This is the Accepted Manuscript version of an article published in the journal *Nature Health* in 2026, which is available at: https://doi.org/10.1038/s44360-026-00151-9
Cite as: Yap W, Duarte F, Zheng Y, Jang KM, Luo P, Vineis P, Ratti C, Biljecki F (2026): Urban motifs associated with population health. *Nature Health*.

Urban Motifs Associated with Population Health

Winston Yap[1,2], Fábio Duarte[1], Yu Zheng[1], Kee Moon Jang[1,3], Peng Luo[1,4], Paolo Vineis[5], Carlo Ratti[1,6], Filip Biljecki[2,7,*]

[1]*Senseable City Laboratory, Massachusetts Institute of Technology, USA*
[2]*Department of Architecture, National University of Singapore, Singapore*
[3]*Thrust of Urban Governance and Design, Hong Kong University of Science and Technology (Guangzhou), China*
[4]*School of Earth, Environment, and Sustainability, University of Iowa, USA*
[5]*School of Public Health, Imperial College London, UK*
[6]*Department of Architecture, Built Environment, and Construction Engineering, Politecnico di Milano, Italy*
[7]*Department of Real Estate, National University of Singapore, Singapore*

Abstract

Where we live profoundly shapes our health, with urban environments playing a critical role in shaping population health outcomes. As health disparities persist within and between cities, ensuring equitable urban design has become critical to advancing population wellbeing. Yet most studies focus on case studies of single cities and overlook differences between general, physical, and mental health dimensions, limiting our understanding of how urban factors shape health outcomes at scale. To address this gap, we integrate census tract-level health data, crowdsourced geospatial information, and deep learning to identify urban features associated with general, physical, and mental health across the most populous urban areas in the United States. Our analysis reveals distinct associative relationships through which urban contextual and socioeconomic factors shape health outcomes. We identify the ranked importance of urban determinants for each health dimension, along with cross-cutting factors that consistently matter. Our findings suggest that urban service enhancements in low-income neighborhoods is associated with 100-462% greater health gains over high-income areas. Furthermore, we find strong links between the heterogeneity of urban spatial patterns and both health and income inequalities. Overall, our findings highlight strong associations between equitable access to urban services and coherent city planning with observed patterns of population health inequalities across cities.

Keywords: Graph Deep Learning, Healthy Cities, Urban Analytics, Urban Wellbeing, Sustainability

Introduction

Cities, often seen as engines of economic growth and innovation, display uneven benefits, contributing to significant health disparities among their residents [1–4]. For example, in the US, lifestyle-related chronic diseases dominate healthcare expenditures, costing US$4.9 trillion in 2023 [5]. Globally, deaths caused by non-communicable diseases have risen sharply, accounting for 73.9% of total deaths in 2019, up from 59.5% in 2000 [6]. The convergence of

*Corresponding author
Email address: filip@nus.edu.sg (Filip Biljecki)

unhealthy urban environments, rapid urbanization, and climate change exacerbates these disease burdens, as dramatically evidenced during the COVID-19 pandemic [7]. These disparities not only impact health but also influence long-term happiness, socioeconomic mobility, and climate resilience, ultimately affecting progress toward the Sustainable Development Goals (SDGs) [8–13].

The Healthy Cities movement of the 1980s marked a pivotal shift in acknowledging the broader responsibilities of urban environments in promoting health [14]. Driven by the realization that economic growth alone does not guarantee better health outcomes, the movement underscored the value of targeted urban planning and policy interventions in shaping healthier living conditions. Although cities in wealthier nations generally show better health indicators, economic growth does not automatically benefit all residents [15–17]. Low-income populations, in particular, often face an "urban penalty", with poorer health outcomes than both affluent urbanites and rural populations [18–20]. Factors such as urban sprawl, social segregation, and gentrification exacerbate these inequalities within cities [21–23].

As planners have worked toward designing healthier cities, it has become increasingly clear that sustainability and population health are deeply interconnected. For instance, initiatives that promote active transportation and expand access to clean energy can deliver substantial co-benefits for overall population well-being. These measures can improve air quality, which in turn helps reduce child mortality from acute respiratory infections [24, 25]. At the same time, growing evidence indicates that environment–health interactions are highly context dependent and shaped by cultural and social norms as well as economic conditions [26], underscoring that health gains cannot be assumed to translate uniformly across settings. This contextual complexity helps explain why health extends beyond SDG 3 (good health and wellbeing) and is explicitly linked to multiple SDGs, including SDG 1 (no poverty), SDG 10 (reduced inequality), and SDG 11 (inclusive, safe, resilient, and sustainable cities) [27–29]. Together, these linkages reflect a broader paradigm shift in which urban health is recognized as a cross-sectoral outcome, requiring coordinated, multi-stakeholder action to address health disparities and strengthen long-term resilience.

Being able to measure progress towards health policy targets and alignment with public health evidence is cornerstone for achieving healthy and sustainable outcomes [30]. In recent years, cities have increasingly sought to understand population health determinants across multiple scales, from global comparative studies to place-based analyses of intra-city disparities. Intercity efforts reveal how city-specific characteristics like spatial structure, size, and density shape disease prevalence [31, 32]. At the same time, factors such as walkability and urban greenery were found to have a largely universal positive influence on health outcomes across different cities. The important International Physical Activity and Environment Network (IPEN) study, covering 14 cities in 10 countries, found a curvilinear relationship where population (5,665 people km^2) and intersection (98 intersections per km^2) density thresholds were associated with more than an 80% probability of walking for transport [33]. Similarly, [34] found that unequal access to greenspace is associated with poorer health outcomes in over 1,000 cities globally, and evidence from 93 European cities suggests that expanding tree-canopy coverage to 30% could substantially reduce heat-induced premature mortality [35].

At the same time, the growing availability of high-resolution geospatial data has enabled data-driven approaches to examine socio-spatial health inequalities within cities [3]. These studies underscore the multifaceted ways in which urban environments and social determinants of health collectively shape population health outcomes [36–43]. For example, walkable neighborhoods are associated with better cardiovascular health, particularly in socioeconomi-

cally disadvantaged areas [26, 44], while zoning practices and residential segregation continue to entrench health inequities [45]. More time spent in urban nature, such as parks, rivers, or ponds, show strong links to better mental wellbeing [46]. On the other hand, neighborhoods characterized by high intersection density and limited greenery are linked to unhealthy diets and higher obesity rates [47, 48]. Urban features such as access to green space, shorter commuting times, and moderate residential density are generally associated with better mental health outcomes, although these effects are context-dependent and can vary by social and cultural settings [49–51]. Among young adults, urban design priorities extend beyond the physical environment to include access to safe public spaces, stable employment, and participatory planning processes [52].

Although substantial research links urban environments to health disparities, large-scale patterns in health-environment interactions remain poorly understood. Most high-resolution studies focus on single cities, which may suffer from geographic bias and limit the broader applicability of their findings to heterogeneous urban contexts [53, 54]. A key challenge lies in disentangling the effects of population characteristics from those of the built and natural environment, as well as consistently quantifying these relationships [55, 56]. Furthermore, while recent large-scale and fine-grained population studies represent important progress [57–61], most do not explicitly link urban environmental factors to health specific indicators or systematically compare effects across physical, mental, and general health. As a result, it remains unclear whether the association of urban contextual factors with health outcomes are consistent across different urban settings. For example, communities with a higher prevalence of poor mental health conditions may be associated with different urban design priorities than those dealing primarily with poor physical health. This lack of differentiation limits the development of targeted, actionable strategies to address specific crises at the neighborhood level.

To address these challenges, we conduct a comprehensive analysis that integrates high-resolution multimodal geospatial data with census tract-level indicators of general, physical, and mental health from the Centers for Disease Control and Prevention (CDC). We focus our analysis to the most populous urban areas within the contiguous US, encompassing 28,323 census tracts with high-resolution geospatial data coverage. Figure 1 provides an overview of our conceptual and methodological framework.

Our conceptual framework is guided by the approach of [62], modeling the nexus between city planning and population health by conceptualizing population health outcomes as the interplay between urban contextual factors (such as urban form, environment, accessibility, and perception) and compositional factors (poverty, education, financial stability, demographics, and living arrangements). We ground our selection of social determinants of health in the CDC PLACES framework, which identifies nine key measures, derived from 5-year estimates of the American Community Survey (ACS), to ensure policy relevance, comparability across contexts, and alignment with established public health frameworks. Our selection of urban contextual factors reflects indicators aligned with Global Burden of Disease (GBD) risk domains (e.g., air pollution, healthcare access, and food environments), but focuses on variables that are widely operationalized in urban planning practice and not in medicine [63]. To validate our approach, we perform extensive benchmarking and feature ablation (see SI Appendix, Table S1 for feature ablation and benchmarking results). Consistent with previous urban population health studies [64], our study examines three dimensions of urban health by measuring the absence of poor health in the general, physical, and mental categories. We obtain health outcome indicators from the 2023 PLACES dataset provided by the US Centers for Disease Control and Prevention (CDC). The PLACES health dataset employs small-area estimation with multilevel

regression and poststratification to provide prevalence of more than 40 chronic disease, including cardiovasular health, obesity, asthma, communicable diseases, and other health-related measures for adults aged 18 years and older at the census tract level. These include three health outcome measures of the proportions of adults reporting poor general, physical, or mental health for more than 14 days in the past year.

To overcome the limitations of conventional statistical and machine learning approaches, which often rely on coarse spatial aggregations and struggle to capture relational structures between urban elements, we employ a graph deep learning framework. This allows us to learn spatially explicit representations of the urban environment while achieving state-of-the-art predictive performance. Our graph deep learning model outperforms prior machine learning and deep learning approaches, achieving substantial improvements across all evaluation metrics (see Supplementary Table 1). Compared to the best-performing alternative, it reduced mean squared error (MSE) by 12.37% respectively for general health ($R^2 = 0.971$); by 11.10% for physical health ($R^2 = 0.948$); and by 11.50% for mental health ($R^2 = 0.920$). Through graph explainability methods, we reveal spatially generalizable patterns of urban feature importance and effect size, demonstrating the applicability of our findings across diverse urban contexts.

Our findings strengthen the evidence base showing how social determinants and urban contextual features are jointly associated with multiple dimensions of population health. Using urban greenery as a thematic example, we show that different forms of greenery support different health outcomes: street greenery is most strongly associated with general health, highlighting the value of everyday exposure during routine travel; local climate zones characterized by dense tree cover are most strongly associated with physical health, likely because they enable recreational activities such as walking, hiking, and nature-based exercise; and parks and other green and blue open spaces are most strongly associated with mental health. Across census tracts, we also identify a widespread health multiplier effect, where improving access to urban services in low-income tracts is associated with up to 462% larger health benefits than comparable improvements in higher-income tracts. Moreover, this multiplier effect increases with state-level per capita income (General Health: $p < 0.001$; Physical Health: $p < 0.001$; Mental Health: $p < 0.001$), suggesting that even affluent states can contain deeply deprived neighborhoods. Collectively, these findings underscore the importance of research-informed, evidence-based approaches that account for social inequality and tailor strategies to the differentiated needs of communities to improve population health.

Results

Nexus of Health Dimensions, Urban Factors, and Social Determinants of Health

Despite their interconnectedness, general, physical, and mental health represent distinct dimensions. Here, we systematically compare and establish the statistical relationships among general, physical, and mental health. We ground our selection of social determinants of health indicators in the CDC PLACES framework, which identifies nine key measures, derived from 5-year estimates of the American Community Survey. Using the inverse rating, we report the estimated percentage of adults (aged 18 and above) in good health, defined as those not experiencing poor health for 14 or more days within a year.

