Senseable City Lab :.:: Massachusetts Institute of Technology

This paper might be a pre-copy-editing or a post-print author-produced .pdf of an article accepted for publication. For the definitive publisher-authenticated version, please refer directly to publishing house's archive system

SENSEABLE CITY LAB

PNAS RESEARCH ARTICLE | SOCIAL SCIENCES

Exploring the social life of urban spaces through AI

Arianna Salazar-Miranda[a,1,2], Zhuangyuan Fan[b,1], Michael Baick[c], Keith N. Hampton[d], Fabio Duarte[e], Becky P. Y. Loo[b], Edward Glaeser[f,2], and Carlo Ratti[e]

Affiliations are included on p. 8.

Contributed by Edward Glaeser; received November 25, 2024; accepted June 16, 2025; reviewed by Nathaniel Baum-Snow and Diego Puga

We analyze changes in pedestrian behavior over a 30-y period in four urban public spaces located in New York, Boston, and Philadelphia. Building on William Whyte's observational work from 1980, where he manually recorded pedestrian behaviors, we employ computer vision and deep learning techniques to examine video footage from 1979–80 and 2008–10. Our analysis measures changes in walking speed, lingering behavior, group sizes, and group formation. We find that the average walking speed has increased by 15%, while the time spent lingering in these spaces has halved across all locations. Although the percentage of pedestrians walking alone remained relatively stable (from 67% to 68%), the frequency of group encounters declined, indicating fewer interactions in public spaces. This shift suggests that urban residents increasingly view streets as thoroughfares rather than as social spaces, which has important implications for the role of public spaces in fostering social engagement.

computer vision | streetlife | pedestrian behavior

Significance

Urban public spaces have traditionally served as places for gathering and social connection, shaping the social fabric of cities. This study reveals important shifts in pedestrian behaviors over a 30-y period in four US public spaces. By using AI and computer vision to analyze historical and contemporary video footage, we observe an increase in walking speed and a decrease in time spent lingering, along with fewer group encounters. This trend suggests a growing perception of city streets as corridors for movement rather than spaces for social interaction. These findings highlight a changing urban dynamic, where efficiency increasingly shapes public space usage, potentially impacting social connections and the community-building role of these environments.

Do city sidewalks serve primarily for pedestrian mobility, or do they also serve as social spaces to connect people? Has their function changed over time, either because increasing urban wages have raised the opportunity cost of lingering on the street or because the demand for social interaction has gone up? In this paper, we observe pedestrian behaviors in four urban public spaces, analyzing videos taken 30 y apart. Using these data, we measure changes in walking speed and social interactions.

We use computer vision and deep learning to analyze archival video footage collected by William Whyte from 1979–80 (hereafter called 1980) and contemporary footage from 2008–10 (hereafter called 2010), focusing on public spaces in New York, Boston, and Philadelphia. This approach builds on the visual observation traditions of urbanists like Lynch, Jacobs, and Gehl, and Whyte (1–4). The pioneering observational work of Whyte (4) is the direct intellectual ancestor of this paper. Whyte documented pedestrian movement, lingering, and group formation using time-lapse video and manual coding. His work highlighted the importance of everyday behaviors, such as pausing to talk, walking in groups, or forming spontaneous interactions, in shaping the social dynamics of public spaces. While we look at the same spaces and even the same videos as Whyte (4), our analysis is essentially automated, which reduces the labor-intensive and subjective nature of manual video analysis.

In our application, we document the evolution of social dynamics in two public spaces in New York (outside of Bryant Park and the Metropolitan Museum), Boston (Downtown Crossing), and Philadelphia (Chestnut Street). These sites were initially analyzed by Whyte (4), and we had to use these locations in order to anchor our observations of changes over 30 y. We use these films to measure changes in group behaviors, focusing on walking speed, lingering behavior, group sizes, and group formation.

To calculate walking speed and lingering behavior, we first measure the speed for everyone in the film. We then define individuals who are moving at less than 0.5 m per second as lingerers. Everyone else is defined as a walker. We find that in all four locations, the share of lingerers decreased by 14 percent, and walking speeds increased by 15 percent, which suggests that urbanites are using streets more as walkways than as social spaces.

