International Cartographic Association
Association Cartographique Internationale

The Roaring ‘20s Project: Mapping Pre-Renewal Built Environments in Chicago from Sanborn Fire Insurance Maps Using a Computational Workflow

Parker Otto[a,*] and Yue Lin[a]

[a] *University of Chicago, Parker Otto – parkerotto@uchicago.edu, Yue Lin - liny2@uchicago.edu*

* Corresponding author

Keywords: urban planning, historical preservation, Sanborn fire insurance, digitization, digital humanities

Abstract:

Sanborn Fire Insurance maps are some of the most complete records of the historical built environment that researchers have access to, with building-level data for over 12,000 North American cities contained in large atlases that date back more than a century. These maps have become invaluable resources for digital humanities research, helping to visualize, disseminate, and interpret the history of urban environments (Ross, 1971; Krafft, 1993). Figure 1 shows an example of one such map. A key challenge in working with these maps, however, is ensuring the preservation and accessibility of their information through digitization, which has traditionally been largely contained to tedious hand-tracing methodologies. While there is existing research to extract building information through machine learning, the methodology is costly in both time and processing power due to the need to train large datasets, which can limit its scalability to broader areas (Tollefson et al., 2021; Lin et al., 2023).

Figure 1. Example Map from Library of Congress Collection Sanborn Atlas of Chicago Volume 16 Sheet 4 (1926).

This paper contributes a novel, open, and scalable computational workflow for extracting building footprint information from Sanborn maps and transforming them into tidy vector polygons that will work universally for all Sanborn atlas sheets. The workflow includes four steps. First, maps are downloaded from the Library of Congress collection and are georeferenced using QGIS. Second, the georeferenced maps are processed in Python by detecting light RGB pixels and are cleaned using raster manipulation techniques including binary opening, closing, dilation, and erosion to produce a binary image that is separated into footprints and background. Next, property lines and building outlines are isolated by detecting dark RGB pixels and are filtered to create stencils that serve to remove as much noise from the binary image as possible and to ensure that buildings are split apart. Finally, the images are polygonised in QGIS and run through a post-processing workflow including hole removal, buffering, simplification, orthogonalization, and minor manual corrections (the manual step is necessary to ensure that wall-to-wall buildings are properly split). The full source code of the workflow is available at https://github.com/parkerotto/R20. The computational workflow is applied to 33 Sanborn maps for Hyde Park, Chicago from 2 atlases.

Abstracts of the International Cartographic Association, 10, 220, 2025
32nd International Cartographic Conference (ICC 2025), 17–22 August 2025, Vancouver, Canada.
https://doi.org/10.5194/ica-abs-10-220-2025

Beginning in the late 1940s, Hyde Park underwent extensive urban renewal, which involved the destruction of hundreds of buildings deemed “blighted” under racialized policies. The Sanborn maps used in this study, retrieved from the Library of Congress’s scanned digital archive, date back to 1925-6 (prior to the renewal efforts). Figure 2 shows the building footprints extracted from these maps using the proposed computational workflow. Both visual and quantitative inspections suggest that the proposed workflow extracts footprints with high accuracy, given an average F-Score of 0.9385 when comparing model-extracted footprints to manually labeled ones. This is a large improvement in accuracy compared to leading methods in Sanborn building footprint extraction via machine learning pioneered by Lin et al. (2023) which averaged an F-Score of 0.5851 on the same data. Our method is also significantly faster than using machine learning—processing a high-resolution (8,248 x 5,706 pixels) Sanborn map in 3 minutes vs. more than 2 hours. The speed and accuracy of this method can be used to model entire neighborhoods, which we demonstrate by mapping all of Chicago’s Hyde Park Community Area as it was in 1925-6. Figure 3 further compares the extracted commercial footprints from 1920 with contemporary ones, illustrating the impacts of university-led urban renewal the neighborhood. This comparison uncovers a severe decline in commercial space and dense housing throughout Hyde Park because of urban renewal’s sweeping demolitions and restructuring of the built environment.

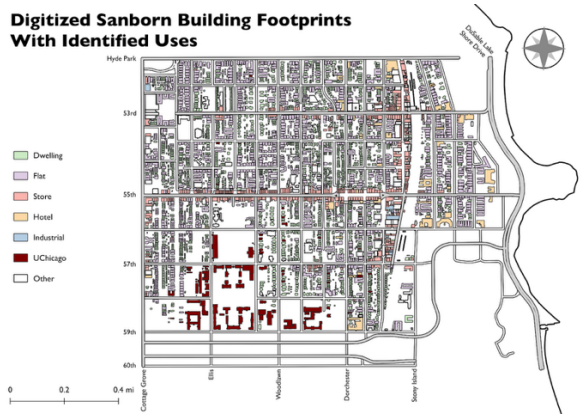

Figure 2. 1925-6 Hyde Park building footprints and uses.

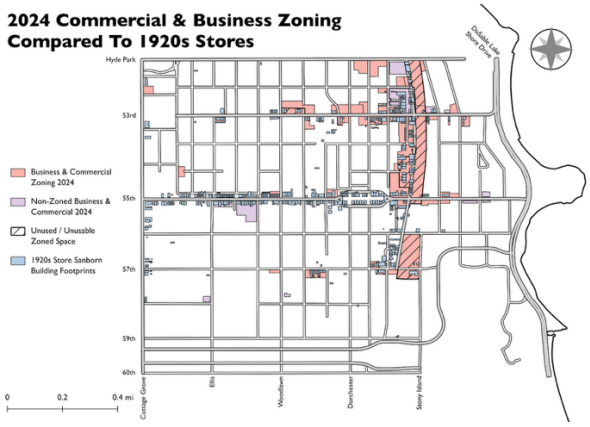

Figure 3. 1925-6 Hyde Park store building footprints compared to 2024 zoning for commercial and business uses.

References

Krafft, T. (1993). Reconstructing the North American urban landscape: Fire insurance maps - an indispensable source. *ERDKUNDE*, *47*(3), 196–211.

Lin, Y., Li, J., Porr, A., Logan, G., Xiao, N., & Miller, H. J. (2023). Creating building-level, three-dimensional digital models of historic urban neighborhoods from Sanborn Fire Insurance maps using machine learning. *PLoS ONE*, *18*(6), e0286340.

Ross, S. H. (1971). The central business district of Mexico city as indicated on the sanborn maps of 1906. *Professional Geographer*, *23*(1), 31–39.

Tollefson, J., Frickel, S., & Restrepo, M. I. (2021). Feature extraction and machine learning techniques for identifying historic urban environmental hazards: New methods to locate lost fossil fuel infrastructure in US cities. *PLoS ONE*, *16*(8), e0255507.

Abstracts of the International Cartographic Association, 10, 220, 2025
32nd International Cartographic Conference (ICC 2025), 17–22 August 2025, Vancouver, Canada.
https://doi.org/10.5194/ica-abs-10-220-2025