REGION
The Journal of ERSA
Powered by WU

Volume 13, Number 2, 2026, 53–80 journal homepage: region.ersa.org
DOI: 10.18335/region.v13i2.660 ISSN: 2409-5370

Neighborhood assets as enablers of urban-suburban innovation capacities: Uncovering patterns and divides in Massachusetts and New York using traditional and API-mined data

Eleni Oikonomaki[1]

[1] Aristotle University of Thessaloniki: Thessaloniki, Greece

Received: 15 December 2025/Accepted: 31 May 2026

Abstract. Large, open-source, and API-mined data have revolutionized the analysis of urban and suburban city neighborhoods with respect to their innovation capacities. Such data complements traditional census and government data collection methods and sources to help explain significant variations in invention activity across neighborhoods within a city. This paper combines diverse data sources to assess urban-suburban capacities and divides in the generation of patent-based innovation. First, it identifies patterns in innovation outcomes using USPTO patent records and illustrates spillovers across all ZIP codes in New York and Massachusetts and compares them with data from Google Maps to obtain a more granular and broader understanding of the concentration of innovation spaces. The findings reveal variations in the distribution of activity when comparing traditional patent records and API-mined data. Second, the study examines how economic, social, and spatial characteristics are associated with high levels of granted patents across neighborhoods in New York and Massachusetts. Drawing on a composite dataset assembled from multiple sources, the research interrogates the density of correlations between neighborhood variables and patent activity. Third, a comparative design spanning the two US states with distinct urban contexts deploys difference-in-means testing to systematically identify both shared patterns and divergences in how place-level conditions relate to high patent performance. Beyond the empirical findings, the study contributes a replicable methodology for assembling neighborhood asset indicators from heterogeneous data sources and offers a grounded interpretation of results for planning practice and innovation policy. The findings suggest that while traditional datasets provide long-term insights for formal study, alternative data-driven approaches reveal dynamic information that is important for understanding the evolving nature of innovation landscapes and informal innovation activities. Ultimately, this study advocates for a hybrid model that integrates both traditional and API-mined data paradigms to promote urban intelligence and equity. It advocates considering alternative innovation measures when designing innovation policies.

1 Introduction

Innovation is a key driver of economic growth, yet its distribution across regions is highly uneven. Countries and regions exhibit significant geographical variation in innovation activities. Although skill-based segregation across metropolitan areas remains modest, we have transitioned from a scenario in which skills were strikingly evenly distributed geographically to one in which metropolitan areas are increasingly differentiated by their levels of human capital (Berry, Glaeser 2005). This concentration heightens the demand for geographical proximity, rendering such knowledge more "sticky" in an area and less spatially mobile (Leyshon, Thrift 1997, Feldman 2001, Cortright, Mayer 2002). Urban cores serve as hubs of technological advancement, benefiting from digital ecosystems, research institutions, and well-connected industries with more complex activities and most strikingly, radical and disruptive innovations, those with the potential to redefine industries, almost exclusively being concentrated in large cities with a high density of activities, with this concentration in city cores growing over time (Holl et al. 2024, Balland et al. 2020, Mewes, Broekel 2022, Balland, Rigby 2017). In contrast, suburban regions frequently lag, facing challenges such as limited access to innovation networks and fewer knowledge spillovers, with evidence indicating that more sophisticated products are located in a densely connected urban core, whereas less sophisticated products are located in a less connected periphery (Hidalgo et al. 2007).