We do not, however, find any change in the share of pedestrians who are traveling alone. In 1980, 67 percent of pedestrians were walking by themselves, and this number only changed to 68 percent in 2010. That seeming stasis, however, masks two significant changes in individual sites. Bryant Park experienced an increase in the share of dyads walking (15 percentage points), likely due to its development as a more prominent destination. Conversely, Downtown Crossing showed a decline in dyads (10 percentage points),

Author contributions: A.S.-M., Z.F., F.D., and C.R. designed research; A.S.-M. and Z.F. performed research; A.S.-M., Z.F., K.N.H., B.P.Y.L., and C.R. contributed new reagents/analytic tools; A.S.-M., Z.F., and E.G. analyzed data; and A.S.-M., Z.F., M.B., and E.G. wrote the paper.

Reviewers: N.B.-S., University of Toronto Rotman School of Management; and D.P., Centro de Estudios Monetarios y Financieros.

Competing interest statement: Dr. Glaeser and Dr. Baum-Snow coauthored two introductory-style papers in 2023.

Copyright © 2025 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]A.S-M. and Z.F. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: arianna.salazarmiranda@yale.edu or eglaeser@harvard.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2424662122/-/DCSupplemental.

Published XXXX.

PROOF NOT FINAL EMBARGOED

possibly reflecting its more utilitarian role focused on shopping and quick errands rather than social or leisurely activities.

Finally, we look at changes in group formation. Do people run into people on the street and then interact with them? We find a substantial decrease in the amount of group formation, meaning that there is less spontaneous interaction and/or less use of urban spaces as prearranged meeting spots.

We find that the nature of interactions has shifted significantly over time. People are now spending less time in public spaces and moving through them at a faster pace. These shifts reflect a change in how public spaces are used, potentially affecting their role in facilitating social engagement and everyday social connections.

Sidewalks, Transportation, and Entertainment

City streets have long been places of movement, entertainment, and social connection. People watching, window shopping, and running into friends have long been pleasures created by a stroll through a dense urban area. But what determines the balance between simply passing through and lingering in space, and what would cause that balance to shift over time?

The function of the sidewalk is shaped by both users' demand for speed and the inherent interest in the space. In our spaces, the Metropolitan Museum of Art (Met), in New York, stands out as a particularly unusual space, one which typically attracts large numbers of tourists who are typically not in a hurry for whom the space is particularly novel. We expect to see more lingering and lower speeds at the Met than at our other locales.

Downtown Crossing in Boston, is both a busy transportation hub, where people shift from walking to public transit and an area that is well endowed with retail shops with somewhat interesting windows (precisely because it has so many pedestrians in transit). Consequently, it is unclear whether lingering increased or decreased in Downtown Crossing between 1980 and 2010.

Chestnut Street in Philadelphia serves as an "in-between" space, functioning primarily as a pedestrian transit corridor between destinations. We anticipate higher walking speeds and lower lingering in this location, similar to the transit-focused aspect of Downtown Crossing.

One plausible explanation for changes over time is that significant increases in urban incomes have led to a greater demand for speed because time has become more valuable. According to County Business Patterns data for 2010, average annual earnings per employee were over 70,000 in the zip codes associated with both Downtown Crossing and Bryant Park, suggesting these areas might exhibit particularly fast walking speeds. In contrast, per-employee earnings were closer to 50, 000 in the other two zip codes associated with Chestnut Street and the Metropolitan Museum of Art, possibly indicating lower walking speeds and more lingering in these areas.

An alternative hypothesis is that an increasingly prosperous and safer urban environment has enhanced the appeal of public spaces, making them more engaging and attractive for pedestrians. Of our four areas, Bryant Park stands out as a location that has seen a marked improvement in safety over the 30-y period, potentially contributing to changes in how it is used (5). Chestnut Street has also gentrified significantly over the 30-y period (6).

Analyzing Video Footage of Public Spaces

Our dataset combines archival and contemporary video footage. For the archival component, we sourced 1980 videos from the Project for Public Spaces (PPS). This nonprofit organization was founded by Fred Kent, who worked closely with the renowned urbanist William "Holly" Whyte (4). This footage, consisting of Super 8 films captured by Whyte and his students from PPS, was the basis of Whyte's pioneering work analyzing cities with time-lapse videos. They were made available to us in a digital form thanks to the work done by Keith Hampton and his students.