Figure 2A shows that health outcomes across tracts follow consistently left-skewed normal distributions, indicating that most areas are healthier than average. On average, most areas report better physical health outcomes ($\mu = 88.96$, $\sigma = 3.23$) than general health ($\mu = 83.39$, $\sigma = 7.12$) or mental health ($\mu = 83.86$, $\sigma = 3.66$). In addition, the variance in general health

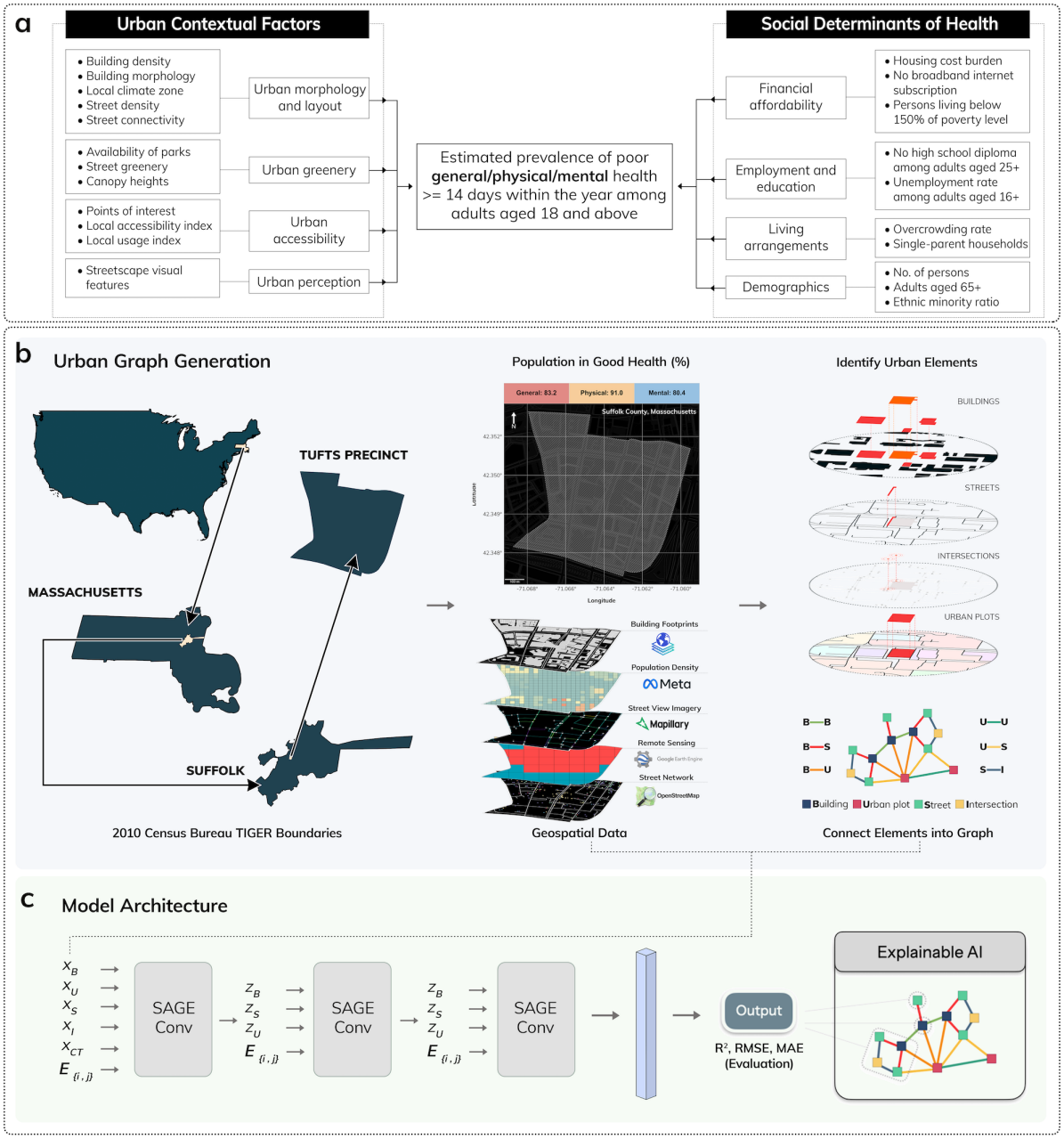

Figure 1: Schematic illustration of the data, methodology, and model architecture with an example of one of the 28,323 analysed census tracts. A) Conceptual framework informed by [62], depicting population health outcomes as arising from the interplay between urban contextual factors and socioeconomic characteristics such as poverty, education, financial stability, demographics, and household structure. B) Multi-modal data, including local climate zone maps, building footprints, street view images, population density maps, and urban points of interest (POI), are harmonized to generate urban analytical features for urban graph nodes [65]. Computational graphs are generated by connecting urban elements based on their spatial and topological relationship. As an example, urban plots that share a common street are connected as neighbors. Similarly, buildings are connected to their nearest adjacent street. B) We specify a three layer heterogeneous graph neural network with GraphSAGE convolutional encoding [66]. We implement our model with the Deep Graph Library [67] which provides implementation for the graph neural network explainer (GNNExplainer) [68] on heterogeneous graphs. (c) Basemap: Mapbox. Data sources: Centers for Disease Control and Prevention, OpenStreetMap contributors, Overture Maps Foundation, Mapillary, and Meta.

is noticeably larger, suggesting sharper urban inequalities across tracts. The systemic difference between health dimensions is further validated by pairwise Kolmogorov-Smirnov tests, which show statistically significant differences between general and physical (KS = 0.383, $p < 0.001$), general and mental (KS = 0.200, $p < 0.001$), and physical and mental (KS = 0.581, $p < 0.001$) health distributions. Figure 2B demonstrates a strong positive correlation between general and physical health ($r = 0.962$, $p < 0.001$), while the correlation between general and mental health, though positive, is weaker ($r = 0.757$, $p < 0.001$). This suggests that physical health is more closely associated with general health, whereas mental health exhibits more complex and heterogeneous relationships across urban contexts. Furthermore, we find significant disparities across health dimensions when factored by sociodemographic groups. Using paired independent sample t-tests, Figure 2C reveals significant health differences ($p < 0.0001$, denoted as ****) across key factors including poverty and minority ratios, levels of urbanization, and rates of aging. These observations confirm the importance of socioeconomic factors in evaluting population health outcomes. Figure 2D shows the spatial coverage of the most populous tracts and the overall distribution of health outcomes (average of general, physical, and mental health). We observe significant spatial clustering in health outcomes across tracts (Moran's I: 0.705, $p < 0.001$), using Queen contiguity spatial weights. This clustering is especially evident in cities such as Detroit and Los Angeles, where areas with poor health are spatially concentrated. Overall, our findings underscore the importance of recognising the interconnectedness as well as systemic differences between health dimensions, and highlight the need for differentiated strategies to improve overall health outcomes.

Guided by our conceptual framework (illustrated in Figure 1A), we select a comprehensive set of urban contextual indicators derived from urban analytics literature (see Supplementary Table 2 which lists these indicators along with their corresponding references). To assess the relative importance of both urban contextual and social determinants of health in shaping general, physical, and mental health outcomes across census tracts, we applied state-of-the-art deep graph explainability methods [68]. Figure 3A-C shows the ranked feature importance of urban factors and social determinants of health for general, physical, and mental health. For tractability, we present only the top 30 factors within each health dimension. Each variable is further categorized according to its urban component to aid comparison between health dimensions. In what follows, we denote each factor as G (General Health), P (Physical Health), or M (Mental Health), followed by a number indicating its ranked importance for each health dimension.

Urban form stands out as one of the most influential factors for population health. Among the top-ranked features for general health are Plot Rectangularity (G-1), Plot Corners (G-2), Plot Building Elongation (G-4), and Plot Building Perimeter Spread (G-8), which reflect how rectangular plots and elongated buildings can enhance commercial foot traffic, storefront visibility, and access to amenities-a pattern visible in places like Midtown Manhattan and Boston's Back Bay. In relation to physical health, features such as Plot Building Coverage (P-1), Plot Building Rectangularity Spread (P-2), and Plot Squareness (P-5) suggest that denser, more compact urban forms promote walkability and active lifestyles. Mental health appears most closely linked to factors like Street Clustering Coefficient (M-1), Plot Building Fractal Dimensionality Spread (M-2), Plot Building Squareness Spread (M-3), Plot Circular Compactness (M-7), and Plot Building Convexity (M-12), indicating that areas with diverse, organic forms and highly interconnected street networks can foster visual interest and less monotonous environments.

Urban greenery also plays a nuanced and essential role across all health dimensions. Green View Factor (G-3), which captures street-level greenery, emerges as an important contributor

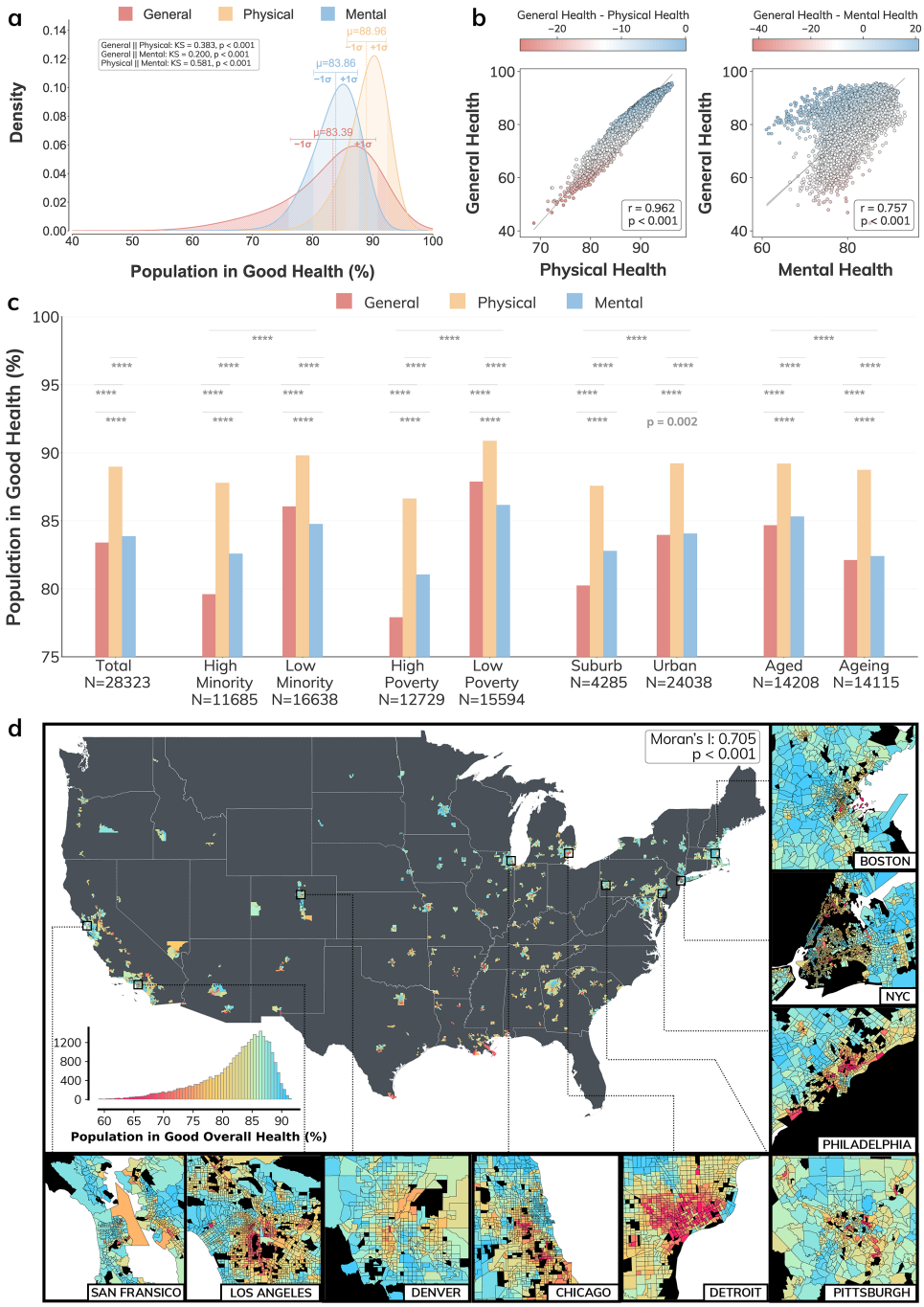

Figure 2: Statistical relationships and systematic comparison across health dimensions. a) Density distributions of general, physical, and mental health outcomes indicated by standard error across census tracts. Pairwise two-sample, two-sided Kolmogorov-Smirnov test for goodness of fit indicate statistically significant distributions between health dimensions. b) Two-sided pearson correlations show statistically significant positive relationships between general health and both physical and mental health with P-values less c) Paired two-sided t-tests show statistical significance across key socioeconomic factors ($P < 0.0001$, ****). For urban settings, statistically significant differences were also observed between general and mental health outcomes ($P = 1.55 \times 10^{-3}$). Categories based on minority ratio [69], poverty ratio [70], urbanization degree [71], and ageing ratio [72]. d) Spatial distribution of overall population health across the most populous census tracts, with significant Moran's I spatial autocorrelation test. Side panels highlight major cities. Data source: (c)US Centers for Disease Control and Prevention.

to general health, reflecting the everyday impact of urban greenery. The presence of Dense Trees coverage (P-3) is strongly linked to physical health, likely because areas with dense greenery create more opportunities for recreational activities such as hiking and nature walks. When considering mental health, features like Canopy Height Spread (M-6) and Parks (M-11) underscore the benefits of biodiversity and accessible public green spaces for psychological well-being. Collectively, these findings reinforce the need for cities to incorporate a diverse range of green infrastructure to support holistic health.

