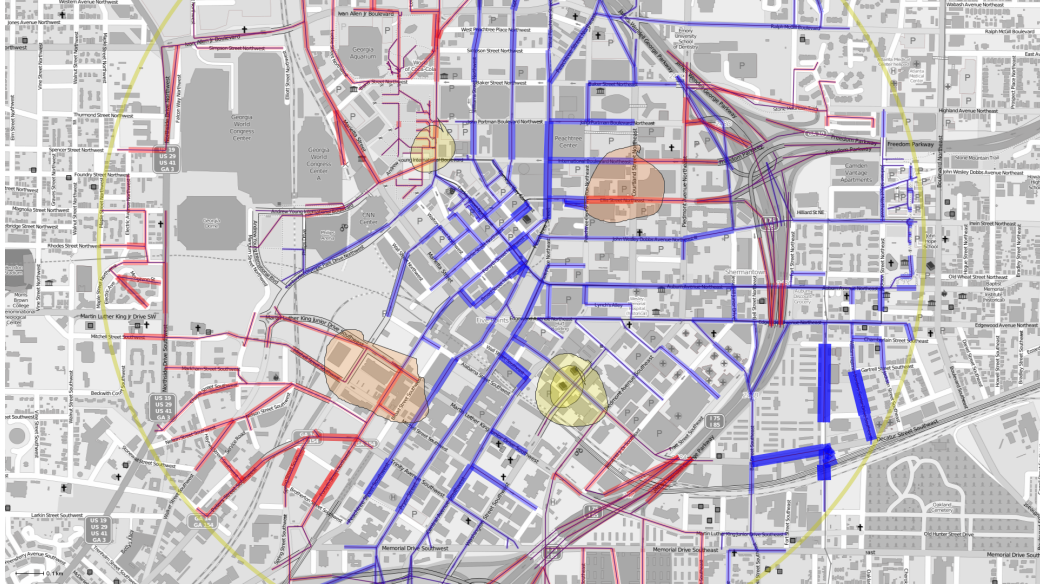


Figure 8: Preliminary map visualization incorporating geospatial features and freehand annotations. Sketched paths in blue have been fused to the closest match on the road network from an origin to multiple destinations. Freehand closed shapes have been used to update and *stain* the paths that pass through the shapes in red.

obtained as road data from OpenStreetMap, is used by SpaceSketch to snap drawn paths to actual roads.

Using a stylus, a path can be drawn from one location to another on the map. The interface supports three modes of path sketching: (1) shortest path; (2) user-specified path; and (3) radial exploration. In the first mode, the system generates the shortest path between the endpoints using the reachable road dataset and ignores the shape of the path drawn by the user. The shortest path mode can also be used to extend a path indefinitely, moving between different waypoints as the user draws each leg of the trip on the map. In the user-specified path mode, the user traces a path from a location. The system follows the traced line to construct a path that follows it in real time, even if another shorter path exists. Finally, in the radial exploration mode, a user investigates the area around a central waypoint by dragging the stylus from that location to any distance around it. While the stylus remains down, a circle is drawn with the radius of the distance from the starting point to the stylus. When the

stylus is lifted from the display, the system identifies all reachable locations within the circle from the central waypoint and creates a path to them. This can be useful if a user wants to explore the neighborhood within a certain walking distance around a central location.

SpaceSketch also provides the basic capability to define districts on the map. A user drags the stylus around the defined area that is automatically closed when it is lifted from the display. The shading of the district is mapped to the number of spatial events (e.g., crimes) that occur within that district. The current rendering could be adapted to highlight the deviation of each area from the median number of internal crimes, distance of contained crimes from the shape centroid, and more. An additional capability of the bounded districts is that they can be used to “paint” paths that pass through the defined area. If, for example, a user specifies a radial exploration with a long distance, the number of paths constructed will be quite high. By examining the rendering style of the paths, the user can identify areas with a higher number of spatial events along each path, and draw a bounded district around those area. The paths that move through that area will then be shaded with a different hue, allowing the user to determine if alternate paths are available to reach the affected destinations.

While SpaceSketch was initially a very promising approach for interacting with maps, the primary limitation was that it was designed primarily for a fairly powerful computer with a multitouch surface, pressure-sensitive stylus, and erasing capabilities. While compelling, this relatively narrow use case limited the application to community participants who did not have access to this type of hardware. Consequently, I began to focus more heavily on techniques that could still utilize touch and sketch controls through a mouse, but take advantage of touch when available in the browser.

3.2 Sketched Roads and Police Patrols

In a paper presented at the Hawaii International Conference on Systems Sciences (HICSS 2017), I explored the use of sketched paths within spatiotemporal models for improved policing [47]. There are many approaches to proactive policing, but broadly, these practices are based on showing police presence, engaging with the community to learn their concerns, and analyzing historical crime reports to identify locations and people that are currently at increased risk of a crime. By moving through a police beat and creating a visible police presence, officers remind the public that they are nearby, discouraging potential violations while encouraging the lawful use of public space. By becoming involved with the community, an officer makes it easier for the residents to willingly participate in policing efforts. Finally, the use of crime analysis techniques and software can enable short-term tactical planning (e.g., which neighborhoods to patrol over the next hours and days) and long-term strategic efforts (e.g., drug market intervention programs that rehabilitate non-violent first term offenders).

While I was primarily interested in this last aspect for the research described in this section, it is important to note that it facilitates the previous two. Commercial software for the analysis and prediction of criminal activity has seen steady deployment throughout the country, bolstering and often improving on the capabilities of existing criminal analysts within departments. PredPol, for example, has been deployed in both Los Angeles and Atlanta to help officers determine at what location they should patrol [12]. Like many systems of this kind, it is not intended to support analysis of the path that an officer can take to get between destinations. Whether by car or on foot, hotspot analysis systems do not typically support the exploration of routes between locations.

While proactive policing is the intent, the reality is that many officers will spend a great portion of their day responding to 911 calls that take them away from their current location. Predictive crime systems can help reduce the response time for

these calls by attempting to position officers in close proximity to areas that receive calls. This reactive aspect changes the nature of the plan, however, in that it puts the officer in a new location and causes time to elapse before a new plan can be created. It is no surprise that in many urban locations, the time of day can have a significant effect on the volume and types of crimes that are expected to occur. As the shift progresses and the unit responds to calls, the initial static analysis of the crime reports generated by many systems become stale. Police units could benefit greatly by a mobile system that allows them to view an updated analysis of crime hotspots based upon their changing location and time.

In this section, I present my research to couple interactive hotspot exploration with rapid route planning and analysis. The primary components are:

- A sketch-based approach to dynamic route planning, which allows a police officer to rapidly specify a path through the city without typing and review the volume and types of crime that occur along that route.
- A spatiotemporal hotspot approach that takes into account time of day, location, season, and recent event volume.

The primary contribution is the novel pairing of these components in an initial prototype, HotSketch, designed for mobile use on tablets. I present the details of this prototype and provide validation through a set of use cases within the domain of police patrol route planning in the city of Atlanta.

HotSketch is designed to allow police officers to more rapidly utilize predictive models for crime in their neighborhood while away from the precinct, where expensive crime prediction software typically resides [90]. Rather than providing a static report, HotSketch is designed to allow officers to explore the data in their area through loose sketches that enable them to query the data and update the dashboard.

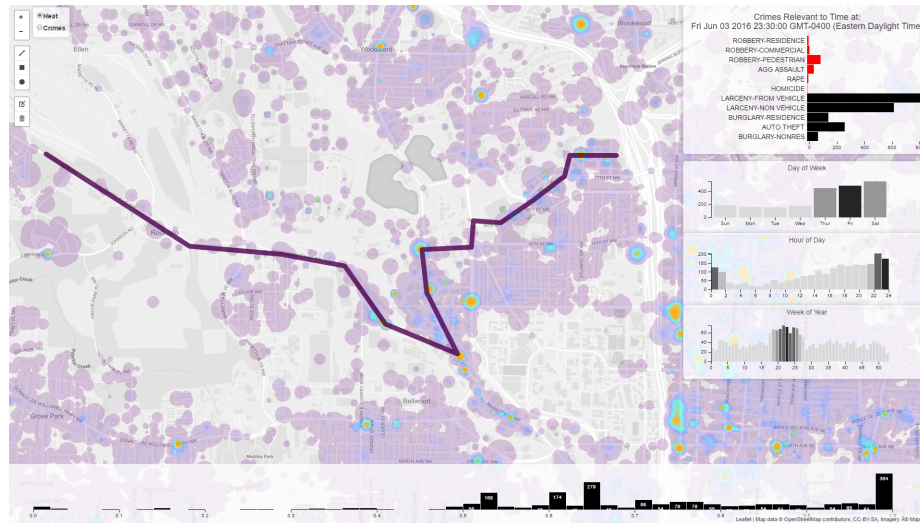


Figure 9: The HotSketch dashboard. A route has been drawn through a commercial shopping area on a Friday night at 11:30pm. The type of crimes along this route and at this time are presented in the category panel in the upper right. Below that, panels show the distribution for crimes by day of week, hour of day, and season (week of year) that are contributing to the count in the category panel. At the bottom of the dashboard, a histogram shows the distribution of the crimes along the planned route

3.2.1 HotSketch

Our initial prototype is designed to run in the browser on tablets and laptops, and has been created using d3.js. The primary element is a 2D map centered on the current location of the user. Additional elements are available in hovering windows around the periphery of the map and are populated with data as the user interacts with the system. Figure 9 depicts the system interface. When the system is loaded, the officer is immediately provided with a heatmap of the crimes in the area based upon the current time of day and the geospatial location of the user. This heatmap consists only of the filtered events that occur within a relevant timeframe to the current time, which is determined through the approach described in Section 3.2.1.1. Then, the officer can begin sketching potential routes through the area between any locations. As described in Section 3.2.1.2, these sketched paths can be used to query the filtered crimes to determine the types and locations of crimes that occur in proximity to the path. Finally, the officer can dig into the details of any particular crime by switching

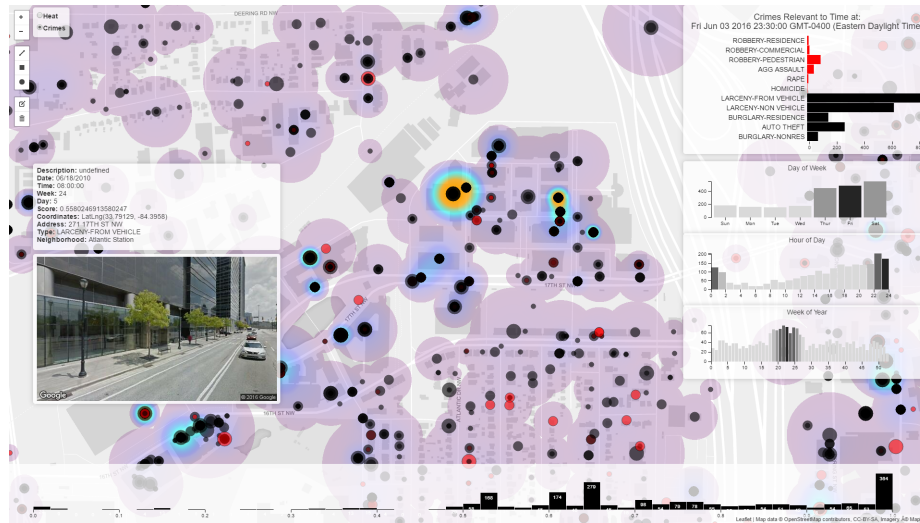


Figure 10: Detailed view of crimes in the dashboard after zooming into the shopping center at the end of the path in Figure 9. Violent crimes are shown as red circles; non-violent crimes are rendered as smaller black circles. The heatmap is displayed behind the event circles to retain context. By tapping or clicking on the location of a crime, a detailed panel on the left reveals additional information. Below that, a street view is shown for the crime location.

to the dot map view (Figure 10), which is described further in Section 3.2.1.3.

3.2.1.1 Hotspot Analysis

Our approach to hotspot analysis contends that there are observable patterns in the overall volume and types of crime that are dependent on when the analysis is being performed. For example, certain types of crimes occur more often at night than during the day, and the overall volume of crime increases during the hotter summer months. Similarly, the day of the week can influence this observable change in crime frequency, particularly between weekend and weekdays. Finally, as time progresses, the locations of hotspots will naturally migrate in response to the growth of the city and the effects of police intervening to alter this growth. To account for these factors, I use a weighted summation function that scores the temporal relevance of a crime to the current time and date when an officer opens HotSketch. While a crime analyst would benefit from the ability to manually browse through range of potential parameters, I focus on providing functionality in the context of an operational environment in which the

officer will want rapid answers rather than capabilities for model construction. The rationale is that if an officer opens the system on a Sunday morning in the middle of July, locations should be flagged as more relevant to the current geospatial crime landscape if they occurred in a similar temporal context. An officer opening the system on a Friday night in January should naturally see a different view.

To calculate the difference between the time a crime occurred and the present, I utilize a kernel function with a bandwidth parameter, in this case the bisquare function (Equation 1). For example, if a crime occurred on a Thursday and the current day is Friday, I want that crime to be considered less relevant than crimes that also occurred on Friday. Crimes occurring on more distant days of the week should be considered sparingly, if at all. I determine the difference between them as the number of days d_w and decide on a bandwidth size for which crimes can be considered relevant (h_w). In most of the examples provided in this section, I use ($h_w = 2$ days), and since Equation 1 increases monotonically as d_w decreases, this will heavily favor events that have occurred on the same week day while still including events that occur on the day before or the day after.

$$w_i(u, h_g) = \begin{cases} [1 - \frac{d_i^2}{h_g^2}]^2, & \text{if } d_i < h_g \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Continuing with this approach, I choose meaningful bandwidths for our other factors: time of day d_t , and difference in week of year, or season d_s . Like day of week, these measures are cyclical, so determining the difference parameter d must be done with care before using Equation 1. In most of the examples provided in this section, I use ($h_t = 3$ hours), as the day is divided into four overlapping shifts of six hours each. I choose ($h_s = 6$ weeks), which favors events that occur in roughly the same season as the current date. To determine the relevance that an incident has for the current time context, I create a weighted summation of the individual kernels.

This function can be expressed as $f(x_i, x_j)$ in Equation 2, in which x_i is the date of the incident and x_j is the current date. This process bears a resemblance to the Seasonal Trend decomposition based on Loess (STL) approach utilized by Malik et al. [78], though there are some noticeable differences. In this example, I specify the individual weights to increase the importance of an event that occurs close to the same day of the week and time of day while de-emphasizing the importance of the seasonal bandwidth. This weighting scheme favors our stated use case of an officer looking for highly contextual information based on current location and time. These parameters could be altered to reflect other use cases, however, such as determining routes of safe passage for school children throughout a season. Unlike Malik et al., I primarily rely on this summation approach to filter out irrelevant data from the heatmap and linked visualizations.

$$f(x_i, x_j) = \frac{2}{5}k(d_t, h_t) + \frac{2}{5}k(d_s, h_s) + \frac{1}{5}k(d_w, h_w) \quad (2)$$

Once the equation has been applied to each event, the value for that event is added to a heatmap layer that is displayed on top of a 2D map (e.g., Figure 9). As the criminal incidents are not counted equally, this leads to variation in the distribution of events from a standard heatmap and movement of hotspot locations from a straightforward count. Instead, this new distribution reflects both the location of events and the temporal context in which the request was made, allowing an officer to see the distribution of events for their area and throughout their shift.

3.2.1.2 *Sketching Patrol Paths*

Paths are drawn onto the map using editing controls in the top left of the dashboard. Using a mouse or by tapping on the map, an officer creates a path by specifying the control points that define a polyline. Once the path has been specified, the user can choose to edit the path by dragging the control points or to delete it entirely. Once

finalized, a completed path can be explored in the context of the data by clicking on it. HotSketch will then compare the line segments of the path to the crime incidents in the area to determine which ones are closest. In addition, I derive the point on the line that is closest to each point within a minimum distance specified by the user. In the examples presented in this section, this threshold is set to 200m, which is sufficient to include incidents within one or two blocks of the path drawn by the user.

By determining the point along the line that is closest to each crime, I can reconstruct the order in which an officer would pass by these crimes from the beginning of the patrol path to the end. In Figure 9, for example, the officer has drawn an initial route from a location into a commercial district with numerous shopping options. Once the details of the route have been loaded into the dashboard, the officer can review the types of crimes that occur along the route. The events are filtered and weighted based on the approach described in the previous section, emphasizing those crimes that occur not only in proximity to the line sketched by the officer but in relevance to the temporal context in which the path is created.

That is, the time and date in which the user is drawing these lines affects the types of data shown (subject to the approach described in the previous section). For example, in Figure 9 the panel in the upper right indicates the relative distribution of crimes along this route by category. This panel depicts only the crimes that have a non-zero value from Equation 2. The panels below indicate the frequency of events according to day of week, time of day, and week of year. Since Equation 2 is additive, events that receive a low or zero score for any of the individual kernel components can still appear as long as they have a positive value for one of the other kernels. The emphasis of the kernels is evident, however, in the peaks of the distributions for each of the detailed panels. I also shade the fill color of the bars within the individual histograms to indicate the current day, hour, and week (Figure 11). The darkest bar indicates the current time, and a monotonically decreasing grayscale indicates the

potential score of events coinciding with the other bars for each kernel.

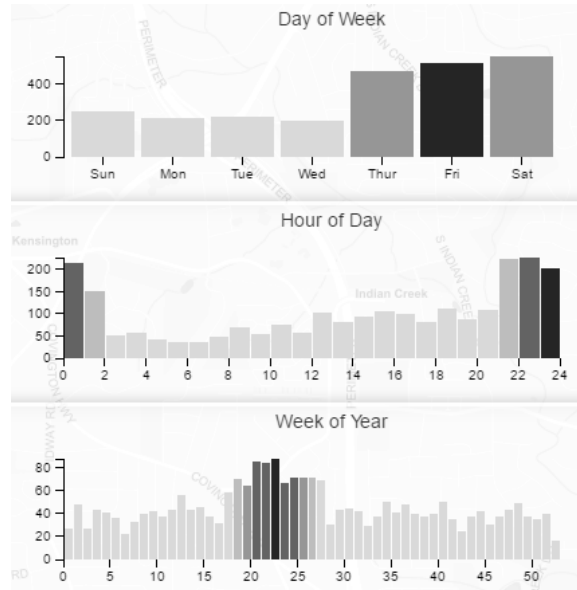


Figure 11: Panels provide a detailed look at the distribution of relevant events through histograms. For each panel, the current time is filled in black; other bars are colored by the score for individual kernel windows (darker bars score higher than light bars).

3.2.1.3 Event Details

If the officer wants to review additional details, she can switch to a dot map that shows the exact location of all the crime events within the area that have not been filtered out by the approach described in Section 3.2.1.1. In Figure 10, for example, an officer is examining relevant crimes that have occurred in the proximity of a commercial shopping area. In the figure, the most common type of incident is a larceny from an automobile, which likely results from the massive parking lot within the facility and the propensity for shoppers to leave valuables visible and unattended in their cars. When the officer clicks on one of the event circles on the map, linked displays to the left provide details for the selected event. The top pane provides text details on the event as they are available, such as the type of crime, the date, the time of day, and a description if one is available. To help provide additional visual context, a secondary pane on the map provides street-level imagery pointed at the crime location. This

imagery is obtained through the Google Street View Image API, and reflects the most up-to-date picture available at that location regardless of the date or time the crime occurred.

3.2.2 Usage Scenario

Non-violent property crimes such as larceny from a vehicle account for a massive portion of the crimes throughout the city. While the police cannot be everywhere at once, maintaining a visible presence in areas of significant property crime can reduce the number of incidents at that location. In our use case, an officer is moving through her beat as part of her normal patrol. Using a wayfinding app, she is given a relatively quick route from her current location to her destination within the eastern edge of her beat. However, by quickly drawing the route on HotSketch, she sees that there is an area directly to the south of her that appears to have a high volume of activity (Figure 12). The area contains a collection of single family homes and small apartment buildings with a history of assault, robbery, and larceny. While this new location is a little out of the way for her, it is also an opportunity to establish a presence and spend some time in an area within her patrol beat without being summoned there for a call. She quickly sketches a few alternatives, deciding on a route that detours through a section of the neighborhood before returning to her original route (Figure 13). By taking this detour, however, she is substantially increasing her dwell time in proximity to areas of historically high crime, and the timeline panel along the bottom of the dashboard reflects a gain in nearby incidents towards the middle of her trip.

To familiarize herself with the problem areas along her new route, the officer switches to the details view in the dashboard (Figure 14). In this image, she has zoomed into a residential area where a high volume of relevant crimes have occurred. Clicking on one of the violent crimes provides her with a description an aggravated assault at the address, one of many in the vicinity. The street view panel has imagery,

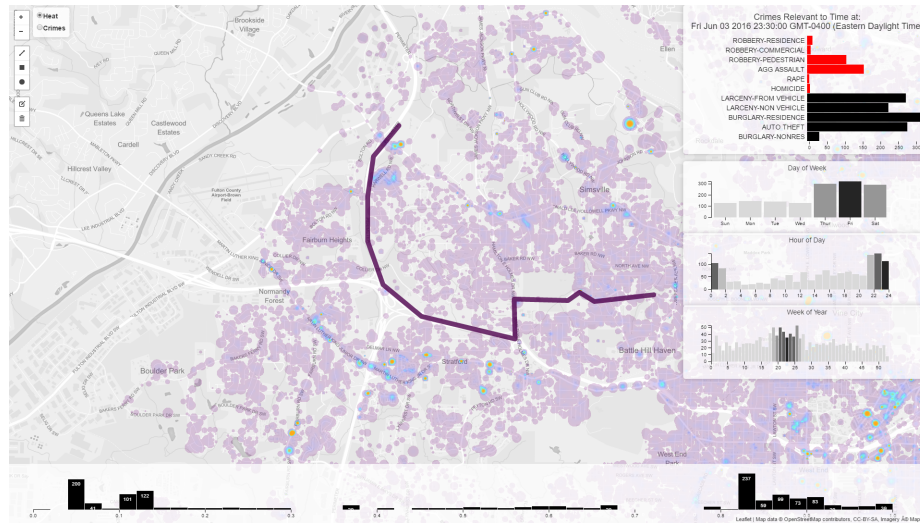


Figure 12: An officer has drawn an initial route through her beat based on the driving directions given by a wayfinding app. There is a history of activity at the start and end of her route, but few incidents along the route. The distribution of violent crimes in this area are high, especially compared to the commercial shopping area in Figure 10. To provide police presence in high activity areas, she extends part of her trip through a busy area to the south.

so she is able to see the building in which many of these incidents are occurring. She now has a route, and knows what to look for on her way through. She will keep an eye on this location in the future, and look for opportunities to get to know residents in the area so that her presence as an officer of the law will be recognized and respected.

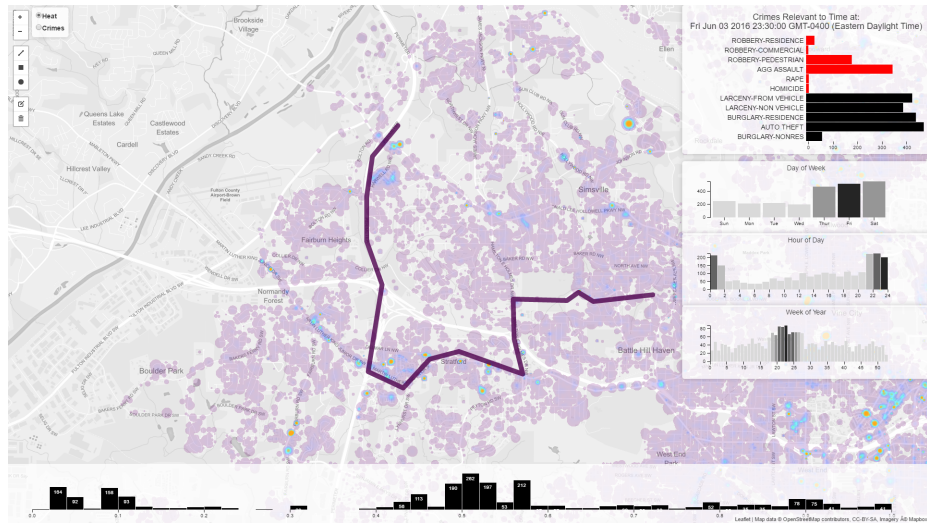


Figure 13: While it does increase the length of her drive to her destination, the addition of a detour through a busy street has significantly increased her dwell time in areas of high activity. The trip histogram at the bottom now includes a new collection of activity for the detour. The distribution of violent crimes along this route are also high, particularly for aggravated assault and pedestrian robbery. While this distribution is similar to that of her original route, the volume of historical crimes along the new route is nearly double that of the former.

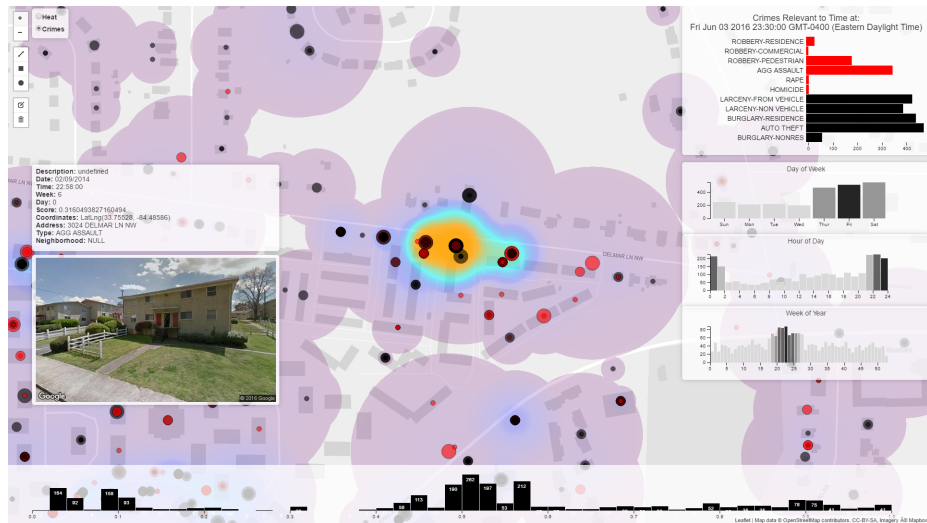


Figure 14: The officer switches views to get a detailed look at a neighborhood with a history of frequent criminal activity. By clicking on the dots representing the relevant crimes, the officer becomes familiar with the details of past incidents and the context of the locations in which they occurred.

CHAPTER IV

REACHABILITY MAPS OF URBAN SPACES

In this chapter, I describe a visualization system that uses a combination of paths and districts to create choropleth maps based on the structure of urban spaces. This model relies on the reachability of each district, or the path structure within a district and connecting it to other districts, as a basis for comparing the relative distribution of events between districts (Section 4.1).