This has intensified polarization and fueled discontent, leaving many poorer regions and entire countries stuck in a low-income trap, unable to adapt to economic changes or transitions toward more integrated, open economies (Ezcurra, Rodríguez-Pose 2013, Diemer et al. 2022). Knowledge diffusion from cores to peripheries - or from more developed to less developed regions - faces significant barriers due to the absence of effective and well-functioning transmission channels (Boschma 2005, D'Este et al. 2013, Iammarino 2018). This situation implies that income disparities across regions are more likely to be reinforced rather than reduced by innovation and diversification processes (Pinheiro et al. 2022). The inequalities arise from a range of complex, localized factors that can be difficult for centralized policies to address effectively. Empirical research has identified common patterns in left-behind regions and countries, including a smaller manufacturing sector, higher dependency ratios, lower educational attainment, and weaker innovation capacity, which are evident across regions with varying income levels (Diemer et al. 2022). The distribution and effects of national R&D subsidies vary across regions with different levels of innovation activity (Herrera, Nieto 2008), while the differences in access to R&D subsidies often persist across localities despite national and/or regional policy efforts.

This current unbalanced growth raises questions about the effectiveness of regional policies, which aim to address widening disparities in development across regions. For example, while the Smart Specialization Strategy (S3) in the EU aims to guide regional innovation and diversification, it has struggled to provide practical solutions for regions facing diverse socioeconomic and geographical challenges. Efficiency-driven approaches have not lifted most left-behind territories out of their disadvantaged state, resulting in persistent within-region inequalities in both advanced and emerging economies (Rodríguez-Pose et al. 2024). These variations in innovation activities are evident not only between regions but also within cities, with substantial differences observed at a lower spatial scale (Holl et al. 2024). Such tendencies are shaped and sustained by disparities in socioeconomic infrastructure and other contextual factors. The emergence of 'left-behind places' identified by several scholars is an inherently territorial phenomenon (Storper 1995).

As a result, over the last years, top-down economic development strategies have increasingly been replaced by place-based approaches that align with the needs and preferences of local actors (Rodríguez-Pose, Gill 2003), as local policies can better address their efforts to spur innovation and promote equitable economic growth. These initiatives move beyond agglomeration-focused strategies that have often exacerbated polarization and left-behindness (Rodríguez-Pose et al. 2024). Place-based economic strategies aim to leverage local potential by promoting economic activities that align with and depend on the region's unique economic conditions and comparative advantages (Pike et al. 2006). In its simplest form, place-based economic development involves identifying, mobilizing, and utilizing a locality's unique potential (Vazquez-Barquero 1999).

Despite growing knowledge about the 'geography of innovation,' there remains a gap in the specific local capacities and mechanisms that drive differences in innovation outputs. Our understanding of the territorial and other significant factors enabling "local innovation capacity" remains limited. The dominant mode of knowledge production about innovation districts is overwhelmingly qualitative and case-driven. Whether examining Kendall Square in Cambridge, Triangle Research Park in North Carolina, 22@ in Barcelona, or the King's Cross development in London, the field has built its understanding through rich narrative accounts of how specific places evolved, tracing the decisions of anchor institutions, the role of public investment, the gradual accumulation of firms and talent, and the cultural conditions that made collaboration possible. These case studies are invaluable for capturing the path-dependent character of district formation, but they lack generalization. This highlights the importance of assessing neighborhood assets that drive disparities in innovation capacities.

The typologies discussed in Section 2 emerge from pattern-matching across cases rather than from systematic measurement, meaning that concepts like "innovation" or "anchor-plus" districts or "business /corporate parks" are descriptive categories grounded in a handful of well-documented examples rather than empirically validated classes derived from large-sample analysis. The result is a body of literature that is deeply insightful about particular places but structurally limited in its ability to say, with any rigor, which physical, infrastructural, and socioeconomic characteristics are present in innovation success. Innovation stems from the combined effects of individual characteristics and geographically rooted socioeconomic factors, yet disparities in innovation generation and in neighborhood regeneration efforts to spur innovation ecosystems with patent-based activities across localities remain only partially understood.