Local climate zone (LCZ) coverage demonstrates a strong and intuitive convergence with the earlier findings, further illuminating the interplay between urban form, greenery, and population health. LCZ classes, by definition, reflect combinations of built structure and vegetative cover, and the most important LCZ features for each health dimension align closely with patterns observed in urban form and greenery factors. Compact Lowrise (G-7) and Sparsely Built (G-14) coverage, which rank highly for general health, mirror the importance of accessible, well-structured environments interspersed with open areas and street greenery. In the case of physical health, Compact Highrise (P-8) and Heavy Industry (P-15) emerge as key features, highlighting how higher density and industrial land uses can influence opportunities for physical activity and exposure to environmental stressors. Open Highrise (M-4) and Water (M-5) are critical for mental health, aligning with the psychological benefits of open, green, and blue spaces within urban settings. LCZ findings provide further intuition into how the structure and character of the built environment, together with natural elements, jointly shape health across multiple dimensions.

Socioeconomic factors are also critical in shaping health. The significance of Minority Ratio (G-6) and Females per Plot (G-11) for general health highlights how racial and gender inequalities in neighborhoods can affect access to care, social support, and well-being. Features such as Internet Access (P-4), Population Size (P-13), and Females per Plot (P-17) influence physical health by improving access to health information, enhancing social resources, and contributing to urban vibrancy. Persistent gender disparities may affect opportunities for physical activity and access to health-promoting environments. In the context of mental health, Unemployment Rate (M-16) and Ageing Rate (M-19) are especially important, as higher unemployment and larger elderly populations are potentially linked to social isolation, psychological distress, and increased need for social support. These results underscore the importance of addressing social determinants alongside physical and green infrastructure to promote health equity in cities.

The availability and accessibility of diverse urban services play a critical role in shaping population health outcomes. Cultural institutions (G-5) and restaurants (G-13) are closely linked to general health, as these venues foster social engagement, leisure, and cultural interaction within communities. For physical health, both POI Accessibility (P-10) and the presence of healthcare facilities (P-14) are important, ensuring residents have access to essential medical services and amenities that promote active mobility. With respect to mental health, the availability of drugstores (M-10) stands out, suggesting the critical role of accessible first and last point health for mental health screening and medical advice.

Recognizing these differences, Figure 3D presents a systematic overview across health domains, highlighting the factors (in bold) that are important for all health dimensions. POI accessibility, urban compactness, heavy industry land cover, and cultural institutions consistently stand out as key influences across general, physical, and mental health. Policymakers concerned with population health may draw on these consistently salient factors as a coherent empirical reference when informing neighborhood-level planning and policy deliberations.

Our results characterize how different combinations of urban contextual factors and social determinants are associated with general, physical, and mental health outcomes. Depending on the prevailing health conditions in a given neighborhood, different sets of factors may be more relevant when considering neighborhood-level planning and health-related decision-making.

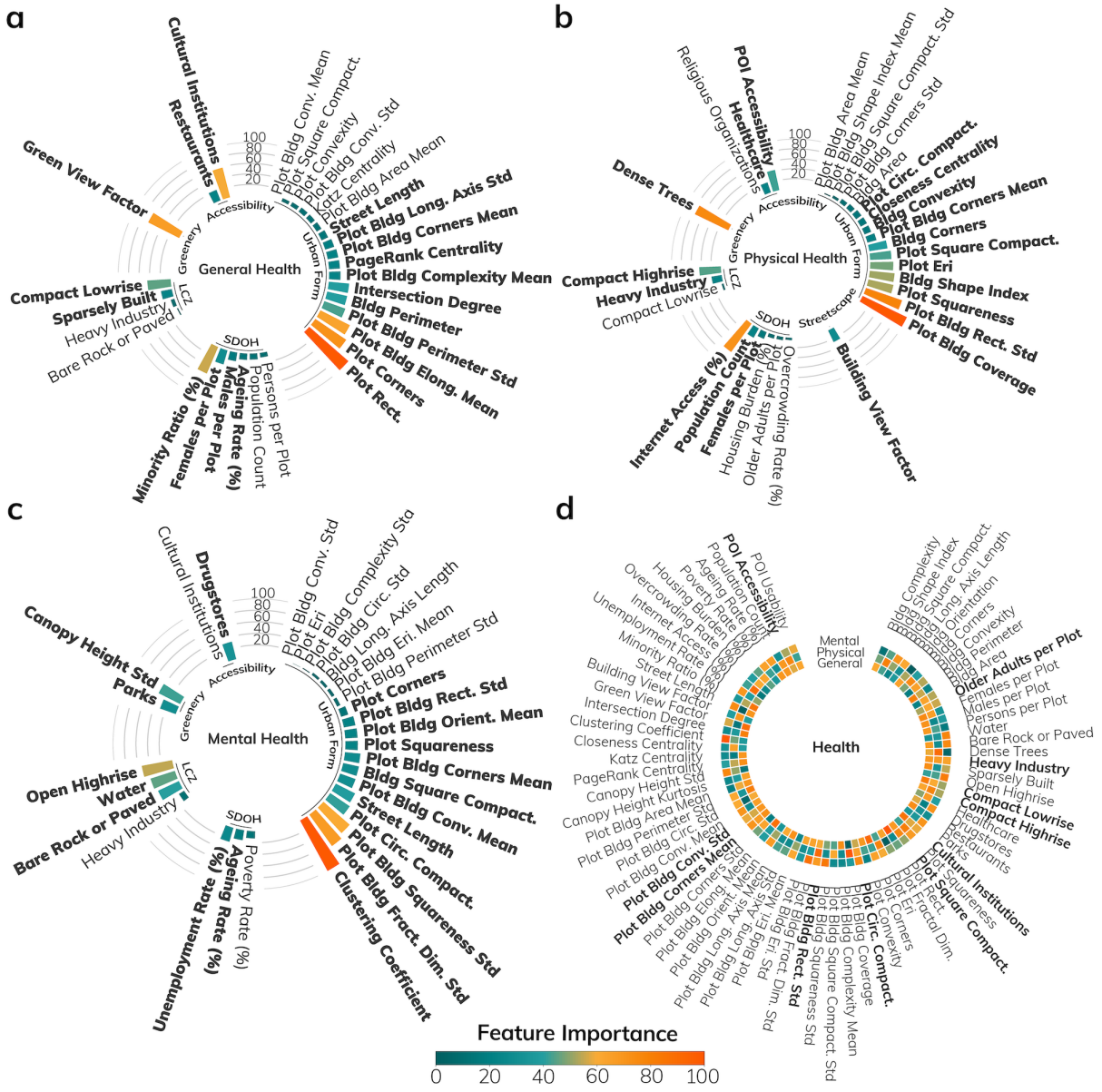

Figure 3: Ranked feature importance of the graph deep learning model, computed with GNNExplainer [68]. We show the top 30 urban features across all census tracts, categorized by their importance for explaining general, physical, and mental health outcomes. Importance scores are rescaled from 0 to 100 to reflect the relative contribution of each node feature to the model's predictions. The top 20 features are highlighted in bold. a) Factors relating to urban form and street-side greenery are important for general health. b) Essential factors for physical health include POI accessibility, healthcare amenities, compact urban form, and dense tree coverage, all of which enhance walkability and active mobility. c) Mental health is associated with the presence of open, green, and blue spaces, park availability, connected street networks, and consistent urban morphology. d) Stacked feature importance across health dimensions highlights universally critical features. Features with feature importance score exceeding 50 across general, physical, and mental health are bolded.

Posthoc Assessment of Associated Population Effects

While our previous findings highlight the differentiated importance of urban contextual and socioeconomic factors for general, physical, and mental health, tangible improvements to population health require more than identifying key determinants. For pragmatic design, urban planners and policymakers must translate these insights into actionable strategies in the face of persistent health inequities and limited resources. Here, our goal is to identify the factors that have the greatest impact in underserved neighborhoods by examining how targeted changes to urban features can influence health outcomes across census tracts with varying socioeconomic contexts. Specifically, we leverage our heterogeneous graph neural network model trained on all urban contextual and social determinant of health indicators (general health: $R^2 = 0.971$; physical health: $R^2 = 0.948$; mental health: $R^2 = 0.920$) to conduct a posthoc assessment, estimating how changes in urban features could improve or reduce health outcomes across census tracts with varying median incomes. This assessment involved applying graph perturbations to each tract and measuring the resulting changes in health outcomes for individual features (see Methods for further details and sensitivity analysis).

Our analysis shows that the greatest disparities in health outcomes across tracts with different income levels are driven by changes in accessibility to urban points of interest. Figure 4A shows the relationship between expected tract health prevalence effect per 10,000 persons and median income across different POI categories. We observe a multiplier effect where the associated effect of health improvements in areas with lower median incomes can exceed those in more affluent areas by up to 462%. Where *ME* denotes the multiplier effect, *CI* the 90% confidence interval, *HI* high-income tracts, and *LI* low-income tracts, the largest improvements in general health are associated with closing service gaps in healthcare (*ME* : 390%, *LI-CI* : $[4.3\times10^{-3}, 4.7\times10^{-3}]$, *HI-CI* : $[9.3\times10^{-4}, 1.4\times10^{-3}]$), schools (307%, $[5.8\times10^{-3}, 6.2\times10^{-3}]$, $[1.7\times10^{-3}, 2.2\times10^{-3}]$), POI accessibility (258%, $[6.5\times10^{-2}, 6.8\times10^{-2}]$, $[2.4\times10^{-2}, 2.8\times10^{-2}]$), and drugstores (203%, $[4.7 \times 10^{-3}, 5.3 \times 10^{-3}]$, $[2.1 \times 10^{-3}, 2.8 \times 10^{-3}]$). For physical health, the greatest benefits come from improving POI accessibility (*ME* : 250%, *LI-CI* : $[4.8 \times 10^{-2}, 5.0\times10^{-2}]$, *HI-CI* : $[1.8\times10^{-2}, 2.1\times10^{-2}]$), drugstores (228%, $[2.2\times10^{-3}, 2.6\times10^{-3}]$, $[8.6\times10^{-4}, 1.3\times10^{-3}]$), restaurants (170%, $[1.7\times10^{-3}, 1.9\times10^{-3}]$, $[9.4\times10^{-4}, 1.2\times10^{-3}]$), and healthcare (168%, $[9.4 \times 10^{-4}, 1.2 \times 10^{-3}]$, $[5.0 \times 10^{-4}, 7.6 \times 10^{-4}]$). For mental health, improvements to cultural institutions (*ME* : 462%, *LI-CI* : $[4.4 \times 10^{-3}, 4.7 \times 10^{-3}]$, *HI-CI* : $[8.1 \times 10^{-4}, 1.2 \times 10^{-3}]$), parks (446%, $[1.2 \times 10^{-3}, 1.6 \times 10^{-3}]$, $[-7.9 \times 10^{-4}, -3.7 \times 10^{-4}]$), religious organizations (327%, $[9.1\times10^{-4}, 1.1\times10^{-3}]$, $[-9.0\times10^{-4}, -6.9\times10^{-4}]$), and restaurant (174%, $[7.2\times10^{-3}, 7.7\times10^{-3}]$, $[4.0\times10^{-3}, 4.5\times10^{-3}]$) have the strongest positive impact. Overall, these findings reveal the extent to which urban inequalities are closely associated with variations in health patterns across communities.

Figure 4B shows a significant positive correlation between state-level per capita income and the health multiplier effect derived from improved POI accessibility for general health ($R^2 = 0.524$, $p < 0.001$), physical health ($R^2 = 0.560$, $p < 0.001$), and mental health ($R^2 = 0.507$, $p < 0.001$). Intuitively, this indicates that states with higher per capita incomes tend to exhibit greater health disparities between affluent and poorer areas, thereby producing a larger multiplier effect.

