Comparisons of Chicago Neighborhood Boundaries from Crowdsourced Resident Drawings

Crystal J. Bae[1]
The University of Chicago, IL, USA

Lydia Wileden
The University of Chicago, IL, USA

Emily Talen
The University of Chicago, IL, USA

Abstract

The idea of the urban neighborhood has long been of interest to residents, planners, and scholars. We describe a project focused on Chicago neighborhood mapping and pose related questions about the analysis of crowdsourced neighborhood boundary drawings. To gain insight into Chicago residents' cognitive maps and the relationship between those internal representations and existing administrative boundaries, the authors launched the Chicago Neighborhood Project (CNP), which invited Chicago residents to draw their own and other neighborhoods within the city using an online mapping interface. The goal of CNP is to examine variation in how neighborhoods are defined by residents and use that variation to inform how policymakers, planners, and researchers create, implement, and measure place-based policies. Because the project had a goal of collecting a large sample of neighborhood map drawings, the project took a crowdsourced approach, recruiting responses via email to community groups, social media, targeted web advertisements, flyering, collaborations with news media, and word of mouth. This paper describes our data collection methodology, resulting in over 5,000 responses, as well as decisions related to initial data cleaning and analysis. We present early findings from the project in relation to understanding Chicago residents' cognitive boundaries of the "neighborhood."

TL;DR: We present preliminary results of the Chicago Neighborhood Project (CNP), which collected over 5,000 drawings of neighborhood areas from residents, making it the largest such effort to elicit an understanding of neighborhood regions in Chicago.

2012 ACM Subject Classification Applied computing → Sociology; Human-centered computing → Geographic visualization; Applied computing → Psychology

Keywords and phrases cognitive regions, urban neighborhoods, boundary mapping, sketch mapping

Digital Object Identifier 10.4230/LIPIcs.COSIT.2024.26

Category Short Paper

Funding Thank you to the Mansueto Institute of Urban Innovation at the University of Chicago for their funding support in the form of an Urban Innovation Grant for 2023–2024.

Acknowledgements The authors would like to acknowledge the contributions of Divij Sinha for assistance with web programming and database management, as well as our research assistants on the project: Serena Bernstein, Michael Reznik, Victor Quadros, and Keyan Dunmore.

[1] Corresponding author

16th International Conference on Spatial Information Theory (COSIT 2024).
Editors: Benjamin Adams, Amy Griffin, Simon Scheider, and Grant McKenzie; Article No. 26; pp. 26:1–26:10
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

The work presented here focuses on understanding resident perceptions of urban neighborhood boundaries through survey methodology by assessing results of a large web-based survey of Chicago residents in 2023–2024. This builds on prior methodological work in cognitive boundary mapping [1, 6, 14], and is inspired by similar approaches for other large cities such as for Boston [21], Los Angeles [18], and New York City [19]. Prior crowdsourced neighborhood boundary projects such as Bostonography [21] have allowed city residents to communicate their perceptions of neighborhood boundaries and names beyond those delineated on official administrative maps. More recently, *The New York Times* launched a large-scale crowdsourced neighborhood boundary project for the city of New York City, largely drawing on the far-reaching reader base of the major newspaper. Although Chicago is the third largest city and metropolitan area in the United States, and is often referred to as a “city of neighborhoods,” it has no official neighborhood designations in use today. Instead, the City relies on the definition of Chicago Community Areas (CCAs), first established in and largely unchanged since the 1920s [3, 15], to serve as a proxy for the “neighborhood” in many communications and practical planning contexts.

Cognitive mapping approaches to understanding cognitive regions, especially the urban *neighborhood*, have received prominent attention in the sociological and geographic literature [4, 5, 10]. One dominant approach to eliciting the features and structure of human cognitive regions is through sketch mapping. Sketch mapping is a common method by which researchers have elicited spatial knowledge of and beliefs about the environment [7], such as for understanding the microgenesis of spatial knowledge in response to moving to a new environment [8], or for comparing newcomers’ and long-term residents’ cognition of the city [16]. In the case of neighborhood boundaries, sketch map drawings may serve as a form of knowledge production where boundary representations point to cognitive understanding of neighborhood regions within the city, independent of other regions [12]. However, there is limited academic work comparing the relationship between these crowdsourced approaches to neighborhood delineation by residents with the prior literature on cognitive regions and urban cognitive maps [13].