Beginning in 2008, Hampton digitized parts of the Whyte archive and took new footage of the same sites during similar times of the day to guarantee comparability (7). To ensure temporal consistency between periods, we adjusted the frame rates of the archival 1980 videos to roughly match those of the 2010 videos (details in *SI Appendix*). The archival footage was taken from rooftops and building windows. The 2010 footage was taken from a lower altitude but still at an elevated view. Fig. 1 shows a sample of the 1980 and 2010 videos for the four sites in our analysis and highlights the location of these public spaces.

Summary statistics describing the duration, time of day, and date for these videos are in Table 1. Our analysis draws on 638 total minutes of footage. We only included frames where the camera was stationary and unobstructed. To ensure comparability across videos, we normalized all results to the frame level since the length of individual videos varied from 30 min to 3 h. Most of the videos were filmed at midday, except for Chestnut Street, which was captured in the morning.

Measuring Group Behaviors Using Computer Vision and Deep Learning

We combined computer vision and deep learning methods to detect and analyze group behaviors. Our methodology consists of three components: 1) pedestrian tracking and classification, 2) group detection, and 3) identification of encounters.

Pedestrian Tracking. To analyze the movement of pedestrians, we track individuals' trajectories across frames. We first detect pedestrians using a Convolutional Neural Network-based object detection model called YOLO (version 5) (8). We then track individuals across frames using a Deep-SORT framework (9). This allows us to trace pedestrians' trajectories and compute their speed. The YOLO model detects pedestrians in each frame, while Deep-SORT maintains the continuity of pedestrians across frames. This combination has also been previously used for pedestrian tracking (10–12).

YOLO is trained using the COCO 2017 dataset (13), which contains over 1.5 million labeled object instances (one for each time an object appears in the data) across 80 categories, including people, cars, and bicycles. In this study, we focus exclusively on detecting pedestrians, one of the object categories in COCO. To evaluate the performance of the YOLO model in detecting pedestrians, we use three metrics: precision, recall, and mean average precision (mAP). Precision measures the proportion of detected bounding boxes that correctly identify pedestrians, minimizing false positives (e.g., boxes drawn around nonpedestrian objects or empty areas). Recall measures the proportion of actual pedestrians in the frame that are correctly detected, minimizing false negatives (e.g., pedestrians present but not detected). Together, these metrics provide complementary assessments of the model's accuracy: Precision ensures we are identifying pedestrians, while recall ensures we are not missing any.

The pretrained YOLO model achieves a state-of-art mean average precision of 0.68 at IoU = 0.5 (mAP@0.5) for pedestrian identification in the 2010 videos. However, the model was less

PROOF. NOT FINAL EMBARGOED

Fig. 1. Historical and contemporary videos for the four public spaces. (*A*) Frame extracted from videos in 2010. (*B*) Frames extracted from videos in 1980. (*C*) Location of the study sites and their corresponding boundary. Boundaries are defined using the common field of view in both the archival and the contemporary videos in each location.

NOT FINAL
EMBARGOED

accurate when analyzing 1980 videos with lower resolutions. To improve its performance, we fine-tuned YOLO with 260 labeled images (containing 4,792 labeled pedestrians) from the 1980 videos. This fine-tuning process improved the model's pedestrian detection performance, increasing its mAP@0.5 from 0.435 to 0.781, precision from 0.673 to 0.902, and recall from 0.84 to 0.93 in the test data (See *SI Appendix*, section 1 for details).

Our study distinguishes between pedestrians who are in transit and those who are stationary. We classify pedestrians in these categories based on their speed. A pedestrian is labeled as "lingering" if their movement speed falls below 0.5 m/s for more than 5-s (results are also reported for different time thresholds in *SI Appendix*, Fig. S5*B*). Our choice of the 0.5 m/s speed threshold to mean stationary is based on a mean comfortable gait speed ranging from 1.2 m/s to 1.47 m/s (14, 15). This choice is also supported by the distribution of speed in our dataset, which is bimodal and has a clear mass of people (the people lingering) with speeds below 0.5 m/s. (see *SI Appendix*, Fig. S5*A* for speed distribution). This definition differs from previous work by Hampton et al. (7), who manually coded people as lingerers if they are in the same place in two consecutive frames separated by 15 s. Their more stringent definition results in a lower share of lingerers.