Relationship between Spatial Configuration and Dimensions of Urban Health

In this section, we further examine how spatial configurations are linked to health outcomes and socioeconomic status. Health outcomes are shaped not only by the characteristics of urban

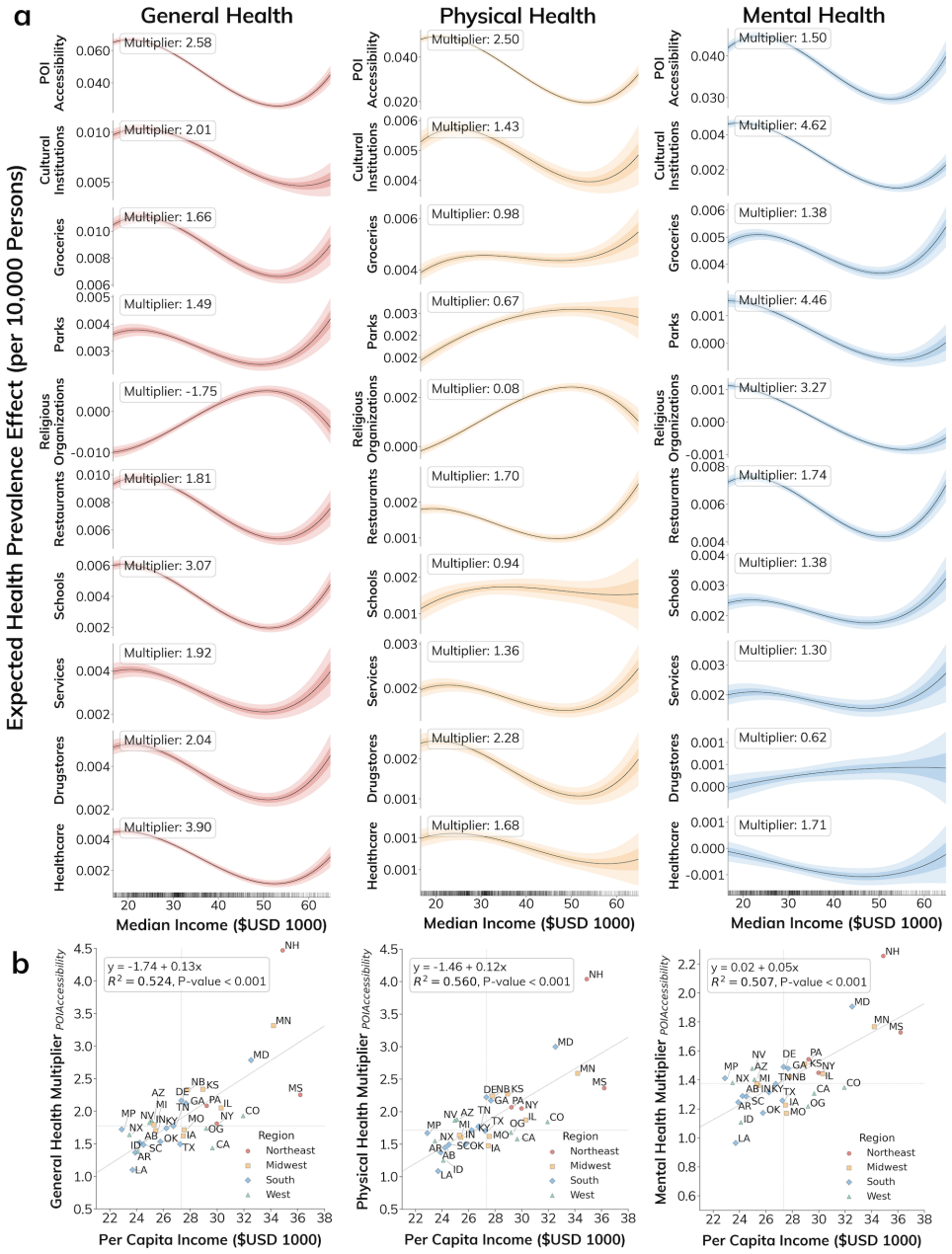

Figure 4: Relationship between expected health prevalence effect (per 10,000 persons), estimated via posthoc perturbation analysis, and median household income across census tracts. (A) A consistent health multiplier effect—defined as the ratio of expected health prevalence effects between the 75th and 25th percentiles of median household income across census tracts—is observed for POI accessibility and urban services. Each line plot is color-coded to show the 50% (darker shade) and 90% (lighter shade) confidence intervals around the mean response. Based on mean estimates, lower-income areas experience health improvements from urban services that are 100–462% larger than those observed in higher-income areas. (B) Scatterplots showing two-sided ordinary least-squares linear regressions between census-tract–level estimates and state-level aggregates of the health multiplier across census tracts. A statistically significant positive correlation is observed between the overall health multiplier associated with POI accessibility and state-level per capita income (General Health: $P = 4.13 \times 10^{-06}$, Physical Health: $P = 1.31 \times 10^{-06}$, Mental Health: 6.97×10^{-06}). Healthcare amenities similarly exhibit a strong positive multiplier effect on general health outcomes. All statistical tests were conducted as specified above. Because analyses were hypothesis-driven and limited in number, no adjustment for multiple comparisons was applied.

environments themselves, but also by the way different spatial entities, such as plots, buildings, and streets, are arranged and positioned in relation to each other.

To systematically capture these spatial patterns, we use the concept of graph motifs which are defined as small, recurring subgraphs that quantify the structural elements of the urban fabric. Intuitively, graph motifs represent characteristic connection patterns that appear frequently within a larger network. A familiar example is triadic closure, which occurs when two people who share a mutual friend are likely to become friends themselves, forming a triangle of social ties. These motifs serve as the fundamental building blocks of complex networks, and their frequency of occurrence has been widely used to characterize different types of systems and their underlying organizational principles [73]. In urban settings, for instance, triangular closed-loop street segments prevent cul-de-sacs and enable continuous movement, contributing to a walkable and permeable urban fabric where circulation can occur along multiple routes.

Urban graph motifs capture important spatial arrangements within cities, reflecting the layout and configuration of plots, buildings, and streets. For tractability, we compute graph motifs not on the entire graph, but on a smaller, connected subgraph for each tract, consisting only of the most important nodes identified by our model. We categorize motif types according to their cluster labels, using the sequence of plot type, dominant building type, and dominant street type. For instance, C-1-1 refers to a motif consisting of a plot of type C (Suburban Settlement), a dominant building of type 1 (Compact Blocks), and a dominant street of type 1 (Desolate Terrain). A detailed description of each cluster and the procedure for constructing graph motifs is detailed in Supplementary Figure 1.

Figure 5A illustrates the relationship between urban spatial motifs and both health outcomes and social determinants of health. Our analysis shows that areas characterized by balanced living, suburban settlements, and dense urban forms consistently exhibit moderate positive associations with health indicators and negative associations with adverse socioeconomic factors. For example, motif A-2-4, (balanced living, large complexes, and expansive landscapes) is positively associated with higher general and physical health scores. We also find that individual graph motifs often co-occur within a given tract, suggesting that it is not only the presence of specific motifs but also their composition and interaction that shape health outcomes. In fact, some motifs appear in both healthier and less healthy areas, indicating that pairwise association alone cannot fully explain variations in urban health.

To better understand the role of motif composition, we quantified motif diversity within each census tract using Shannon entropy. As shown in Figure 5B, motif entropy across tracts approximates a normal distribution, with some areas displaying much higher or lower diversity than expected. This pattern may reflect the outcomes of urban planning policies designed to promote uniformity or, conversely, permit greater variation in the built environment. Figure 5C demonstrates a clear association between motif diversity and health and socioeconomic outcomes. In particular, we find that tracts with lower motif diversity tend to have better health outcomes and lower proportion of population living below the poverty line. This finding suggests that consistency and cohesion in the urban fabric may play an important role in population health and social well-being, highlighting the value of coordinated planning regulations, such as zoning and design guidelines, that foster stable neighborhood development. On the other hand, urban landscapes characterized by high motif diversity or mixed and contrasting development patterns may signal uncoordinated growth or underlying social disparities, which in turn could negatively impact population health. Overall, our results highlight the central role of spatial configuration in shaping both health and socioeconomic trajectories in urban environments.

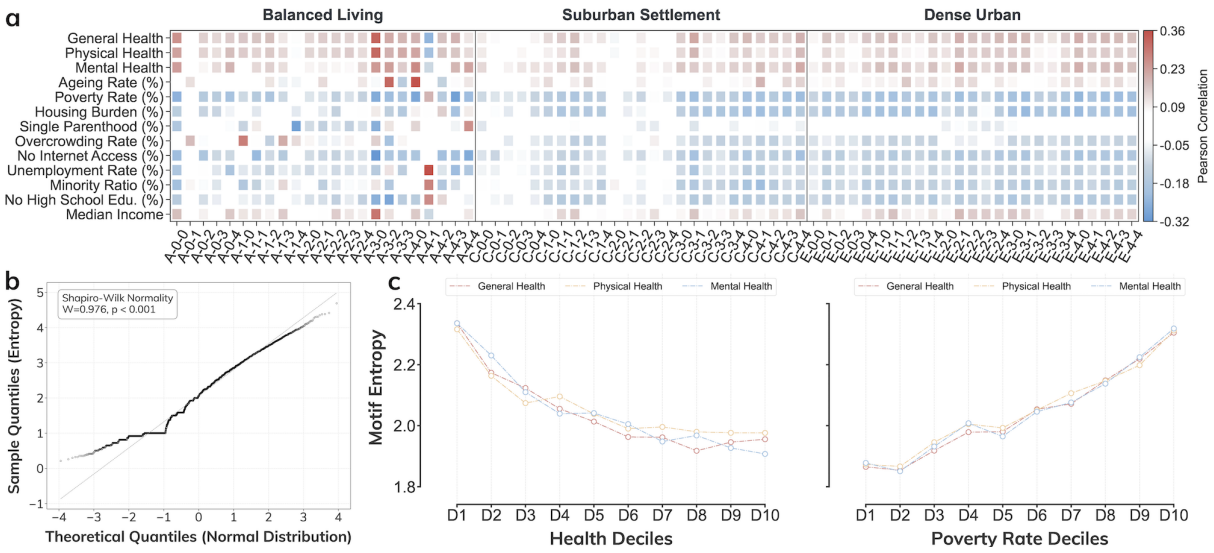

Figure 5: Relationship between the distribution and linear association of urban graph motifs with health outcomes and socioeconomic variables. a) Correlation matrix between urban graph motifs with health outcomes and social determinants of health. Urban areas characterised by balanced living, lowrise settlements, and dense urban show positive associated with health outcomes and negative association with poverty, unemployment, and housing. b) Motif entropy intuitively measures how diverse or uniform the mix of urban motifs is in an area—low entropy indicates highly coordinated, homogeneous development patterns, while high entropy reflects a more heterogeneous urban structure. Quantile–quantile plots and Shapiro-Wilk normality one-tailed test indicate that motif entropy deviates from a normal distribution, with a pronounced left-skewed tail. This pattern is consistent with the presence of urban areas characterized by more coordinated development patterns. c) Motif composition quantified by information entropy shows that more homogeneous and consistent urban development patterns are positively linked to better health outcomes and lower levels of urban poverty.

Discussion

The principle of "leaving no one behind" is a central commitment of the 2030 UN Agenda for Sustainable Development [6]. Yet substantial gaps remain in ensuring that systemically marginalized and underserved populations can fully exercise their rights and reach their potential. In the United States, urban health burdens tend to fall disproportionately on healthcare systems, making improvements to urban environments a sustainable pathway for advancing population health [74]. Achieving population-level health improvements ultimately requires cohesive and integrated urban planning, as demonstrated by international experience from well-planned cities such as Amsterdam, Vienna, and Singapore, as well as local examples in the United States, including Chicago's long-standing investments in public transit and pedestrian-oriented infrastructure and Seattle's efforts to advance 15-minute city planning. While our ranking framework offers a practical starting point for municipalities with constrained resources to prioritize investments, encompassing and inclusive urban health policy that target improvement of all indicators should remain the end goal. In this context, our analysis is intended to complement holistic governance frameworks by identifying empirically grounded patterns and associative leverage points that can inform the consideration and alignment of planning priorities within broader urban transformation agendas, rather than to substitute for comprehensive planning efforts.

As rapid urban growth continues to outpace the provision of essential infrastructure, transformative initiatives at local and state levels are required to disrupt entrenched urban health disparities and drive systemic change. Our analysis quantifies the health gains associated with reducing urban inequalities and identifies a widespread health multiplier effect across the most

populous urban areas in the contiguous United States. These findings align with established public health evidence showing that interventions tend to generate larger benefits for populations with elevated baseline risks and for socioeconomically disadvantaged groups [75, 76]. While the broad presence of this multiplier effect is encouraging, it simultaneously underscores the urgency of reallocating resources toward communities with the greatest unmet needs. However, existing neighborhood-level resource allocation mechanisms within US states constrain the extent to which investments from wealthier areas can be redirected to less advantaged communities. Addressing these structural limitations will require systemic policy reforms at both local and state levels to overcome persistent forms of resource redlining in marginalized neighborhoods.

Collaborative, community-embedded urban health initiatives provide an important foundation for advancing equitable city systems, but scaling them requires stronger spatial evidence and systems-aware planning. Programs such as the CDC's High Obesity Program (HOP) and Racial and Ethnic Approaches to Community Health (REACH) demonstrate how public health, urban planning, and community engagement can be integrated to support vulnerable populations. Consistent with built environment recommendations from the Community Preventive Services Task Force, our findings show that compact, activity-supportive urban form, service co-location through neighborhood hubs, and improved last-mile access to care are associated with better population health outcomes. When embedded within flexible, locally grounded frameworks, such as the Austin Healthcare Council's coordination of home-based early-life interventions through community health workers, these principles can strengthen community capacity while remaining adaptable to diverse urban contexts.