To this end, the Chicago Neighborhood Project (CHP) aims to update the present-day understanding of Chicago’s neighborhoods from the perspective of its residents. The project collected over 5,000 neighborhood boundary drawings from Chicago residents, which is to date the largest such effort to capture neighborhood conceptions in the city of Chicago. This paper outlines the data collection process and initial data analytic decisions made with regards to crowdsourced boundary drawings, which we treat as simple sketch maps of neighborhood areas, to reveal both the challenges and the opportunities associated with a large-scale survey of residents’ neighborhood conceptions. Preliminary results also demonstrate the potential for applying boundary summary methods [1] and shape compactness measures [11] to the study of urban neighborhood representations.

Research Questions

We present the following questions in this short paper:

- What are the challenges associated with using crowdsourced data as provided through a web interface as a source of primary data collection of residents’ cognitive neighborhood boundaries?
- How do approaches to polygon summarization, such as through a raster-based method, reveal differences in residents’ agreement and understanding of cognitive regions such as neighborhoods?

2 Method

Survey responses were recorded through a website, The Chicago Neighborhood Project, located at `https://chicago-neighborhoods.com`, during the project data collection phase. The survey was restricted to current Chicago residents who were at least 18 years of age and gave their consent (IRB Exempt; number anonymized for review).

2.1 Survey Design

The web-based survey used for data collection had three main steps. A response was considered fully complete (and included in analysis) if the respondent submitted at least one neighborhood boundary drawing in Step 2. Respondents could end their participation at any time, and partial responses were recorded.

For the map-based steps of the survey, we used a built-in web map tool from the *Leaflet.draw* library [20] that allowed our participants to add a home location point marker in Step 1, and in Step 2 to draw a polygon using vertices, which were visually connected by lines as they were drawing, and complete the polygon by clicking on the starting vertex to close the shape. The web map used a Mapbox Streets template as the basemap, which we customized to remove the default Points of Interest (POIs) as well as any neighborhood labels typically displayed on the basemap. However, labels for street and highway features, transit stops, and water bodies were displayed to help participants orient themselves on the map.

The steps of the survey were as follows:

1. Respondent reported their home location by dropping a point marker on the map of Chicago. We did not collect home addresses nor exact locations, only a recorded coordinate location based on the location of the dropped point marker.
2. Respondent was prompted, "Where are the boundaries of your current neighborhood?" and drew a polygon representing the boundary of their neighborhood using the interactive map drawing tool. This involved adding vertices of a polygon to draw a boundary line, then enclosing the polygon area by clicking again on the first vertex point added. They were also asked "What do you call your neighborhood or the area where you live?" and gave a name corresponding to what they called their neighborhood (selected from a pre-populated list or provided as text by the respondent). This step could be repeated after the first neighborhood drawing to optionally draw additional neighborhoods.
3. Respondent completed optional residential and demographic information questions, which comprised questions asking about their length of residence in the city of Chicago, length of residence in the current neighborhood, whether they own or rent their current residence (or "other" tenancy arrangement), gender identity, age, and racial/ethnic background. Respondents were informed that they could answer or skip any of these questions.

2.2 Distribution of Survey and Data Collection

We took a multi-pronged approach to survey dissemination to maximize the ways in which we reached potential respondents (all current Chicago residents). Although the survey was administered through a custom-built website, advertising and distribution spanned both online digital and traditional marketing methods. These methods included email distribution lists including those of the Chicago alderpersons (Chicago City Council) who agreed to reach their wards, news outlets including local newspapers and TV stations,[2] in-person tabling at local events and flyering at spatially distributed locations throughout neighborhoods (such as public library branches), public talks on the project, word-of-mouth distribution, web advertising and social media, and through the assistance of the [University Survey Lab; anonymized for review] to reach local community organizations through further in-person and mailing-based outreach focused on under-represented areas of the city in our survey.