4.1 Districts and Paths

Districts are most commonly represented visually through choropleth maps, which encode a single hue for each district that maps directly to that district's value for some data attribute. However, aspects of choropleth maps can make them ill-suited for analysis at local scales, specifically in urban contexts. For example, the modifiable areal unit problem (MAUP) has been well-studied in geospatial contexts, but continues to be a challenge in the design of choropleth maps [84, 42]. As the boundaries for each region serve as a fixed inclusion criteria for querying the underlying data, items just outside the boundaries are not counted towards summary statistics that characterize a region. There are many contexts in which this is not ideal. For example, aggregating crime statistics within political boundaries provides the mistaken impression that adjacent regions are unaffected by the crimes that occur just on the other side of the boundary. Approaches exist for surmounting this challenge through geographically weighted (GW) statistics [18, 31], but they use Euclidean distance and do not take into account travel distance along roads or accessibility. These challenges are unfortunate given the practical benefits of choropleth maps. By aggregating data into discrete regions, it becomes much easier to assign administrative and legislative

solutions to regions that are overseen by existing governmental structures. By providing a method for comparing regions, stark contrasts between them can become apparent that can either be used to highlight best practices or to identify areas that need improvement.

I addressed these shortcomings through 1) a technique that combines analysis of event reachability at local and global scales; and 2) a system implementing this technique through multiple coordinated views. First, I describe a road-based sampling technique for point data that can help estimate the event density at local scales throughout a region and can be used as an aggregate value for choropleth regions that is affected by reachability. Second, I describe a global weighting technique for adjusting the value of a region by taking into account the values of nearby regions that are reachable from the region, not just close by Euclidean distance. Third, I describe my implementation of these techniques within a system and my application of them to an investigation of crime data in the city of Atlanta. This system allows for the exploration of the additional metadata generated by applying these techniques to a region, such as the relative rank of one region compared to others and the reachability from one region to its surroundings. Finally, I compare the described techniques to alternative approaches for GW.

4.1.1 Reachability

I use the term *reachability* to refer to the distance between locations based on the distance between them along paths. For event reachability, this is the distance along paths from that event to any location that can be reached by paths (i.e., road network distance). For districts, a district is more reachable if there are many paths that cross its boundaries from its interior to neighboring regions. The techniques described in this section offer a generalizable approach for joining this topological analysis to the districts that are typically used to make administrative decisions. For example,

emergency responders often have a prescribed jurisdiction for responding to accidents quickly, but depending on the current workload and incident severity it is common for responders to provide assistance to neighboring regions. When determining the annual budget for emergency responders within a region, it would be useful to go beyond a simple count of accidents within each region to incorporate the district's internal topology as well as the topology of nearby reachable districts that receive assistance. This additional level of detail would reveal the true "reach" of the responders within and outside a district, adjust the level of funding they receive accordingly, and identify nearby districts that are likely to receive additional emergency support.

4.1.2 Reachability Weighted Statistics

The use of Reachability Weighting (RW) as an alternative distance metric for geographically weighted (GW) statistics creates several interesting design challenges. First, events must be aggregated in some way to measure the internal score of a region before it can affect, and be affected by, the scores of neighboring regions. A region that has many events very near to its paths should have a higher score than a region that has few events that are far from its paths. Second, the weighting of one region and the influence that region has on neighboring reachable regions should reflect the underlying reachability of those regions with respect to one another. Regions that are easily reachable from each other in one or more ways should be weighted more heavily than regions that are difficult to reach in fewer ways. Finally, care should be taken to allow the user to explore how variation in window bandwidth affects the weighting. Dykes and Brunsdon emphasized the importance of bandwidth on the analysis facilitated by GW using traditional distance measurements [31]. Unlike traditional GW, increased window size in RW does not necessarily cause the adjusted values of each region to tend to the initial global mean, and alternative representation and interaction methods are necessary to explore this variation. In this section, I discuss

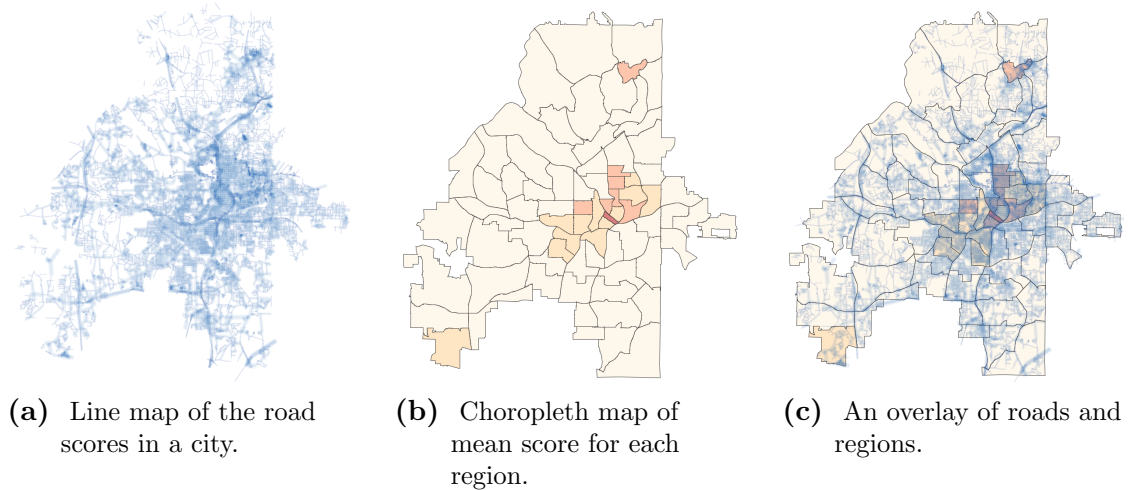


Figure 15: Representations of events in proximity to the road topology of an urban area. In Figure 15a, road scores have been joined to the underlying road network, revealing the infrastructure that occurs in close proximity and frequency to the event data. In Figure 15b, the regions have been colored based on the mean road score for the segments contained in that region. In Figure 15c, the two representations have been overlaid to explore the correlation between road sampling and region score. Region colors have been selected using Color Brewer [55].

my approach to each of these challenges: (1) Sample roads to generate a region score; (2) Weight regions relative to reachability; and (3) Explore alternative bandwidths interactively.

4.1.3 Road Sampling and Region Score

First, I sample the number of events close to a given road segment to determine that road’s score. I follow the road network topology structure of primal road-centerline-between-nodes graphs in which each street is reduced to the endpoints from one intersection to another [88]. Taking the pieces of road between any two pairs of intersections as road segments, I count the number of events that occur within a local window bandwidth h_l . For each of these events e_1, e_2, \dots, e_n , I determine the minimum distance d_i from that event e_i to any part of the road segment. Using the kernel density estimation function (Equation 3), I can then incorporate these distances to estimate the density of events for that particular road segment [22, 94]. Following

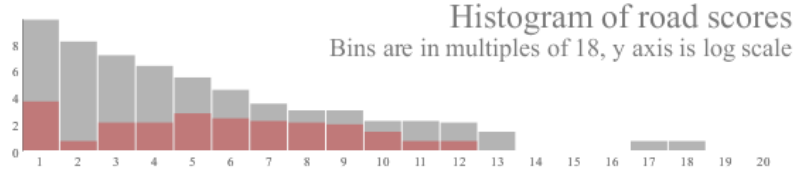


Figure 16: Histogram of the road-event scores in a city. Scores are represented in log-scale due to the inverse power distribution. The scores for a region selected in the map are shown in red.

the example of Kim et al.[64], I apply an Epanechnikov kernel as the windowing function (Equation 4).

$$f(x) = \frac{1}{nh_l} \sum_{i=1}^n K\left(\frac{d_i}{h_l}\right) \quad (3)$$

$$K(u) = \begin{cases} \frac{3}{4}(1 - u^2), & \text{if } \|u\| \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

The local result of applying these equations is that a road that has events that occur close to that segment will score more highly than a road that has fewer events at greater distance. At this scale, the bandwidth h_l is intended to be relatively small to reduce the potential effects of landscape features that would affect reachability (e.g., rivers, mountains). If I render the road segments with a mapping from these road-event scores to line thickness, I can then generate representations such as Figure 15a. In this figure, areas of the map can be compared based upon the probability that an event has occurred in close proximity to the road network. Road features can be discerned but are more salient in the areas in which they frequently coincide with events. Reviewing a histogram of the road scores for the events indicates that they follow an inverse power distribution. In Figure 16, the distribution of road scores is depicted with the y-axis at log-scale, allowing for more thorough examination of the long tail of values.

The line map in Figure 15a is a useful first step in aggregating the events into

a reachability-based representation, but much like a typical heatmap, it does not allow us to directly compare bounded political regions. Using an existing scheme of boundaries (e.g., police beats, census districts), I can aggregate the road segments that each region contains rather than the events themselves. Then, I can map the mean road score of each region to an ordinal color scale to more easily compare them [55]. An example of this mapping is shown in Figure 15b, in which the mean score of several regions in the center of the map are visibly higher than the surrounding regions. I can compare the regional mean road score representation to the line map through an overlay, which shows the correlation between the density of the road and region scores (Figure 15c).

4.1.4 Region Weighting

The approach described in the previous section presents a promising first step towards incorporating reachability into interactive choropleth maps by sampling roads to generate an reachability weighted (RW) region score. But it is not enough, as the regions are still subject to the constraints imposed by their artificial boundaries, and consequently, the modifiable areal unit problem. Moving a boundary in any direction would alter the inclusion or exclusion of road segments and affect the mean value calculated for that region. To move forward, I present a general overview of the underlying concepts necessary to understand geographically weighted (GW) statistics and my contributions regarding reachability.

The essential purpose of GW can be linked to an observation by Walter Tobler in the analysis of urban growth, “*Everything is related to everything else, but near things are more related than distant things*” [100]. Known as Tobler’s first law, this understanding leads us to move towards methods of spatial analysis in which areas on a map are considered to be more related to those areas that are closer than those areas that are farther away. For choropleth maps, this takes shape in a spatial smoothing

in which the value calculated for each region is combined with the predetermined values of nearby regions to derive a localized mean. In the approach proposed by Brunson et al., a bandwidth h_g is used to determine the window size for each region and, subsequently, the regions that will affect its local score [18]. For any location, or region u , I can then determine the local GW mean at that point through Equation 5.

$$M(u, h_g) = \frac{\sum_{i=1}^n r_i w_i(u, h_g)}{\sum_{i=1}^n w_i(u, h_g)} \quad (5)$$

In this equation, the weight $w_i(u)$ is the weight applied to the value of some other region u_i when computing the value for u . This is determined through the use of a distance function that yields higher values for nearby objects and decreases monotonically as the distance increases. The bisquare equation is utilized for these comparisons (Equation 1) in Brunson et al.'s original proposal for GW mean [18], and I use it in kind here. For this comparison, the distance d_i between a remote location u_i and a region u is used to determine a weight w_i , typically taken as the distance between region centroids. This weight is then applied to the value of that region r_i and, along with the weighted values for all the other regions u_1, u_2, \dots, u_n reachable from u , is summed. This includes the value for u , which is included in the summation and weighted at $w = 1$ given the solution at $d = 0$ for Equation 1.

$$w_i(u, h_g) = \begin{cases} [1 - \frac{d_i^2}{h_g^2}]^2, & \text{if } d_i < h_g \\ 0, & \text{otherwise} \end{cases} \quad (1 \text{ revisited})$$

Once all the values are summed, they are then divided by the sum of the weights to derive the GW mean that characterizes a location u . This process is repeated for all other locations u_1, u_2, \dots, u_n . It should be noted that the bandwidth size h_g utilized for the GW mean, while similar in purpose to the h_l utilized for the kernel density estimation in Equation 3, should typically be much larger. This is due to the different purposes to which they are each intended: h_l is a local window size

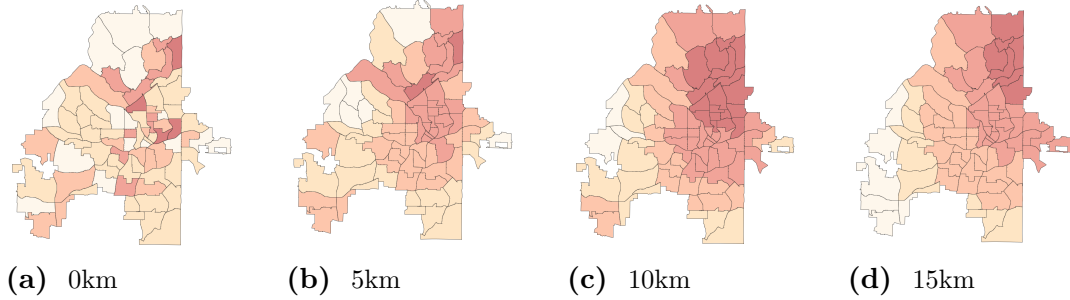


Figure 17: Geographically weighted mean applied at a range of scales. In Figure 17a, events are counted within regions. In subsequent figures, h_g is increased in increments of 5km.

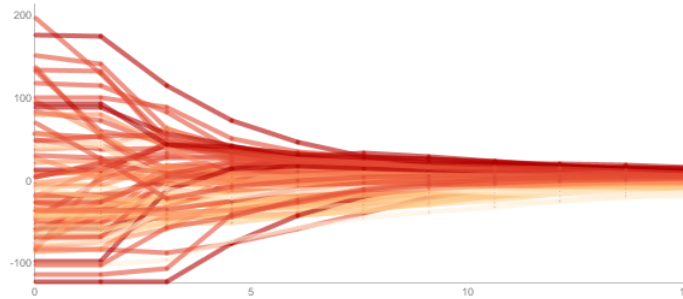


Figure 18: Scalogram of the relative change in region score as the bandwidth h_g is gradually increased from 0 to 15km in Figure 17; lines are colored by their score at 15km. The scores exhibit substantial volatility from 0-5km that stabilizes from 5-10km. By 10km, most regions maintain their ordering and tend towards the mean.

that is used to control the bandwidth for aggregating events to roads, while h_g is a global window size that is used to control the bandwidth for aggregating bounded regions into locally weighted means. While each can be derived statistically through examination of the distribution of event and region distances, they can also be set to meaningful values with respect to the domain under consideration.

4.1.4.1 Reachability Weighting and Events

When applied to multivariate statistical values at a range of scales, geographical weighting can be used to great effect [31]. As discussed in previous sections, however, the role of reachability between bounded regions is somewhat different in many domains than the role of linear distance between region centroids. To provide a comparison, I present a GW mean analysis in Figure 17 of the events from previous

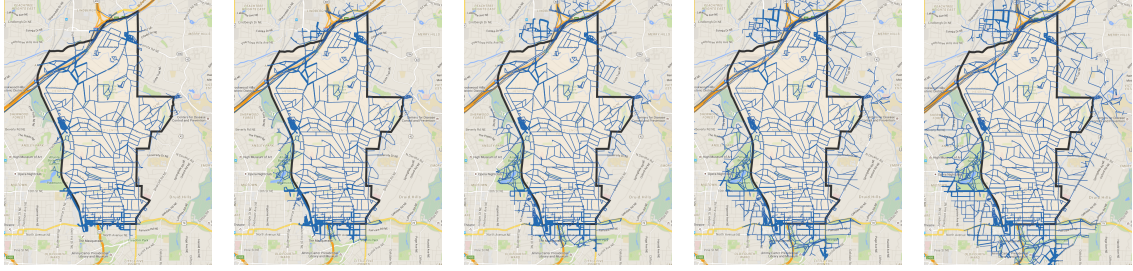


Figure 19: Reachable locations in the periphery of a bounded region are included in its score. From left to right, the maximum distance is gradually increased to illustrate the non-uniform reachability of the region to nearby locations. While there are many reachable locations to the north, southwest, and south, other areas around the periphery of the region remain more sparse.

examples. As stated before, the initial count of events for each region in the map can be misleading due to the lack of adjustment due to population data or region size (Figure 17a). Once the smoothing is applied at a bandwidth of $h_g = 5km$ in Figure 17b, some of the outliers tend toward the locally derived means and a more complete picture of the underlying distribution is visible. As the bandwidth is increased to $h_g = 10km$ in Figure 17c and again to $h_g = 15km$ in Figure 17d, this smoothing becomes more coarse and appears to indicate the presence of an overall heightened event density in the north-northeast regions that diminishes towards the west-southwest.

As the size of the bandwidth increases, however, these differences are extremely minimal and all the regions have begun to approach the global mean of the initial values for all visible regions. This is a known limitation when increasing the bandwidth in GW statistics, which in part motivates the careful selection of h_g . Dykes & Brunson proposed the use of scalograms to examine the variability of regions with respect to the global mean as the bandwidth h_g is increased [31]. Using a scalogram to examine the range of values for the event data in Figure 17, I can see that there is substantial volatility as the bandwidth is initially increased from 0 to 5km (Figure 18). This is to be expected, as the jump from no GW to even a minimal window will smooth out many of the existing outliers. As the bandwidth is increased from left to right in Figure 18, many of the regions with a higher value dip significantly while

regions with low values begin to rise.

While this method presents a substantial improvement over the relatively basic analytic questions facilitated by the simple count of Figure 17a, this approach is still missing a big piece of the urban puzzle. As described in Section 4.1.1, several domains are heavily affected by the reachability afforded by road topology. Also missing in this approach is the consideration of topography that can divide areas that would otherwise be considered to be in close proximity (e.g., rivers). In consideration of housing markets, for example, Lu et al. found that incorporating topology and travel time into the functions utilized for GW made a significant difference in the derived values [73]. Specifically, the division created by the Thames between sections of London causes Euclidean distance to be a less accurate distance function when analyzing pricing data across multiple variables.

There are several feasible methods for incorporating reachability into the distance function used for determining the local mean of each region. For example, the region calculations described in Section 4.1.4 could be used to generate the preliminary score for each region. Then, a GW mean could be applied to smooth the outliers and derive more accurate regional means. However, this approach leads to a similar outcome as outlined in Figure 17: the volatility swiftly declines as the scale increases and the region scores tend towards the global mean. One of the challenges of this approach is that it only incorporates one possible route between the centroids of each region, which ignores the dense interconnection that may be present between two more distant regions or the lack of accessible roads between two regions in close proximity. Similarly, the centroid of one region may not be easily reachable from a region though many other road segments are nearby. Even if travel time or hop distance between centroids were substituted for Euclidean distance, these concerns would remain.

4.1.4.2 *Peripheral Road Topology*

To account for the differences in connectivity due to road topology, an alternative geographic weighting method is required. The road sampling and region scores described in Section 4.1.4 provide the foundation for a method that accounts for the differences in accessibility from one place to another in terms of the distance between them along surface roads and the number of routes that are available. I propose the following method:

1. Determine the preliminary region scores by sampling the internal roads and their proximity to events through a local distance function.
2. Extend the search beyond the region boundary to include the road segments that are reachable within the limits of a travel bandwidth h_g .
3. Combine the internal and peripheral road segments scores using an extension of Equation 6.

To perform the second step, I first identify each of the road segments that cross the region boundary using a simple line intersection test. Then, I explore the road segments that are connected to these road segments but that do not lie inside the boundary of the region. As I explore the topology of the road network surrounding the region, I maintain a data structure of the accessible regions and the shortest distance that was used to reach them from the boundary. A breadth-first search of the peripheral topology ensures coverage of the many roads that surround the region, which is unlikely to be consistent in every direction. In Figure 19, for example, the progressive deepening of the search around an enclosed region reveals that the region is more tightly connected to the south, southwest, and north. While there are connections to the regions in other directions, they are more sparse, which would not have been reflected in a traditional smoothing by Euclidean distance between region centroids.

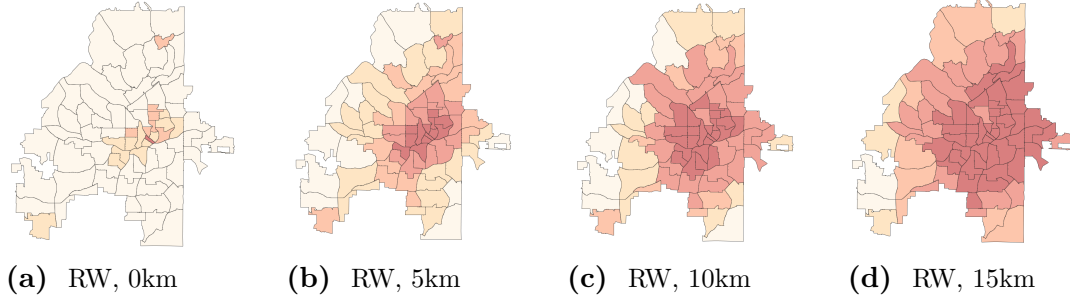


Figure 20: Reachability weighted (RW) mean applied to an event data set at a range of scales. In Figure 20a, no smoothing is applied between regions and the mean event proximity to roads is displayed. In subsequent figures, h_g is increased in increments of 5km. A lightly-shaded outlier is present in the center of the map, warranting further scrutiny.

To perform the final step, I revisit Equation 5. In this circumstance, however, I must first perform a slight modification to Equation 1. Rather than solely comparing the region values, I instead evaluate the road segment scores both within the boundaries of the region and those reachable paths explored by the breadth-first search outside the boundaries. Extending Equation 1, I assign a weight $w_i = 1$ if the road segment falls within the boundaries of the region and follow the bisquare evaluation for road segments in the periphery.

$$w_i(u, h_g) = \begin{cases} 1, & \text{if segment lies within region} \\ [1 - \frac{d_i^2}{h_g^2}]^2, & \text{if } d_i < h_g \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

Using Equation 6, I can then carry out Equation 5 on the road segments that are reachable from the region but outside its boundaries. As before, I weight the external values against the internal value and divide by the total weight to determine the local mean of a region. This new value, defined as the reachability weighted mean (RW mean), exhibits a different behavior as the scale of h_g varies compared to the traditional GW mean using Euclidean distance.

4.1.4.3 Procedure Summary

In this section, I provide a summary list of the steps necessary to reproduce my technique for RW mean and a concise discussion of algorithmic complexity. Given a set of point-based events $E = \{e_1, e_2, \dots, e_n\}$, a connected graph of road segments $R = \{r_1, r_2, \dots, r_n\}$, and a set of bounded regions $U = \{u_1, u_2, \dots, u_n\}$:

1. For each region u_i , identify the road segments $\{R_l \subset R\}$ that are at least partially contained within the boundaries of the region.
2. Of the road segments within $\{R_l\}$, identify the subset $\{R_b \subset R_l\}$ that intersect the boundary of the region, as these lead to other nearby regions.
3. For each road segment within $\{R_b\}$, perform a breadth-first traversal to identify road segments $\{R_g\}$ that are reachable within the global bandwidth h_g but not in $\{R_l\}$. Record the minimum distance necessary to reach each segment.
4. For each road segment within $\{R_l \cup R_g\}$, identify the set of point-based events $\{E_l\}$ that are within the local bandwidth h_l . Record the minimum distance necessary to reach each event. Using Equations 3 and 4, assign a score to each road based on these proximity values.
5. For each region, aggregate the scores of road segments from $\{R_l \cup R_g\}$ using Equation 6.

Reducing the complexity of this approach depends heavily on the use of supporting data structures. For example, Steps 1–4 benefit substantially from the use of a quad tree, r-tree, or other hierarchical spatial data structure to reduce the number of comparisons that are necessary for determining the subset of items that are in proximity to another item (e.g., events in proximity to a road). This reduces the complexity of Steps 1 & 4 each from $O(nm)$ to $O(n \log m)$ where $n = |U|, m = |R|$

for Step 1 and $n = |\{R_i \cup R_g\}|$, $m = |E|$ for Step 4. The subset created in Step 2 can be performed in parallel with Step 1. Comparisons of distance from individual road segments to events can be conducted in parallel, as can the overall process for each bounded region.

For Step 3, in which a breadth-first traversal is conducted starting from each of the road segments on the boundary of a region, there can be significant overlap between the explored regions of nearby but distinct boundary roads. During this traversal, the most direct way to track the distance from the boundary road to each of the visited road segments outside the boundary of the region is through a hash table. These traversals can also be conducted in parallel, but I advocate the use of a hash table designed for concurrent access shared by each of the individual processes so that road segments in the periphery are not unnecessarily revisited. The traversal of each periphery can be considered within $O(|V| + |E|)$, where $|V|$ consists of the road endpoints reachable within the bandwidth h_g from the boundary segment and $|E|$ consists of the road segments connecting them. Many of these vertices are not explored, as they are within the boundary of the region and thus within $\{R_i\}$.

It should be noted that this process overall takes considerably longer on a single computer than the time required for GW mean, which consists of a single $O(n^2)$ weighted aggregation between regions. As stated previously, many of the steps in this process are highly parallelizable, which could significantly reduce the time required for completion. Also, a practical implication of running this procedure for a given set of parameters for h_l and h_g also yields the data necessary for deriving the aggregate region scores for lower values of those parameters. This is put to use in the construction of the scalogram (Figure 18), for example, by specifying a maximum h_g and then using the collected data to determine the relative region scores for ten sub-intervals of that score.

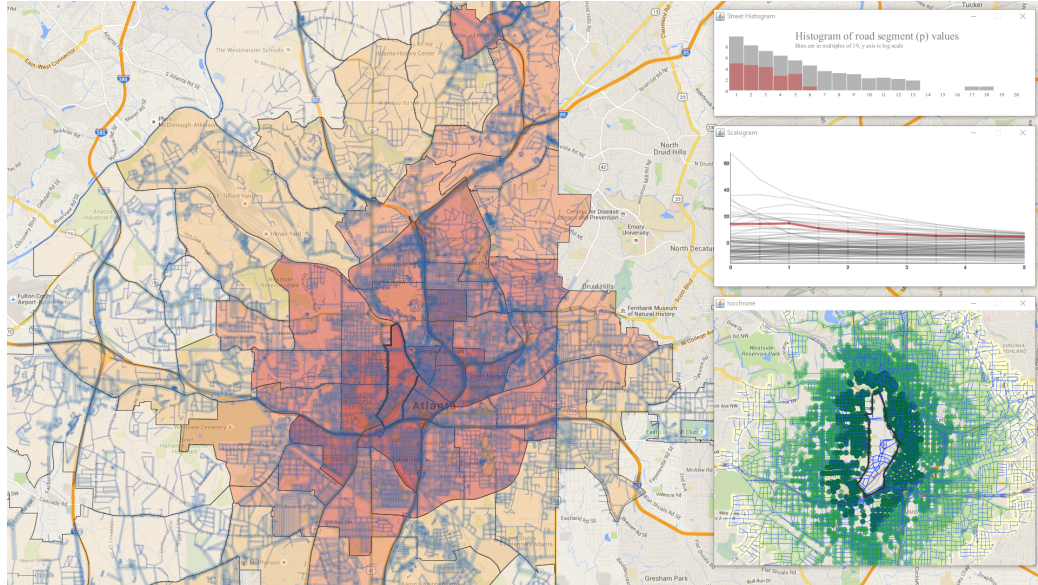


Figure 21: Exploration of bandwidths through coordinated views. An overview of all regions can be used to compare reachability weighted mean scores overlaid with road scores. Linked views provide (from top to bottom): details for road score distribution; a scalogram of variation in region score as bandwidth increases; and a detail isochrone map for a selected region.