Transfer into practice, i.e. city planning, of our observations requires some additional steps. First, our data are cross-sectional, i.e. city characteristics and health data to the most recent available data. This implies that reverse causation cannot be excluded; that is, based on housing market values, poorer strata of the population, who also have worse health status, tend to be relegated to city areas with less advantageous profiles (typically, less green). This can be addressed with longitudinal investigations in future analyses. Second, we have provided a broad picture of the urban motifs that are associated with health status. Fragmented development patterns are consistently linked to poorer outcomes, while cohesive and internally consistent urban configurations—whether predominantly residential or consistently mixed-use—are associated with better health. These spatial patterns often cluster across adjacent tracts, highlighting the need for cross-neighborhood coordination. To establish causality requires access to intermediate variables to complete the causal pathways. The associations with urban motifs are open to multiple interpretations: motifs can be characterized by uneven environmental exposures such as air pollution or urban heat exposure [78], or can represent a distribution of population subgroups with unhealthy behaviours and less access to healthcare services.

Further causal assessment is needed before our findings become mature for policy choices. Translating these insights into scalable frameworks demands systems-aware and context-sensitive approaches, as human behaviour and health outcomes are highly dependent on local conditions [26, 79, 80]. Although we identify a broad set of urban and socioeconomic determinants, their relative importance varies across neighborhoods and likely operates through intermediate pathways. The broad picture we have provided from a city planning perspective needs to be complemented with plausible intermediate pathways, such as: (a) role of the market and market value of houses (including speculation), implying potential for reverse causality; (b) role of behavioural factors and commercial determinants of health (including e.g. diet); for example, obesity is one of the main drivers of city dwellers health,

and is related to a complex interplay of personal behaviours, commercial determinants of health and urban motifs [81]; (c) access to healthcare services. To make our observations actionable for city planners, longitudinal studies that include variables for these pathways are needed [76]. However, we believe that our broad picture provides an overarching scaffolding for the understanding of the social and material determinants of health and can guide analytical studies. This has been partially addressed in our analyses with graph perturbations, that can be interpreted as thought experiments.

Several limitations also warrant consideration. The 1978 WHO Declaration of Alma-Ata defines health as a state of complete physical, mental, and social wellbeing, not merely the absence of disease. Accordingly, our focus on general, physical, and mental health captures only a subset of the broader, multidimensional concept of health shaped by social and environmental determinants. Future research could extend this framework to incorporate outcomes such as chronic disease incidence, adaptive capacity, and health equity. In addition, our analysis is conducted at the census tract level, which inevitably masks within-tract health heterogeneity captured in more fine-grained studies and does not always align with administrative boundaries used in urban health planning. Nonetheless, tract-level analyses offer spatially resolved insights that are well suited for comparative analysis and, when aggregated, can support examination across service areas or policy jurisdictions. Finally, our study does not establish causal relationships between urban or socioeconomic factors and health outcomes. We further acknowledge that the associations identified between urban form and health outcomes, including their disproportionate impacts on systematically disadvantaged groups, are consistent with a well-established body of evidence accumulated over the past two decades. Accordingly, the contribution of this work lies not in establishing these relationships anew, but in spatially extending and quantitatively characterizing them at scale. Future work could investigate agglomerative effects arising from inter-tract composition and spatial configuration across entire urban landscapes. Integrating longitudinal data such as the National Health and Nutrition Examination Survey (NHANES) Longitudinal Study and large-scale human mobility information within a systems-based, graph-modeling framework represents a promising pathway to move beyond cross-sectional associations toward causal inference and the development of transferable foundation models for urban health analysis.

Methods

Our work applies a graph deep learning framework to 28,323 census tracts in the most populous urban areas of the contiguous United States, integrating tract-level health data, crowdsourced geospatial information, and sociodemographic variables to model associations between urban context and population health outcomes. To maintain alignment with our urban research focus, we excluded approximately 20,000 predominantly rural census tracts characterized by low population density and limited urban built-environment features. Including these tracts would introduce substantial heterogeneity in infrastructure, environmental conditions, and mobility patterns that are not well captured by the urban measures examined here. Consistent with prior urban-focused work [82], the resulting sample remains geographically broad and representative of urban populations across the contiguous United States. Table 1 presents summary statistics of our dataset for each state.

Model Architecture, Benchmarking, and Feature Ablation

Our prediction task involves an inductive graph-level regression task to predict the proportion of population in good general, physical, and mental health for each census tract. Here we

clarify that our objective is not to build a model that generalizes to completely new regions or countries, but rather to accurately learn the relationships between urban contextual and social determinants of health and observed health outcomes for urban populations within the contiguous US. Therefore, training and validating on spatially disjoint but proximate tracts is a design decision in our particular use case. Instead, it is more important to ensure that the model does not overfit to the specific features of the training set, allowing robust interpretation of health determinants across different areas.

We use random stratified sampling within US states and split the set of tracts (N=28,323) into training (80%), validation (10%), and test (10%) sets. The graph convolutional branch uses a three-layer heterogeneous GraphSAGE model [66] that integrates building features (15 dimensions), plot features (85 dimensions), street features (6 dimensions), intersection features (8 dimensions), and tract features (13 dimensions). At each layer, the graph neural network aggregates and encodes information from neighboring nodes based on adjacency relationships, generating a 128-dimensional representation. The final tract node embedding is then passed through two fully connected layers of 1024 dimensions each to predict the health outcome (one dimension) for each tract. For interpretability, we train three separate models that correspond to general, physical, and mental health outcomes respectively.

Each model configuration was trained for 50 epochs with hyperparameter tuning of hidden dimension size (64/128/256); learning rate (10^{-3}, 10^{-4}, 10^{-5}); batch size (16, 32, 64). We employ Adam optimizer with default parameters with default beta values of 0.9 and 0.999. All models were developed with the Deep Graph Library (DGL) framework [67]. We benchmark our approach against several state-of-the-art algorithms and feature set ablation results (see SI Appendix, Table S1). To compute model explainability, we run GNNExplainer [68] and posthoc assessment on the best-performing model based on the lowest mean squared error (MSE) and highest variance explained (R^2). On hardware, we utilize an NVIDIA RTX 4090 GPU for model training and evaluation.

GNN Explainability and Posthoc Perturbation

To explain our predictions, we use the widely validated and popular GNNExplainer [68] with heterogeneous convolutional implementation from DGL [67]. GNNExplainer, although computationally intensive, has been shown to identify consistent and compact subgraphs that maximize similarity between the predictions of the original and explained graphs. Model explanations are generated by optimizing the objective function:

$$l(y, \hat{y}) + \alpha_1 \|M\|_1 + \alpha_2 H(M) + \beta_1 \|F\|_1 + \beta_2 H(F),$$

where $l(\cdot, \cdot)$ is the loss function, M and F are edge and feature masks, y is the original model prediction, $\hat{y}$ is subgraph prediction, and $H(\cdot)$ denotes the entropy function. We use the default parameter settings for the heterogeneous GNNExplainer, namely, $\alpha_1 = 0.005$, $\alpha_2 = 1.0$, $\beta_1 = 1.0$, and $\beta_2 = 0.1$. Intuitively, the approach works by prioritizing shorter masks which results in more compact graphs, and chooses features that reduces uncertainty in prediction. GNNExplainer returns both node feature and edge connection importance. Node feature importance scores range from 0 to 1 with higher value corresponding to higher importance. We scale original feature importance scores between urban features to range between 0 to 1 to highlight relative differences. Node feature importance are obtained through masking based on thresholding of edge importance scores and retaining nodes.

To quantify the impact of different features, we utilize a graph perturbation algorithm that generates variations of the original graph by adjusting node feature values. Specifically, we

create a perturbed graph (P) by increasing the target node feature by a marginal value (1%; 5%; 8%; 10%) compared to the original graph (G). We find that our results remain robust across different magnitudes of marginal feature perturbation, indicating the consistency of the observed effects and show results based on the largest step size. For features that represent social determinants of health and are already expressed as percentages, we apply a smaller increment of one percent. We then analyze how these modifications affect the model's predictions by comparing outcomes between the perturbed and the original graphs. By applying this method across all graphs and node features, we calculate an expected feature effect value for each node feature.

To calculate the multiplier effect across various census tract median income levels, we fit a univariate polynomial linear regression for each feature. We then compute the ratio of the predicted values $\hat{y}$ between the 10th and 90th percentiles of median income:

$$\text{Multiplier Effect Ratio} = \frac{\hat{y}_{10\%}}{\hat{y}_{90\%}}$$

Urban Graphs Generation

Urban graphs provide a structured approach to modeling the complex, non-linear interactions among different urban elements. Following the approach of [83], we construct heterogeneous urban graphs for each census tract. These graphs include multiple types of nodes consisting of buildings (B), urban plots (P), streets (S), street intersections (I), and tract (T) to represent the diverse elements of the urban environment. A summary of the nodes and edges across US states is detailed in Supplementary Table 3.

To support population health modeling at scale, we substantially extend this urban graph framework beyond prior implementations. Health outcomes and many social determinants are typically reported at the level of administrative or planning units, rather than at the level of individual urban elements. To address this mismatch, we introduce a hierarchical graph architecture that explicitly links fine-grained urban entities (buildings, plots, streets, and intersections) to census tract–level nodes, allowing tract-level socioeconomic and health indicators to be integrated consistently with detailed urban morphology and accessibility features. This hierarchical design enables multi-modal and multi-scalar data integration while preserving the structural relationships between urban form and population-level outcomes. In addition, to ensure interpretability and compatibility with state-of-the-art graph explainability methods that attribute importance at the node-feature level, the graph representation was carefully redesigned to avoid redundancy and double-counting of urban attributes across layers (e.g., aggregating building morphology measures to the plot level). Finally, given the national scale of the analysis, we introduced large-scale geospatial data engineering optimizations, including cloud-native data formats, spatial tiling, memory caching, and parallel computation, to enable efficient and scalable graph construction across thousands of census tracts.

We use the Urbanity Python package [65] to generate a feature-rich embedding for each network node and its adjacent links. For our use case, we define urban plots that share a common street as neighbors, and buildings are connected to their nearest adjacent street. To capture spatial relationships between buildings, we create edges between those whose centroids are within 100 meters of each other, considering them neighbors.

Guided by our conceptual framework (detailed in Figure 1A), we select a comprehensive set of urban analytical features commonly used in urban planning practice-spanning network topology, building and plot morphology, street view indicators, land cover, population density, urban services, and mobility—to capture urban contextual characteristics (see SI Appendix,

Table S2). Studies have shown that street view features are strongly tied to urban well-being and can help estimate disease prevalence and mental health outcomes in cities [84]. Similarly, urban mobility factors, such as service distribution, accessibility, and intersection density, have been consistently linked to active travel behaviors and mental health outcomes [33, 82, 85]. Urban form factors such as local climate zones [86] and urban morphology [87, 101] influence pedestrian behavior, thermal comfort, and spatial segregation, all of which play crucial roles in shaping health outcomes. Informed by the Global Burden of Disease (GBD) framework, our feature selection also includes spatial indicators for key risk factors such as air pollution and spatial availability to healthcare access and nutrition [63].

Although the link between air pollution and health outcomes is well established, our study is limited by the lack of large-scale, high-resolution air pollution data. Previous research that identifies associations between air pollution and health typically relies on hyperlocal air quality measurements and focuses on fine-grained health outcomes, making it difficult to translate these findings to broader, population-level studies. This challenge is compounded by the time-lagged effects of air pollution on health outcomes [88], which are difficult to model at large spatial scales. Our preliminary analyses using large-scale, global $PM_{2.5}$ data [89] at 1 km grid size showed minimal regional variation across US census tracts and little correlation with general, physical, or mental health outcomes (see Supplementary Figure 3 for details). As such, we acknowledge the omission of high resolution air pollution data as a limitation of our study.

Non-communicable diseases can also be shaped by social network dynamics, as influential work has shown that conditions such as obesity may diffuse through interpersonal ties [90]. Social interactions therefore constitute an important component of the urban exposome. However, large-scale data on urban social network structures are not only scarce but also subject to significant geospatial biases, with certain communities systematically underrepresented. As a result, we do not explicitly model interpersonal ties. Instead, we approximate social exposure using aggregated, activity-based mobility patterns derived from mobile phone data, which capture co-location and movement but not underlying social relationships. We acknowledge this as a limitation and opportunity for future research.

Without loss of generality, we define a heterogeneous network as $G = \{N, E, R, T\}$:

$$n_i \subset N$$

$$n_i, r, n_j \subset E$$

$$T(n_i)$$

$$r \subset R$$

where N, E, R, T, refers to the set of nodes, set of edges, edge relation type, and node relation type respectively. In our case, node types correspond to $T = \{B, P, S, I\}$ and edge types $R = \{B_B, B_S, B_P, P_P, P_S, S_I\}$.

Urban Graph Motifs

To construct urban graph motifs, we first use an unsupervised k-means clustering approach to categorize key urban elements consisting of plots, buildings, and streets into distinct classes (see Supplementary Figure 1 for methodological details and workflow illustration) [77]. We then create urban graph motifs by examining each plot and its connections to the most common types of buildings and streets associated with it. Intuitively, these motifs capture typical configurations of urban spaces that define the character and function of different areas and collectively represent the fundamental building blocks that make up the urban fabric of cities.