Group Detection. Our group detection algorithm is executed in three steps. First, we employ the DBSCAN (16) algorithm for every frame to cluster individuals who are in close proximity. Second, we calculate the vector representing the trajectory of

Table 1. Summary statistics of the selected sites

	(I)		(II)		(III)		(IV)	
	Bryant park		Chestnut street		Downtown crossing		MET	
City	New york city		Philadelphia		Boston		New york city	
Type	Crosswalk		Street		Street plaza		Grand stairs	
Date	1980-4-10	2008-12-05	1979-8-30	2010-05-19	1979-9-28	2010-05-21	1979-6-2	2010-08-08
Time of day	13:17:00	14:19:44	10:30	8:33:43	9:50	11:57:55	12:00	12:01:18
Total duration (mins)	101	77	31	202	43	63	85	36
Mean # pedestrian (frame)	3.78	17.21	1.38	3.09	20.67	10.32	11.10	28.87

Table reports the summary statistics of the video footage for all locations (column 1), Bryant Park (column 2), Downtown Crossing (column 3), and the MET (column 4). Bryant Park videos were recorded at the corner of 45th Street and 6th Avenue. Chestnut Street videos were recorded between 10th and 11th Streets. Downtown Crossing videos were recorded next to Macy's. MET videos correspond to the main stairs of the Metropolitan Museum of Art. The reported video length here only contains the time when the camera is stable and no obstruction in front of the camera. The mean # pedestrian per frame counts the pedestrian observed within the study area for more than 2 s.

Fig. 2. Lingering behavior in public spaces. (*A*) Percentage of people lingering per frame for all four sites (*Left*) and broken down by location (*Right*). (*B*) Percentage of singles lingering per frame for all four sites (*Left*) and by location (*Right*). See *SI Appendix*, Table S6 for more details.

each individual, comparing their movements between the current frame and the preceding one. The final step classifies individuals into groups. Two individuals are considered part of a group if they are positioned no more than 1.9 m apart for more than 2 s and I) their trajectory vectors are positively correlated at the time, indicating similar movement directions; or II) both maintain speeds below 0.5 m per second, suggesting that they are stationary.

The selection of the 1.9-m distance threshold is based on Edward T. Hall's sociological theory on personal distance. Hall argued that individuals within 1.2 m are in their "personal spaces" (17). The 1.9-m threshold we use accounts for the fact that the width of each person's shoulders is typically 35 cm.

Our method assigns a unique identifier to each pedestrian within each time frame and site when they are classified as part of a group. This approach also tracks changes in group membership as pedestrians join, leave, or stay within the group over the duration of the video.

To validate the group detection methodology, we randomly selected 80 frames (with over 3,000 pedestrians) from each location from 1980 and 2010 (See *SI Appendix*, Table S1 for details). A researcher reviewed each frame to determine whether individuals were alone or in groups. The performance of our approach is in line with results reported in previous studies (18, 19), with an overall accuracy rate (compared to the codings made by the researcher) of 87.91%.

Identifying Group Encounters. Our analysis of group formation distinguishes between preformed groups (i.e., those established prior to entering the public site) and group encounters, which may be either spontaneous encounters or preplanned interactions (i.e., groups that form during the observation period as people meet through the frames). We classify a group encounter as occurring if its members entered the site alone at least 5 s before they formed the group. The 5-s threshold accounts for the possibility that people in a preformed group may enter the video at either a slightly different time. *SI Appendix*, Fig. S4 provides a visual representation of this classification.

Results

This section describes the three main results from our analysis.

Result 1: People Linger Less in the Public Spaces Analyzed and Walk Through These Sites Faster Than in 1980. Fig. 2*A* reports the share of people lingering per frame for all four sites. We estimate that 43% of the people in these public sites were classified as lingering in 1980 based on their temporal speed. This share has since declined to about 26%, indicating that people are lingering less. *SI Appendix*, Fig. S5*B* shows that using a higher time threshold to define lingering (above 5 s at a speed below <0.5 m/s) results in a lower number of lingering pedestrians but continues to show a decline in their share over time.

NOT FINAL
EMBARGOED

Fig. 3. Speed in public spaces. (*A*) walking speed of pedestrians for all sites in 1980 and 2010. (*B*) variance of speed across people for all sites in 1980 and 2010. (*C*) walking speed variance across people by location.