4.1.5 System of Coordinated Views

Throughout the previous sections, I have generated additional metadata for further analyzing the quality of each weighting scheme: the relative distribution of road-event scores within a city (Figure 16), the relative change in regions score as the bandwidth is changed (Figure 18), and the reachable regions inside the boundary and within the periphery of a bounded region that are included in its score (Figure 19). This metadata is generated for both GW and RW means, and the visual representations provide additional context for gaining insights into the underlying data. To provide access to this context, I have developed a system (Figure 21) combining a primary map display with three linked panels displaying (from top to bottom) the histogram, scalogram, and an isochrone map. Selecting a region in the primary map display highlights the road-event score within that region in the histogram, the line of the changing relative value within the histogram as the bandwidth is increased, and the roads that are on the periphery of the bounded region and contributing to its score.

The user can also draw a selection box in the scalogram to select one or more of the lines, which updates the other views and highlights the region in the primary map. This interaction can be useful for finding groups of regions that score similarly at different bandwidths.

Within the isochrone map, dark green roads are close to the boundary of the selected region and exert a higher influence on its weighted score, while more lightly shaded roads exert a smaller influence. Roads inside the region are drawn in blue and not rendered with the isochrone shading so that the internal structure of the region can be more easily examined. As evident in Figure 19, increased window size does not uniformly distribute the effects on a region in all directions, and these alternative representation and interaction methods are necessary to explore this variation. Unlike the weighting maps provided by Dykes & Brunson [31], the contribution of surrounding areas to the local mean is highly dependent on the number of roads that pass through them and the distance to reachable road segments. There are many approaches to directly representing the reachability of a region, employed typically to analyze the movement of people and things through a space as trajectories [1, 6, 101] or to investigate travel time on public transportation [116]. Given that my focus in reachability weighting is the accessibility of regions rather than exploration of movement data or optimal routing, the isochrone map idiom is ideal for depicting the road segments that are most likely to influence the score of a region due to accessibility [34, 85].

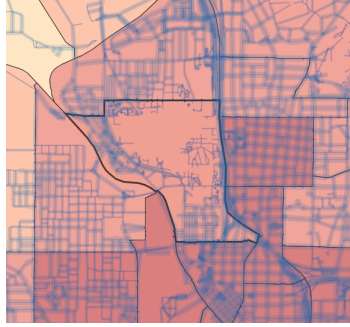
For example, one of the most striking features of Figure 20 as the distance is gradually increased is a local outlier near the center of the map that continues to have a comparatively low score in relation to the surrounding regions. Examining the overview reveals an area in the region where no roads are depicted (Figure 22a). Investigating the isochrone map in Figure 22b reveals that there are a number of reachable roads for the region both within its boundaries and outside. This indicates

that incidents are either not occurring in this area, or that perhaps they have been excluded from the original data set for some reason. Subsequent investigation later revealed that it is the latter, as the area falls on a college campus that keeps an independent record of events that are not distributed within the common data set for the entire city. This insight, while readily apparent as an outlier in the RW images of Figure 20, is absent from the GW image and the original choropleth maps of event counts.

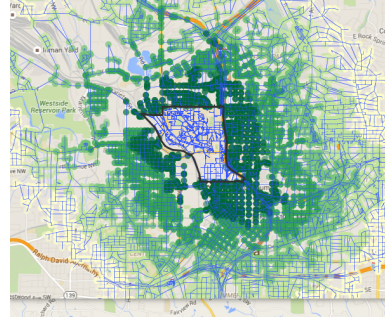
Taken together, these views allow a user to interactively explore the choropleth data to understand why a region has been assigned a certain score, rather than forced to accept the static relative ordering generated by many maps. Through the three linked panels, a user can determine if the internal score is the most significant component of the region's ranking compared to others, or if it's due to its proximity to other high scoring regions. Similarly, the user can determine if the events are uniformly distributed around a region, or if the influence within the bandwidth is due primarily to one or more neighboring regions along its border. Finally, a user can examine the affect that the parameters of the model itself is having, and if the current setting for the bandwidth distance is causing regions to jump or plummet in relative score in comparison to other possible settings.

4.1.6 Evaluation

To evaluate the benefits of reachability weighted statistics for the analysis of event data within bounded regions, I analyzed the relative density of crime incidents in the city of Atlanta across three primary schemes of modifiable areal unit: police beats, neighborhood planning units, and census districts. I compared the results generated by RW means to a choropleth map and geographically weighted means to determine how the modifiable areal unit problem (MAUP) affected each.



(a) Roads and region scores.



(b) Isochrone of the reachable roads.

Figure 22: A detailed look at a region of the map that has a lower score than other regions. Investigation of the overlay in Figure 22a reveals an area where no events have been sampled through roads, even though roads are visible in the isochrone of Figure 22b. This indicates that event data may be missing for the area.

4.1.7 Data

The data used for this evaluation consists of three parts: events in an urban area, a road network, and alternative methods for categorizing the area into separate bounded regions. In this section, I describe each, how they were obtained, and implications for analysis by reachability weighted mean.

4.1.7.1 Crime Data

The data collected for this evaluation consisted of crime incidents occurring in the city during a one-year period of time. Incidents fell within one of the major categories reported to the FBI as part of the uniform crime reporting (UCR) initiative, and included both violent and non-violent offenses (Table 1). This data is made available to the public by the Atlanta Police Department, and is routinely updated with new incidents. In addition to category, each incident has a recorded date, time, and location. Incidents are collected from the city's central records, and locations are obtained from the computer aided dispatch system. Figures throughout this section have been populated using the crime incidents described here so that they would be more familiar to readers at this point.

Table 1: Incidents of crime in Atlanta for one year

Nonviolent	Count	Violent	Count
Larceny(from vehicle)	9,798	Aggravated Assault	2,159
Larceny(not from vehicle)	7,234	Robbery(Pedestrian)	1,798
Auto Theft	4,469	Robbery(Commercial)	233
Burglary(residential)	3,988	Robbery(Residence)	209
Burglary(non-residential)	862	Rape	144
		Homicide	75
Total	26,351		4,618

4.1.7.2 Road Topology

Road topology is obtained from Open StreetMap through the Overpass API. All roads are downloaded for the bounding region that encapsulates the crime data; no distinction is made for pedestrian roads. Roads are stored as a sequence of nodes with fixed locations and unique IDs. A list of nodes is created and a tally is maintained as it appears in subsequent roads. Nodes that appear more than once are considered *intersections*. Once all roads have been processed, a second pass identifies occurrences of multiple intersections within the same road and stores an edge for them. An intersection may have multiple edges, but this direction is not stored and redundant edges are discarded. As the nodes have a location given by a latitude and longitude, edge weights can be derived for the physical distance between the endpoints. This allows for more fine-grained calculation of path distances, but could be replaced with an estimation of travel time for more accurate values. The described approach does exclude dead ends and simplifies longer curvy roads without intersections. Still, this process is fairly accurate, and is easily repeatable across other cities for which crime data can be obtained. For the city of Atlanta, the network consists of 71,751 nodes and 195,946 edges.

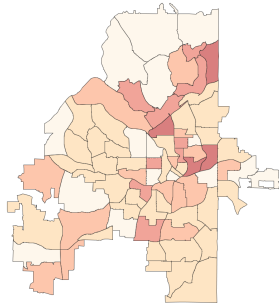
4.1.7.3 Modifiable Areal Units

Depending on context, there are several alternative methods of organizing the city into bounded regions. First, the city of Atlanta is divided into six major public safety

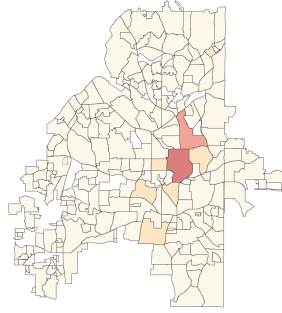
zones, which are further individually subdivided into dozens of police beats. During a shift, each beat is primarily assigned to a single unit that responds to dispatched calls. Beat and zone boundaries are redrawn periodically to reflect the changing population and trends in criminal activity. Beats are drawn so that a unit responding to a call can generally reach any portion of their beat within a short amount of time. The city can also be divided into several major neighborhood planning units (NPU), which are used primarily by neighborhood planners to coordinate with citizens and civic organizations to decide zoning and policy issues for the city. NPUs are frequently divided into one of any number of neighborhoods depending on who is drawing the lines, which can reflect trends in the housing market and the desirability of commercial real estate. One example neighborhood parceling scheme is utilized in this section, which is further complicated by its incompleteness. Several gaps exist in the boundary file where neighborhoods are missing or were never originally encoded. Finally, the city can also be divided into the regions that are used in the federal census. Census tracts are determined at the federal level, and can vary widely in area and population. In this section, I examine the most recent census tracts available from 2010.

4.1.8 Comparison of Techniques

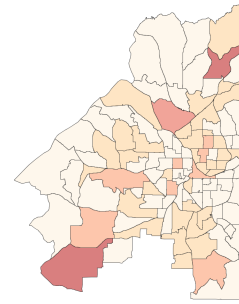
In Figure 23, I compare three techniques for analyzing spatial events: regional count, geographically weighted (GW) mean, and my reachability weighted (RW) mean. For each technique, I generate alternative mappings using the three different areal units described in Section 4.1.7.3. Each of the smaller figures has been assigned a color mapping with a normalized scale independent of the other figures so that the relative distribution of events within the mapping can be compared. This is necessary given that each weighting method generates substantially different numbers. Darker regions correspond to a higher score in each image, regardless of the method used to determine each score.



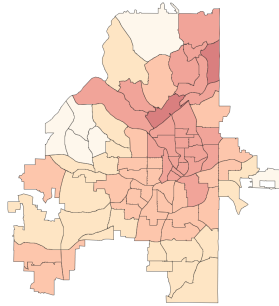
(a) Event count within police beats.



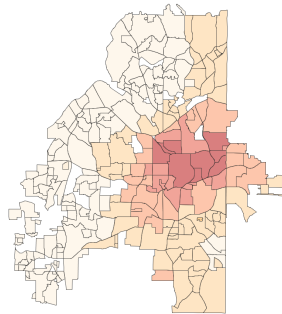
(b) Event count within neighborhoods.



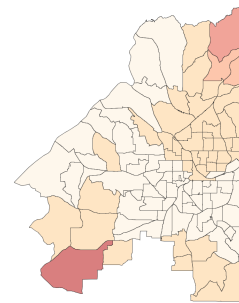
(c) Event count within census districts.



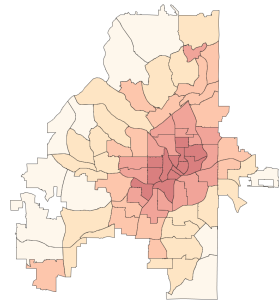
(d) GW mean of police beats with $h_g = 5km$.



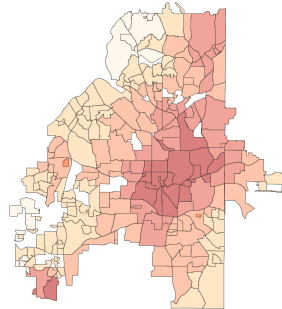
(e) GW mean of neighborhoods with $h_g = 5km$.



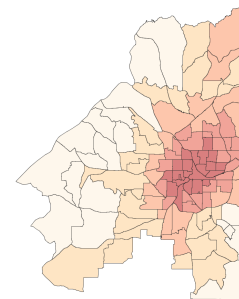
(f) GW mean of census districts with $h_g = 5km$.



(g) RW mean of police beats with $h_g = 5km$.



(h) RW mean of neighborhoods with $h_g = 5km$.



(i) RW mean of census districts with $h_g = 5km$.

Figure 23: Comparison of weighting methods. In the first row, events are counted without smoothing by three different modifiable areal units: police beat, neighborhoods, and 2010 census districts. In the middle row, geographically weighted (GW) mean has been applied to these regions with a bandwidth of $h_g = 5km$, however, the choice of boundaries yields significantly different results. In the bottom row, reachability weighted (RW) mean has been applied to the regions with the same bandwidth. Unlike the other methods, the color encoding in RW is relatively stable across the mapping schemes despite the changes in region size, shape, and count.

For the first comparison, each of the areal units are aggregated by the count of the number of events that occur within the boundaries of each region. In the first row of Figure 23, it is clear that the shape and placement of the boundaries is highly influential on the relative shading of the region. For the police beats and census districts, the color mapping exhibits no discernible pattern. In the neighborhood mapping, the larger regions in the center of the map have collected more events, which brings the representation more in line with the true distribution of events (Figure 25). In the second comparison and middle row of Figure 23, GW has been applied to each of the bounding schemes with a bandwidth $h_g = 5km$. For the police beats, the higher number of crime incidents in the center and throughout the north has been emphasized, but to a lesser degree than in the mapping of the neighborhoods in the central column. While somewhat similar to each other, both figures differ substantially from the census district mapping. The census district mapping fosters a substantially different interpretation of the data than the other mappings, drawing attention to the relatively large southwest and northernmost regions as hotspots (Figure 26).

The variation in distribution apparent in the three images underlies an important consideration in the use of GW: the selection of an appropriate bandwidth h_g . As the scale of the bandwidth changes, outliers in the data will alter position and new distributions will become apparent. As reinforced by Figure 23, the variability in scale is dependent on the shape, number, and area of the regions in the mapping scheme. This is problematic when the bandwidth is meaningful within the domain. For example, a bandwidth of $h_g = 5k$ represents a walking distance outside of a neighborhood that can be covered in about an hour. This is a substantial distance to cover in each direction to obtain groceries, but in sprawled cities is not uncommon. If alternative mapping schemes are used to analyze the relative placement of new grocery locations, the meaningful bandwidth should not be altered to fit the mapping scheme. In the final row of Figure 23, TWR is applied to each of the region mapping

schemes. Unlike the other approaches, distribution of scores for the regions appears relatively stable despite variability in boundary positions and sizes among the three mapping schemes (Figure 27).

To more directly compare the score variation between the three different weighting methods, I conducted an analysis of variance in assigned scores for a collection of points distributed throughout the map. Using Bridson’s method for generating random samples in a Poisson-disc distribution [15], I created a set of 713 sample locations spaced at 0.5km or more from each other (Figure 24). Only samples located within the boundaries of a region in all three mapping schemes were used for comparison. In each trial, I recorded the normalized score of the bounding region that the sample was located within. I then obtained a standard deviation of the assigned values across the mapping schemes for each location. My hypothesis was that the deviation between values for each location would be significantly lower for RW than for region counts or GW.

A one-way between subjects ANOVA was conducted to compare the effect of weighting method on normalized region score in RW, region count, and GW conditions. There was a significant effect of weighting method on normalized region score at the $p < .05$ level for the three conditions [$F(2, 2136) = 150.02, p < .001$]. Post-hoc comparisons using Tukey’s HSD indicated that the mean score for RW ($M = 0.08, SD = 0.03$) was significantly different than both the region count ($M = 0.15, SD = 0.10$) and GW ($M = 0.15, SD = 0.09$). The region count method, however, did not significantly differ from GW. Taken together, these results suggest that the RW assigns region scores more consistently across different areal mapping schemes. This means that insights derived from analysis of an event data set with one mapping scheme are less likely to significantly differ from those that are derived using an alternative mapping scheme, which represents a useful step in mitigating the MAUP in situations where the reachability from one region to another is important for analysis of

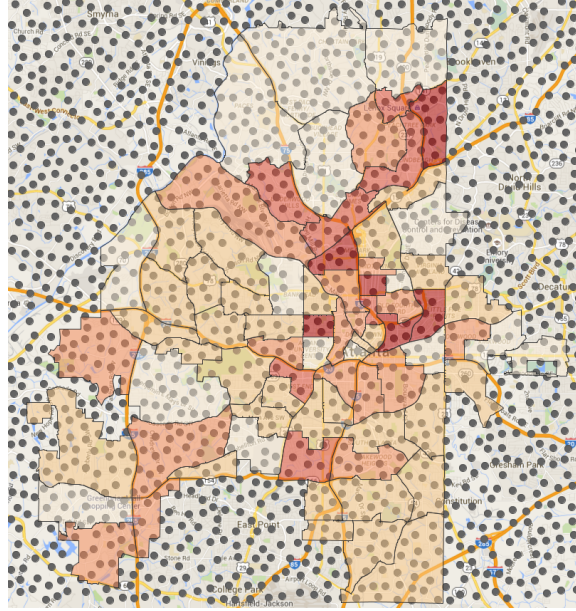


Figure 24: Random samples used to compare the variation in assigned score across areal unit schemes. Sample point locations are generated using Bridson’s method [15] with a minimum spacing of 0.5km.

the data. In situations where reachability is not important, such as weather analysis, alternative methods for interpreting the data at multiple scales and mitigating the MAUP would still be necessary.

4.1.9 Discussion

In this section, I explored methods for analyzing geographical event data at local and global scales based on the structure of urban spaces. I have described a method for reachability weighted means of these districts to better understand the local distribution of events around a region and how travel distance and access to nearby regions can affect this distribution at the global scale. I have presented coordinated views for interactive exploration of the boundaries of a region across varying bandwidths. Finally, I evaluated these methods within the context of the modifiable areal unit problem, and found that the contributed method is significantly less volatile in the scores assigned to regions across alternative areal units.

This research provides one important component for the application of image

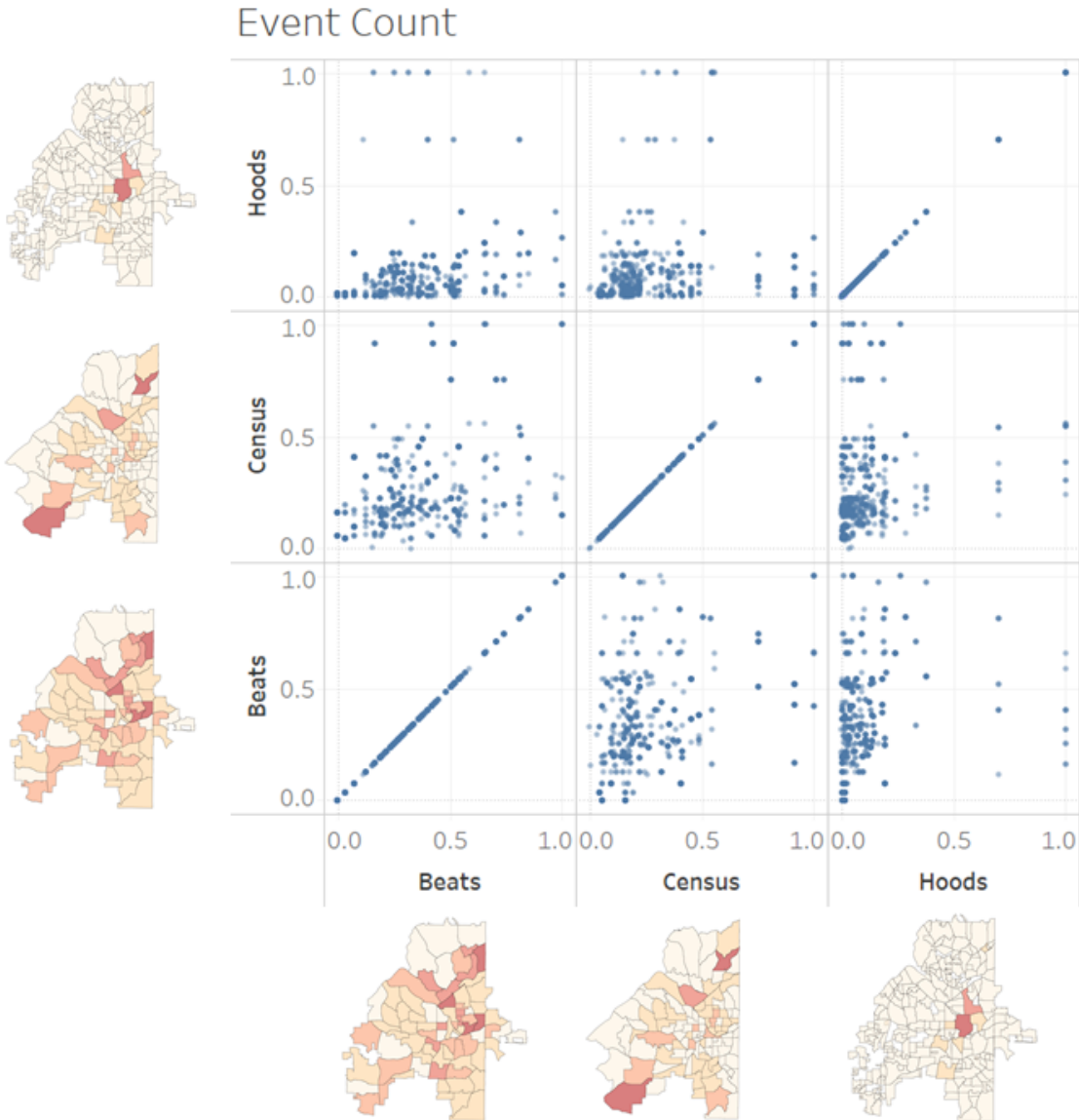


Figure 25: Comparison of sample point assignment as a consequence of modifiable areal units. In this figure, percentile rankings have been assigned based upon the number of events within the region that the sample point is located within. Correlation between areal unit schemes indicates that there is significant variability and little consistency, leading to very different outcomes in choropleth map interpretation and insights.

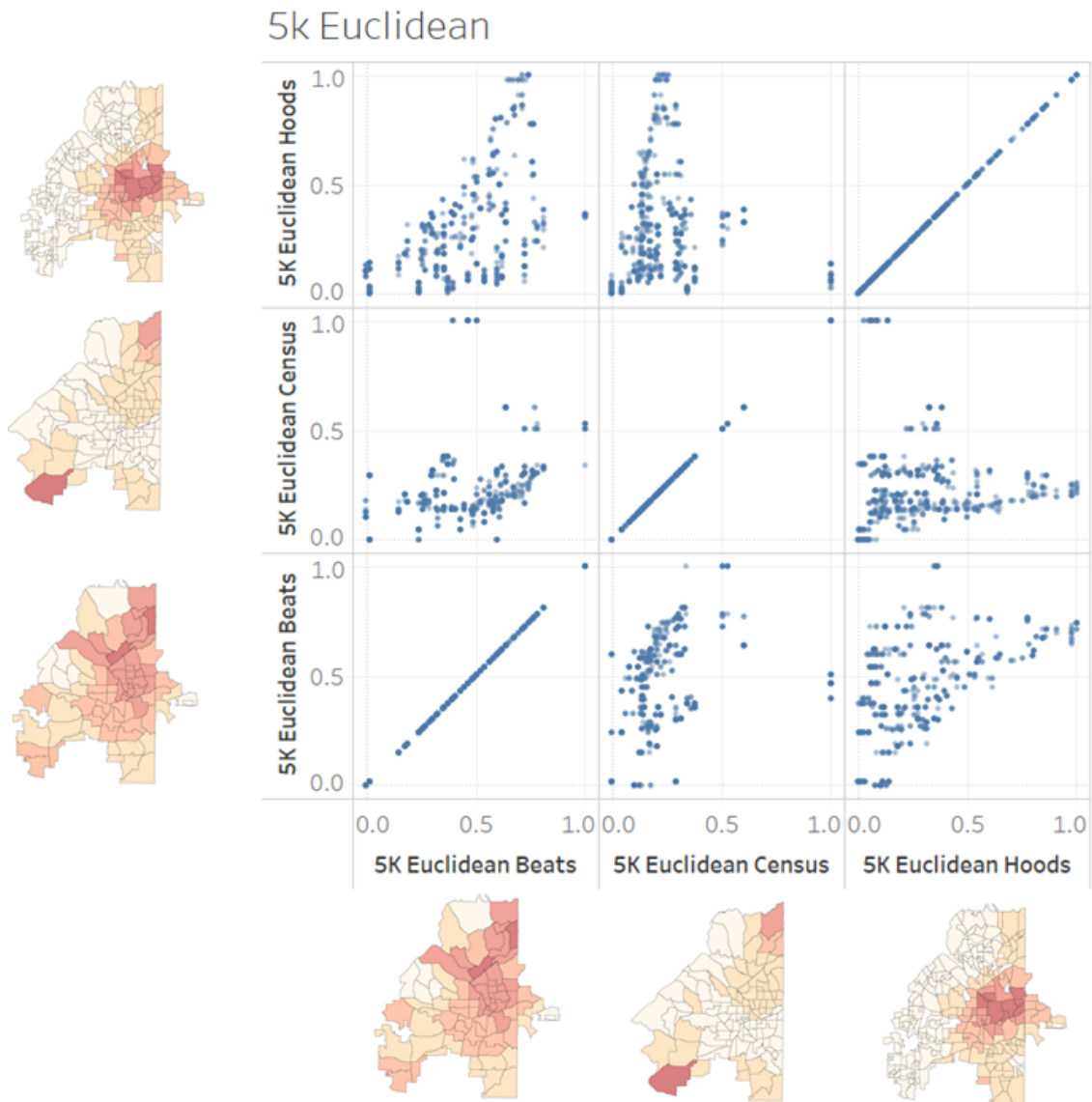


Figure 26: Comparison of sample point assignment as a consequence of modifiable areal units. In this figure, percentile rankings have been assigned based upon the geographically weighted mean (using Euclidean distance) of the region that the sample point is located within. Correlation between areal unit schemes indicates that there is significant variability but some consistency between beats and neighborhoods, leading to different outcomes in choropleth map interpretation and insights depending on the binning strategy used.