US Health and Demographic Data

Our study adopts the approach of [64], defining health as the absence of poor general, physical, and mental health. We use the 2023 PLACES dataset from the Centers for Disease Control and Prevention (CDC), which integrates data from the Behavioral Risk Factor Surveillance System, the Census, and the American Community Survey, all aligned with 2010 Census Bureau TIGER boundaries and population counts. PLACES provides 40 chronic disease and health-related measures for adults aged 18 and older at the census tract level using small-area estimation with multilevel regression and poststratification. Among these are three health outcomes consisting of the proportion of adults reporting poor general, physical, or mental health for more than 14 days in the past year. For consistency, we also use the 2010 US Census Bureau TIGER boundaries and population counts, along with sociodemographic data from the 2015–2019 American Community Survey 5-Year Estimates.

To account for potential confounding effects among correlated social determinants of health, we conduct further experiments by modeling the annual changes in reported health metrics and these indicators from 2021 to 2023. We then use local aggregate effect plots to illustrate how specific features influence different dimensions of health (see Supplementary Figure 2). For instance, higher rates of overcrowding and older populations are associated with declines in physical and mental health, aligning with established urban epidemiological findings [91, 120].

Crowdsourced Data

Urban infrastructure comprises the physical spaces that support urban activities, including buildings, urban services, streets, green and blue spaces, and land use and form. We obtained urban road networks from OpenStreetMap (OSM), which is available under the Open Data Commons Open Database License (ODbL). Using the Pyrosm API to access daily updated raw OSM data from GeoFabrik, we simplified the road networks to a primal planar form for network analysis.

For building footprints and POI, we relied on Overture Map Foundation's Building and Places data themes. Launched by the Linux Foundation in December 2022, the Overture Maps Foundation complements crowdsourced OSM by prioritizing community-contributed data and scaling with state-of-the-art machine learning. Building footprints are available under the ODbL license, while the Places theme is combined from Facebook and Microsoft urban services data and released under the CDLA Permissive 2.0 license. To obtain building footprints and POI for each census tract, we use Overture's Python command-line tool which supports spatial query and data format conversion. In line with community recommendations [92], we exclude POI rows with confidence score lower than 0.6. Following [82], we focused on nine essential amenities consistent with the 15-minute city model [93]: restaurants, schools, parks, health care, drugstores, cultural institutions, grocery stores, services, and religious organizations.

We obtain street view imagery from Mapillary, a crowdsourced platform offering extensive global coverage under a CC-BY-SA 4.0 license. We selected all images within each census tract, linked them to the nearest street, and excluded any images located more than 10 meters from the road centerline. Eligible images were then processed with an image segmentation workflow, following [94], to extract semantic features. To address missing imagery in areas with sparse street view coverage, we implemented a spatial imputation strategy that aggregates semantic features at the tile level (Mapillary zoom level 14, 2.4 km at the equator) and fills missing segments with corresponding tile-level averages. Prior studies typically adopt either

a global imputation strategy, filling missing values using census-tract–level averages, which is computationally efficient but may obscure local heterogeneity, or local network-based approaches, which estimate missing values from nearby segments to better capture fine-grained spatial variation but are computationally intensive for large-scale studies. This approach balances computational efficiency with spatial specificity and outperforms global tract-level imputation in high-heterogeneity contexts such as greenery. Supplementary Figure 4 presents a comparison of imputation strategies across tracts of varying population sizes.

To obtain population estimates for each urban plot, we utilize Meta's high-resolution (1-arcsecond, 30-m) population estimates [95] hosted on the Humanitarian Data Exchange.

Google Earth Engine Raster Maps

Local Climate Zones (LCZs) [86] offer a universal typology for analyzing urbanization typologies, categorizing landscapes into 17 urban and rural classes based on combinations of micro-scale land covers and physical properties. This study utilizes a global LCZ map with 100-meter spatial resolution, available under the Creative Commons Attribution 4.0 International license [96]. We compute the proportion of each typology:

$$P_i = \frac{n_i}{\sum_{j=1}^{18} n_j}, \quad i = 1, 2, \ldots, 18. \tag{1}$$

where P_i is the proportion for i-th typology over 18 (including null class) possible local climate zone categories.

High-resolution (1-m) tree canopy height maps are vital for assessing global carbon stocks, forest structure, and biodiversity coverage. We used the canopy height maps developed by the World Resources Institute and Meta, accessed via Google Earth Engine [97], of which approximately 80% originated from imagery acquired between 2018 and 2020. Built using AI models applied to high-resolution Maxar satellite imagery covering over half the globe, this dataset enables precise canopy height predictions with a mean absolute error of 2.8-meters.

We adopt a hierarchical approach to compute raster statistics for high resolution canopy height maps. For each census tract, we download canopy height data with Google Earth Engine and store the numeric values and spatial information as a compressed array. Next, we rasterize urban plot shapefiles to obtain a categorical raster mask. We then use the raster masking operation to obtain the overlapping values with canopy height raster and compute aggregate statistics. To overcome memory constraint limitations for large urban plots, we split them into equal sized grids, compute aggregate statistics in each of them, before merging them together.

US-15 Mobility Data

We use measures of local, consumption-related trip behavior derived from 40 million US mobile GPS traces spanning 420 US cities [82]. Local Usage Index reflects the degree to which local POI drive neighborhood-level movement and indicates how effectively they meet the service needs of nearby residents. By contrast, the Local Access Index measures the availability of amenities within a given area, combining both the number of amenities and their relative importance (based on the number of trips they generate). Mathematically, these indices are defined as follows:

$$\text{usage}_A = \frac{\sum_{a \in A}(\text{usage}_a \cdot \text{population}_a)}{\text{population}_A}. \tag{2}$$

$$\text{access}_A = \frac{\sum_{a \in A}(\text{access}_a \cdot \text{population}_a)}{\text{population}_A}. \tag{3}$$

where usage_a and access_a refer to the census block group estimate, and population_a refers to the population of the census block group. The population_A refers to the aggregated population for each census tract.

To account for potential sampling bias and heterogeneity in device coverage [82], we supplement models with tract-level sociodemographic covariates (age, gender, race, and income) from the 2015–2019 American Community Survey and high-resolution population counts and subgroup estimates [95].

Data Availability

The model-based health outcome data at the census tract level are available at: `https://data.cdc.gov/500-Cities-Places/PLACES-Local-Data-for-Better-Health-Census-Tract-D/em5e-5hvn/about_data`. OpenStreetMap daily data extracts are available at `https://www.geofabrik.de/data/download.html`. The crowdsourced street view imagery can be obtained via Mapillary API with a registered developer key at: `https://www.mapillary.com/developer/api-documentation`. The Meta High Resolution Population Density Maps are available at `https://ai.meta.com/ai-for-good/datasets/high-resolution-population-density-maps/`. The Global Map of Local Climate Zone are available under Google Earth Engine API at: `https://developers.google.com/earth-engine/datasets/catalog/RUB_RUBCLIM_LCZ_global_lcz_map_latest`. High Resolution Global Canopy Height Maps can be obtained from `https://registry.opendata.aws/dataforgood-fb-forests/`.

We release data from our study on Figshare at `https://doi.org/10.6084/m9.figshare.31094953`. The repository contains data files for feature importance scores, census tract socioeconomic and health indicators, and graph motif association scores necessary to reproduce the reported results.

Code and Software Availability

The analysis was conducted using Python. The code to generate city graphs and reproduce main results are available at the following GitHub repository (`https://github.com/winstonyym/largescale_health`). All study data are included in the article and/or SI Appendix.

Acknowledgements

The authors gratefully acknowledge the contributions of the open-source community. We thank the members of the MIT Senseable City Lab and NUS Urban Analytics Lab for the discussions. The authors thank all members of the MIT Senseable City Lab Consortium: FAE Technology, Dubai Future Foundation, Sondotécnica, Seoul AI Foundation, Arnold Ventures, Sidara, Woven by Toyota, Abu Dhabi's Department of Municipal Transportation, A2A, UnipolTech, Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Hospital Israelita Albert Einstein, KACST, KAIST, AMS Institute, and the municipalities of Amsterdam, Laval, and Rio de Janeiro.

Funding Statement

This research was supported by the Singapore-MIT Alliance for Research and Technology (SMART), funded by the National Research Foundation Singapore under its Campus for

Research Excellence and Technological Enterprise (CREATE) programme. The first author thankfully acknowledges the NUS Overseas Immersion Award and NUS Graduate Research Scholarship granted by the National University of Singapore. This research is part of the project Large-scale 3D Geospatial Data for Urban Analytics, which is supported by the National University of Singapore under the Start-Up Grant R-295-000-171-133.

Author Contributions

Winston Yap 1) Methodology Conceptualization and Design; 2) Conceive study; 3) Methodology Development; 4) Data Acquisition and Analysis; 5) Data Testing and Validation; 6) Wrote Original Manuscript.

Fábio Duarte 1) Methodology Conceptualization and Design; 2) Conceive study; 3) Revise and Review Manuscript; 4) Research Supervision.

Yu Zheng 1) Methodology Conceptualization and Design; 2) Revise and Review Manuscript.

Kee Moon Jang 1) Methodology Conceptualization and Design; 2) Revise and Review Manuscript.

Peng Luo 1) Methodology Conceptualization and Design; 2) Revise and Review Manuscript.

Paolo Vineis 1) Revise and Review Manuscript; 2) Research Conceptual Refinement.

Carlo Ratti 1) Revise and Review Manuscript; 2) Research Conceptual Refinement.

Filip Biljecki 1) Methodology Conceptualization and Design; 2) Revise and Review Manuscript; 3) Conceive study; 4) Research Supervision; 5) Project Funding.

Competing Interests

The authors declare no competing financial or non-financial interests.

Tables

Table 1: Overview of dataset across states in the US.

State	#Tracts[1]	#Buildings	#Street View Images	#POI	Per Capita Income
Alabama	359	747,230	301,631	74,170	31,118
Arizona	1,083	1,983,069	314,301	165,854	34,206
Arkansas	123	274,145	26,173	30,967	30,837
California	5,971	9,521,973	1,222,519	948,732	38,986
Colorado	847	1,633,101	299,960	233,553	39,436
Delaware	147	225,328	52,036	25,994	35,257
Georgia	1,178	2,490,528	551,899	247,782	35,606
Idaho	103	343,293	21,490	31,614	31,510
Illinois	1,844	2,953,513	316,779	277,816	38,704
Indiana	622	1,341,840	144,780	112,010	32,826
Iowa	228	455,519	53,540	44,983	34,792
Kansas	326	562,190	99,995	49,243	37,905
Kentucky	346	698,909	73,693	59,153	32,820
Louisiana	510	907,106	99,818	91,531	29,457
Maine	57	126,384	35,062	15,066	36,716
Maryland	1,148	2,012,591	438,633	154,198	44,562
Massachusetts	1,133	2,029,550	353,506	200,150	42,351
Michigan	1,693	2,764,898	399,338	204,213	32,901
Minnesota	678	1,172,650	216,887	103,694	41,659
Mississippi	135	337,314	71,810	34,257	30,793
Missouri	622	1,120,810	141,124	109,121	35,123
Nebraska	234	389,109	52,343	33,442	35,923
Nevada	502	762,745	144,671	78,951	33,113
New Hampshire	139	270,720	69,240	27,970	39,651
New Mexico	167	321,885	28,678	29,442	30,517
New York	2,900	3,212,793	436,489	384,972	39,519
North Dakota	43	101,019	8,245	11,013	36,282
Oklahoma	380	659,417	71,201	60,153	32,628
Oregon	400	880,531	106,833	84,721	35,793
Pennsylvania	1,896	3,131,729	503,758	289,587	35,490
South Carolina	564	1,364,323	213,779	121,303	31,612
South Dakota	18	47,468	4,986	4,584	33,172
Tennessee	703	1,319,394	371,272	133,870	33,405
Texas	1,224	2,330,576	809,279	237,695	37,004
Total	28,323	48,493,650	8,055,748	4,711,804	30,445

[1] Census tracts correspond to the 2010 Census Bureau TIGER geographic boundaries to ensure alignment with the CDC PLACES 2023 dataset.

References

1. For Health Development, W. H. O. C. *Hidden cities: unmasking and overcoming health inequities in urban settings* 2010 (World Health Organization, 2010).
2. Sarkar, C., Webster, C. & Gallacher, J. *Healthy cities: public health through urban planning* 2014 (Edward Elgar Publishing, 2014).
3. Boeing, G. *et al.* Using open data and open-source software to develop spatial indicators of urban design and transport features for achieving healthy and sustainable cities. *The Lancet Global Health* **10,** e907–e918. 2022 (2022).
4. Mazzucato, M. & Ghebreyesus, T. A. Advancing the economics of health for all. *The Lancet.* 2024 (2024).
5. Centers for Medicare & Medicaid Services. *National Health Expenditure Data: Historical* 2024. 2024.