The survey ran for 5 months from November 2023 to March 2024. In that time, a total of 10,366 initial responses were recorded. Of those, 5,937 respondents who began the survey reached completion, which was defined as when respondents submitted a polygon drawing of their neighborhood boundary (Step 2). From respondent feedback, we note that the much of the response drop-off between starting the survey and completing the polygon drawing could attributed to user difficulties with the web mapping tool, as many reported that it was difficult to draw a polygon on the interactive map.[3] After submitting an initial boundary drawing, respondents could either submit additional boundary drawings, such as for other neighborhoods they were familiar with, or move onto optional residential and demographic information questions in Step 3.

To capture diverse perspectives, the online survey was translated into Spanish, Polish, and Mandarin, and web advertisements were placed in both English and Spanish. To incentivize responses, respondents were entered into a drawing for multiple $50 gift cards. Recognizing that the selected convenience sample approach could produce bias in responses, the research team monitored survey completions throughout the field period for demographic and geographic bias. In response to over representation of white respondents and respondents on Chicago's North Side, efforts were made to target recruitment to low-response areas on the South and West Sides based on counts of responses by ZIP Code Tabulation Areas and Community Areas. These further efforts resulted in a more even representation of respondents from across Chicago's racially and ethnically segregated landscape, in terms of both demographics and spatial location, than in initial waves of responses. Details of spatial coverage and demographics of respondents are given in Section 4 below.

3 Data Cleaning

We describe our general data cleaning process and corresponding decisions based on the raw polygons in the crowdsourced survey responses. This was the process for the initial treatment of drawn polygons:

A. Classify invalid geometries for correction or removal. We checked the validity of drawn polygons in the open-source Geographic Information Systems software QGIS. We dropped invalid polygons containing fewer than 3 points, as they were unusable as areas for our polygon analyses. We further classified invalid polygons with self-intersection errors to determine which could be corrected for use. Examples of those polygons that could be corrected based on the original invalid geometries are described below, and the process for correction is shown in Figures 1 and 2.

[2] Including interviews and featured reporting through Crain's Chicago Business, Block Club Chicago, WTTW News, and the 'On The Block' program on WCIU-TV Chicago.

[3] Largely this was the case for those who were using a mobile device to complete the survey rather than a laptop or desktop computer; unfortunately we did not collect data on the device used to respond to the survey, so we are unable to determine the exact proportion of respondents who used a mobile device.

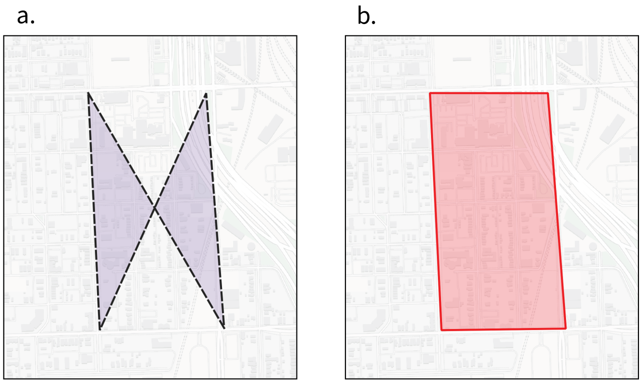

Figure 1 a. Example of invalid polygon based on incorrect sequence of vertices drawn; classified as a *bow-tie shape*. b. Corrected polygon based on re-ordering the original vertices.