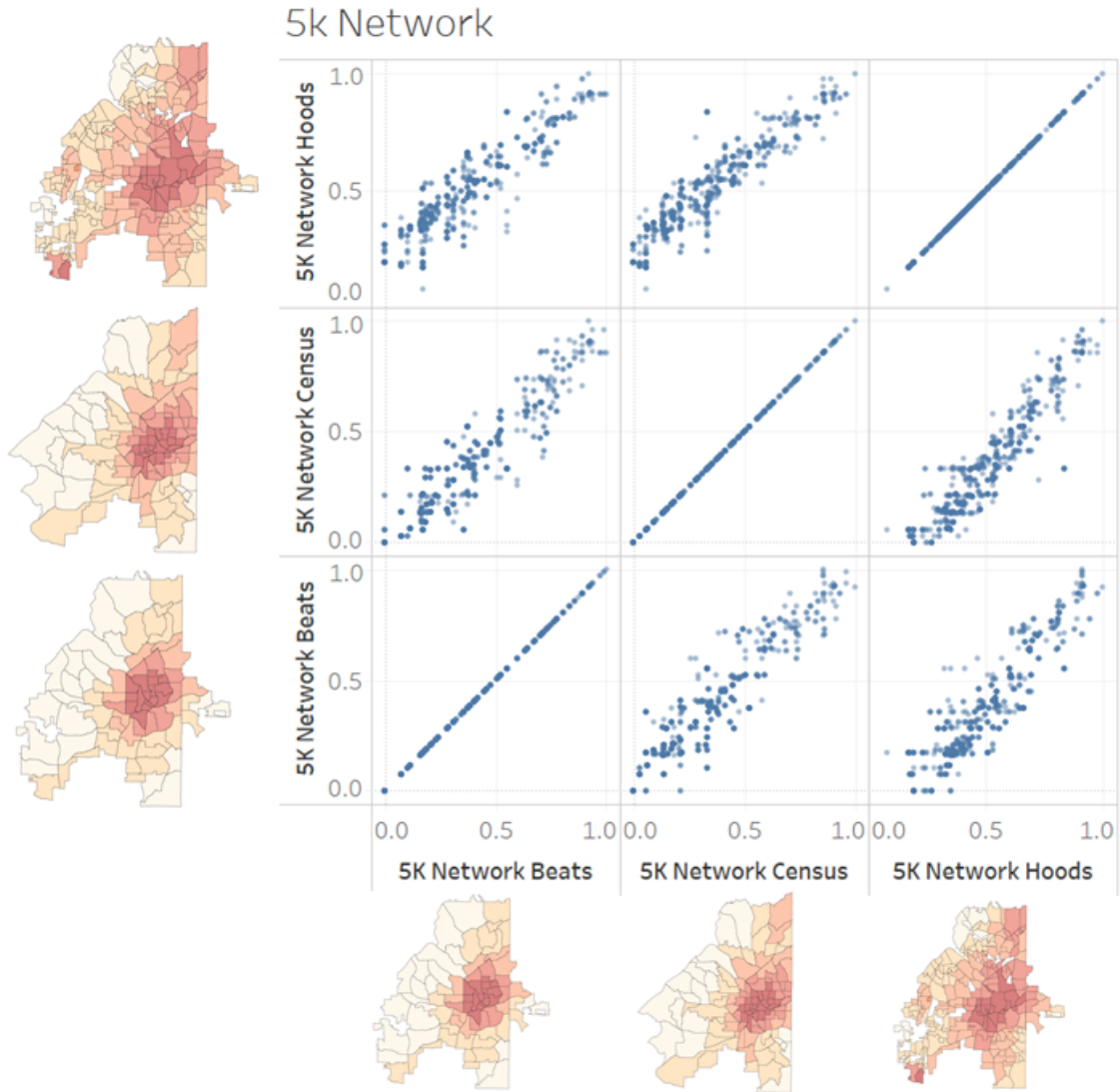


Figure 27: Comparison of sample point assignment as a consequence of modifiable areal units. In this figure, percentile rankings have been assigned based upon the reachability weighted mean (using path network distance) of the region that the sample point is located within. Correlation between areal unit schemes indicates that there is significant consistency and little variability, leading to more consistent outcomes in choropleth map interpretation and insights.

elements in a mental map of urban spaces. By incorporating paths and districts in the interpretation of spatial data, I can allow for the use of various alternative approaches to separating the districts of a city without accidentally imposing artificial partitions. The representation of each district is most influenced by the data contained within that region, but can also still be affected by neighboring data. This allows for the designation of a district that has a definitive *inside or outside* without walling it off from the other elements of the city. More importantly, this also allows many different participants to generate alternative boundary schemes for the city while still retaining the ability to analyze the same data set and have a collaborative discussion about a shared physical environment with distinct perspectives.

CHAPTER V

ANALYZING CIVIC DATA AND URBAN SPACES

In this chapter, I describe applications of these techniques to analyzing civic data and urban spaces across three cities. The first application, a community dashboard developed through a participatory design effort, incorporates community assets as node elements (e.g., libraries, schools, senior centers) within a public safety visualization. This allows community members to discern which nodes are in proximity to high crime areas (Section 5). The next section of this chapter describes my work to formally incorporate important activity nodes and barrier edges within a model of event distribution (Section 5.1). The final section of this chapter describes an experiment to collect mental maps of Atlanta citizens and compare the representation and interaction techniques that are most effective for analyzing urban data in this context.

Individual neighborhoods within large cities can benefit from independent analysis of public data in the context of ongoing efforts to improve the community. For example, communities in the Westside of Atlanta have been changing their neighborhoods for the better by organizing amongst themselves and collaborating with organizations that have a local presence. Georgia Tech has partnered with Westside communities via the Westside Community Alliance (WCA), a communications network started in 2011 by the Ivan Allen College of Liberal Arts in collaboration with the College of Design and the Office of Government and Community Relations. The WCA works to build and sustain relationships among constituencies located in West Atlanta to strengthen partnerships around issues of common concern. In a paper presented at the Bloomberg Data for Good Exchange, I explored these issues with members of

the WCA (Katie O’Connell, Mackenzie Madden), student fellows from the Data Science for Social Good program (Yeji Lee, Shawn M. Staudaher, Firaz Peer), and Ellen Zegura of Georgia Tech [82].

Since its inception in 2011, the WCA recognized the importance of data dissemination as a tool for community development. In February 2016 the WCA launched the WCA Data Dashboard. This on-line website is designed to be a one-stop data shop with information presented in locally recognized and meaningful geographies. Rather than census tracts or zip codes, data is presented by neighborhood or Neighborhood Planning Unit (NPU). In the Atlanta area, neighborhoods and NPUs represent the unit of local community organization as well as local identity and pride. The Dashboard developers have gathered previously siloed data sets and integrated them into one platform to support examining data relationships. Central to the Dashboard’s design is regular engagement with local organizations and community groups both for design feedback and for data literacy training. The Data Dashboard is organized into portals that correspond to community concerns. Currently there are five portals—community profile, education, historic data, historic timeline, and resource library—with several more in the pipeline [83].

In Summer 2016, the WCA served as a client and partner in the Atlanta Data Science for Social Good (DSSG-ATL) summer internship program. DSSG-ATL is modeled after the DSSG summer fellowship program started at the University of Chicago in 2013; students work full-time in teams on projects that come from local partners, with a focus on supporting understanding and decision-making based on data, in domains of social importance. DSSG-ATL started in 2014 [115].

In this section, I describe our participatory approach to building the public safety module of the WCA’s Data Dashboard, with some key insights into how one can approach similar projects in the future. We conducted interviews with key community stakeholders and participated in local government meetings to understand the needs

of our users. These needs include the ability to locate where crimes have occurred, a metric to quantitatively evaluate the efficiency of their public safety programs, and a means to learn how other factors such as education, age, transportation, housing, and more are related to crime. To satisfy these needs, we introduce a mapping tool with the capability to locate current and historic hotspots of criminal activity. This tool includes a variety of overlays that allow users to spatially correlate features of the built environment, such as code violations with criminal activity as well as crime prevention efforts. We also present a statistical model that highlights correlations between crime and other socio-economic factors specific to particular neighborhoods in Atlanta.

Parts of the public safety module have been built with Tableau and web-based mapping libraries like CartoDB and Leaflet.js to help maintain consistency with the existing Data Dashboard. The paper was submitted for publication during the final stages of developing the first version of the tool, and the public release of the dashboard was completed in the fall of 2016.

5.0.1 Design Goals

Because the Data Dashboard was developed within the WCA, it shares design goals and values with the overarching organization. In particular, the WCA is a partnership between a university (Georgia Tech) and local communities. The structure and priorities of the WCA are arrived at via community engagement. Staff in the WCA live in these communities and are regular participants in community meetings and events. Long-standing relationships with the community are created and sustained through regular and myriad interactions. The Data Dashboard was designed to be accessible to citizens in ways that fit how citizens think about their communities, useful for citizen information gathering and advocacy, and integrated so that citizens do not need to navigate and synthesize data from disparate sources.

For example, would it be possible for churches to input the different youth programs they run, into a tool to determine how it has impacted juvenile crime in their neighborhoods? As part of Operation Shield, the Police Foundation, with the help of a \$1.2 million grant from Invest Atlanta, has recently installed about 80 cameras in different parts of the neighborhood that are considered hotspots for drug related crimes. Can we overlay these camera locations on a map to see what impact they may have had on these hotspots? The Westside also has a high number of vacant properties, which many residents believe is the cause for high crime in their neighborhoods. Can we numerically determine the relationship between crime and code violations, so residents can make a case with the city to demolish specific properties that have a high correlation with crime? These questions are representative of the kinds we hope we can answer through our tool. But before we could build any of it, our first task was to understand the people we were working with, to make sure that the tool we build is the one they need. Community feedback and design participation—described in more detail in the next section—are key to achieving accessibility and usefulness.

5.0.2 Approach

In addition to weekly meetings between the WCA and the DSSG team, student members of the research team also participated in NPU meetings. These are public planning meetings attended by residents and other interested stakeholders within the community. The meetings provide a place for residents to interact with community and city leadership. These tend to be highly contested spaces, as committee members share updates, residents hear about and vote on specific changes they would like to see implemented, and organizations/researchers get buy-in for various initiatives they have in the pipeline. Attending these NPU meetings gave us a good sense of the issues residents were currently grappling with as well as a first-hand exposure to the dynamics of community governance.

Some high priority issues have their own committees and meeting schedules; public safety has recently emerged as an area worthy of committee instantiation. The Vine City Public Safety Committee meets on a regular basis to discuss the status of the many public safety programs, and their meetings are attended by police officers and residents alike. We attended one instance of this meeting, where we described our preliminary tool, along with some screenshots of the kinds of visualizations we had in mind, to seek feedback.

The attendees at the public safety meeting seemed excited about the prospect of having free and open access, along with the ability to analyze crime and code violations data within their neighborhoods. We spoke with one of the code enforcement officers present at the meeting, who was interested in talking with us further about ways his department could use our data visualizations in their day-to-day operations. This was significant, as public safety officers were one of the user groups we were hoping to design for. He introduced us to one of the senior analysts at the Code Enforcement division of the Atlanta Police Department, who gave us an overview of exactly how his team goes about collecting and reacting to code violation complaints in their jurisdiction. This meeting also helped clarify many questions we had about the code violations dataset we were working with.

5.0.3 Community Asset and Crime Mapping Tool

The third module in the dashboard is the spatial visualization of the crime data, code violations, and community assets. The primary goal of these visualizations is to provide users with a means to precisely pin-point where crimes occur, and to give them the ability to learn how these locations change by crime-type and over time. Additionally, the map allows users to focus on specific geographic areas to examine the relationship between crime, code violations, and community assets.

The crime data is visualized with a hex-based heat map (Figure 28). This map is

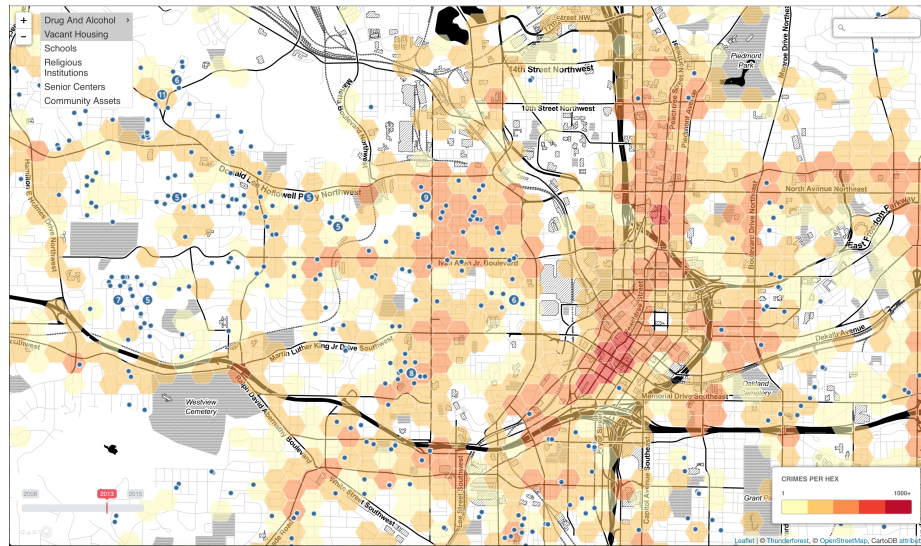


Figure 28: Map Module - The map displays drug and alcohol related crime (colored hexes) and vacant houses selected (blue circular pins).

constructed by counting the number of crimes within each hex and assigning one of five colors based on a logarithmic scale (i.e. the first color represents a single crime per hex, the second color represents 2-10 crimes per hex, the third color is 11-100 crimes per hex, and so on). This logarithmic aggregation of crime into colored hexes naturally highlights hot-spots, where small regions have significantly higher numbers of crime than average. In addition, crimes may be selected by their Uniform Crime Reporting Code, or into larger categories of crime including drugs and alcohol, sex crime, theft, or violent crime. Furthermore, a specific time period may be selected by specifying a date and time span (all data, a year, or a month).

The code violation data is sparser than the crime data, and a heat map would not be an appropriate visualization. Instead, these data are shown as circular pins on the map, with clusters of points represented by a larger pin with the number of points in the cluster inscribed. These data may also be selected by time in the same manner as the crime data. Community assets are visualized with pins as well, but without clustering as these data are sparse enough to not require aggregation. In addition, further information specific to the asset type is displayed on mouseover.

The goal of this map is not to highlight areas of high crime in a negative manner, but to be a tool the community may use to coordinate their crime prevention strategies. For instance, an NPU public safety chair may be leading a drug prevention campaign. They may use the heat map to locate areas of drug use in their neighborhoods, and then overlay drug prevention programs from other sources on the map. They may then choose to concentrate their efforts on an area with drug usage and without another active drug prevention program. In another case, a public safety chair may use the map to examine historic data. They may zoom to a location where they have been active in the past and learn how crime has changed over time in the specific area that they work.

The aggregated data and correlation modules are generated with the Tableau data visualization software. The map is created with CartoDB, a database and geospatial visualization portal. The data are hosted with CartoDB's PostgreSQL server, and the visualization and navigational tools are created with the JQuery and CartoDB JavaScript libraries. All three modules were hosted on the WCA data-dashboard. Since that time, unfortunately, the dashboard has been taken offline due to restructuring within the WCA.

5.1 Nodes, Paths, and Edges

In a paper presented at EuroVis 2017, I proposed a novel technique for applying mental maps based on nodes, paths, and edges to spatial data [48]. I describe my implementation of this technique in a sketch-based system that can highlight the disparities between mental maps, and provide preliminary findings from the application of my technique to property crime data in three U.S. cities: Baltimore, Atlanta, and Chicago. This technique provides a method for identifying disparities in mental maps of urban spatial data and provides insight into the effects these maps have on the analysis of urban data, which is necessary for making decisions that affect all

members of a community regardless of their image of the city.

Though competent mental maps can be drawn from any of the five elements [108], I chose to focus on nodes, paths, and edges for this technique. Brantingham and Brantingham have demonstrated that these elements are necessary, if not sufficient, components in a framework for the analysis of the spatial distribution of crime [14]. Spatial data visualization represented along paths has been previously explored by Xie and Yan for traffic accidents [113, 114], Wong et al. for power grids [109], and Kim et al. for crime [64]. Wood et al. have also demonstrated abstract hierarchical representation that depict the connectivity between regions on a map, though this is primarily for trajectory data rather than joining of spatial data to paths [110, 112]. Nodes, or areas of high activity, can most directly be compared to the hotspots derived during more spatial analysis of point-based data. However, unlike Euclidean [77] or grid-based [89] approaches, our technique allows users to actively specify nodes rather than try to passively detect them. Our technique is novel in its incorporation of edges, or barriers, which are not present in other approaches to model spatial data.

I do not include landmarks, as they are often visible at great distances (e.g., tall buildings and landscape features), requiring the capability to model observability and occlusion in a 3D space. This was a significant feature of Glander and Döllner’s system [45], but it presents additional complexity when the landmarks are user-specified. As our intent was to provide a technique to facilitate sketch-based specification of urban mental maps in a 2D-space—similar to a hand-drawn sheet of paper—this was outside the scope of our work. I also do not include districts, or bounded regions, as considerable issues arise in both model construction and interpretation due to the modifiable areal unit problem [84]. Considerable effort has been put forward to overcome this problem in planar spatial representations, notably through geographically weighted regression [31] and more recently through Bayesian weighting [25]. However, it remains an open problem to meaningfully incorporate path-based data models

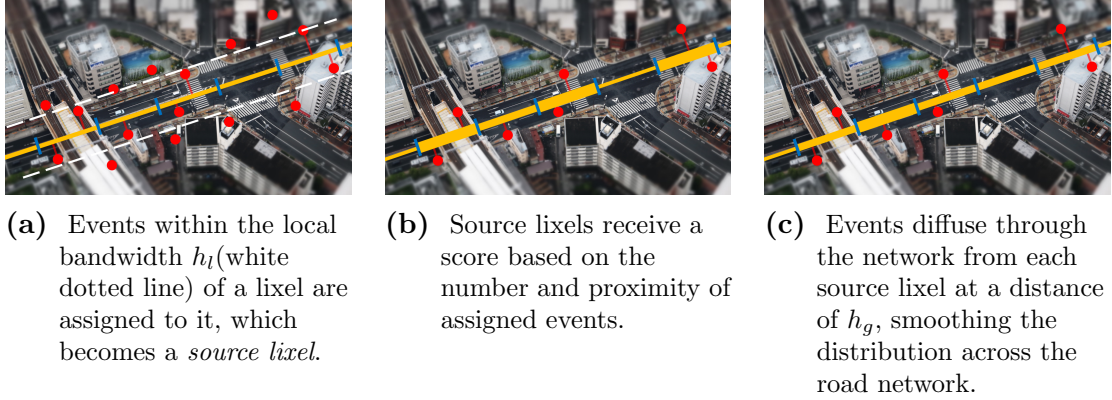


Figure 29: Events are assigned to the closest lixel in the network before diffusing to nearby lixels. The local bandwidth h_l represents the maximum distance that a crime should be considered to be associated with a representative street location. The global bandwidth h_g represents the maximum distance that connected and nearby streets are affected by nearby crimes.

(e.g., [113]) and bounded regions. Our technique described in Section 4.1 represents one promising approach.

5.1.1 Paths Revisited

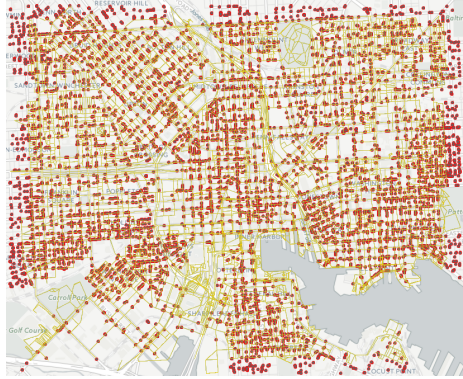
I utilize paths as the primary urban imaging element in our technique. Building from previous work in network-based Kernel Density Estimation (KDE), I utilize a path network consisting of lixels [113]. This approach differs from standard planar KDE in that the distance between points on the map are not measured in Euclidean space, but based upon network distance. This is similar in many ways to the technique described previously, however, in that the lixels are more fine-grained and consistent in length than the primal approach described in Section 4.1 and adapted from Porta et al. [88]. In our approach, I obtained road-level data from OpenStreetMap(OSM) and constructed a course graph between intersections. As an open data source that can be modified by the public, OSM is ideal for constructing the backdrop for analysis of community-oriented data. I acknowledge, however, that this differs from sketches or explicitly-defined paths created by users. For this technique, I include roads accessible by car and footpaths accessible by pedestrians.

I begin by dividing the road topology into *lixels*, or linear pixels. Lixels consist of linear road segments of equal length. In terms of KDE, lixels are similar to selection of a pixel resolution for the planar space. The selection of lixel length in network KDE is, as with the selection of pixel resolution in planar KDE, an important consideration affecting the variation details of spatial patterns (Figure 31). Once road segments have been divided into lixels, I then assign each spatial data item in the set (e.g., crimes) to the nearest lixel. Each lixel with one or more assigned data items is a *source lixel*, and serves as the point of origin for the network KDE within the network topology. I follow the approach of Kim et al. [64] rather than Xie and Yan, and assign scores to each lixel based on a weighted kernel function (Equation 3) and a minimum detection bandwidth rather than using a count of nearby items. For each of these events e_1, e_2, \dots, e_n , I determine the minimum distance d_i from that event e_i to any part of the lixel (Figure 29). For this approach, the choice of kernel function does not affect the results as much as the choice of bandwidth h_l , which should be chosen carefully based upon the domain. I utilize the Epanechnikov kernel, depicted in Equation 4.

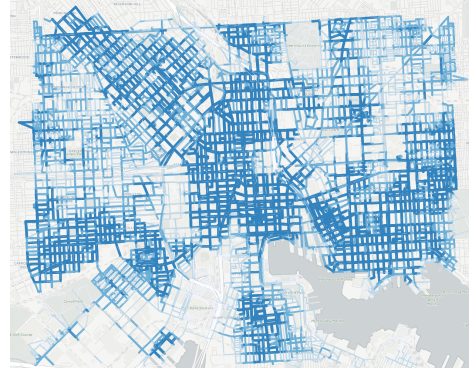
$$f(x) = \frac{1}{nh_l} \sum_{i=1}^n K\left(\frac{d_i}{h_l}\right) \quad (3 \text{ revisited})$$

$$K(u) = \begin{cases} \frac{3}{4}(1 - u^2), & \text{if } ||u|| \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (4 \text{ revisited})$$

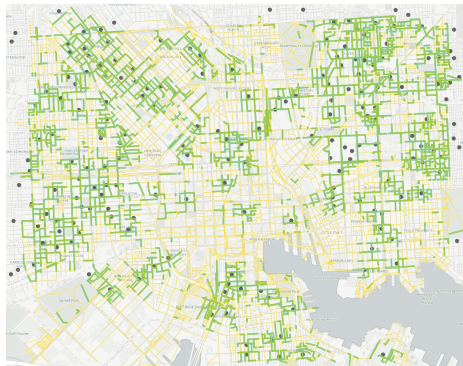
Once the spatial data items have been assigned to the correct lixel, I iterate through the source lixels to determine each of the other lixels that are visible from their position given the current h_l . This visibility is determined by network distance rather than Euclidean distance. Each other lixel that a source lixel can “see” receives a score adjustment relative to the network distance between them. The resulting score for any lixel represents the network KDE score aggregated from all of the other



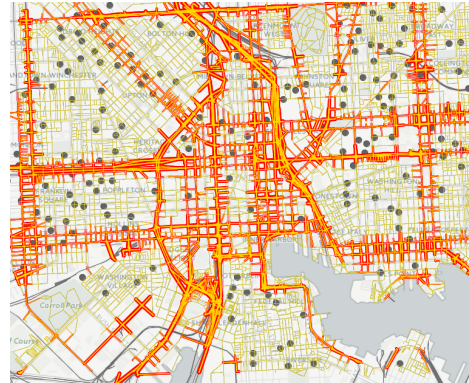
(a) Dotmap of 2015 Property Crimes in Baltimore.



(b) Line map of 2015 Property Crimes.



(c) Church and School Nodes



(d) Highway, Water, and Train Edges

Figure 30: 2015 Property Crimes in Baltimore. In Figure 30b, the path KDE is rendered as line thickness and saturation, with thicker and darker lines indicating a higher network score. In Figure 30c, the path network is shown in yellow, with green highlights indicating the lixels that have been amplified due to node proximity (church and school nodes are black circles). In Figure 30d, the red highlights indicate lixels that have been dampened due to edge proximity (highway, water, and train edges are black lines mostly occluded by the path highlights).

lixels that can reach it (Figure 30b). This score is aggregated using Equation 3, though I substitute a bisquare equation for the kernel (Equation 1). This equation utilizes a different bandwidth, h_g , than the preceding detection bandwidth, as I am calculating the density of the lixel with respect to the nearby lixels in the network topology rather than aggregating nearby crimes. This larger bandwidth must also be chosen carefully, as it determines the default distance that a source lixel can reach other lixels and affect their scores (Figure 32).

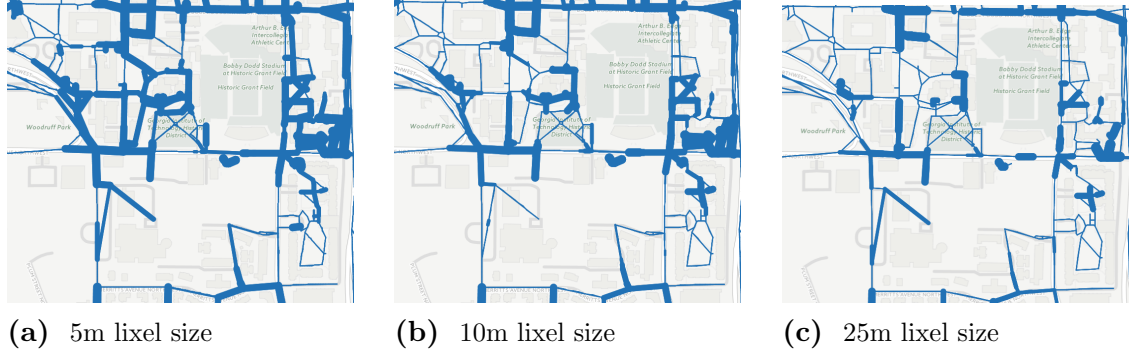


Figure 31: The paths of an urban area are used to represent spatial data through network KDE with varying lixel size. Paths with a nonzero score are drawn in blue; path thickness is mapped to each lixel’s net KDE percentile score for the map. Bandwidth h_g sizes for all images are 50m. Reducing lixel length increases the fidelity of the model but increases the computation cost for deriving the network KDE.

$$K_i(u, h_g) = \begin{cases} \left[1 - \frac{d_i^2}{h_g^2}\right]^2, & \text{if } d_i < h_g \\ 0, & \text{otherwise} \end{cases} \quad (\text{Eq. 1 revisited})$$

5.1.2 Nodes

As elements of an urban mental map, nodes represent areas of high activity: community, shopping, education, worship, and more. However, what represents a node to one person, or one group of people, might differ significantly. This difference can be influenced by where a person lives, where they work, what they do for a living, or their cultural and ethnic background. Nodes are “the concentration of a district, over which their influence radiates and of which they stand as a symbol” [75].

To incorporate node elements into a cognitive model of the city, I update the existing path model to modify the *sight distance* of a source lixel. This modification provides an incremental boost to the h_g of all source lixels within the radius of the node subject to the weighted distance obtained by Equation 1. We default nodes to an activity radius of half the current h_g , though other parameters would yield alternative results (Figure 30c). For example, if $h_g = 100m$, then the activity radius would be

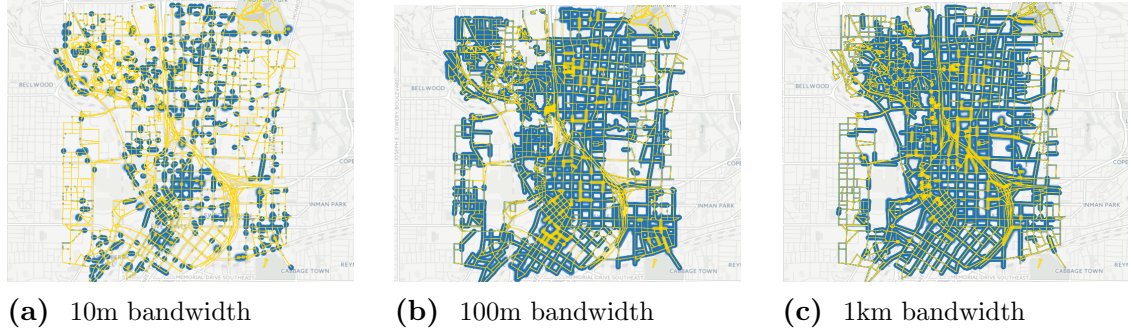


Figure 32: The paths of an urban area are used to represent spatial data through network KDE with varying bandwidth size. Lixel size for all images is 10m. Paths with a nonzero score are highlighted in blue; highlight thickness is mapped to each lixel’s net KDE percentile score for the map; yellow paths for each line are thicker for lixels in the top 2%. In Figure 32a, the bandwidth h_g is set at 10m, which reveals only the most localized trends in path pattern variation. In Figure 32b, h_g is increased to 100m, which allows for identification of larger patterns throughout the network. Finally, in Figure 32c, h_g is set at 1km, which depicts the overall trend of the map but obscures fine-grained patterns in this data.