6. United Nations. *Sustainable Development Goals Report 2024* https://unstats.un.org/sdgs/report/2024/. Accessed: 2025-01-06. 2024. 2024.

7. Miller, I. F., Becker, A. D., Grenfell, B. T. & Metcalf, C. J. E. Disease and healthcare burden of COVID-19 in the United States. *Nature medicine* **26,** 1212–1217. 2020 (2020).

8. Glaeser, E. L., Resseger, M. & Tobio, K. Inequality in cities. *Journal of regional science* **49,** 617–646. 2009 (2009).

9. Pollock, K. Policy: urban physics. *Nature* **531,** S64–S66. 2016 (2016).

10. Sampson, R. J. Urban sustainability in an age of enduring inequalities: Advancing theory and ecometrics for the 21st-century city. *Proceedings of the National Academy of Sciences* **114,** 8957–8962. 2017 (2017).

11. Hong, B., Bonczak, B. J., Gupta, A. & Kontokosta, C. E. Measuring inequality in community resilience to natural disasters using large-scale mobility data. *Nature communications* **12,** 1870. 2021 (2021).

12. He, C. *et al.* The inequality labor loss risk from future urban warming and adaptation strategies. *Nature communications* **13,** 3847. 2022 (2022).

13. Meehan, K., Jurjevich, J. R., Everitt, L., Chun, N. M. & Sherrill, J. Urban inequality, the housing crisis and deteriorating water access in US cities. *Nature Cities,* 1–11. 2024 (2024).

14. Ashton, J., Grey, P. & Barnard, K. Healthy cities—WHO's new public health initiative. *Health promotion international* **1,** 319–324. 1986 (1986).

15. Programme, U. N. H. S. *State of the world's cities 2010/2011: Bridging the urban divide* 2010 (Earthscan, 2010).

16. Singh, G. K. & Siahpush, M. Widening rural–urban disparities in life expectancy, US, 1969–2009. *American journal of preventive medicine* **46,** e19–e29. 2014 (2014).

17. Vilar-Compte, M. *et al.* Urban poverty and nutrition challenges associated with accessibility to a healthy diet: a global systematic literature review. *International journal for equity in health* **20,** 1–19. 2021 (2021).

18. Marmot, M. & Wilkinson, R. *Social determinants of health* 2005 (Oup Oxford, 2005).

19. Marmot, M., Friel, S., Bell, R., Houweling, T. A. & Taylor, S. Closing the gap in a generation: health equity through action on the social determinants of health. *The Lancet* **372,** 1661–1669. 2008 (2008).

20. Sverdlik, A. Ill-health and poverty: a literature review on health in informal settlements. *Environment and urbanization* **23,** 123–155. 2011 (2011).

21. Frumkin, H. Urban sprawl and public health. *Public health reports.* 2002 (2002).

22. Kimani-Murage, E. W. *et al.* Vulnerability to food insecurity in urban slums: experiences from Nairobi, Kenya. *Journal of urban health* **91,** 1098–1113. 2014 (2014).

23. Roux, A. V. D. Neighborhoods and health: what do we know? What should we do? *American journal of public health* **106,** 430. 2016 (2016).

24. Haines, A. *et al.* Public health benefits of strategies to reduce greenhouse-gas emissions: overview and implications for policy makers. *The Lancet* **374,** 2104–2114. 2009 (2009).

25. Bassolas, A. *et al.* Hierarchical organization of urban mobility and its connection with city livability. *Nature communications* **10,** 4817. 2019 (2019).

26. Salvo, D. *et al.* Physical activity promotion and the United Nations sustainable development goals: building synergies to maximize impact. *Journal of physical activity and health* **18,** 1163–1180. 2021 (2021).

27. Organization, W. H. *et al. Stronger collaboration, better health: global action plan for healthy lives and well-being for all: strengthening collaboration among multilateral organizations to accelerate country progress on the health-related sustainable development goals* 2019 (World Health Organization, 2019).

28. Tonne, C. *et al.* Defining pathways to healthy sustainable urban development. *Environment international* **146,** 106236. 2021 (2021).

29. Ramaswami, A., Russell, A. G., Culligan, P. J., Sharma, K. R. & Kumar, E. Meta-principles for developing smart, sustainable, and healthy cities. *Science* **352,** 940–943. 2016 (2016).

30. Lowe, M. *et al.* City planning policies to support health and sustainability: an international comparison of policy indicators for 25 cities. *The Lancet Global Health* **10,** e882–e894. 2022 (2022).

31. Gomez-Lievano, A., Patterson-Lomba, O. & Hausmann, R. Explaining the prevalence, scaling and variance of urban phenomena. *Nature Human Behaviour* **1,** 0012. 2016 (2016).

32. Bettencourt, L. M. The origins of scaling in cities. *science* **340,** 1438–1441. 2013 (2013).

33. Cerin, E. *et al.* Determining thresholds for spatial urban design and transport features that support walking to create healthy and sustainable cities: findings from the IPEN Adult study. *The Lancet Global Health* **10,** e895–e906. 2022 (2022).

34. Chen, B. *et al.* Contrasting inequality in human exposure to greenspace between cities of Global North and Global South. *Nature Communications* **13,** 4636. 2022 (2022).

35. Iungman, T. *et al.* Cooling cities through urban green infrastructure: a health impact assessment of European cities. *The Lancet* **401,** 577–589. 2023 (2023).

36. Dasgupta, P. *Human well-being and the natural environment* 2001 (OUP Oxford, 2001).

37. Alcock, I., White, M. P., Wheeler, B. W., Fleming, L. E. & Depledge, M. H. Longitudinal effects on mental health of moving to greener and less green urban areas. *Environmental science & technology* **48,** 1247–1255. 2014 (2014).

38. Fisher, J. C. *et al.* Human well-being responses to species' traits. *Nature Sustainability* **6,** 1219–1227. 2023 (2023).

39. Maes, M. J. *et al.* Benefit of woodland and other natural environments for adolescents' cognition and mental health. *Nature sustainability* **4,** 851–858. 2021 (2021).

40. Marmot, M., Allen, J., Bell, R., Bloomer, E. & Goldblatt, P. WHO European review of social determinants of health and the health divide. *The Lancet* **380,** 1011–1029. 2012 (2012).

41. Sieck, C. J. *et al.* Digital inclusion as a social determinant of health. *NPJ digital medicine* **4,** 52. 2021 (2021).

42. Frank, J. *et al.* Work as a social determinant of health in high-income countries: past, present, and future. *The Lancet* **402,** 1357–1367. 2023 (2023).

43. Kohli, K. *et al.* The digital divide in access to broadband internet and mental healthcare. *Nature Mental Health* **2,** 88–95. 2024 (2024).

44. Sarkar, C., Webster, C. & Gallacher, J. Neighbourhood walkability and incidence of hypertension: Findings from the study of 429,334 UK Biobank participants. *International journal of hygiene and environmental health* **221,** 458–468. 2018 (2018).

45. Lens, M. Low-Density Zoning, Health, and Health Equity| Health Affairs Brief. *Culture of Health. Washington, DC: Health Affairs.* 2021 (2021).

46. Fisher, J. C. *et al.* Exploring how urban nature is associated with human wellbeing in a neotropical city. *Landscape and Urban Planning* **212,** 104119. 2021 (2021).

47. Hilmers, A., Hilmers, D. C. & Dave, J. Neighborhood disparities in access to healthy foods and their effects on environmental justice. *American journal of public health* **102,** 1644–1654. 2012 (2012).

48. Anza-Ramirez, C. *et al.* The urban built environment and adult BMI, obesity, and diabetes in Latin American cities. *Nature communications* **13,** 7977. 2022 (2022).

49. Mouratidis, K. & Yiannakou, A. What makes cities livable? Determinants of neighborhood satisfaction and neighborhood happiness in different contexts. *Land use policy* **112,** 105855. 2022 (2022).

50. Berdejo-Espinola, V., Fuller, R. A. & Zahnow, R. Well-being from nature exposure depends on socio-environmental contexts in Paraguay. *Nature Cities* **1,** 335–345. 2024 (2024).

51. Lutz, W. *et al.* Years of good life is a well-being indicator designed to serve research on sustainability. *Proceedings of the National Academy of Sciences* **118,** e1907351118. 2021 (2021).

52. Collins, P. Y. *et al.* Making cities mental health friendly for adolescents and young adults. *Nature* **627,** 137–148. 2024 (2024).

53. Ala-Mantila, S., Heinonen, J., Junnila, S. & Saarsalmi, P. Spatial nature of urban well-being. *Regional Studies* **52,** 959–973. 2018 (2018).

54. Brelsford, C., Lobo, J., Hand, J. & Bettencourt, L. M. Heterogeneity and scale of sustainable development in cities. *Proceedings of the National Academy of Sciences* **114,** 8963–8968. 2017 (2017).

55. Nieuwenhuijsen, M. J. Urban and transport planning, environmental exposures and health-new concepts, methods and tools to improve health in cities. *Environmental Health* **15,** S38. 2016 (2016).

56. Dyer, G. M. *et al.* Commentary: A road map for future data-driven urban planning and environmental health research. *Cities* **155,** 105340. 2024 (2024).

57. Montana, F. *et al.* Building a Healthy Urban Design Index (HUDI): how to promote health and sustainability in European cities. *The Lancet Planetary Health* **9,** e511–e526. 2025 (2025).

58. Higgs, C. *et al.* Policy-relevant spatial indicators of urban liveability and sustainability: Scaling from local to global. *Urban Policy and Research* **40,** 321–334. 2022 (2022).

59. Frank, L. D. & Wali, B. Monitoring changes in walkability over time: An environmental exposure change detection framework with implications for equity and social justice. *Sustainable Cities and Society* **117,** 105808. 2024 (2024).

60. Braun, L. M., Fox, E. H. & Frank, L. D. Is walkability healthy for all? Using the National Environmental Database to examine equity in the environmental health characteristics of pedestrian-supportive neighborhoods in the US. *Social Science & Medicine* **374.** 2025 (2025).

61. Dyer, G. M. *et al.* Exploring the nexus of urban form, transport, environment and health in large-scale urban studies: A state-of-the-art scoping review. *Environmental Research* **257,** 119324. 2024 (2024).

62. Giles-Corti, B. *et al.* City planning and population health: a global challenge. *The Lancet* **388,** 2912–2924. 2016 (2016).

63. Murray, C. J. & Lopez, A. D. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *The Lancet* **349,** 1436–1442. 1997 (1997).

64. Rydin, Y. *et al.* Shaping cities for health: complexity and the planning of urban environments in the 21st century. *The Lancet* **379,** 2079–2108. 2012 (2012).

65. Yap, W., Stouffs, R. & Biljecki, F. Urbanity: automated modelling and analysis of multidimensional networks in cities. *npj Urban Sustainability* **3.** 2023, (10.1038/s42949-023-00125-w) (2023).

66. Hamilton, W., Ying, Z. & Leskovec, J. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems* **30.** 2017 (2017).

67. Wang, M. *et al.* Deep graph library: A graph-centric, highly-performant package for graph neural networks. *arXiv preprint arXiv:1909.01315.* 2019 (2019).

68. Ying, Z., Bourgeois, D., You, J., Zitnik, M. & Leskovec, J. Gnnexplainer: Generating explanations for graph neural networks. *Advances in neural information processing systems* **32.** 2019 (2019).

69. US Federal Housing Finance Agency. *Underserved Areas Data* `https://www.fhfa.gov/data/underserved-areas-data`. Accessed: 2025-01-18. 2024. 2024.

70. US Department of Agriculture. *Poverty Area Measures* `https://www.ers.usda.gov/data-products/poverty-area-measures`. Accessed: 2025-01-18. 2025. 2025.

71. US Census Bureau. *Redefining Urban Areas following the 2020 Census* `https://www.census.gov/newsroom/blogs/random-samplings/2022/12/redefining-urban-areas-following-2020-census.html`. Accessed: 2025-01-18. 2022. 2022.

72. Organization for Economic Cooperation and Development. *Health at a Glance: Asia/Pacific 2022* `https://www.oecd.org/en/publications/health-at-a-glance-asia-pacific-2022_c7467f62-en/full-report.html`. Accessed: 2025-01-18. 2022. 2022.

73. Barthélemy, M. Spatial networks. *Physics Reports* **499,** 1–101. 2011 (2011).

74. Ding, D. *et al.* The economic burden of physical inactivity: a global analysis of major non-communicable diseases. *The Lancet* **388,** 1311–1324. 2016 (2016).

75. Nyberg, S. T. *et al.* Health benefits of leisure-time physical activity by socioeconomic status, lifestyle risk, and mental health: a multicohort study. *The Lancet Public Health* **10,** e124–e135. 2025 (2025).

76. Ribeiro, A. I. *et al.* Association of neighbourhood disadvantage and individual socioeconomic position with all-cause mortality: a longitudinal multicohort analysis. *The Lancet Public Health* **7,** e447–e457. 2022 (2022).

77. McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426.* 2018 (2018).

78. Hsu, A., Sheriff, G., Chakraborty, T. & Manya, D. Disproportionate exposure to urban heat island intensity across major US cities. *Nature communications* **12,** 2721. 2021 (2021).

79. Salvo, D., Jáuregui, A., Adlakha, D., Sarmiento, O. L. & Reis, R. S. When moving is the only option: the role of necessity versus choice for understanding and promoting physical activity in low-and middle-income countries. *Annual review of public health* **44,** 151–169. 2023 (2023).

80. Strain, T. *et al.* Levels of domain-specific physical activity at work, in the household, for travel and for leisure among 327 789 adults from 104 countries. *British Journal of Sports Medicine* **54,** 1488–1497. 2020 (2020).

81. Behrens, P. *et al.* Obesity and climate change: co-crises with common solutions. *Frontiers in Science* **3,** 1613595. 2025 (2025).

82. Abbiasov, T. *et al.* The 15-minute city quantified using human mobility data. *Nature Human Behaviour* **8,** 445–455. 2024 (2024).

83. Yap, W., Wu, A. N., Miller, C. & Biljecki, F. Revealing building operating carbon dynamics for multiple cities. *Nature Sustainability,* 1–12. 2025 (2025).

84. Fan, Z., Zhang, F., Loo, B. P. & Ratti, C. Urban visual intelligence: Uncovering hidden city profiles with street view images. *Proceedings of the National Academy of Sciences* **120,** e2220417120. 2023 (2023).

85. Chen, Y.-R., Hanazato, M., Koga, C., Ide, K. & Kondo, K. The association between street connectivity and depression among older Japanese adults: the JAGES longitudinal study. *Scientific Reports* **12,** 13533. 2022 (2022).

86. Stewart, I. D. & Oke, T. R. Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society* **93,** 1879–1900. 2012 (2012).

87. Elzeni, M. M., ELMokadem, A. A. & Badawy, N. M. Impact of urban morphology on pedestrians: A review of urban approaches. *Cities* **129,** 103840. 2022 (2022).

88. Mork, D., Braun, D. & Zanobetti, A. Time-lagged relationships between a decade of air pollution exposure and first hospitalization with Alzheimer's disease and related dementias. *Environment International* **171,** 107694. 2023 (2023).

89. Wei, J. *et al.* First close insight into global daily gapless 1 km PM2. 5 pollution, variability, and health impact. *Nature Communications* **14,** 8349. 2023 (2023).

90. Christakis, N. A. & Fowler, J. H. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine* **357,** 370–379. 2007 (2007).

91. Pearson-Stuttard, J. & Davies, S. C. The Health Index: a framework to guide health-driven prosperity. *The Lancet* **405,** 777. 2025 (2025).

92. Observable. *Overture Places Quality Analysis* `https://observablehq.com/d/9847c08c46f56ed6`. Accessed: 2025-01-17. 2025. 2025.

93. Moreno, C., Allam, Z., Chabaud, D., Gall, C. & Pratlong, F. Introducing the "15-Minute City": Sustainability, resilience and place identity in future post-pandemic cities. *Smart cities* **4,** 93–111. 2021 (2021).

94. Cheng, B., Misra, I., Schwing, A. G., Kirillov, A. & Girdhar, R. *Masked-attention mask transformer for universal image segmentation* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), 1290–1299. 2022.

95. Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University. High Resolution Settlement Layer (HRSL). Source imagery for HRSL © 2016 DigitalGlobe. Accessed 1 Aug 2023. 2016 (2016).

96. Demuzere, M. *et al.* A global map of Local Climate Zones to support earth system modelling and urban scale environmental science. *Earth System Science Data Discussions* **2022,** 1–57. 2022 (2022).

97. Tolan, J. *et al.* Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. *Remote Sensing of Environment* **300,** 113888. 2024 (2024).

98. Hu, Y., Dai, Z. & Guldmann, J.-M. Modeling the impact of 2D/3D urban indicators on the urban heat island over different seasons: A boosted regression tree approach. *Journal of environmental management* **266,** 110424. 2020 (2020).

99. Li, Y., Schubert, S., Kropp, J. P. & Rybski, D. On the influence of density and morphology on the Urban Heat Island intensity. *Nature communications* **11,** 2647. 2020 (2020).

100. Tikhonova, O. & Beirão, J. A tale of two cities-A comparative study of historical urban cores. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability* **13,** 448–465. 2020 (2020).

101. Biljecki, F. & Chow, Y. S. Global Building Morphology Indicators. *Computers, Environment and Urban Systems* **95,** 101809. 2022 (2022).

102. Basaraner, M. & Cetinkaya, S. Performance of shape indices and classification schemes for characterising perceptual shape complexity of building footprints in GIS. *International Journal of Geographical Information Science* **31,** 1952–1977. 2017 (2017).

103. Labetski, A., Vitalis, S., Biljecki, F., Arroyo Ohori, K. & Stoter, J. 3D building metrics for urban morphology. *International Journal of Geographical Information Science* **37,** 36–67. 2023, (10.1080/13658816.2022.2103818) (1 2023).

104. Dibble, J. *et al.* On the origin of spaces: Morphometric foundations of urban form evolution. *Environment and Planning B: Urban Analytics and City Science* **46,** 707–730. 2019 (2019).

105. Feliciotti, A., Romice, O. & Porta, S. Urban regeneration, masterplans and resilience: the case of Gorbals, Glasgow. *Urban Morphology* **21,** 61–79. 2017 (2017).

106. Huang, J., Lu, X. X. & Sellers, J. M. A global comparative analysis of urban form: Applying spatial metrics and remote sensing. *Landscape and urban planning* **82,** 184–197. 2007 (2007).

107. Gil, J., Beirão, J. N., Montenegro, N. & Duarte, J. P. On the discovery of urban typologies: data mining the many dimensions of urban form. *Urban morphology* **16,** 27. 2012 (2012).

108. Schirmer, P. M. & Axhausen, K. W. A multiscale classification of urban morphology. *Journal of Transport and Land Use* **9,** 101–130. 2016 (2016).

109. Fleischmann, M., Feliciotti, A., Romice, O. & Porta, S. Morphological tessellation as a way of partitioning space: Improving consistency in urban morphology at the plot scale. *Computers, Environment and Urban Systems* **80,** 101441. 2020 (2020).

110. McGarigal, K. & Marks, B. J. Spatial pattern analysis program for quantifying landscape structure. *Gen. Tech. Rep. PNW-GTR-351. US Department of Agriculture, Forest Service, Pacific Northwest Research Station* **10.** 1995 (1995).

111. Lima, F. T., Brown, N. C. & Duarte, J. P. A grammar-based optimization approach for walkable urban fabrics considering pedestrian accessibility and infrastructure cost. *Environment and Planning B: Urban Analytics and City Science* **49,** 1489–1506. 2022 (2022).

112. Liu, H., Gou, P. & Xiong, J. Vital triangle: A new concept to evaluate urban vitality. *Computers, Environment and Urban Systems* **98,** 101886. 2022 (2022).

113. Szarka, N. & Biljecki, F. Population estimation beyond counts—Inferring demographic characteristics. *Plos one* **17,** e0266484. 2022 (2022).

114. Gauvin, L. *et al.* Gender gaps in urban mobility. *Humanities and Social Sciences Communications* **7,** 1–13. 2020 (2020).

115. Wang, R. *et al.* Perceptions of built environment and health outcomes for older Chinese in Beijing: A big data approach with street view images and deep learning technique. *Computers, Environment and Urban Systems* **78,** 101386. 2019 (2019).

116. Ha, H.-H. & Thill, J.-C. Analysis of traffic hazard intensity: A spatial epidemiology case study of urban pedestrians. *Computers, Environment and Urban Systems* **35,** 230–240. 2011 (2011).

117. Kruse, J., Kang, Y., Liu, Y.-N., Zhang, F. & Gao, S. Places for play: Understanding human perception of playability in cities using street view images and deep learning. *Computers, Environment and Urban Systems* **90,** 101693. 2021 (2021).

118. Mouratidis, K. Urban planning and quality of life: A review of pathways linking the built environment to subjective well-being. *Cities* **115,** 103229. 2021 (2021).

119. Clark, W. A., Yi, D. & Huang, Y. Subjective well-being in China's changing society. *Proceedings of the National Academy of Sciences* **116,** 16799–16804. 2019 (2019).

120. Mah, S. M., Rosella, L. C., Kivimäki, M. & Carmeli, C. Overcrowded housing during adolescence and future risk of premature mortality: a 28-year follow-up of 556,191 adolescents from Switzerland. *The Lancet Regional Health–Europe* **31.** 2023 (2023).

121. Klinkhardt, C. *et al.* Using OpenStreetMap as a Data Source for Attractiveness in Travel Demand Models. *Transportation Research Record* **2675,** 294–303. 2021 (2021).

122. Weiss, D. *et al.* Global maps of travel time to healthcare facilities. *Nature Medicine* **26,** 1835–1838. 2020 (2020).

123. Liu, X., Wu, M., Peng, B. & Huang, Q. Graph-based representation for identifying individual travel activities with spatiotemporal trajectories and POI data. **12,** 1–13. 2022 (2022).

124. Liu, X. & Long, Y. Automated identification and characterization of parcels with OpenStreetMap and points of interest. *Environment and Planning B: Planning and Design* **43,** 341–360. 2016 (2016).

125. Zhou, J. & Yang, Y. Transit-based accessibility and urban development: An exploratory study of Shenzhen based on big and/or open data. *Cities* **110,** 102990. 2021 (2021).

126. Liu, S. *et al.* A generalized framework for measuring pedestrian accessibility around the world using open data. *Geographical Analysis* **54,** 559–582. 2022 (2022).

127. Wang, Z., Ma, D., Sun, D. & Zhang, J. Identification and analysis of urban functional area in Hangzhou based on OSM and POI data. *PLoS one* **16,** e0251988. 2021 (2021).

128. Li, X. Examining the spatial distribution and temporal change of the green view index in New York City using Google Street View images and deep learning. *Environment and Planning B: Urban Analytics and City Science* **48,** 2039–2054. 2021 (2021).

129. Middel, A., Lukasczyk, J., Maciejewski, R., Demuzere, M. & Roth, M. Sky View Factor footprints for urban climate modeling. *Urban climate* **25,** 120–134. 2018 (2018).

130. Ki, D. & Lee, S. Analyzing the effects of Green View Index of neighborhood streets on walking time using Google Street View and deep learning. *Landscape and Urban Planning* **205,** 103920. 2021 (2021).

131. Dong, G., Yan, Y., Shen, C. & Wang, H. Real-time high-performance semantic image segmentation of urban street scenes. *IEEE Transactions on Intelligent Transportation Systems* **22,** 3258–3274. 2020 (2020).

132. Yap, W., Chang, J.-H. & Biljecki, F. Incorporating Networks in Semantic Understanding of Streetscapes: Contextualising Active Mobility Decisions. *Environment and Planning B: Urban Analytics and City Science.* 2022 (2022).

133. Huang, Y., Hong, T. & Ma, T. Urban network externalities, agglomeration economies and urban economic growth. *Cities* **107,** 102882. 2020 (2020).

134. Xue, J. *et al.* Quantifying the spatial homogeneity of urban road networks via graph neural networks. *Nature Machine Intelligence* **4,** 246–257. 2022 (2022).

135. Prieto-Curiel, R., Schumann, A., Heo, I. & Heinrigs, P. Detecting cities with high intermediacy in the African urban network. *Computers, Environment and Urban Systems* **98,** 101869. 2022 (2022).

136. Boeing, G. Street network models and indicators for every urban area in the world. *Geographical Analysis* **54,** 519–535. 2022 (2022).

137. Ozuduru, B. H., Webster, C. J., Chiaradia, A. J. & Yucesoy, E. Associating street-network centrality with spontaneous and planned subcentres. *Urban Studies* **58,** 2059–2078. 2021 (2021).

138. Kirkley, A., Barbosa, H., Barthelemy, M. & Ghoshal, G. From the betweenness centrality in street networks to structural invariants in random planar graphs. *Nature communications* **9,** 1–12. 2018 (2018).

139. Agryzkov, T., Tortosa, L., Vicent, J. F. & Wilson, R. A centrality measure for urban networks based on the eigenvector centrality concept. *Environment and Planning B: Urban Analytics and City Science* **46,** 668–689. 2019 (2019).

140. Curado, M., Tortosa, L., Vicent, J. F. & Yeghikyan, G. Analysis and comparison of centrality measures applied to urban networks with data. *Journal of Computational Science* **43,** 101127. 2020 (2020).

141. Jia, C., Du, Y., Wang, S., Bai, T. & Fei, T. Measuring the vibrancy of urban neighborhoods using mobile phone data with an improved PageRank algorithm. *Transactions in GIS* **23,** 241–258. 2019 (2019).