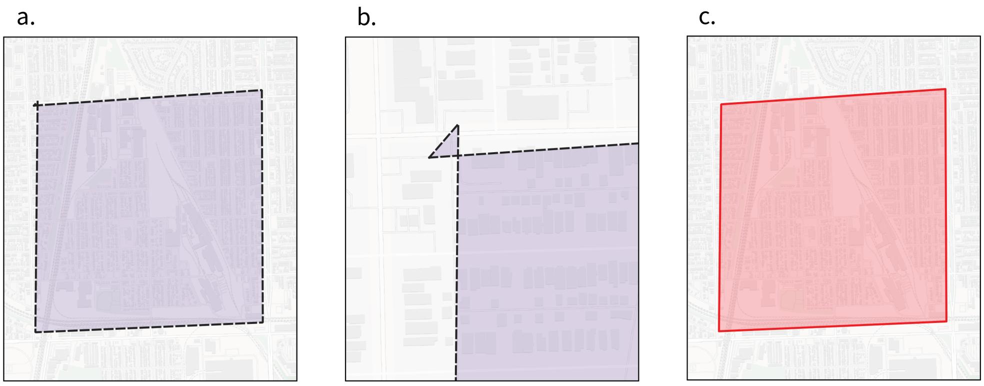

Figure 2 a. Example of invalid polygon based on a minor self-intersection issue near one vertex; classified as a *hanging geometry*. b. Close-up view of the problem area demonstrating polygon self-intersection for a small corner in the upper-left. b. Corrected polygon based on dropping the small "hanging" part of the geometry.

1. *Bow-tie shapes*, indicating an incorrect input sequence of polygon vertices during drawing. We infer here that the respondent's intent was to create a rectangular representation of the neighborhood in these cases, rather than two triangular sections. In these cases, the polygon could be corrected by swapping the order of the vertices which were incorrectly ordered (e.g. not sequentially ordered in either the clockwise or counter-clockwise direction). In practice, this correction was approximated by creating a minimum bounding rectangle around the original "bow-tie" shape. See Figure 1 for an example of an invalid geometry drawn by a polygon classified as "bow-tie" and corrected.
2. *Hanging geometries*, where there was an extra area captured in the polygon drawing due to multiple clicks near one vertex location that caused a small area to be enclosed outside of the main polygon. This type of issue caused the full polygon drawing to be classified as "invalid" due to violating the self-intersection condition. These "hanging geometries" could be treated spatially via application of a negligible distance buffer around the original geometry, which drops the smaller problem area. See Figure 2 for an example of an invalid geometry classified as a "hanging geometry" and corrected.

B. Remove duplicate entries. Erroneous duplicate entries were identified through matching unique user ID and same timestamp for entry, showing a duplicated entry induced by the survey tool.

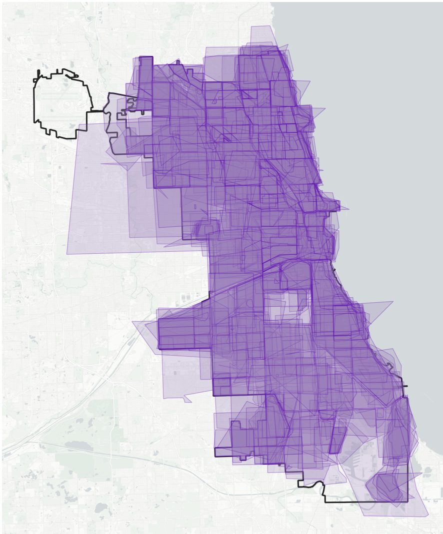

Figure 3 Visual summary of all polygons drawn by respondents and retained after data cleaning.

C. Remove boundaries which fall mostly outside of the City of Chicago. We compared overlap between each drawn area and the City of Chicago, removing those that had an overlapping area of less than 50%.

Each of the described data cleaning steps were accompanied by manual check by the researchers, through the visual inspection of dropped cases, to confirm that relevant responses were not incorrectly removed. After the initial data cleaning process was applied, we had a usable dataset of drawn boundaries (n = 5,507) from the original dataset (n = 5,937; difference of 430) represented as polygons that mostly overlapped with the City of Chicago.

4 Preliminary Findings and Discussion

4.1 Spatial Coverage and Respondent Demographics

To our first research question about challenges posed by the data collection method, we examine how this approach to crowdsourcing large numbers of drawings of the cognitive boundaries of neighborhoods works for an understanding of where commonly-understood neighborhoods in the city are located. Further planned work examines the relative size of neighborhoods, their connection to existing and historic administrative boundaries and physical barriers in the city, and spatial characteristics in relation to normative theories of the urban neighborhood.

Here we examine response coverage and gaps in representation. By looking at spatial coverage and areas of low response, we can make initial descriptions of our survey responses as a whole, and assess the representativeness of this survey by comparing our responses to overall population trends and spatial patterns within Chicago and its neighborhoods. Because the survey was administered through an online website, there was minimal additional information that could be used to interpret respondents' boundary drawings. This posed a challenge in the interpretation of the invalid geometries recorded, for which we took a conservative approach and were able to make inferences about the intended shapes for only 178 of 303 (58.7%) incorrectly recorded geometries. Those decisions are described above in Section 3 Data Cleaning. Figure 3 provides a visual summary of the polygons drawn by respondents, which correspond to areas drawn using the mapping tool in the Chicago Neighborhood Project survey and retained after the data cleaning steps above were applied.