50m. A source lixel within 25m of the activity node would receive an improved sight of 28.125m in addition to the default sight of 50m. A source lixel that occurs in proximity to many activity nodes could receive additional sight modifications. The result of these modifications is that source lixels that occur in proximity to activity nodes within the map have a pronounced effect on the overall scores for the map.

5.1.3 Edges

Edges represent barriers to movement: highways and train tracks that hinder pedestrians, water features that bottleneck traffic, or other features of the environment. Edges may also be more abstract, such as the dividing line between two neighborhoods. It may still be possible to move across an edge, but it can be difficult, depending on the type of barrier it represents. Criminal activity is often higher at edges, as they often provide a place where land use provides opportunities for mischief and strangers go unchallenged [14].

In our technique, edge elements modify the lixel topology by artificially inflating lixel length (Figure 30d). When an edge is added to the map, it increases the artificial length of all lixels that are within its activity radius, subject to Equation 1. As with

nodes, the default activity radius is half the current h_g . As source lixels are modifying the scores of reachable nodes, they observe inflated lixels as longer than they actually are. Consequently, any destination lixel that receives an increased network KDE score does so at a reduced value.

5.1.4 Crowd-Sourced Mental Maps of Property Crime

To understand the effects of mental image elements on quantitative spatial data, I analyzed the property crime in three US cities using an alternative mental map. I obtained property crime data for Atlanta, Baltimore, and Chicago, three cities with significant difference in layout and navigability. Atlanta, for example, is notable for its wide sprawl and relatively sparse downtown and midtown. It also contains a relative absence of water features, such as rivers or lakes, but is conspicuously divided from north to south by a major US highway. In contrast, Baltimore is a harbor city that contains a relatively dense street grid within its interior. Like Atlanta, however, the population density is low within the downtown, as inhabitants largely commute to the interior to work, shop, or dine. Chicago also lies on the edge of a significant water feature, Lake Michigan, and contains an incredibly dense network topology within its inner loop. All three cities are known for their ongoing efforts to curtail significant property crime.

I utilized road network topology data obtained from OpenStreetMap (OSM) to form the paths within each city. I incorporated all types of paths accessible on foot or by car, though the inclusion of alternative types of paths would help create a more nuanced model (e.g., only including footpaths or only including major highways utilized by out of town commuters). To account for this, I included a second set of features as edges: major interstate highways, train tracks (excluding subway features), and water features (rivers and streams). The inclusion of these features still allowed for the inclusion of the highways as aspects of the cognitive model of each city, but

skewed the perspective towards inhabitants of the urban that might use a highway sparingly while relying on shorter streets and sidewalks to move around on a daily basis.

For each city I create a distinct mental maps to represent community perception. These are intended to represent inhabitants of the city that are attempting to understand the spatially distributed property crimes. This crime consists primarily of nonviolent larceny, auto theft, and nonresidential burglary but also includes pedestrian and residential robbery when it occurs. The data set is limited to only the events that occurred in 2015. For the mental maps, the image of the city is composed of church and school locations, as these represent important components of the community and family. Node locations for the mental maps were obtained from the crowd-sourced tags on OSM. To limit the effects that clusters of relevant nodes might have on the model, I apply a hierarchical aggregation scheme to cluster together nodes that occur within h_l , the activation bandwidth, of each other. Cities were analyzed with a lixel size of 25m to provide a resolution of approximately four lixels per city block. Local bandwidth was set at $h_l = 100m$ so crimes would be associated with a lixel at the nearest block but no further. Finally, the global bandwidth was set at $h_g = 200m$ to limit the propagation to a maximum distance of two blocks.

I then created visual representations of the disparities between the original data distributions and the new distributions. Red lines indicate neighborhoods where the distribution is more dense in the mental map, while blue lines indicate neighborhoods where the distribution is more dense in the original data. For the city of Atlanta, the greatest disparity between the mental map and the raw data occurs around West Midtown and the intersection between the major highways in the center of the map near downtown (Figure 33). For the city of Baltimore, the largest disparity occurs with high mental map distributions in East Baltimore in the neighborhoods of Oliver, Dunbar-Broadway, and Middle East (Figure 34). Many other neighborhoods, notably

downtown and Fells Point, are lower, though Harlem Park to the west and the stadium area to the south also have higher property crime distributions. Finally, for the city of Chicago, most of the disparities in the mental map exist on the edge between the Loop and South Loop neighborhoods (Figure 35). To the south, neighborhoods along the lake and to the southwest along the highway also have an increase, as well as the northern neighborhoods around Goose Island. For the neighborhoods that are drawn in red, the presence of nodes and edges may indicate that the crime density is higher than what is captured in the official data source, or that inhabitants of those neighborhoods perceive crime as being higher because it occurs close to important nodes in their community. In either case, programs that are designed for reducing the amount of crime in those neighborhoods would be well-advised to include citizens of those neighborhoods in planning sessions.

5.1.5 Discussion

One of the primary benefits of incorporating mental maps is to identify disparities between perception of a quality within the urban landscape and the quantitative data about that quality. For example, members of a community might be fearful because it feels to them that crime is increasing, when the authoritative data of crime distributions indicates no significant change. Recognizing these differences in mental maps is just the first step to understanding how they affect perceptions of quantitative data. In this section, I proposed a novel technique for applying mental maps based on nodes, paths, and edges to spatial data. Finally, I provided preliminary findings from the application of our technique to property crime data in three U.S. cities: Baltimore, Atlanta, and Chicago. This technique provides a method for identifying disparities in mental maps of urban spatial data and provides insight into the effects these models have on the data perception, which is necessary for making decisions that affect all members of a community regardless of their image of the city.

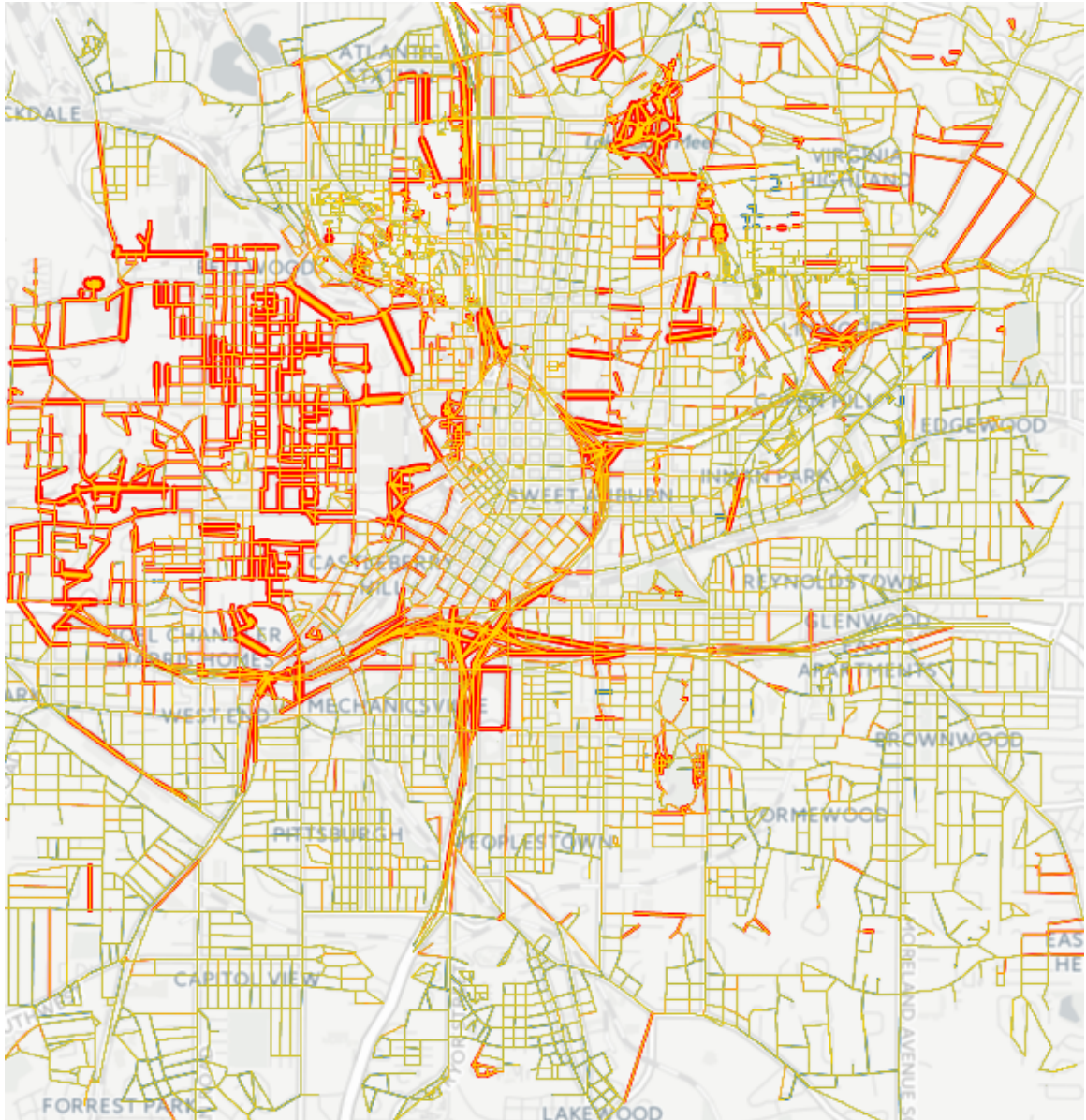


Figure 33: The disparities between the crime map and the church and school mental map. Areas where church and school lixel values are higher are in thick red lines, while areas where the original map is higher are in thick blue lines. The greatest disparity between the mental map and the raw data occurs around West Midtown and the intersection between the major highways in the center of the map near downtown.

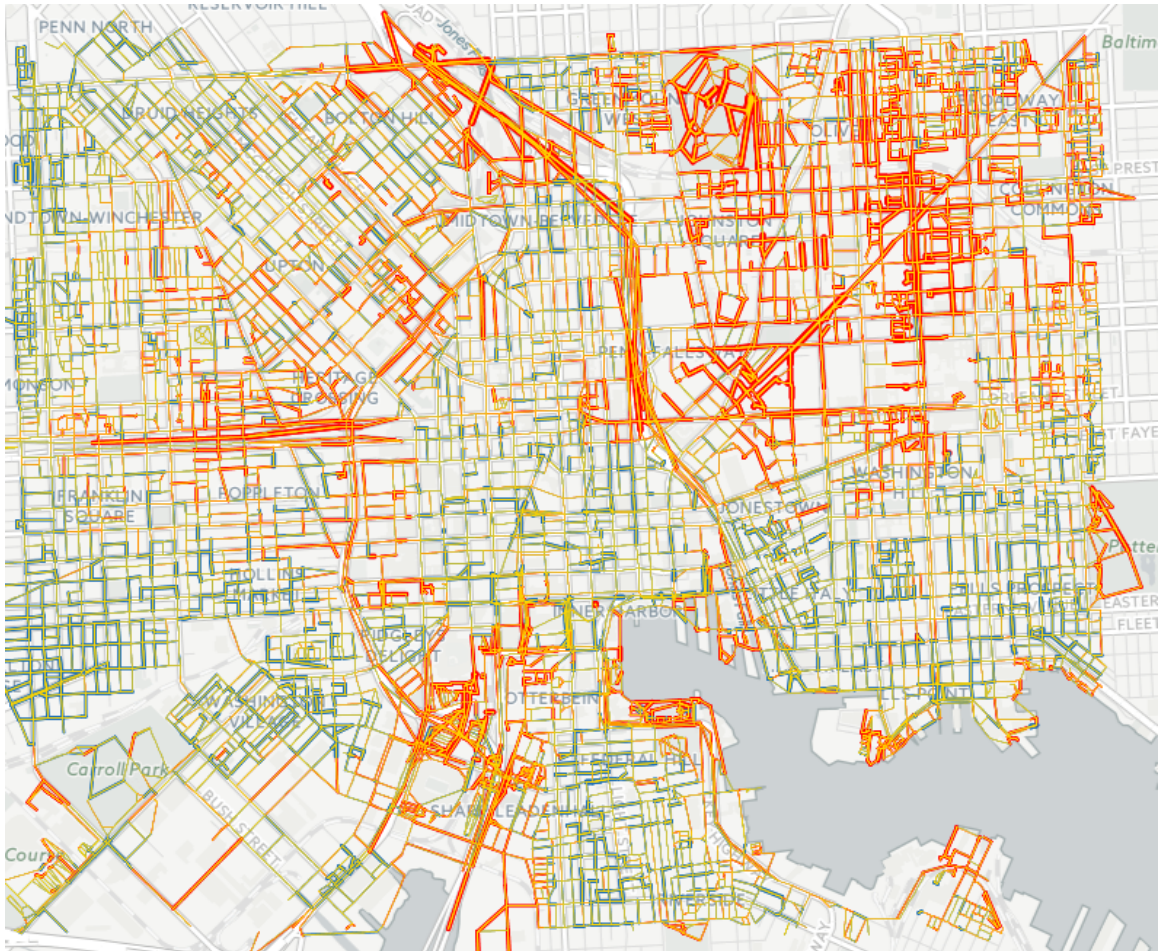


Figure 34: The disparities between the crime map and the church and school mental map. Areas where church and school lixel values are higher are in thick red lines, while areas where the original map is higher are in thick blue lines. the largest disparity occurs with high mental map distributions in East Baltimore in the neighborhoods of Oliver, Dunbar-Broadway, and Middle East. Many other neighborhoods, notably downtown and Fells Point, are lower, though Harlem Park to the west and the stadium area to the south also have higher property crime distributions.

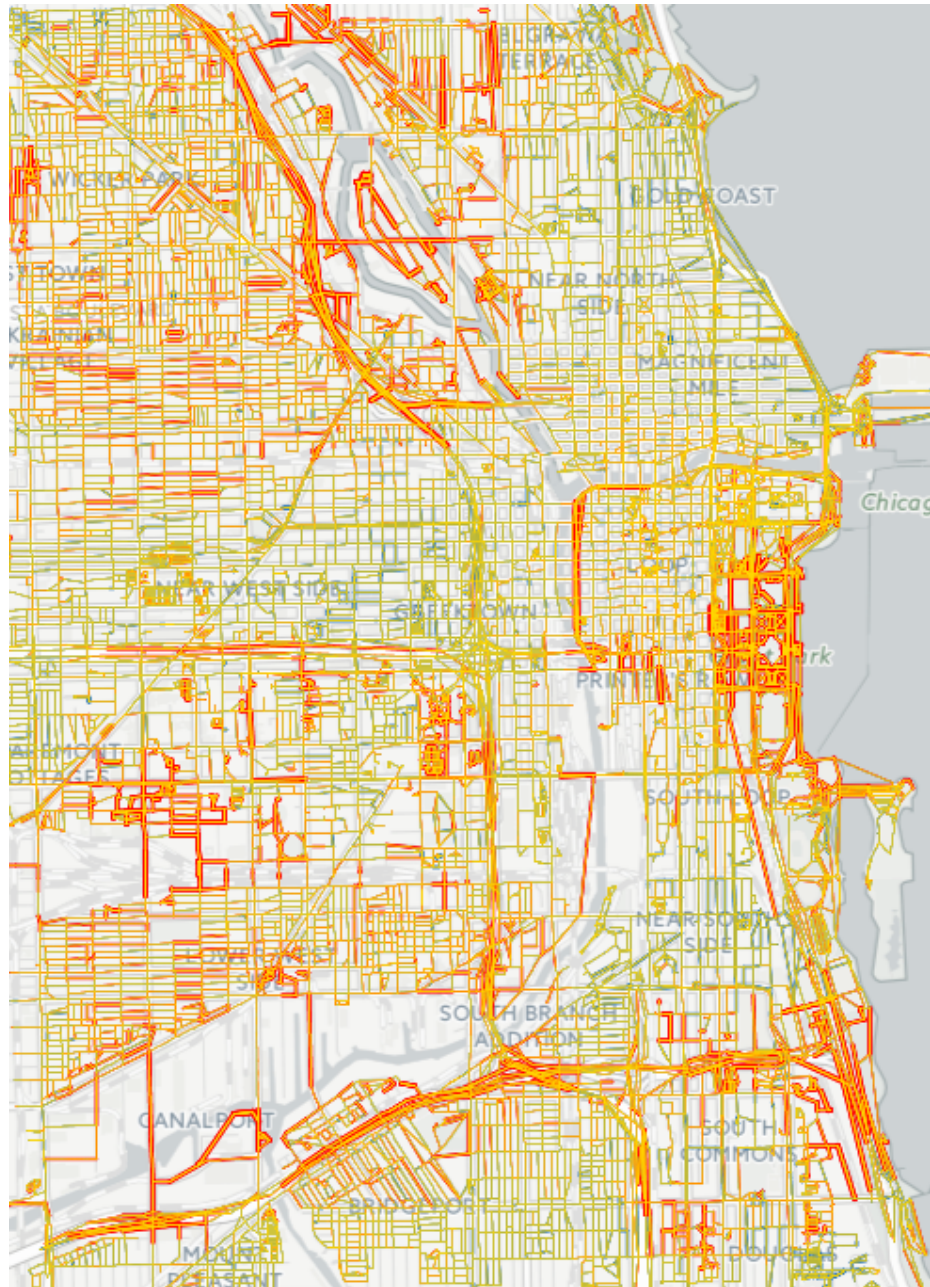


Figure 35: The disparities between the crime map and the church and school mental map. Areas where church and school lixel values are higher are in thick red lines, while areas where the original map is higher are in thick blue lines. For the city of Chicago, most of the disparities in the mental map exist on the edge between the Loop and South Loop neighborhoods. To the south, neighborhoods along the lake and to the southwest along the highway also have an increase, as well as the northern neighborhoods around Goose Island.

5.2 Online Collection and Interpretation of Atlanta Mental Maps

Many of the prototypes used in the examples described in this work were constructed in Java using Processing [43] and several associated rich-client libraries. This approach was originally taken because the tools available within a rich-client implementations allowed for both more expressive capture of sketch and multi-touch interactions and more responsive visualizations of map-based data than were possible in the browser. However, this imposes a significant limitation on the availability of the benefits of the proposed techniques to those capable of running the existing software: namely, people who already possess or can afford touch-screen laptops and tablets that they can install the software onto. An approach was needed to study the same or similar capabilities in a browser so that it they are accessible to as wide an audience as possible. As a thin-client solution, a browser application also makes it easier to determine the external validity of the proposed cognitive mapping techniques.

Improvements to the core d3.js framework [10] as well as the availability of several supporting libraries (e.g., Leaflet [3], Turf [57]) have made the possibility of a web-based sketching and visualization application much more feasible. These components have already been incorporated into a prototype for one of the systems described in this work, HotSketch (see Section 3.2), which is intended for use on tablets and laptops through a browser. The development of HotSketch has shed light on the capabilities that would be necessary for a more open-ended sketch-based analytic application.

One of the most significant requirements for a successful browser application was a backing database for the path networks of roads, highways, and streets. Neo4j was a potential solution, as it allows a graph structure to be accessed through a REST API, HTTP API, or through plugins for javascript and node.js. Ultimately, an approach based on smaller and more localized neighborhood street networks in geojson and PostgreSQL was determined to be more efficient and more easily implemented. This

also has the benefit of being easier to deploy and maintain by community organizations. Heroku was chosen as a hosted server for instances of each study prototype, allowing for wide availability at minimal cost. The data used to populate these prototypes was obtained from OpenStreetMap through custom but easily modifiable scripts. As the intent would be for multiple cities to be accessible through the application, the system could eventually scale to keep the individual path networks and associated data separate and allow the user to select the city of interest through the user interface of the application, but for the studies described subsequently the focus is primarily on the data around the core of Atlanta.

A secondary requirement was for the storage of spatial data sets consisting of location records and events. Most of the previous research conducted to date has been focused on crime events, which could be included within a PostgreSQL instance at larger scales but were deployed here through geojson. As the data is updated, however, updates would need to be made to the dataset. The city of Atlanta, for example, updates the Federal UCR data posted publicly every two weeks (or in the words of the officer who manages the data, “whenever I get around to it”). A Cron job or similar structure would be helpful to track the data posted to the Atlanta Police public site and pull the most recent data when it is made available. In addition, several server-side scripts could be developed to detect changes to previously uploaded data or the data model itself, which have both occurred with the Atlanta data as new information on previous crimes has come to light or new UCR formatting requirements have been handed down.

Finally, it was beneficial to store the elements of the cognitive models sketched by each user as geojson feature collections within the PostgreSQL instance. This way, it would be possible for an individual user to construct a set of important nodes, paths, edges, and districts and store that model within the system for future use as the spatial data set is updated over time. The elements themselves can be

stored as map features (e.g., points, lines, polygons) within a PostgreSQL table for review in subsequent studies. The basic capability supporting the matching of a table to a user was accomplished through stored cookies in the browser, but a more robust capability of login authentication may be necessary to allow for users who are borrowing computers or using them temporarily (e.g., in a public library) to maintain access to developed models across sessions.

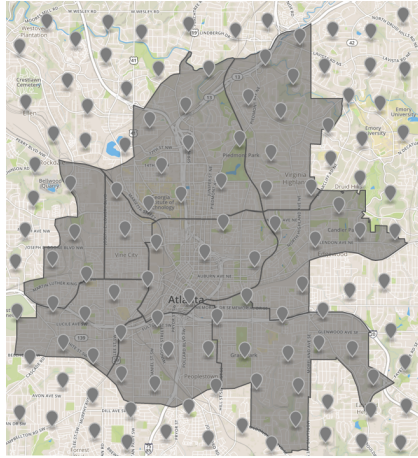
But even with this infrastructure in place—questions remain. How can interactive visualization improve current methods for capturing mental maps at the neighborhood and city levels? For the city level, what interaction methods best facilitate the creation of desirability contours? For the neighborhood level, what interaction methods best facilitate the creation of nodes, paths, and edges? What separates a “good” mental map from a “bad” one? Does heightened interactivity and visualization improve the fidelity of these mental maps, or do they inhibit accurate capture? To answer these questions, I conducted a study with participants drawn from the area around midtown and downtown Atlanta.

5.2.1 Motivation

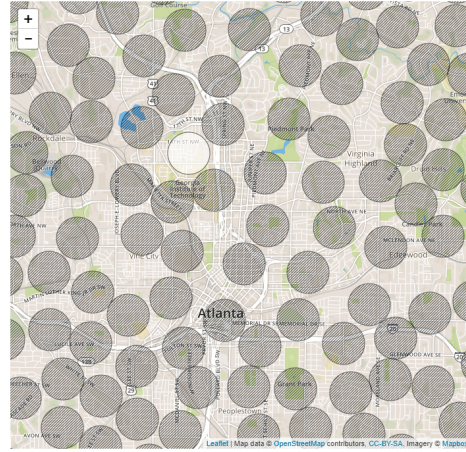
What is the ideal interaction strategy for rapidly and accurately specifying mental maps of spatial data? This study sought to determine how interactive visualization can improve current methods for capturing mental maps at the neighborhood and city levels. Currently, it is unclear what the optimum interaction paradigm is for capturing mental maps. For example, when users create maps of the city in studies based on the work of Lynch [75] and Appleyard [8], they use blank sheets of paper to rapidly construct off-the-cuff models of the city that best represent the elements that matter the most to them. Then, a group of researchers typically codes the elements that are present in a drawing, which is confirmed by additional reviewers if possible. Is the best canvas for specifying a mental map of the city a blank page, or it more

effective to provide a simple ranking interface similar to the maps provided by Gould and White [52]? Does the presence of dynamic feedback in the system alter the effectiveness of each approach?

Through the following online study, I sought to determine the differences in interaction strategies for rapidly and accurately specifying mental maps. Two groups were assigned from a pool of everyday citizens that live in the nine central neighborhood planning units (NPU) of Atlanta. The Simpler Interaction Group utilized representation and interaction techniques that closely mirrored the techniques of Lynch, Appleyard, Gould, and White for eliciting mental maps on paper. These techniques were intended to be faster, easier to understand, easier to implement, and have less potential for error. The Complex Interaction Group utilized representation techniques that were more expressive and provided nuanced feedback to users in the form of dynamic modifications to the map representation. These techniques benefited more from the digital format of the online study, but were more complex to explain and carry out than the more straightforward Simple versions of tasks. All participants were asked to indicate which neighborhood they live in, and which neighborhoods they feel the most fear from crime and which are the most expensive. This is to establish a baseline of indicators that can be correlated both with the stated desirability of the neighborhoods and the actual data available on crime statistics and housing prices. People then participated in two tasks. The first was designed to mimic the larger-scale desirability studies of Gould and White by eliciting broad preferences across the entire city [52]. The second was designed to mimic the smaller-scale imageability studies of Lynch and countless others by challenging participants to remember and record the features that they feel are most important in characterizing the city [75].



(a) Random samples in the context of central Atlanta NPUs.



(b) Areas enclosed by each sample location.

Figure 36: Random samples used to group participants to the closest “home” neighborhood and to elicit rankings of the areas in the city of Atlanta. After seeding each NPU with a single sample point at its centroid, sample point locations were generated using Bridson’s method [15] with a minimum spacing of 1km. In the home selection interface, a textured circle, rather than a point marker, is used to convey that the area nearest to the center of the circle is also included in that region.

5.2.2 Method

I created 118 sample points around central Atlanta spaced at 1km apart (Figure 36a). Participants were drawn from current residents of the nine neighborhood planning units (NPUs) around downtown and midtown Atlanta. The study was advertised through email lists, Craigslist, NextDoor, Facebook, Twitter, and through fliers in NPU meetings. People participated in the study through a browser by following the link in the advertisements. Participants certified that they had lived at an address within one of the NPU boundaries containing the sample points for at least six months.