PROOF NOT FINAL EMBARGOED

The decline in people's propensity to linger in public spaces is sizable and visible across all locations analyzed. For example, 54% of people in Downtown Crossing in 1980 lingered there, while only 14% lingered in 2010.

The decline in people's propensity to linger is also visible for singles (Fig. 2*B*), which shows that the observed decline is not simply a function of changes in group composition over time. Single pedestrians lingered 39% of the time in 1980 (as compared to a lingering rate of 54% for all people). Their lingering rate declined to 24% in the recent 2010 videos, which is a similar decline to that seen for all groups.

The decline in lingering time was accompanied by an increase in walking speed among those who do not linger. Fig. 3*A* reports statistics for average and median speed in 1980 and 2010, pooling all sites. These statistics are computed only for people who are classified as movers (moving at speeds above 0.5 m/s). In 1980, movers traversed the public sites in our sample at an average speed of 1.25 m/s (median speed of 1.22 m/s). These speed levels are comparable to those reported in the Nationwide Personal Transportation Survey from 1977. The survey asked respondents to report the distance and time spent on recent walking trips. There is a total of 80 trips reported for Boston, NY, and Philadelphia of more than 5 min and where people traveled 2 km or less, which should be roughly comparable to the trips in our video sample. The average speed (computed as distance over time) for these trips is 1.34 m/s (median 1.24 m/s).

More importantly, comparing these historical numbers to the ones estimated from recent videos on the same sites, we find that, by 2010, speed increased by 15% to 1.41 m/s (median speed 1.41 m/s). This increase in speed was also accompanied by an increase in the variance of speed across people (*B*), with the SD in speed increasing from 0.32 in 1980 to 0.4 m/s in 2010. This change is visible across all the public sites in our sample (*C*). Fig. 3*C* also shows that the increase in walking speed and its dispersion rose in all four public sites by a similar amount. Among the four sites, pedestrians on Chestnut Street had the highest average walking speed in both 1980 and 2010 (1.33 m/s and 1.50 m/s, respectively). On the contrary, the Met has seen the lowest average walking speed in both years (1.20 m/s and 1.32 m/s).

Result 2: Public Spaces Are Dominated by Single Pedestrians. The average share of single pedestrians across the sites studied was 67% in the historical videos from 1980 and remained high, at 68%, in the recent videos from 2010.

This pattern is visible across all sites analyzed and is shown in Fig. 4*A*. In Chestnut Street, for example, the share of single pedestrians remained high at 79% both in the 1980 and 2010 videos (*B*). Even in the Met, we estimate that 48% of pedestrians are singles, with this share remaining constant in time. To ensure that our results are not driven by our choice of time threshold, we tested a range of thresholds from 0.1 to 10 s to define groups. *SI Appendix*, Table S3 and Fig. S3 show that singles continue to make up the majority across all sites, though at the Met, group proportions slightly increase with thresholds below 2 s.

Most of the remaining people in public spaces are in groups of two (dyads), accounting for about 27.7% in the 1980s and 26.7% in the 2010s, while a slim minority forms triads (3.53% in the 1980s, 3.49% in the 2010s) or larger groups (1.40% in the 1980s, 1.82% in the 2010s). The high prevalence of single-group pedestrians is in line with previous findings by Hampton et al. (7). These patterns are further summarized in *SI Appendix*, Table S2, which reports the share of people in groups of different sizes separately for each site.

Result 3: Decline in Group Encounters. We now ask whether groups are actually forming on these sidewalks. These groups may have either been preplanned meetings or spontaneous encounters.

One critical challenge in measuring group formation is to separate people who were together but entered the film at slightly different times from people who genuinely met in this space. We report the share of people that form at least one group s seconds after they enter the location and vary the value of s to allow for more stringent methods of measuring true meetings rather than straggling friends.