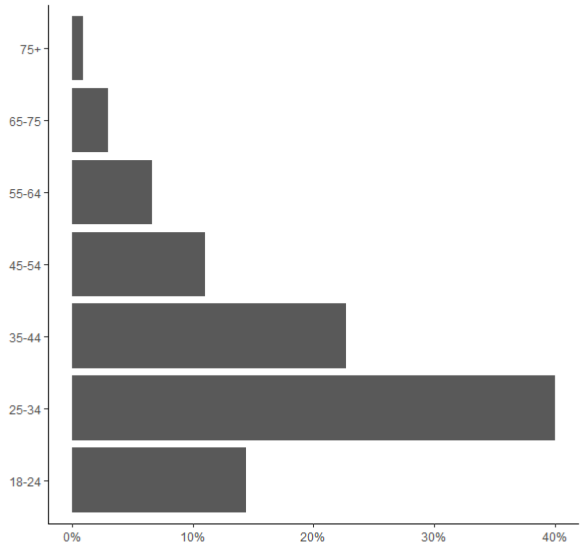

Figure 4 Age of respondents, excluding 1.4% of respondents who gave no response.

Table 1 Race and ethnicity identity of survey respondents. Chicago residents statistics included for comparison, from Census 2018-2022 ACS estimates. [2]

Race and Ethnicity	Survey Respondents	Chicago Residents
White	78.4%	75.1%[1]
Hispanic / Latino	10.2	29.0
Asian	6.3	7.0
Black / African American	6.0	28.8
Multi-racial / Two or more races	4.0	9.7
Middle Eastern / North African	1.2	Not reported[2]
Other	1.2	Not reported[2]
American Indian or Alaska Native	0.4	0.7
Native Hawaiian or Pacific Islander	0.3	0.1

[1]When "White alone" and "White alone, not Hispanic or Latino" are combined; as our survey does not split these categories.

[2]Not reported in Census estimates.

In initial exploration of the full cleaned dataset, we note that spatial extent of the drawn polygons cover almost the entire area of the City of Chicago (95.1%) with the most obvious gaps in coverage over the O'Hare airport region and the sparsely-populated Riverdale region[4] of the far South Side. In terms of respondent demographics, which were each optional questions in the survey, we report on tenancy, gender identity, age, and racial/ethnic identity.

[4] Total population of the Riverdale Community Area was 7,262 residents as of 2020 [9].

For tenancy, 49.0% owned their residence, 45.6% rented, and 5.4% reported some “other” housing arrangement. This is comparable to the overall demographics for tenancy in the City of Chicago, although slightly higher than the city-wide owner-occupied housing rate of 45.6% (as of ACS 2018-2022 estimates [2]). Our respondents’ mean length of residency in the current neighborhood was 8.2 years (median = 4.0, *sd* = 10.7); the mean length of residency in Chicago overall was 16.0 years (median = 11.0 years, *sd* = 15.0 years). For gender identity, our respondents were 56.1% male, 37.5% female, 3.8% non-binary, and 0.7% other, with 1.9% giving no response; overall this skews more male than expected.

In terms of age and ethnicity, Figure 4 displays the proportion of respondents within each age range and racial/ethnic category; 1.4% of respondents gave no response for age. For race and ethnicity categories, Table 1 shows the race and ethnicity proportions for our respondents, with a comparison to Chicago residents overall; more than one response could be selected for race and ethnicity, so percentages add up to greater than 100%.

Further planned comparisons involve looking at average shape and area of polygons in order to compare to normative theories about urban neighborhoods such as typical population and size, as well as correspondence to services and physical boundaries [17]. The polygon areas collected in this study are likely to have important shortcomings for interpretation. One critical point is that respondents are potentially much less likely to draw circular boundaries due to the nature of the polygon drawing tool. This is due to the user interface in our mapping tool, as respondents had to construct their polygon through adding sequential vertices.