I asked all participants which point is the closest to where they currently live, but did not ask for their exact address of residence or distance to that point (Figure 36b). Participants were divided into two groups: participants who use a simpler capture method (Group A: *Simple*) or a complex method with dynamic information visualization feedback to interactions (Group B: *Complex*). Participants in both groups participated in three tasks. Participants were counterbalanced by assigning them to

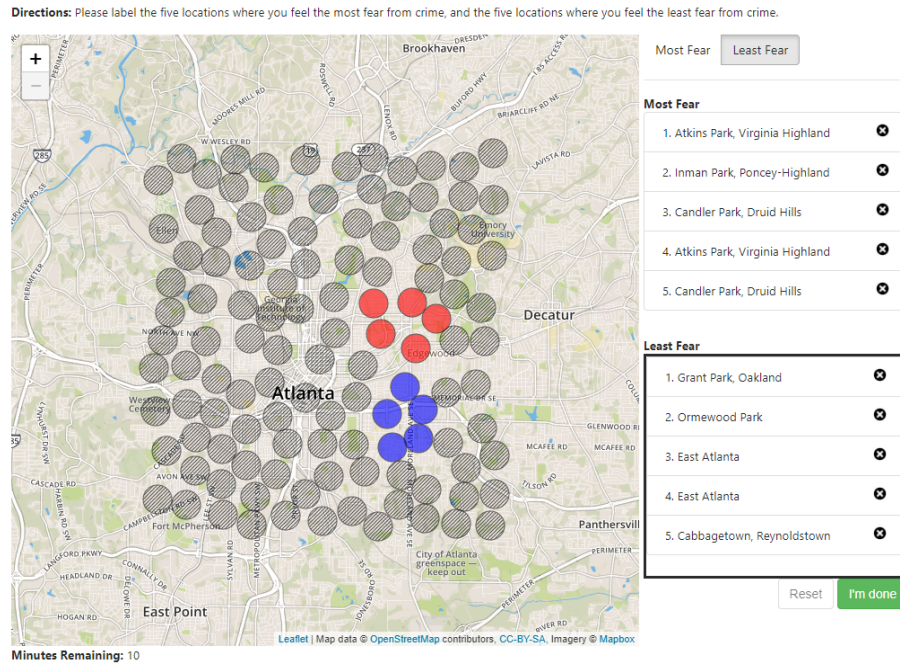


Figure 37: All participants were prompted: *Of all of the neighborhoods in Atlanta, which five do you feel the most fear from crime? Which five do you feel the least fear from crime?* By clicking on neighborhoods in the map, locations were added to reorderable lists. Red markers are ranked high, blue markers are ranked low. A similar interface was used to elicit responses to the question: *Of all of the neighborhoods in Atlanta, which five do you feel is the most expensive to live in? Which five do you feel is the least expensive to live in?*

the group (A or B) that currently had the fewest number of participants within the NPU that their neighborhood was located in.

In the initial task, all participants in both groups were given a map with the labeled sample circled areas and asked to click the five areas where they feel the most fear from crime, and the five sample areas where they fear the least fear from crime. Participants were also asked to click the five sample areas where they feel homes are the most expensive, and the five sample areas where the homes are the least expensive (Figure 37). I collected this data to facilitate direct comparison between the individual rankings for crime and price as component features of the overall desirability for each neighborhood. This data also allowed us to directly compare the perceived rankings for these specific attributes to publicly available quantitative statistics on crime and housing price throughout the area.

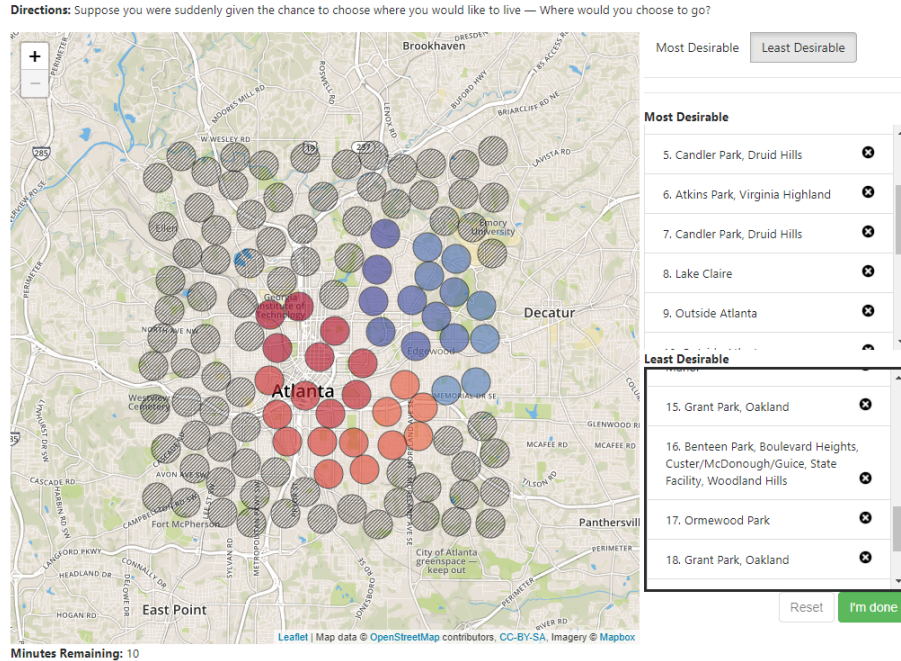
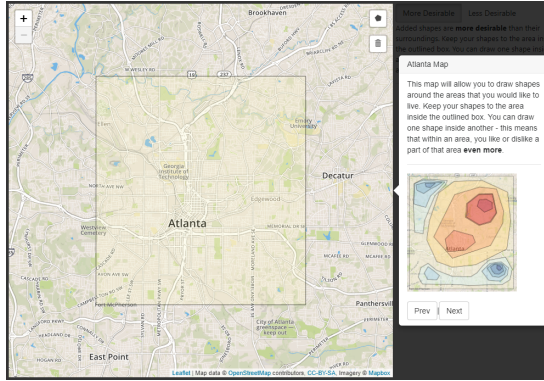


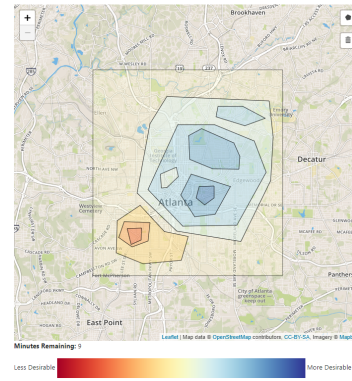
Figure 38: Participants in the Simple Group were prompted: *Suppose you were suddenly given the chance to choose where you would like to live- an entirely free choice that you could make quite independently of the usual constraints of income or job availability. Where would you choose to go?* By clicking on neighborhoods in the map, locations were added to reorderable lists. Blue markers are ranked highly desirable, red markers are ranked less desirable. The task is a digital version of the map ranking task used by Gould and White [52].

In the next task, all participants in both groups were prompted with the following question: "Suppose you were suddenly given the chance to choose where you would like to live- an entirely free choice that you could make quite independently of the usual constraints of income or job availability. Where would you choose to go?" Participants had ten minutes to provide a response through the mechanism specific to their group.

Simple Group Participants clicked on neighborhood circle areas to rank them from highest to lowest desirability, which appeared in a sorted reorderable list in the interface. At any time, participants could switch from labeling most desirable points to least desirable, and back. Participants had to choose at least ten points to continue. Locations were colored using the d3 "RdYlBu" diverging color scheme, with



(a) Participants in the Complex Group were asked to draw non-intersecting shapes to represent their desire to live in locations around Atlanta.



(b) Shapes were given a hue to represent their desirability as new shapes were added to the map.

Figure 39: Complex group participants were prompted: *Suppose you were suddenly given the chance to choose where you would like to live- an entirely free choice that you could make quite independently of the usual constraints of income or job availability. Where would you choose to go?* The task is a dynamic version of the blank sketching process used by Lynch and Appleyard to elicit imageability from participants on paper maps [75, 8], and uses the digital interface to dynamically adjust the color in response to the user’s sketched lines.

blue locations representing more desirable locations and red locations representing less desirable ones (Figure 38).

Complex Group Participants were directed to draw non-intersecting closed shapes around areas on the map and indicated through a button whether the shape was *more* or *less* desirable than the area around it. They were able to drag contour lines around once constructed to edit the visualization dynamically. As closed shapes were completed, the colors assigned to the interiors of regions were updated to reflect the current ordinal desirability ranking of regions.

Simple Group participants received the benefit of familiarity with the selection interface for this task, as it was nearly identical to the selection mechanism used for the ranking of crime and housing prices. The disadvantage was that the Simple group was also unable to express more nuanced preferences in the boundaries between the predetermined sample locations. Complex group participants benefited from this expressiveness, and were able to provide a higher-resolution depiction of their perceived

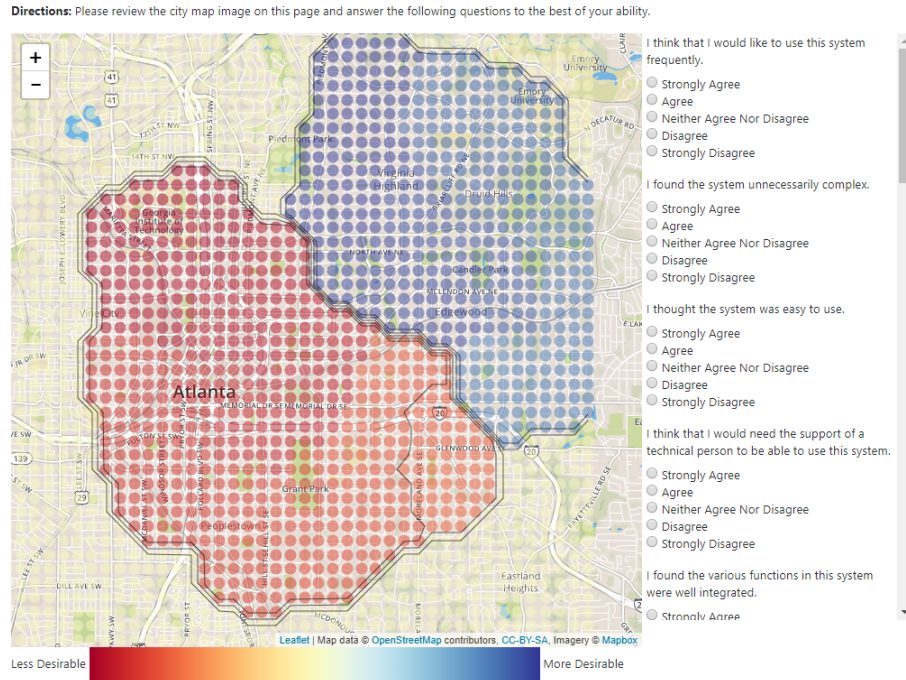


Figure 40: Participants in the simple group were shown this style of map moments after expressing the desirability of the neighborhoods in the city, which has been constructed using a regularly spaced grid of sample locations that are assigned a hue based on proximity to the nearest neighborhood ranked by the participant. For consistency with the Complex group, isolines have been drawn to separate major groups by their desirability.

desirability within the region. One drawback of the study execution, however, was that simple group participants were encouraged to rank no fewer than ten neighborhoods and were encouraged to rank more. Complex group participants were allowed to proceed after drawing even one closed shape for either high or low desirability, which may have negatively affected the level of detail in participant responses by reducing the number of shapes.

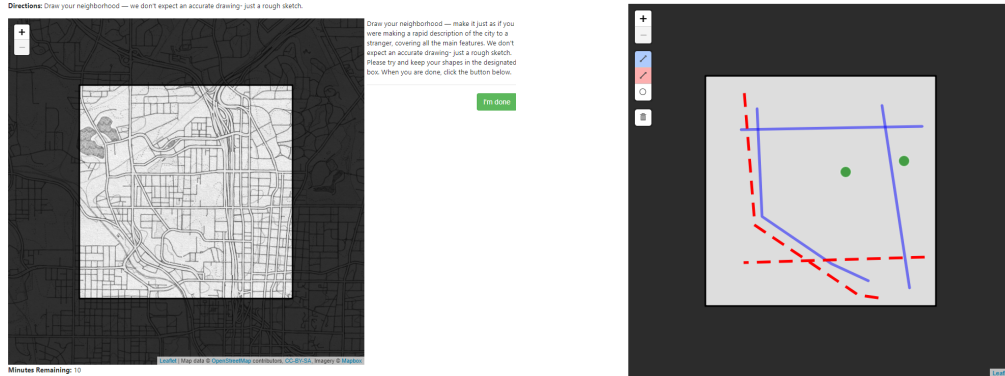
Participants were then shown a digital representation of the mental map constructed from their interaction with the interface in the previous task (Figure 40). They were prompted with: *Please review the city map image on this page. This image has been constructed based on your responses to the previous tasks. For the task you participated in and the image displayed, please answer the following questions to the best of your ability.* They were asked to complete a short survey to establish how

well the representation matches their internal mental map (Section 5.2.3). They were also asked questions about how they believe someone else would interpret the mental map. The survey consisted of ordinal Likert questions drawn from the System Usability Scale [16] and free responses specific to this study.

For participants in the Simple Group, the representation method differed somewhat from the image that was generated as the participant interacted with the system. This was done to more directly mimic the separation between the transcription of a paper-based mental map and the interpretation by a secondary expert user. While the new representation is not drastically different from the desirability map encoded by the participant, it is clear that additional steps of analysis have been performed without explanation. The resulting representation has been constructed through the use of a regularly spaced sample grid of points that are distributed evenly throughout the map area. Each sample point is assigned the ranking of the nearest neighborhood ranked by the user in the previous task. Isolines have been added to the map to separate major breaks between rank bins.

In the second task, participants were prompted to create a sketch of the 2km radius area around the point closest to where they live with the following question: *“Make it just as if you were making a rapid description of the city to a stranger, covering all the main features. We don’t expect an accurate drawing- just a rough sketch.”* Participants had ten minutes to provide a response through the mechanism specific to their group.

Simple Group The simple sketch group was primed briefly for 10 seconds with a map of the area they were to draw in order to get oriented, then given a blank canvas upon which to sketch their mental map of the area around their home sample point. They were able to choose one of four drawing tools: blue line, a circle marker tool, red dashed line, or an eraser. Blue lines represented roads or other paths, circles



(a) Participants in the Simple Group were primed for 10 seconds with an unlabeled image of the map tiles around the neighborhood they selected.

(b) After 10 seconds, tiles were hidden and participants sketched the important features of their neighborhood.

Figure 41: Participants in the Simple Group sketch the important features of their neighborhood. The task is a digital version of the blank sketching process used by Lynch and Appleyard to elicit imageability from participants on paper maps [75, 8].

represented important nodes, and dashed lines represented edges. After drawing any feature, participants were able to type in the name of the feature.

Complex Group The Complex sketch group were given an unlabeled basemap that indicates the location of blocks and major geographic features but few details and no text. They were directed to draw important paths, nodes, and edges directly through three separate drawing techniques. After blue lines were drawn on the map, the road network is queried for paths that are close to and reachable from the drawn path. Circle markers could be placed to represent nodes, and were rendered with a surrounding "area of effect" that amplified the distance of reachable paths from the circle and paths already on the map. Drawn dashed lines were interpreted as edges, and dampened the reachability of nearby paths placed on the maps. As people added features, the map updated to show the changes in the underlying network model through line thickness and hue.

Participants were then shown a digital representation of the mental map constructed from their interaction with the interface. Much like in the previous task,

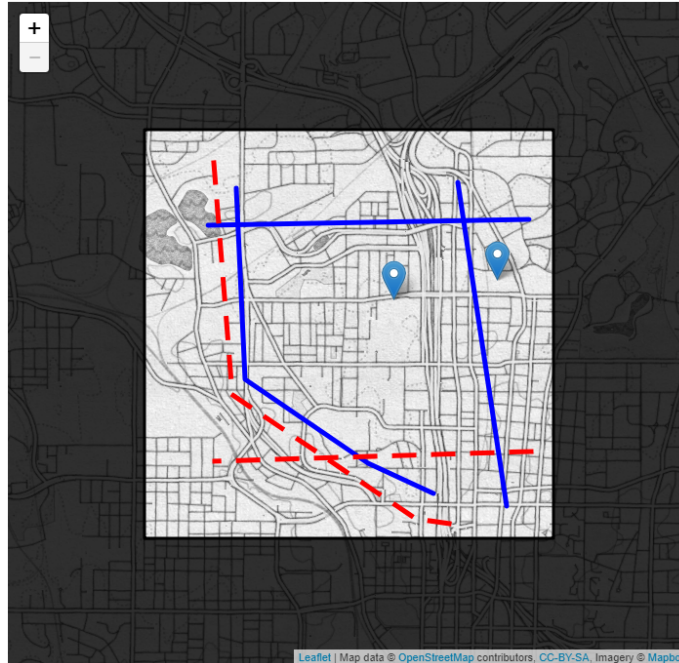


Figure 42: Participants in the Simple Group were shown this style of map after sketching the important features of their neighborhood. It depicts the features that they drew in the task but reveals the underlying map tiles for context.

they were asked to complete a short survey to establish how well the representation matches their internal mental map. They were also asked questions about how they believe someone else would interpret the mental map (Section 5.2.3). The survey consisted of ordinal Likert questions drawn from the System Usability Scale [16] and free responses specific to this study.

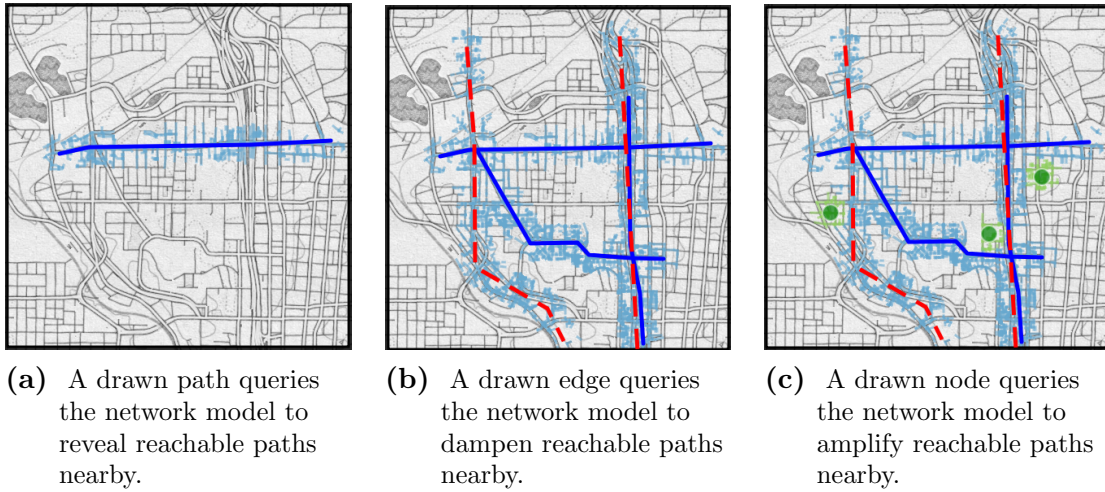


Figure 43: Participants in the Complex Group sketch the important features of their neighborhood. The task is a digital version of the blank sketching process used by Lynch and Appleyard to elicit imageability from participants on paper maps [75, 8], enhanced with dynamic feedback from interacting with the reachability model of connected streets.

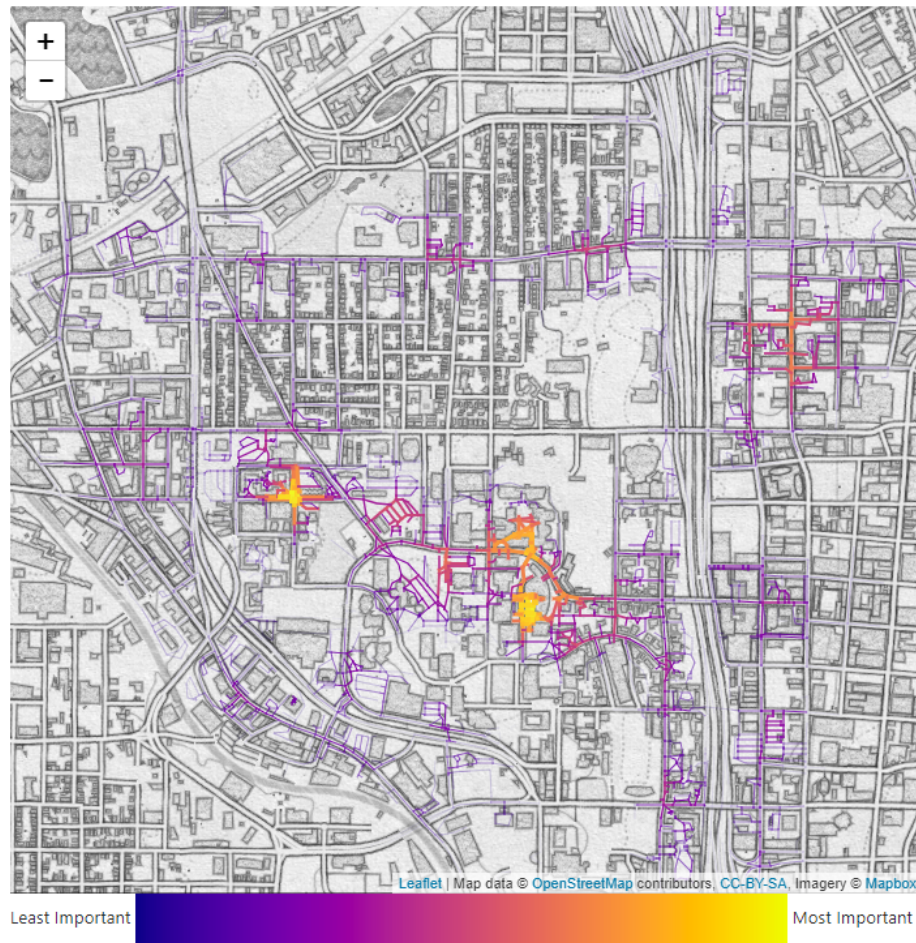


Figure 44: Participants in the Complex Group were shown this style of map after sketching the important features of their local neighborhood. The sketched features have been applied to a network reachability model in which edges dampen the reachability of nearby lixels and nodes amplify nearby lixels. Lixels that are close to any feature sketched by a user are considered *source lixels*, and are assigned a score of 1. In the generated representation, the network KDE score is encoded in the hue and thickness of the lines overlaid on the map.

5.2.3 Questionnaire

These are the questions that were posed to participants at the end of the desirability and feature sketching tasks. The first ten items are based on the System Usability Scale (SUS) questions [16]. Question 11 is designed to determine the perceived accuracy of the created image, and is worded slightly differently for the feature sketching task. The final five questions are designed to elicit qualitative data that can provide context for the rankings and sketched features transcribed during each task.

1. I think that I would like to use this system frequently. (agreement-Likert)
2. I found the system unnecessarily complex. (agreement-Likert)
3. I thought the system was easy to use. (agreement-Likert)
4. I think that I would need the support of a technical person to be able to use this system. (agreement-Likert)
5. I found the various functions in this system were well integrated. (agreement-Likert)
6. I thought there was too much inconsistency in this system. (agreement-Likert)
7. I would imagine that most people would learn to use this system very quickly. (agreement-Likert)
8. I found the system very cumbersome to use. (agreement-Likert)
9. I felt very confident using the system. (agreement-Likert)
10. I needed to learn a lot of things before I could get going with this system. (agreement-Likert)
11. The city image accurately represents my desire to live in the various neighborhoods of the city. (agreement-Likert)

12. Please describe, in your own words, what characteristics are shared by the areas that you **DO** desire to live in? (free response)
13. Please describe, in your own words, what characteristics are shared by the areas that you **DO NOT** desire to live in? (free response)
14. If a city official were to review this map, what conclusions do you think they would reach about the city? (free response)
15. If a city official were to review this map, what conclusions do you think they would reach about the person who created it? (free response)
16. If you were to use the created image to argue for one policy change about the city, what would it be? (free response)

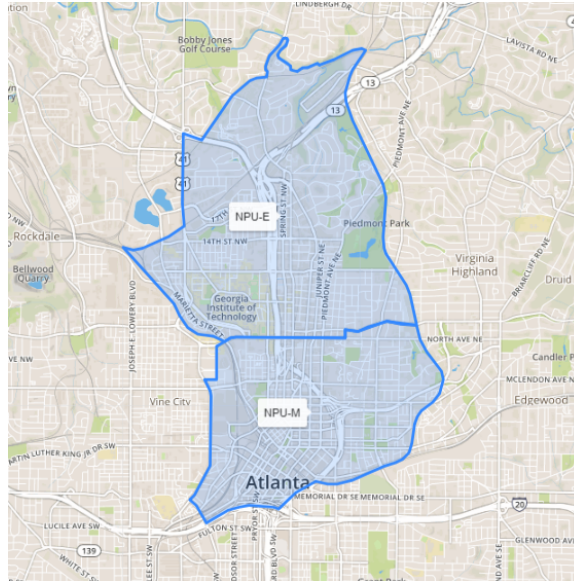


Figure 45: Neighborhood planning units E (midtown) and M (downtown) were the most common home NPUs for participants in the study (NPU-E : 56; NPU-M : 18).

5.2.4 Analysis

Overall, 86 people took part in our study and were relatively balanced across sex (44 female, 40 male, 2 no response). The median age of participants was 27.5 (mean = 31.73, standard deviation = 12.68). The responses were markedly less balanced in terms of race (see Table 2). Participants tended to be white or Asian, which is more reflective of the demographic makeup of the university campus than its surrounding population in the sampled area. Of the nine neighborhood planning units (NPUs) in the survey region, the two most heavily represented are NPU-M (downtown) and NPU-E (midtown) (Figure 45).

However, it should be noted that not all participants completed all tasks; though most participants completed the tasks on reporting fear from crime and perceived housing price few of the participants completed the subsequent tasks. Of the 86 participants who completed any task in the study, 60 participants persisted long enough to complete the questionnaire about the usability of the desirability interface and 44 completed all tasks leading up to and including the final questionnaire regarding the

Table 2: Participant Demographics

Race	Female	Male	No Response	Total
Asian	11	11		22
Black	1	1		2
Hispanic	1	1		2
No Response			1	1
Other	3	2		5
White	28	25	1	54
Total	44	40	2	86

system usability of the map feature sketching interface. Part of the attrition can certainly be attributed to the method of compensation for research participants, a raffle, which under Georgia Law is available to anyone for entry regardless of the degree to which they complete the study.

The responses from the completed SUS questions were aggregated within each group following the methods of Brooke [16] to yield individual SUS scores. A between-groups t-test was performed to determine if a significance difference existed between the perceived system usability of the desirability interface between the simple group (n=33) and the Complex group (n=27). The results of the t-test indicate that there is not a significant difference between the two interaction methods for the desirability task ($f=1.5$, $p=0.135$) which is confirmed by an examination of the confidence intervals (-1.84, 13.18). T-tests of the perceived system accuracy for both the desirability and map sketching interfaces similarly did not indicate a significant difference.