To put different values of s in perspective, Panel *A* in Fig. 5 plots the share of people who exit a public site (defined here as exiting the video frame) s seconds after they enter (with their entry defined by the first time they show on a frame), which we define as the churn rate. In 1980, 80% of the people entering a site exited within 20 s of their arrival. In 2010, this share went up to 92%. This increase in the churn rate is the natural

Fig. 4. Group dynamics over time in public spaces. (*A*) Percentage of different group sizes (singles, dyads, triads, and groups of four or more) observed across all locations in 1980 and 2010. See *SI Appendix*, Table S2 for details. (*B*) Percentage of pedestrians by group size for each location in 1980 and 2010. We also repeat the analysis using different time thresholds to define groups. The full result are reported in *SI Appendix*, Table S3.

PROOF
NOT FINAL
EMBARGOED

consequence of walking faster. However, changing churn rates over time means that it is particularly important to hold the time delay constant for defining group formation.

Fig. 5*B* reports the share of people that form at least one group s seconds after they enter the location—setting higher values for s guarantees that the groups were formed in situ and that the individuals did not arrive together at slightly different times. Using a baseline value of $s = 5$ s, we estimate that in 1980, 5.5% of people (one in every 20 people) entering one of our public spaces formed a group. By 2010, this percentage declined to 2% (one in every 50 people). The same decline is visible when using different thresholds s to define group encounters.

Fig. 5*C* shows this decline is pervasive across all locations. The decline is particularly pronounced in the Met (from 12.5% to 7%) and in Downtown Crossing (from 8% to 1%). While this trend reflects broad changes in how these spaces are used, our data cannot distinguish whether the decline is due to fewer prearranged meetings or reduced spontaneous interactions.

Conclusion

This paper shows how AI methods can analyze videos to study group dynamics in public spaces. This approach builds on visual observation traditions in urban studies while addressing the limitations of manual analysis, such as labor intensity and interobserver reliability. Our approach facilitates differentiation between various group sizes and identification of "lingering" behavior and spontaneous meetings. We apply this method to study four important public sites: Chestnut Street in Philadelphia, Downtown Crossing in Boston, Bryant Park in New York City, and the Metropolitan Museum of Art in New York City.

Our findings reveal several trends in these locations. First, public spaces appear dominated by single individuals and not so much by groups. Second, the nature of interactions has shifted significantly over time: People are now spending less time in public spaces and moving through them at a faster pace. Finally, group formation has declined significantly, suggesting that public spaces are becoming less conducive to both planned and unplanned social encounters.

These patterns are consistent with the idea that people balance their need for speed with the desire for engaging experiences in urban environments. While higher wages might be associated with increased walking speeds and reduced social interactions in so many settings (20), the Metropolitan Museum stands out as an exception, retaining high levels of lingering despite being located in a high-income neighborhood. This suggests that when urban amenities are sufficiently compelling, they can still encourage slower social behavior.

A second potential explanation for our findings considers the role of concurrent societal and technological changes. For example, the widespread adoption of smartphones by 2010 likely altered how people interact in public spaces, reducing the likelihood of encounters by drawing attention to digital screens rather than the immediate surroundings. To assess this, we trained a computer vision model to detect phone users in our videos. In the sites where this could be reliably measured, we found that phone use was relatively low during the 2008/2010 period (1.0% in Bryant Park, 3.4% in Downtown Crossing, and 8.3% in Chestnut Street), suggesting that smartphone adoption alone does not fully explain the observed behavioral shifts. If phone use were the primary driver, we would expect locations with the highest detected phone usage to show the

Fig. 5. Churn and group encounters in public spaces. (*A*) Churn rate, defined as the share of people who exit a public site, measured at different time intervals. (*B*) Share of group encounters, measured at different seconds after they enter the location. (*C*) Share of group encounters by location.

PROOF: NOT FINAL
EMBARGOED

most significant declines in social interactions, yet the behavioral shifts were comparable across sites. These results show that the documented changes in public space utilization preceded widespread smartphone adoption and cannot be attributed solely to this technological shift.

Another trend that may have influenced public space use is the proliferation of coffee shops and other indoor venues. Instead of lingering on sidewalks, people may have moved their social interactions into air-conditioned, more comfortable private spaces. While we do not have direct data to measure this, it is plausible that as indoor venues became more common, they provided alternative gathering spaces that reshaped how people engaged with public areas.