4.2 Neighborhood Boundary Comparisons

The focus of this analysis is on the summarization of drawn polygon representations of neighborhood boundaries. We demonstrate the application of a raster-based summarization method appropriate for understanding agreement areas across urban neighborhood boundary drawings. We observe that there were 263 unique neighborhood names given in conjunction with the polygon drawings that were reported by at least 2 residents. For those responses with shared neighborhood names, we applied a raster summary method described below to understand the proportion of overlap (agreement area) for each neighborhood.

This method of summarizing neighborhood polygon drawings follows the raster method proposed by Bae and Montello [1]. This was implemented in R using a script written to create a raster grid representation of spatial extent for each neighborhood, then calculate the number of drawn boundary polygons that overlapped within the area. For each grid cell containing at least one response polygon, the proportion of overlapping polygons is calculated, creating a visual representation that we label based on percent agreement.

In the example summary of one given Chicago neighborhood in Figure 5, we calculated a raster summary of all drawings that our respondents called “Hyde Park”. When compared to a basemap, this raster summary shows that strong edges are visible in relation to park areas on the west and south, and in relation to the lakeshore to the east. Although general agreement areas can be visually discerned from this simple output, planned next steps involve comparing methods for summarizing the polygons using a vector-based approach [1, 6], and the continued development of methods for calculating and visualizing regions of agreement and contestation between multiple neighborhood definitions.

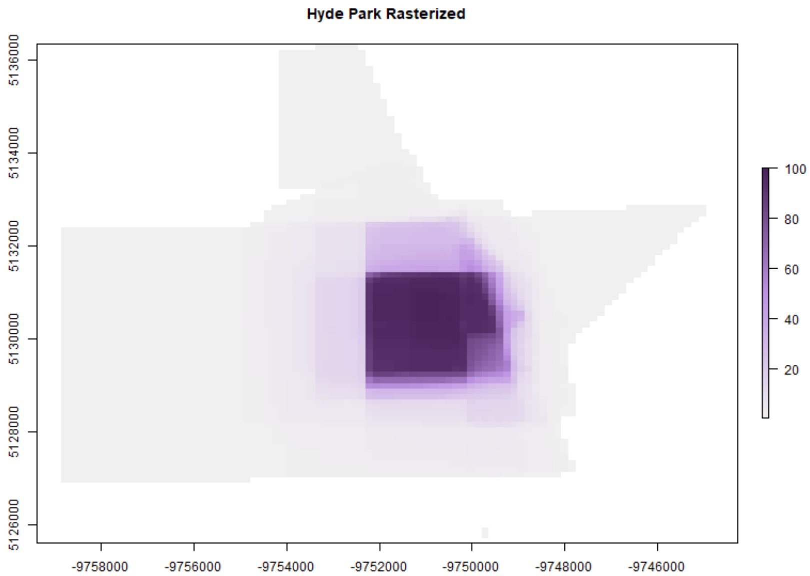

Figure 5 Example raster output for the "Hyde Park" neighborhood, with corresponding degree of consensus on the included neighborhood area expressed as color on a 0 – 100% agreement scale.

5 Discussion

To date, this is the largest-ever effort to collect data on residents' understanding of Chicago neighborhoods. We demonstrate the theoretical contribution of such an effort in terms of assessing spatial areas of agreement about neighborhood location and extent, and argue for the importance of applying novel polygon summarization methods to better understand how relevant data collected through crowdsourcing methods. This should inspire further methodological innovation in how to analyze representations of neighborhoods, such as through assessing the strength of neighborhood edges and areas, or the overlap and nesting between competing neighborhood ideas.

In addition, analyses we plan to pursue with the collected data include connecting these shared cognitive boundaries of Chicago neighborhoods to current and historic administrative and planning boundaries, such as those relevant to zoning and land-use within the city, as well as to physical features and infrastructure which may serve as barriers to interaction and movement in the city. We believe there is great potential for the neighborhood boundary ideas here to be analyzed in relation to residents' place identity and pursue further questions related to sense of place, residential location decisions, and social cohesion within urban neighborhoods.