However, a between-groups t-test was also performed to determine if a significant difference existed between the perceived system usability of the map sketching interface between the simple group (n=26) and the Complex group (n=18). The results of the t-test indicate that there is a **significant difference** between the two interaction methods for the sketching task ($f=-2.46$, $p=0.018$) which is confirmed by an examination of the confidence intervals (-26.99, -2.66). In this case, participants felt that the Complex interface was significantly more usable than participants who were using

the simple interface. Realistically, much of this difference can be attributed to the requirement for participants in the simple group to sketch from memory, a demand that was not required of the participants in the Complex sketching group.

To analyze the constructed mental maps in more detail, I used the data to construct aggregated maps for the overall study population, participants from only NPUs E, and only from M. I collected data from the Atlanta Police Department on crime in the city for 2017, and bolstered it with crime data collected from the Georgia Tech Police Department. I also collected data from the Zillow API for 10,757 for-sale properties across the 13 zip codes represented in the study region. By obtaining data from these secondary sources, I am able to make direct comparisons between the perceptions of the neighborhoods and publicly available quantitative measures of crime and median housing listing price.

For all the residents of Atlanta, this meant a marked shift between the areas of perceived crime and the crime locations (Figure 46). From the dots used to mark perception, it is quite clear that the overall and NPU-E perspective is that a line runs from the northwest of the map to the southeast and that this barrier divides a region of fear from one of relatively safety.

For NPU-M, however, this perception is quite different (Figure 47). Unlike the aggregate view and perception of NPU-E residents, people in NPU-M do not rank the southwest of Atlanta highly as an area that they feel fear from crime. There is apparent clustering still in place, however, with large groupings of perceived safety towards the northeast and fear to the west and immediate south, but the distribution is very different than that found in NPU-E. An interesting aspect of this comparison is that the inhabitants of NPU-E see one location as particularly fearful, while participants in NPU-M perceive fear strongly from several locations in the neighborhood. This disparity deepens if we examine the correlation between the perceived fear and the desirability reported by people from NPU-E (Figure 48). In this figure, we can see

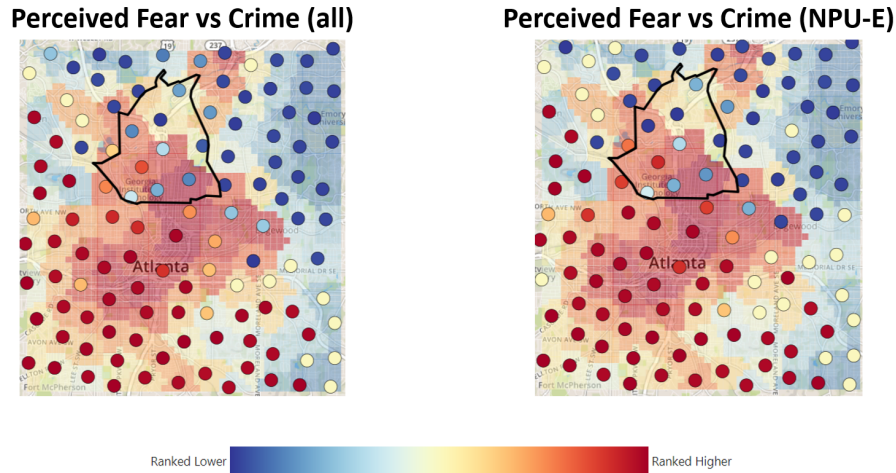


Figure 46: Perception of Crime (dots) vs Crime Locations (background). On the left are perceptions obtained from participants in all neighborhoods, on the right are perceptions obtained only from participants in NPU-E. The foreground dot hue pattern depicting a separating line of fear is in contrast to the background hue pattern depicting crime central to midtown and west of downtown.

that NPU-E resident largely desire to push east and away from the NPU, in particular the neighborhood perceived as high in fear. For residents of NPU-M, however, we see a completely different response despite the perceived fear: participants report a high ranked desire to live in the NPU (Figure 49). While it's possible that the perception of NPU-E residents can be attributed to a high number of student responses and the Clery Safety Act reports of crime in the feared neighborhood, it does little to explain why residents of NPU-M have such a different reaction.

When I examine the aggregated mental maps obtained from the sketching exercise, it becomes immediately clear that the outcome from the Complex sketching exercise is far more legible. The disadvantage is that it takes much longer to compute, requiring more than ten minutes of processing on a moderately powerful laptop and generating enough data to take several seconds just to load into the browser even after processing. Meanwhile, the simple map is computed on the fly and added to the map instantly. With the right browsing functionality to step through and/or hide individual responses, it could prove quite powerful. The aggregated Complex

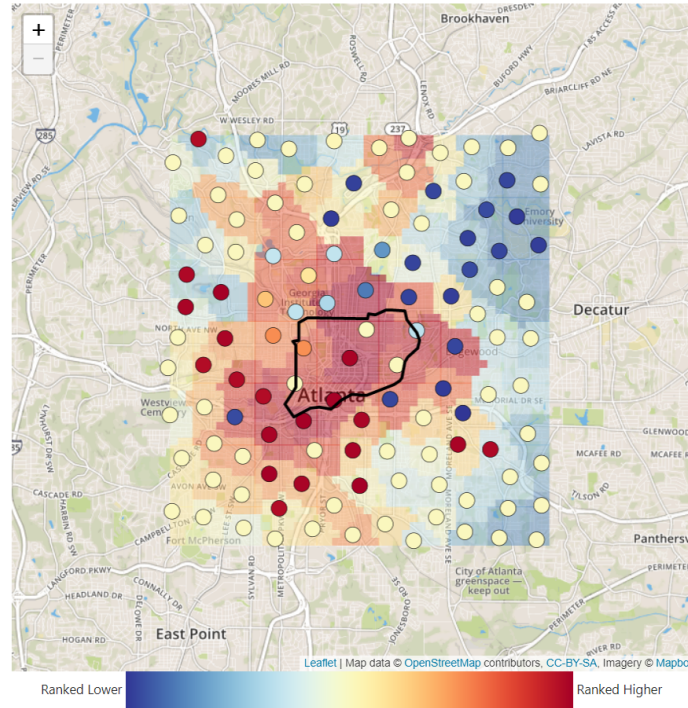
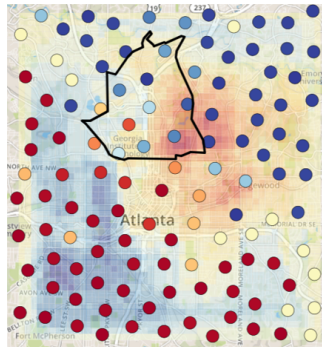


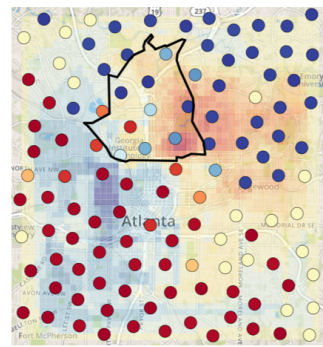
Figure 47: Perception of Crime (dots) vs Crime Locations (background) obtained only from participants in NPU-M. The foreground dot hue pattern is unlike the dot pattern for NPU-E, and is only focused on a small portion of the area around the west of downtown. Notably, a small portion of that area on the west of the NPU is seen as ranked quite low for fear from crime.

map is derived slowly by initially assigning any lixel that is present in the sketched mental maps of participants from that region as a source lixel, then increasing that score for each separate participant that included that lixel. Similarly, the aggregate amplification and dampening values for multiple nodes and edges are aggregated. When complete, the result is a thoroughly detailed map of the city that emphasizes the area that participants thought were most important about the area they live in (Figure 51). In contrast, the aggregated image from the simple features is difficult to read in detail but conveys an overall sense of relative importance within the southeast and east of the region, a somewhat general conclusion that is nonetheless borne out by the desirability maps.

Perceived Fear vs Desire (all)



Perceived Fear vs Desire (NPU-E)



Ranked Lower  Ranked Higher

Figure 48: Perception of Crime (dots) vs Desirability (background). On the left are perceptions obtained from participants in all neighborhoods, on the right are perceptions obtained only from participants in NPU-E. Notably, while NPU-E participants do not perceive most of NPU-E as fearful, they also do not desire strongly to live there, instead showing a preference for the eastern edge of the region and beyond.

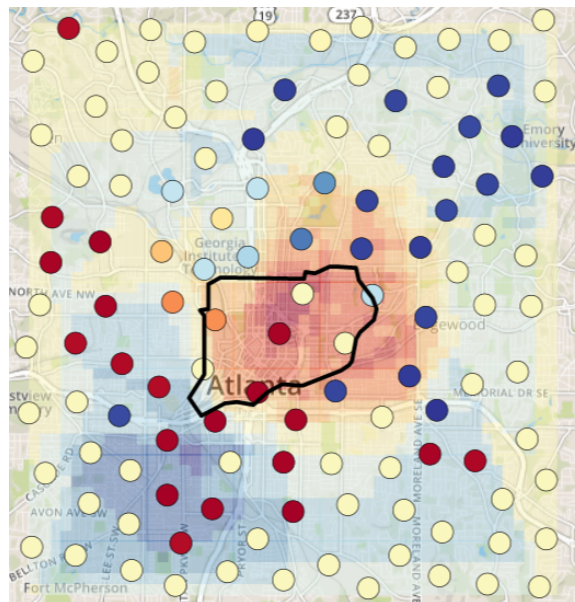
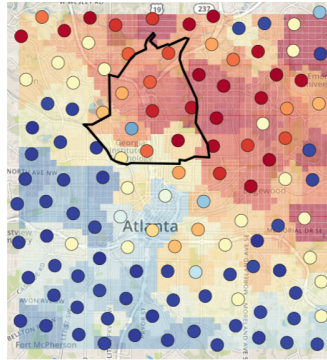


Figure 49: Perception of Crime (dots) vs Desirability(background) obtained only from participants in NPU-M. Unlike the residents of NPU-E, the perceived fear from crime at neighborhoods within the NPU does not dissuade residents from desiring to live there.

Perceived Price vs Price (all)



Perceived Price vs Price (NPU-E)

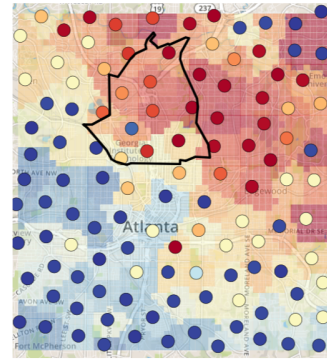


Figure 50: Perception of Price (dots) vs Median House Listing Price (background). On the left are perceptions obtained from participants in all neighborhoods, on the right are perceptions obtained only from participants in NPU-E. Notably, NPU-E participants are fairly accurate at predicting home prices, save for the Home Park area north of campus where perception of crime is similarly different from the available data and the northeast portion of the map which is much more expensive than participants of all groups perceived.

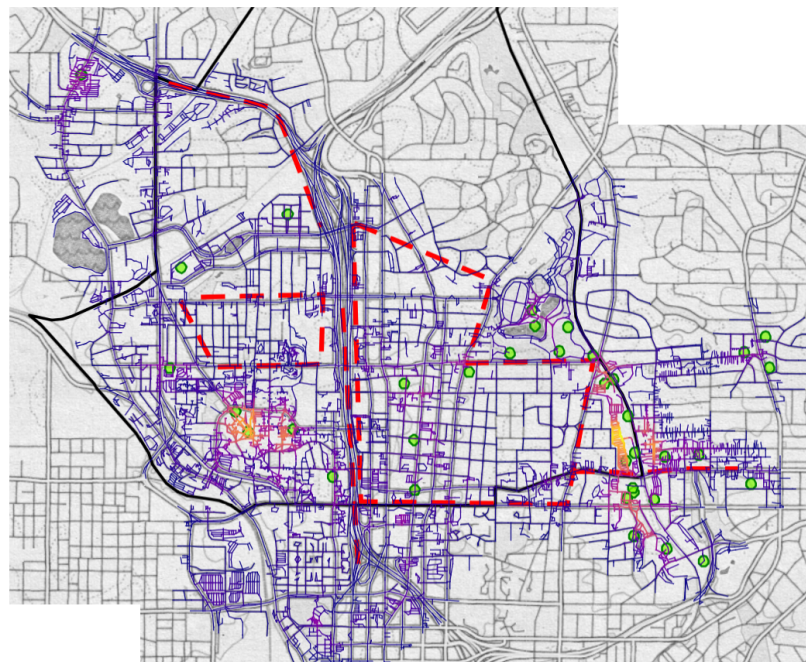


Figure 51: Aggregated mental map features of NPU-E obtained from residents of that neighborhood during a Complex sketching task.

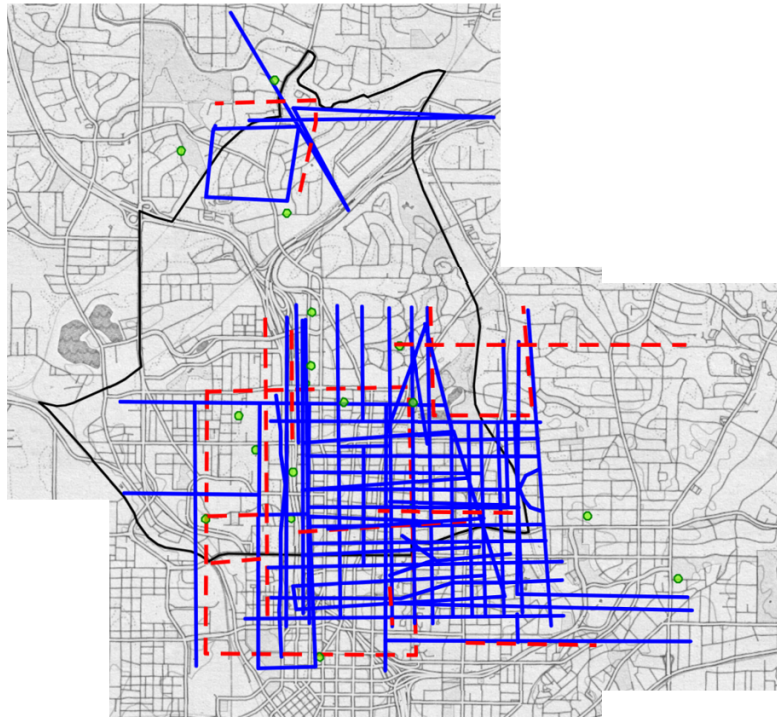


Figure 52: Aggregated mental map features of NPU-E obtained from residents of that neighborhood during a simple sketching task.

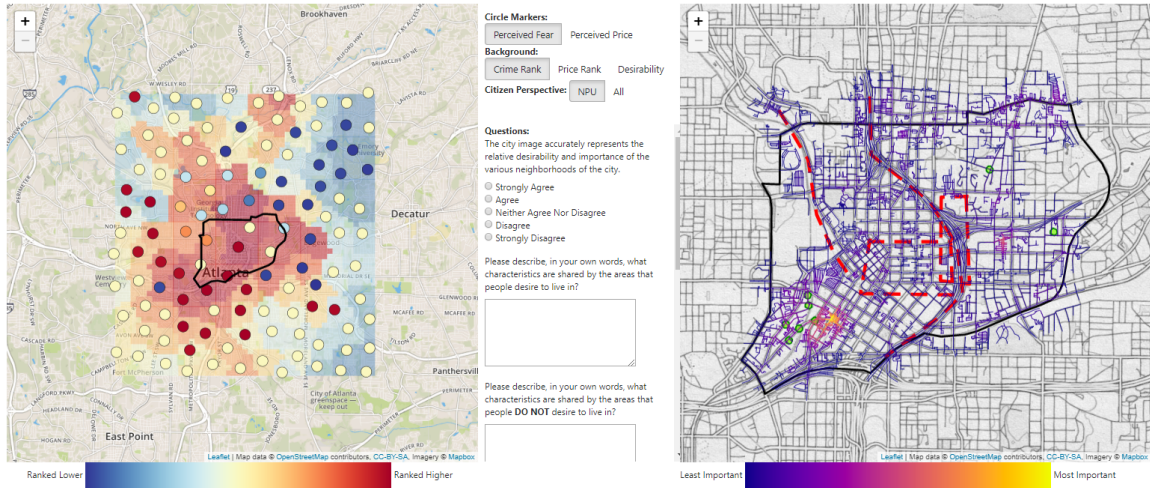


Figure 53: Interface for analyzing the interpretation of mental maps. In this image, the image on the left is aggregated from user reported perceptions of crime, housing price, and Complex desirability as a background layer. The representation on the right depicts the aggregated network features from the Complex group for each of the individual neighborhoods that make up make up the NPU. Controls in the center can be used to select other hues for the foreground or background layers on the left.

5.3 Interpretation of Mental Maps

How can interactive visualizations improve interpretability of mental maps? Is a mental map interpretable by a person who did not draw it? Do the opinions of the sketcher transfer to the reader? Does the representation of the mental map affect this process? How can interaction methods support the interpretation of a mental map?

5.3.1 Method

Participants were drawn from current civic leaders, advertised through Neighborhood Planning Unit (NPU) meetings and through direct mailings to the NPUs and local civic organizations. People participated in the study through a browser by following the link in the advertisements. Participants reviewed anonymous data collected from participants in the previous study (see Section 5.2), analyzed that data, and indicated their findings. Participants were divided into groups based upon the NPU they represent, which affected the data that was used to generate the representations that they interacted with. People participated in two activities (all participants will

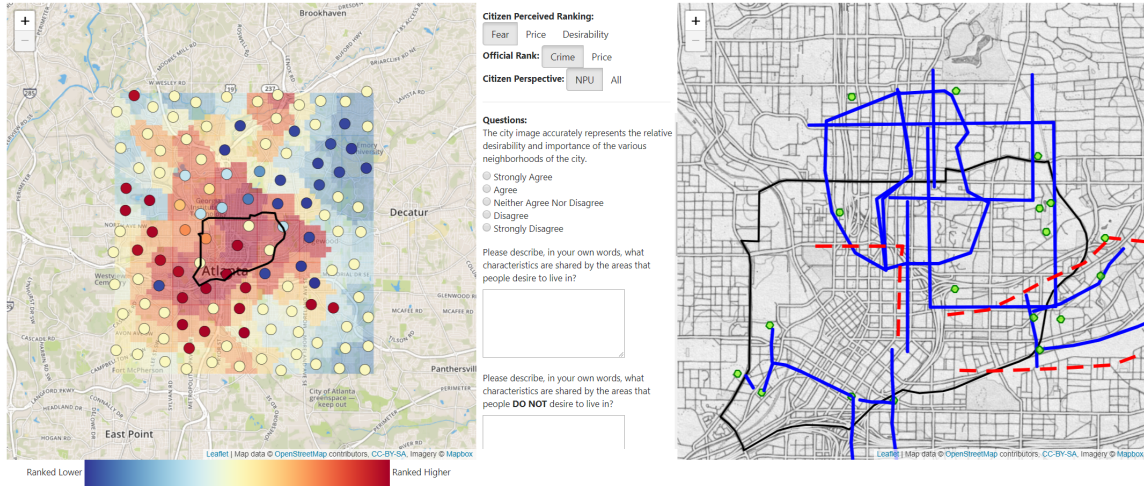


Figure 54: Interface for analyzing the interpretation of mental maps. In this image, the image on the left is aggregated from user reported perceptions of crime, housing price, and simple desirability as a foreground hue for the dots. The representation on the right depicts the aggregated network features from the simple group for each of the individual neighborhoods that make up the NPU. Controls in the center can be used to select other hues for the foreground or background layers on the left.

participate in both, counterbalanced for order).

In each interface, participants were shown representations of maps constructed by people in either the Simple Group or the Complex Group in the previous study (Figure 53). They were able to compare data collected on fear from crime, perceived housing price, and desirability with publicly available records for crime prevalence and median housing prices. They were then given a short survey in which they were asked to interpret the representation in their own words.

Then they reviewed the anonymous free response data from people in the previous study, and rated how closely their own interpretation aligns with the responses from participants. In the second activity, participants were shown representations of maps constructed by people who were using a different interaction technique (whichever they did not review first). They repeated the process of analyzing the maps and repeated the process of answering a questionnaire about their interpretations and insights.

5.3.2 Questionnaire

Participants were directed to answer the following questions while interpreting the spatial data and mental maps:

1. The city image accurately represents the relative desirability and importance of the various neighborhoods of the city. (agreement-Likert)
2. Please describe, in your own words, what characteristics are shared by the areas that people desire to live in? (free response)
3. Please describe, in your own words, what characteristics are shared by the areas that people **DO NOT** desire to live in? (free response)
4. As you review this map, what conclusions do you reach about the city? (free response)
5. As you review this map, what conclusions do you reach about the people who created it? (free response)
6. If you were to use the created image to argue for one policy change about the city, what would it be? (free response)

Participants were directed to answer the following questions while reviewing the free responses obtained during the previous study:

1. After reviewing the responses provided by citizens to these questions, how much do you think your previous responses agree with them? (agreement-Likert)
2. After reviewing the responses provided by citizens to these questions, what do you think you were **most correct** about? (free response)
3. After reviewing the responses provided by citizens to these questions, what do you think you were **most incorrect** about? (free response)

5.3.3 Analysis

Of the three people that have successfully completed the study, two strongly agree that the city image generated from the Complex group accurately represents the relative desirability and importance of the various neighborhoods of the city. These two also agree, while not strongly, that the city images from the Simple group also accurately represent the city. The third respondent is adamant in her decision to neither agree nor disagree with all prompts. One member of NPU-E noted in her response to question 3 that “Northern areas are more desirable than southern areas” when reviewing the Complex group data. This observation is certainly present in the NPU-E data, although one participant noted that it is not entirely uniform, in her response to the question about the **insights** that the thing she was most wrong about was “That lots of South Atlanta is desirable to people.” Although, one participant also noted in his response to a question that “I think that the people creating this map must be really smart because smart developers are sometimes out of touch with simple designs. I had to do too much work to look at the map and answer the questions...” This thoughtful response does speak to a challenge with both forms of interaction, namely, that they are still far more complex and error-prone than a sheet of paper and a pencil. Given that the attrition numbers for the data collection study rose steadily as task complexity increased, one of the most important takeaways for this research may be simply that the simplest solution may be the best, particularly in matters of broad community outreach and participation. Finally, one participant kept primarily to short responses based on pre-existing bias. For example, in response to the question about what elements are shared by the areas that people desire to live in, he merely responded with “the lazy.” Truly, one of the great challenges of public participation in GIS is that not everyone takes participation seriously. I suppose it takes all kinds to make a world.

CHAPTER VI

CONCLUSION

In this document, I have described several distinct visualization research efforts that all work toward simplified interaction techniques for mapping urban data. In Chapter 1, I began with the question: How do we analyze civic data? There were two substantial challenges to urban data analysis that this thesis sought to explore. These challenges led us to these primary research questions:

RQ1 There are many different ways to represent the same underlying data set, and this can lead to misinterpretation and miscommunication. *How can interactive visualizations improve interpretability of representations of mental maps?*

To pursue answers to this question, I reviewed current techniques for mitigating errors across alternative mappings. I also proposed a new technique, reachability weighted mean (rw-mean), for comparing mappings specifically for urban data. The results are promising, and indicate that rw-mean generates consistent visual representations across mapping schemes. It should be noted, however, that there are likely unexplored effects that result from variation in street network density across different areas in the city.

RQ2 People have different values and opinions on what is important about their city: what should be changed, and what should be preserved. *How can interactive visualizations improve current methods for capturing and depicting mental maps at the neighborhood and city levels?*

To pursue answers to this question, I designed a model for joining collected mental maps to official data, and representation methods for depicting mental maps. I also

conducted a study to compare web-based collection techniques. The goal of this work was to provide urban planners, community members, and civic organizations with a simple, expressive set of interaction techniques to analyze urban data. I have shown that dynamic sketch-based mental maps of urban spaces can provide these capabilities to non-experts in GIS, and can help experts to understand differences in qualitative perception for people in the region that they work in. In the course of conducting this research, I have produced the following contributions to the visualization research community:

- A set of techniques for capturing imageability elements (i.e., Nodes, Paths, Edges, and Districts) as mental maps
- A model that joins spatial data to mental maps
- Guidelines for specifying mental maps interactively within sketch-based online tools.
- A data set of mental maps from citizens of the city of Atlanta.
- A sketch-based system for capturing mental maps and analyzing spatial data in a browser.
- An analysis of the effect of interaction technique on spatial data interpretation.

6.1 Future Work

The results of this research have indicated several promising areas for potential future research. Of the many interaction techniques described in this thesis, many are ideal for touch, gesture, and pen-based hardware systems. These systems exist at a range of scales, from smaller hand-held phones and tablets to much larger interactive whiteboards that cover several feet. While the former is ideal for a single person to record a digital mental map and explore urban data, the latter is more optimal for

a collaborative approach that facilitates discussion. I think that, reflecting on the various models of public participation that can exist in a community [71], there are several promising research questions centered on facilitating collaborative discussion of urban data between hand-held devices in a localized setting. As people bring their devices into a shared location for public discussion, how can we enable shared data analysis that supports this discussion? For example, the St. Louis Map Room project by Jer Thorpe discussed in Chapter 2 uses a projected display on top of paper to allow participants to create a shared representation of the built environment by sketching on top of it. It would be interesting to explore methods that allows participants to navigate to a web application on a phone or tablet and digitally sketch in on the projected image as a group. This would make it easier to maintain a digital record of the collaborative activities of the group. This more complex method of interacting with the system may also have downsides, such as a higher maintenance or development cost that makes it less feasible for many communities and PPGIS models.

One of the challenges I have touched on many times in this research is the difficulty curve present in GIS software. An area of open research that is relevant to the questions of PPGIS interaction software is visualization literacy. There have been many attempts to not only define visualization literacy but measure it [11], and there are many implications for civic data analysis. It is a good thing to value the democratic nature of PPGIS and provide a forum for people to engage in the planning of their own community, but how can we ensure that people understand the complexities of the issues that they are debating? Are there representation or interaction techniques that are not equitable for those with a lower visualization literacy? It has long been understood that, just like with charts or even basic statistics, it is possible to create a purposefully misleading map [81]. The question, then, is how do we design map visualization and interaction techniques that are equitable? For those members of the

community that need some additional help understanding complex issues and data, how do we measure their understanding and provide a suitable interface that allows them to still participate in the discussion without getting lost in the complexity?

In Chapter 5, I discussed research that was conducted in part with students in the Data Science for Social Good (DSSG) program at Georgia Tech. In that project, students worked with a core group of civic leaders from a neighborhood in Atlanta and consulted with members of the Westside Communities Alliance to design a web-based interface for analyzing public safety data. I think that a promising area for future research in this area is the development of a broader model of student-community engagement that builds on this premise of students applying their understanding of technical concepts (e.g., data analysis, website design) to problems faced by the community around them. I would like to pursue this research to design a model for civic data labs that can provide students with a steady flow of interesting projects in the community while also providing value to that community. One of the great challenges for this type of work is that the benefits are often one-sided, and I would like to conduct research to determine how these benefits can be made practical and sustainable beyond the short limits of a summer program or project course. I believe that, ultimately, it will require a lot of time to develop trust with a core group of volunteer and civic organizations and projects that persist across multiple years. But, it is my hope to design and deploy this model in my new role as a tenure-track faculty.