A third possible explanation is that the composition of people in some locations changed over time. For example, the area near Bryant Park had some prevalence of homelessness in the 1980s but has since been revitalized. Other areas have become more affluent, attracting different crowds. To explore these possibilities, we trained a computer vision model to detect the presence of homelessness. Both the model and our manual inspection of the footage revealed no evidence of homelessness in our locations in the 1980s, suggesting that this factor does not explain the higher lingering times observed at the time. We also attempted to train a computer vision algorithm to identify affluent visitors based on their clothing, but low image quality made the results unreliable.

Our findings also provide insights into the evolving role of urban public spaces and the challenges of designing environments that foster social interaction. The decline in lingering and group formation we document highlights the tension between creating open, accessible spaces and managing them to support broader use. While thoughtful design can make spaces more vibrant, overly open designs without appropriate management may lead to unintended uses, such as prolonged loitering, which can discourage other users. One potential approach could involve integrating public spaces with nearby businesses—such as adding sidewalk cafes, outdoor seating, or pedestrian-friendly retail areas—to foster interaction while providing natural oversight. Ultimately, urban design must carefully balance physical form and management strategies to ensure that public spaces remain dynamic hubs of activity.

The number of locations we surveyed is relatively limited, but this was intentional to ensure comparability with William Whyte's original 1980 study. While this constraint affects the generalizability of our findings, it enables a focused analysis of changes over 30 y in the same sites. The changes in pedestrian behavior we observe could be due to a variety of factors, including shifts in human behavior, land use, or the character of the space. Additionally, our analysis captures the copresence of different group sizes and frequencies but does not address the quality of these interactions. For instance, our encounter measure cannot distinguish between spontaneous meetings and scheduled ones, friendly encounters versus confrontations, or strangers versus acquaintances. Some factors, such as gender and income, remain challenging even for skilled human coders to identify (7). Future research could expand this work with larger datasets

and more nuanced categories of social interactions, potentially incorporating data from GPS, Bluetooth, and visual imagery.

Ethics Statement

This study involved only secondary analysis of archival video recordings collected in the 1980s and 2010s. No interaction with human participants was conducted so ethical approval was not required.

Data, Materials, and Software Availability. Aggregated data are available in a repository that can be downloaded upon request from the researchers. The code to produce the results is available in a public GitHub repository. Some study data are available: The underlying films cannot be shared partially because they contain recognizable human faces, but all of the data taken from the films will be available. That data can be reanalyzed and checked. Q:11

ACKNOWLEDGMENTS. We thank Yingjing Huang, Kaicheng Zhuang, Ken Lo, Anise Wong, and Daniel Wong for the excellent research assistance and Ricardo Alvarez for helpful early conversations on the topic.

Author affiliations: [a]Yale School of the Environment; [b]The University of Hong Kong; [c]Carlo Ratti Associati; [d]Michigan State University; [e]Senseable City Lab, MIT; and [f]Harvard University Q:12 Q:13 Q:14 Q:15

1. K. Lynch, *The Image of the City* (MIT Press, 1960), p. 194.
2. J. Jacobs, *The Death and Life of Great American Cities* (Random House Inc., New York, 1961).
3. J. Gehl, *Life Between Buildings: Using Public Space* (Island Press, 2011), p. 207.
4. W. H. Whyte *et al.*, The social life of small urban spaces (1980). Q:16
5. Project for Public Spaces, Bryant park (2024). Accessed November 5, 2024.
6. Pew Charitable Trusts, Philadelphia's changing neighborhoods: Gentrification and other shifts since 2000 (2016). Accessed November 5, 2024.
7. K. N. Hampton, L. S. Goulet, G. Albanesius, Change in the social life of urban public spaces: The rise of mobile phones and women, and the decline of aloneness over 30 years. *Urban Stud.* **52**, 1489–1504 (2015).
8. G. Jocher, YOLOv5 by Ultralytics (2020).
9. N. Wojke, A. Bewley, D. Paulus, "Simple online and realtime tracking with a deep association metric" in *2017 IEEE International Conference on Image Processing (ICIP)* (IEEE, 2017), pp. 3645–3649. Q:17
10. A. Brunetti, D. Buongiorno, G. F. Trotta, V. Bevilacqua, Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. *Neurocomputing* **300**, 17–33 (2018).
11. D. Stadler, J. Beyerer, "Improving multiple pedestrian tracking by track management and occlusion handling" in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 10958–10967. Q:18
12. B. P. Loo, Z. Fan, Social interaction in public space: Spatial edges, moveable furniture, and visual landmarks. *Environ. Plan. B Urban Anal. City Sci.* **50**, 23998083231160549 (2023).
13. T. Y. Lin *et al.*, "Microsoft coco: Common objects in context" in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13* (Springer, 2014), pp. 740–755.
14. R. W. Bohannon, Comfortable and maximum walking speed of adults aged 20–79 years: Reference values and determinants. *Age Ageing* **26**, 15–19 (1997).
15. E. M. Murtagh, J. L. Mair, E. Aguiar, C. Tudor-Locke, M. H. Murphy, Outdoor walking speeds of apparently healthy adults: A systematic review and meta-analysis. *Sports Med.* **51**, 125–141 (2021).
16. E. Schubert, J. Sander, M. Ester, H. P. Kriegel, X. Xu, DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Trans. Database Syst. (TODS)* **42**, 1–21 (2017).
17. E. T. Hall, A system for the notation of proxemic behavior. *Am. Anthropol.* **65**, 1003–1026 (1963).
18. W. Ge, R. T. Collins, B. Ruback, "Automatically detecting the small group structure of a crowd" in *2009 Workshop on Applications of Computer Vision (WACV)* (IEEE, 2009), pp. 1–8.
19. Q. Wang, M. Chen, F. Nie, X. Li, Detecting coherent groups in crowd scenes by multiview clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 46–58 (2018).
20. R. Putnam, Bowling Alone: The Collapse and Revival of American Community (2000).