References

1 Crystal J. Bae and Daniel R. Montello. Representations of an urban ethnic neighbourhood: Residents' cognitive boundaries of Koreatown, Los Angeles. *Built Environment*, 44(2):218–240, 2018. doi:10.2148/benv.44.2.218.

2 U.S. Census Bureau. American Community Survey 5-year estimates. Technical report, U.S. Census Bureau, 2022. URL: https://www.census.gov/programs-surveys/acs/data.html.

3 Ernest Watson Burgess and Charles Shelton Newcomb. *Census data of the city of Chicago, 1920.* University of Chicago Press, 1931.
4 Elizabeth Campbell, Julia R. Henly, Delbert S. Elliott, and Katherine Irwin. Subjective constructions of neighborhood boundaries: Lessons from a qualitative study of four neighborhoods. *Journal of Urban Affairs*, 31(4):461–490, 2009. doi:10.1111/j.1467-9906.2009.00450.x.
5 Helen Couclelis, Reginald G. Golledge, N. Gale, and Waldo Tobler. Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology*, 7(2):99–122, 1987. doi:10.1016/S0272-4944(87)80020-8.
6 Nicholas S Dalton and Mark Hurrell. Methods for neighbourhood mapping, boundary agreement. *Environment and Planning B: Urban Analytics and City Science*, 50(2):401–415, 2023. doi:10.1177/23998083221115195.
7 Roger M. Downs and David Stea. Cognitive maps and spatial behavior: Process and products. In Roger M. Downs and David Stea, editors, *Image and Environment: Cognitive Mapping and Spatial Behavior*, pages 8–26. Aldine Publishing Company, Chicago, 1973.
8 Ekaterina Egorova and Crystal J Bae. The image of the city by temporarily displaced children: How place-based citizen science contributes to place discovery. In *PLATIAL'X Symposium Series*, 2023. doi:10.5281/zenodo.8286261.
9 Chicago Metropolitan Agency for Planning. Community data snapshots. Technical report, CMAP, 2023. URL: https://www.cmap.illinois.gov/data/community-snapshots.
10 George Galster. On the nature of neighbourhood. *Urban Studies*, 38(12):2111–2124, 2001. doi:10.1080/00420980120087072.
11 Wenwen Li, Michael F. Goodchild, and Richard Church. An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems. *International Journal of Geographical Information Science*, 27(6):1227–1250, 2013. doi:10.1080/13658816.2012.752093.
12 Andrew Lohmann and Grant McMurran. Resident-defined neighborhood mapping: Using gis to analyze phenomenological neighborhoods. *Journal of Prevention & Intervention in the Community*, 37(1):66–81, 2009. doi:10.1080/10852350802498714.
13 Kevin Lynch. *The Image of the City.* MIT Press, 1960.
14 Daniel R. Montello, Michael F. Goodchild, Jonathon Gottsegen, and Peter Fohl. Where's downtown?: Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition & Computation*, 3(2-3):185–204, 2003. doi:10.1080/13875868.2003.9683761.
15 Chicago Historical Society. Community Areas, The Encyclopedia of Chicago. Technical report, The Newberry Library, 2005. URL: http://www.encyclopedia.chicagohistory.org/pages/319.html.
16 Christopher Spencer and Marie Weetman. The microgenesis of cognitive maps: A longitudinal study of new residents of an urban area. *Transactions of the Institute of British Geographers*, 6(3):375, 1981. doi:10.2307/622295.
17 Emily Talen. *Neighborhood.* Oxford University Press, 2018.
18 The Los Angeles Times. Mapping L.A. neighborhoods, 2009. Accessed on 15 April, 2024. URL: https://maps.latimes.com/neighborhoods/index.html.
19 The New York Times. An extremely detailed map of New York City neighborhoods, 2023. Accessed on 15 April, 2024. URL: https://www.nytimes.com/interactive/2023/upshot/extremely-detailed-nyc-neighborhood-map.html.
20 Jon West, Jacob Toye, and Leaflet. Leaflet.draw. https://github.com/Leaflet/Leaflet.draw, 2017.
21 Andy Woodruff and Tim Wallace. Bostonography. Accessed on 15 April, 2024. URL: https://bostonography.com/.