In Chapter 5, I also discussed a model for joining mental maps to urban data through a modification of network kernel density estimation. I believe this model is a promising direction of research for steering simulations conducted within urban environments. For example, in analyzing traffic throughout the city, many researchers are turning to car data collected from taxi cabs [39]. This type of data has been immensely useful for modeling overall traffic flow, but it only represents one point of view: professional drivers. I think it would be fruitful to model a wider cross-section

of the population by using digital mental maps obtained from other types of drivers along with the traffic data. Much like the comparisons between housing price from listings and perception of housing price from surveys described in Chapter 5, I would imagine the disparity between perceived mental maps of the environment and taxi volume to highlight differences in perception of the navigable environment between professional drivers and other drivers. Understanding this disparity could be useful when trying to design agent-based systems for predicting crowd behavior.

We also know that public perception of data distributions, such as crime, can differ from official records. But how can we predict these differences, or fill in the gaps in official data sources? Are there alternative sources for data that can help us to understand what is missing from the official data? To pursue answers to these questions, I have been working with the Violence Prevention Program at Grady Hospital in Atlanta and the Division of Violence Prevention at the Centers for Disease Control (CDC). They are collecting data from trauma patients at the hospital to determine the rate of unreported violent crimes in the city of Atlanta with the intent of understanding what commercial properties are frequent sources of unreported violent crime. This data, commonly referred to as the Cardiff data (referring to pioneering work for this approach by researchers in Cardiff, UK [41]) contrasts with the official reported data by the Atlanta Police Department (APD). The data collected includes the location for the crime, the type of crime, and the time at which the crime occurred. I have spent much of the past year working to go through the process of being added to the existing legal framework that would permit me to have access to the data and perform research with it, but changes to department and additional protections for Cardiff data have so far prevented this.

In the future, I will continue to work with these researchers to obtain access to the Cardiff data in the hopes of using the mental maps from the other two studies to analyze the Cardiff data for Atlanta. I plan to work with the researchers at Grady and

the CDC to design interaction methods for exploring and combining the mental maps with the Cardiff data. I will compare the Cardiff data to the official data from the APD to determine the differences in distribution between them. I will compare the differences to the mental maps to determine if the mental maps more closely align with the Cardiff data or the APD data. I will compare the participant indicated locations of fear to the Cardiff data and the APD data to determine if the locations of fear more closely correlate with one or the other.

6.2 Discussion and Reflections

I have learned quite a bit over the years spent working on this research. In my earliest projects, I focused entirely on exploring the design space provided by slick, expensive displays: either small or large touchscreens that had a stylus attached. One of the things that became painfully apparent as I gravitated more and more towards community-driven applications of map visualization was how rare these types of systems would be available for public use in that context. While it's certainly true that many larger cities do have systems like a large interactive whiteboard, most do not.

I think this will be a challenge for researchers and students in this area; how to find the best balance between the desire to push technology forward while being sensitive to the social issues that affect adoption of new developments. It is no big revelation that a disparity exists between the availability of the benefits of technology and research not only between cities but also between communities within individual cities. However, I think it has been a challenge for me individually to determine how best to focus my own efforts. Do I try and design something really cool that only a handful of cities could use, or do I follow best practices to design something that a resource-starved community group really needs? Even this dichotomy is an oversimplification, but I think that it captures some of the tension that exists between research that

benefits the technology researcher and research that benefits the community being researched. In the ideal scenario, a novel solution might be found that satisfies the needs of the both, but one of the things that I have come to believe is that there are scenarios in which the best technological solution might be no technology at all.

While designing map interactions that would work for even novice users, I discerned that for many users the act of using a computer interface at all was still a substantial barrier. In these cases, the participant would have potentially been better served by a sheet of paper and something to draw with than all of the software that I had spent time coding. In these circumstances, it is more beneficial to try and push the user to adopt a more advanced system that might not provide the same benefits, or to allow them to use a more expressive and technologically simpler technique? I suspect that there may be no simple answer, but that there is value in taking time to reflect on these questions when conducting research that includes elements of computing and social sciences, as much of my own has.

REFERENCES

- [1] ADRIENKO, N. and ADRIENKO, G., “Spatial generalization and aggregation of massive movement data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 2, pp. 205–219, 2011.
- [2] AFZAL, S., MACIEJEWSKI, R., and EBERT, D. S., “Visual analytics decision support environment for epidemic modeling and response evaluation,” in *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 191–200, IEEE, 2011.
- [3] AGAFONKIN, V., “Leaflet.js.” <http://leafletjs.com/>.
- [4] AL-KODMANY, K., “Using visualization techniques for enhancing public participation in planning and design: process, implementation, and evaluation,” *Landscape and urban planning*, vol. 45, no. 1, pp. 37–45, 1999.
- [5] AL-KODMANY, K., “Visualization tools and methods in community planning: from freehand sketches to virtual reality,” *Journal of planning Literature*, vol. 17, no. 2, pp. 189–211, 2002.
- [6] ANDRIENKO, G., ANDRIENKO, N., DYKES, J., FABRIKANT, S. I., and WACHOWICZ, M., “Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research,” *Information Visualization*, vol. 7, no. 3-4, p. 173, 2008.
- [7] ANDRIENKO, N. and ANDRIENKO, G., *Exploratory analysis of spatial and temporal data: a systematic approach*. Springer Science & Business Media, 2006.
- [8] APPLEYARD, D., *Planning a pluralist city: Conflicting realities in Ciudad Guayana*. MIT Press (MA), 1976.
- [9] ARMITAGE, R., *Crime Prevention Through Housing Design: Policy and Practice*, ch. 6, pp. 123–143. Springer, 2013.
- [10] BOSTOCK, M., OGIEVETSKY, V., and HEER, J., “D³ data-driven documents,” *IEEE transactions on visualization and computer graphics*, vol. 17, no. 12, pp. 2301–2309, 2011.
- [11] BOY, J., RENSINK, R. A., BERTINI, E., and FEKETE, J.-D., “A principled way of assessing visualization literacy,” *IEEE transactions on visualization and computer graphics*, vol. 20, no. 12, pp. 1963–1972, 2014.
- [12] BRANTINGHAM, J. and MOHLER, G., “Predpol.” <http://www.predpol.com/>.

- [13] BRANTINGHAM, P. L. and BRANTINGHAM, P. J., “Mobility, notoriety, and crime: A study of crime patterns in urban nodal points,” *Journal of Environmental Systems*, vol. 11, no. 1, pp. 89–99, 1982.
- [14] BRANTINGHAM, P. L. and BRANTINGHAM, P. J., “Nodes, paths and edges: Considerations on the complexity of crime and the physical environment,” *Journal of Environmental Psychology*, vol. 13, no. 1, pp. 3–28, 1993.
- [15] BRIDSON, R., “Fast poisson disk sampling in arbitrary dimensions.,” in *SIGGRAPH Sketches*, p. 22, 2007.
- [16] BROOKE, J. and OTHERS, “Sus-a quick and dirty usability scale,” *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.
- [17] BRUNSDON, C., CORCORAN, J., and HIGGS, G., “Visualising space and time in crime patterns: A comparison of methods,” *Computers, Environment and Urban Systems*, vol. 31, no. 1, pp. 52–75, 2007.
- [18] BRUNSDON, C., FOTHERINGHAM, A. S., and CHARLTON, M. E., “Geographically weighted regression: a method for exploring spatial nonstationarity,” *Geographical analysis*, vol. 28, no. 4, pp. 281–298, 1996.
- [19] BRUNSDON, C., FOTHERINGHAM, A., and CHARLTON, M., “Geographically weighted summary statistics: a framework for localised exploratory data analysis,” *Computers, Environment and Urban Systems*, vol. 26, no. 6, pp. 501–524, 2002.
- [20] BURNS, R., “Atlantas food deserts leave its poorest citizens stranded and struggling,” *The Guardian*, vol. 17, 2014.
- [21] CABLE, D., “The racial dot map,” *Weldon Cooper Center for Public Service, University of Virginia*, 2013.
- [22] CHAINEY, S., TOMPSON, L., and UHLIG, S., “The utility of hotspot mapping for predicting spatial patterns of crime,” *Security Journal*, vol. 21, no. 1, pp. 4–28, 2008.
- [23] CHANG, R., WESSEL, G., KOSARA, R., SAUDA, E., and RIBARSKY, W., “Legible cities: Focus-dependent multi-resolution visualization of urban relationships,” *IEEE transactions on visualization and computer graphics*, vol. 13, no. 6, pp. 1169–1175, 2007.
- [24] “Neighborhood planning units (NPU) of atlanta.” <http://www.atlantaga.gov/index.aspx?page=739>.
- [25] CORRELL, M. and HEER, J., “Surprise! bayesian weighting for de-biasing thematic maps,” *IEEE Transactions on Visualization & Computer Graphics*, no. 1, pp. 651–660, 2017.

- [26] DANG, G., NORTH, C., and SHNEIDERMAN, B., “Dynamic queries and brushing on choropleth maps,” in *Information Visualisation, 2001. Proceedings. Fifth International Conference on*, pp. 757–764, IEEE, 2001.
- [27] DENNIS, S. F., “Prospects for qualitative GIS at the intersection of youth development and participatory urban planning,” *Environment and Planning A*, vol. 38, no. 11, pp. 2039–2054, 2006.
- [28] DiSALVO, C., NOURBAKHSI, I., HOLSTIUS, D., AKIN, A., and LOUW, M., “The neighborhood networks project: a case study of critical engagement and creative expression through participatory design,” in *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008*, pp. 41–50, Indiana University, 2008.
- [29] DORLING, D., BARFORD, A., and NEWMAN, M., “Worldmapper: the world as you’ve never seen it before,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 757–764, 2006.
- [30] DUNN, C. E., “Participatory GIS—a people’s GIS?,” *Progress in human geography*, vol. 31, no. 5, pp. 616–637, 2007.
- [31] DYKES, J. and BRUNSDON, C., “Geographically weighted visualization: interactive graphics for scale-varying exploratory analysis,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1161–1168, 2007.
- [32] DYKES, J. A., “Exploring spatial data representation with dynamic graphics,” *Computers & Geosciences*, vol. 23, no. 4, pp. 345–370, 1997.
- [33] ECK, J., CHAINEY, S., CAMERON, J., and WILSON, R., “Mapping crime: Understanding hotspots,” tech. rep., National Institute of Justice, 2005.
- [34] EFENTAKIS, A., GRIVAS, N., LAMPRIANIDIS, G., MAGENSCHAB, G., and PFOSE, D., “Isochrones, traffic and demographics,” in *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 548–551, ACM, 2013.
- [35] EGENHOFER, M. J., “Query processing in spatial-query-by-sketch,” *Journal of Visual Languages & Computing*, vol. 8, no. 4, pp. 403–424, 1997.
- [36] EICHER, C. L. and BREWER, C. A., “Dasymetric mapping and areal interpolation: Implementation and evaluation,” *Cartography and Geographic Information Science*, vol. 28, no. 2, pp. 125–138, 2001.
- [37] ENDERT, A., FIAUX, P., and NORTH, C., “Semantic interaction for visual text analytics,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 473–482, ACM, 2012.

- [38] EVANS, A. J. and WATERS, T., “Mapping vernacular geography: web-based gis tools for capturing ‘fuzzy’ or ‘vague’ entities,” *International Journal of Technology, Policy and Management*, vol. 7, no. 2, pp. 134–150, 2007.
- [39] FERREIRA, N., POCO, J., VO, H. T., FREIRE, J., and SILVA, C. T., “Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2149–2158, 2013.
- [40] FISHER, D., “Hotmap: Looking at geographic attention,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1184–1191, 2007.
- [41] FLORENCE, C., SHEPHERD, J., BRENNAN, I., and SIMON, T., “Effectiveness of anonymised information sharing and use in health service, police, and local government partnership for preventing violence related injury: experimental study and time series analysis,” *BMJ*, vol. 342, p. d3313, 2011.
- [42] FOTHERINGHAM, A. S. and WONG, D. W., “The modifiable areal unit problem in multivariate statistical analysis,” *Environment and planning A*, vol. 23, no. 7, pp. 1025–1044, 1991.
- [43] FRY, B. and REAS, C., “Processing.” <http://www.processing.org/>.
- [44] GETIS, A. and ORD, J. K., “The analysis of spatial association by use of distance statistics,” *Geographical analysis*, vol. 24, no. 3, pp. 189–206, 1992.
- [45] GLANDER, T. and DÖLLNER, J., “Abstract representations for interactive visualization of virtual 3d city models,” *Computers, Environment and Urban Systems*, vol. 33, no. 5, pp. 375–387, 2009.
- [46] GODWIN, A. and STASKO, J., “Drawing data on maps: Sketch-based spatiotemporal visualization,” in *Poster presented at the IEEE Conference on Information Visualization*, 2015.
- [47] GODWIN, A. and STASKO, J., “Hotsketch: Drawing police patrol routes among spatiotemporal crime hotspots,” in *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017.
- [48] GODWIN, A. and STASKO, J. T., “Nodes, paths, and edges: Using mental maps to augment crime data analysis in urban spaces,” in *Proceedings of the IEEE VGTC/EuroGraphics Conference on Visualization*, 2017.
- [49] GOODWIN, S., DYKES, J., and SLINGSBY, A., “Poster: Visualizing the effects of scale and geography in multivariate comparison,” in *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 251–252, IEEE, 2014.
- [50] GOODWIN, S., DYKES, J., SLINGSBY, A., and TURKAY, C., “Visualizing multiple variables across scale and geography,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 599–608, 2016.

- [51] GOULD, P., “On mental maps,” tech. rep., Michigan Inter-university Community of Mathematical Geographers, 1966.
- [52] GOULD, P. and WHITE, R., *Mental maps*. Routledge, 2012.
- [53] GREEN, T. M., RIBARSKY, W., and FISHER, B., “Building and applying a human cognition model for visual analytics,” *Information visualization*, vol. 8, no. 1, pp. 1–13, 2009.
- [54] HAKLAY, M. and TOBÓN, C., “Usability evaluation and PPGIS: towards a user-centred design approach,” *International Journal of Geographical Information Science*, vol. 17, no. 6, pp. 577–592, 2003.
- [55] HARROWER, M. and BREWER, C. A., “Colorbrewer. org: an online tool for selecting colour schemes for maps,” *The Cartographic Journal*, vol. 40, no. 1, pp. 27–37, 2003.
- [56] HENDERSON, K., “Flexible sketches and inflexible data bases: Visual communication, conscription devices, and boundary objects in design engineering,” *Science, technology & human values*, vol. 16, no. 4, pp. 448–473, 1991.
- [57] HERLOCKER, M., “Turf.js.” <http://turfjs.org/>.
- [58] HUCK, J., WHYATT, J., and COULTON, P., “Spraycan: A ppgis for capturing imprecise notions of place,” *Applied Geography*, vol. 55, pp. 229–237, 2014.
- [59] INGRAM, R. and BENFORD, S., “Legibility enhancement for information visualisation,” in *Proceedings of the 6th conference on Visualization’95*, p. 209, IEEE Computer Society, 1995.
- [60] INGRAM, R. and BENFORD, S., “The application of legibility techniques to enhance information visualizations,” *The Computer Journal*, vol. 39, no. 10, pp. 819–836, 1996.
- [61] ISENBERG, P., ISENBERG, T., HESSELMANN, T., LEE, B., VON ZADOW, U., and TANG, A., “Data visualization on interactive surfaces: A research agenda,” *IEEE Computer Graphics and Applications*, vol. 33, no. 2, pp. 16–24, 2013.
- [62] JANSEN, Y. and DRAGICEVIC, P., “An interaction model for visualizations beyond the desktop,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2396–2405, 2013.
- [63] JENELIUS, E., PETERSEN, T., and MATTSSON, L.-G., “Importance and exposure in road network vulnerability analysis,” *Transportation Research Part A: Policy and Practice*, vol. 40, no. 7, pp. 537–560, 2006.
- [64] KIM, S., MACIEJEWSKI, R., MALIK, A., JANG, Y., EBERT, D. S., and ISENBERG, T., “Bristle maps: A multivariate abstraction technique for geo-visualization,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 9, pp. 1438–1454, 2013.

- [65] KITCHIN, R. M., “Cognitive maps: What are they and why study them?,” *Journal of environmental psychology*, vol. 14, no. 1, pp. 1–19, 1994.
- [66] KNUDSEN, S., JAKOBSEN, M. R., and HORNBAEK, K., “An exploratory study of how abundant display space may support data analysis,” in *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, pp. 558–567, ACM, 2012.
- [67] KWAN, M.-P., “Feminist visualization: Re-envisioning GIS as a method in feminist geographic research,” *Annals of the association of American geographers*, vol. 92, no. 4, pp. 645–661, 2002.
- [68] LANDAY, J. A. and MYERS, B. A., “Sketching interfaces: Toward more human interface design,” *Computer*, vol. 34, no. 3, pp. 56–64, 2001.
- [69] LANGFORD, M. and UNWIN, D. J., “Generating and mapping population density surfaces within a geographical information system,” *The Cartographic Journal*, vol. 31, no. 1, pp. 21–26, 1994.
- [70] LEE, B., ISENBERG, P., RICHE, N. H., and CARPENDALE, S., “Beyond mouse and keyboard: Expanding design considerations for information visualization interactions,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2689–2698, 2012.
- [71] LEITNER, H., MCMASTER, R. B., ELWOOD, S., MCMASTER, S., and SHEPARD, E., “Models for making GIS available to community organizations: dimensions of difference and appropriateness,” in *Community Participation and Geographical Information Systems*, pp. 63–78, CRC Press, 2002.
- [72] LU, B., CHARLTON, M., BRUNSDON, C., and HARRIS, P., “The Minkowski approach for choosing the distance metric in geographically weighted regression,” *International Journal of Geographical Information Science*, vol. 30, no. 2, pp. 351–368, 2016.
- [73] LU, B., CHARLTON, M., and FOTHERINGHAM, A. S., “Geographically weighted regression using a non-Euclidean distance metric with a study on London house price data,” *Procedia Environmental Sciences*, vol. 7, pp. 92–97, 2011.
- [74] LU, B., CHARLTON, M., HARRIS, P., and FOTHERINGHAM, A. S., “Geographically weighted regression with a non-euclidean distance metric: a case study using hedonic house price data,” *International Journal of Geographical Information Science*, vol. 28, no. 4, pp. 660–681, 2014.
- [75] LYNCH, K., *The image of the city*, vol. 11. MIT press, 1960.
- [76] MACEACHREN, A. M., “Visualizing uncertain information,” *Cartographic Perspectives*, no. 13, pp. 10–19, 1992.

- [77] MACIEJEWSKI, R., RUDOLPH, S., HAFEN, R., ABUSALAH, A. M., YAKOUT, M., OUZZANI, M., CLEVELAND, W. S., GRANNIS, S. J., and EBERT, D. S., “A visual analytics approach to understanding spatiotemporal hotspots,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 2, pp. 205–220, 2010.
- [78] MALIK, A., MACIEJEWSKI, R., TOWERS, S., MCCULLOUGH, S., and EBERT, D. S., “Proactive spatiotemporal resource allocation and predictive visual analytics for community policing and law enforcement,” *IEEE transactions on visualization and computer graphics*, vol. 20, no. 12, pp. 1863–1872, 2014.
- [79] MANAUGH, G., *A Burglar’s Guide to the City*. FSG Originals, 2016.
- [80] MATEI, S., BALL-ROKEACH, S. J., and QIU, J. L., “Fear and misperception of los angeles urban space: A spatial-statistical study of communication-shaped mental maps,” *Communication Research*, vol. 28, no. 4, pp. 429–463, 2001.
- [81] MONMONIER, M., *How to lie with maps*. University of Chicago Press, 2018.
- [82] O’CONNELL, K., LEE, Y., PEER, F., STAUDAHER, S. M., GODWIN, A., MADDEN, M., and ZEGURA, E., “Making public safety data accessible in the westside atlanta data dashboard,” *arXiv preprint arXiv:1609.09756*, 2016.
- [83] O’CONNELL, K., MADDEN, M., and DAVIS-FAULKNER, S., “Westside communities alliance (WCA) data dashboard.” <http://wccadatadashboard.iac.gatech.edu/Home/>.
- [84] OPENSHAW, S., *The modifiable areal unit problem*. Geobooks, 1984.
- [85] O’SULLIVAN, D., MORRISON, A., and SHEARER, J., “Using desktop gis for the investigation of accessibility by public transport: an isochrone approach,” *International Journal of Geographical Information Science*, vol. 14, no. 1, pp. 85–104, 2000.
- [86] PANSE, C., SIPS, M., KEIM, D. A., and NORTH, S. C., “Visualization of geospatial point sets via global shape transformation and local pixel placement,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 749–756, 2006.
- [87] PENDERGRAST, M., *City on the verge: Atlanta and the fight for Americas urban future*. Basic Books, 2017.
- [88] PORTA, S., CRUCITTI, P., and LATORA, V., “The network analysis of urban streets: a primal approach,” *Environment and Planning B: planning and design*, vol. 33, no. 5, pp. 705–725, 2006.
- [89] REY, S. J., MACK, E. A., and KOSCHINSKY, J., “Exploratory space–time analysis of burglary patterns,” *Journal of Quantitative Criminology*, vol. 28, no. 3, pp. 509–531, 2012.

- [90] ROTH, R. E., ROSS, K. S., FINCH, B. G., LUO, W., and MACEACHREN, A. M., “A user-centered approach for designing and developing spatiotemporal crime analysis tools,” in *Proceedings of GIScience*, vol. 15, 2010.
- [91] SCHLAISICH, I. and EGENHOFER, M. J., “Multimodal spatial querying: what people sketch and talk about,” in *Proceedings of the 9th International Conference on Human-Computer Interaction (HCI)* (STEPHANIDIS, C., ed.), pp. 732–736, 2001.
- [92] SCHLOSSBERG, M. and SHUFORD, E., “Delineating ”public” and ”participation” in PPGIS,” *URISA Journal*, vol. 16, no. 2, 2005.
- [93] SIEBER, R., “Public participation geographic information systems: A literature review and framework,” *Annals of the Association of American Geographers*, vol. 96, no. 3, pp. 491–507, 2006.
- [94] SILVERMAN, B. W., *Density estimation for statistics and data analysis*, vol. 26. CRC press, 1986.
- [95] STAR, S. L. and GRIESEMER, J. R., “Institutional ecology, translations’ and boundary objects: Amateurs and professionals in berkeley’s museum of vertebrate zoology, 1907-39,” *Social studies of science*, vol. 19, no. 3, pp. 387–420, 1989.
- [96] SWINK, M. and SPEIER, C., “Presenting geographic information: effects of data aggregation, dispersion, and users’ spatial orientation,” *Decision sciences*, vol. 30, no. 1, pp. 169–195, 1999.
- [97] TALEN, E., “Bottom-up GIS: A new tool for individual and group expression in participatory planning,” *Journal of the American Planning Association*, vol. 66, no. 3, pp. 279–294, 2000.
- [98] TAVERNER, C., “Neo4j spatial.” <https://github.com/neo4j-contrib/spatial>.
- [99] THORP, J., “St louis maproom.” <http://stlmaproom.org/>.
- [100] TOBLER, W. R., “A computer movie simulating urban growth in the Detroit region,” *Economic geography*, vol. 46, pp. 234–240, 1970.
- [101] TOMINSKI, C., SCHUMANN, H., ANDRIENKO, G., and ANDRIENKO, N., “Stacking-based visualization of trajectory attribute data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2565–2574, 2012.
- [102] TURKAY, C., SLINGSBY, A., HAUSER, H., WOOD, J., and DYKES, J., “Attribute signatures: Dynamic visual summaries for analyzing multivariate geographical data,” *IEEE transactions on visualization and computer graphics*, vol. 20, no. 12, pp. 2033–2042, 2014.

- [103] UNWIN, A. and UNWIN, D., “Spatial data analysis with local statistics,” *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 47, no. 3, pp. 415–421, 1998.
- [104] VAN DIJK, T. C., HAUNERT, J.-H., and OEHRLEIN, J., “Location-dependent generalization of road networks based on equivalent destinations,” in *Computer Graphics Forum*, vol. 35, pp. 451–460, Wiley Online Library, 2016.
- [105] VON LANDESBERGER, T., BRODKORB, F., ROSKOSCH, P., ANDRIENKO, N., ANDRIENKO, G., and KERREN, A., “Mobilitygraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 11–20, 2016.
- [106] WALNY, J., HURON, S., and CARPENDALE, S., “An exploratory study of data sketching for visual representation,” in *Computer Graphics Forum*, vol. 34, pp. 231–240, Wiley Online Library, 2015.
- [107] WEINER, D., HARRIS, T., and CRAIG, W. J., “Community participation and geographic information systems,” in *Community participation and geographic information systems* (CRAIG, W. J., HARRIS, T., and WEINER, D., eds.), ch. 1, pp. 3–16, Taylor & Francis, London and New York, 2002.
- [108] WESSEL, G., UNRUH, E., and SAUDA, E., “Heads up: using cognitive mapping to develop a baseline description for urban visualization,” in *Proceedings of the Sixth International Conference on advances in Computer-Human Interactions (ACHI)*, vol. 24, Citeseer, 2013.
- [109] WONG, P. C., SCHNEIDER, K., MACKAY, P., FOOTE, H., CHIN JR, G., GUTTROMSON, R., and THOMAS, J., “A novel visualization technique for electric power grid analytics,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 3, pp. 410–423, 2009.
- [110] WOOD, J., DYKES, J., and SLINGSBY, A., “Visualisation of origins, destinations and flows with OD maps,” *The Cartographic Journal*, vol. 47, no. 2, pp. 117–129, 2010.
- [111] WOOD, J., ISENBERG, P., ISENBERG, T., DYKES, J., BOUKHELIFA, N., and SLINGSBY, A., “Sketchy rendering for information visualization,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 12, pp. 2749–2758, 2012.
- [112] WOOD, J., SLINGSBY, A., and DYKES, J., “Visualizing the dynamics of London’s bicycle-hire scheme,” *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 46, no. 4, pp. 239–251, 2011.
- [113] XIE, Z. and YAN, J., “Kernel density estimation of traffic accidents in a network space,” *Computers, Environment and Urban Systems*, vol. 32, no. 5, pp. 396–406, 2008.

- [114] XIE, Z. and YAN, J., “Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach,” *Journal of transport geography*, vol. 31, pp. 64–71, 2013.
- [115] ZEGURA, E., LE DANTEC, C., and DILKINA, B., “Data science for social good atlanta.” <http://www.dssg-atl.io>.
- [116] ZENG, W., FU, C.-W., ARISONA, S. M., ERATH, A., and QU, H., “Visualizing mobility of public transportation system,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1833–1842, 2014.
- [117] ZHANG, Y. and MACIEJEWSKI, R., “Quantifying the visual impact of classification boundaries in choropleth maps,” *IEEE Transactions on Visualization & Computer Graphics*, no. 1, pp. 371–380, 2017.