PROOF
NOT FINAL
EMBARGOED

Author Query Form	
Query reference	**Query**
Q1	Please note that affiliations and citations have been renumbered to follow sequential order. Please check and confirm.
Q2	Please review 1) the author affiliation and footnote symbols, 2) the order of the author names, and 3) the spelling of all author names, initials, and affiliations and confirm that they are correct as set.
Q3	Please note that the spelling of the following author names in the manuscript differs from the spelling provided in the article metadata: Fabio Duarte and Edward Glaeser. The spelling provided in the manuscript has been retained; please confirm.
Q4	Certain compound terms are hyphenated when used as adjectives and unhyphenated when used as nouns. This style has been applied consistently throughout where (and if) applicable.
Q5	The keywords “streetlife” and “pedestrian behaviors” do not appear anywhere in the article. Please check and update the keywords or the text.
Q6	Please review the author contribution footnote carefully. Ensure that the information is correct and that the correct author initials are listed. Note that the order of author initials matches the order of the author line per journal style. You may add contributions to the list in the footnote; however, funding may not be an author’s only contribution to the work.
Q7	You have chosen to publish your PNAS article with the delayed open access option under a CC BY-NC-ND license. Your article will be freely accessible 6 months after publication, without a subscription; for additional details, please refer to the PNAS site: https://www.pnas.org/authors/fees-and-licenses. Please confirm this is correct.
Q8	If you have any changes to your Supporting Information (SI) file(s), please provide revised, ready-to-publish replacement files without annotations.
Q9	Abbreviations must be used at least twice in the main text; as such, CNN has been deleted.
Q10	PNAS articles should be accessible to a broad scientific audience; as such, please spell out the abbreviation YOLO.
Q11	Note that all data deposited in a publicly accessible database (and therefore not directly available in the paper or SI) must be cited in the text with an entry in the reference list. References must include the following information: 1) author names, 2) data/page title, 3) database name, 4) a direct URL to the data, 5) the date on which the data were accessed or deposited (not the release date). For an example reference entry, visit https://www.pnas.org/author-center/submitting-your-manuscript#manuscript-formatting-guidelines. Please also indicate where the new reference citation should be added in the main text and/or data availability statement. Please add a reference for the following data: GitHub repository.
Q12	Per PNAS style city, state/country, and post codes are required in all affiliations. Please provide the same in affiliations a to f.
Q13	There is no provided division for affiliations a and c. Please provide if this is available.
Q14	Please include the division or department with which the author is associated in affiliations b, d, and f.
Q15	There is no provided division/section/unit for affiliation e. Please provide if this is available.
Q16	Please provide the complete details for refs. 4–6, 8, and 20.
Q17	Please provide the editors name for ref. 9.
Q18	Please provide the editors and publisher name for refs. 11 and 13.