This situation presents two critical imperatives: first, the need to develop more efficient methods, tools, and data to understand both innovation generation processes and their territorial enabling factors, particularly their localized expressions (Cooke et al. 2012). In this context, big, urban, and API-mined data analytics can be utilized to provide insights at low spatial scales, and second, the need to match the innovation policy coming from the national and regional levels to reflect socioeconomic as well as institutional characteristics and prioritize more immediate or localized conditions and realities (Rodríguez-Pose, Wilkie 2017). This way, the paper builds on previous studies that highlighted the need to promote inclusive development, even amid unbalanced growth (World Bank 2009), and examined strategies to foster greater economic inclusivity (OECD 2018). More specifically, the study focuses on providing a better understanding of the impact of territorial factors on patent generation by conducting exploratory statistical analysis that integrates open-source geospatial data to respond to the central research questions guiding this study:

- *RQ1.* Can quantified neighborhood assets – spanning urban morphology, infrastructure provision, and socioeconomic composition – provide a replicable and generalizable framework for differentiating innovation districts from low-performing neighborhoods, beyond what case-study typologies currently allow?
- *RQ2.* How do innovation patterns differ between urban and suburban neighborhoods when analyzed using traditional and alternative API-mined data?
- *RQ3.* How can large urban data help policymakers develop strategies to support innovation in lagging, left-behind regions and areas?

By addressing these questions, the study contributes to ongoing debates on the localization of innovation activities and the role of data and technology in deepening the understanding of innovation geography and furthermore the synergies at localities characterized as 'innovators'. Using US states (Massachusetts and New York), with well-established innovation hubs, it provides an updated understanding of how a locality (measured at the ZIP code level) is constituted through and within the innovation context.

In the context of this study, innovation systems are reconsidered as relational groups of places and districts energized by the interplay of multiple helices, rather than as solely institutional entities. The place as a manifestation of innovation capacity enables the exploration and analysis of a diverse set of variables and datasets across spatial, economic, and social dimensions. Such an approach allows us to contrast territorial trends across urban and suburban contexts within both traditional innovation hubs and emerging neighborhoods, enabling findings grounded in local morphological and infrastructural conditions to inform regional and national innovation policy in a way that is sensitive to, rather than abstracted from, place-specific context.

In terms of its structure, this section introduces contemporary challenges in innovation geography, urban-suburban divides, and innovation policy's role in fostering inclusive innovation ecosystems. Section 2 provides a review of related literature on best paradigms and typologies that articulate innovation, highlighting the role of territories, including morphology and infrastructure, in promoting patent-based innovation and transforming local systems and networks. Section 3 provides a thorough examination of the key methodological approaches adopted in the study, including feature engineering via API data mining to create new variables that capture innovation spaces and activities, as well as big and geospatial data analytics. Section 4 subsequently delivers an analysis of the data and empirical strategies, followed by findings (Section 5) and concluding observations (Section 6).

2 Literature Review on the Evolving Landscape and Geographies of Innovation

2.1 Existing Paradigms of Urban and Architectural Typologies, Critiques and Alternatives

Over the last few years, an increasing part of research has focused on the spatial organization of technological innovation and the associations between urban form, innovation performance, and human experience (Capdevila 2015, Mariotti et al. 2021, Roche et al. 2024). Most of these studies provide a foundation for understanding how morphology, infrastructure, and local forces influence patterns of technological development and diffusion. Along the same lines, this paper examines the evolving geography of innovation, discussing the spatial dynamics of well-established typologies - both ecosystems that require long commutes, and ecosystems located in places that are walkable, bikeable, and connected by transit. It assesses how the innovation landscape has evolved over the last few decades, from the traditional industrial park model that was first introduced in the early 20th century and the tech-focused park pioneered in the 1950s at Stanford University, to the boom of the science and business park model in the 1980s, and today's innovation ecosystems, marked by the establishment of innovation districts.

Industrial parks emerged in the 1920s as industries relocated from crowded urban centers to suburban peripheries. They were designed to accommodate single-story, in-line manufacturing, and they represented a new approach to suburban industrial development. In most cases, a master developer would subdivide and organize the land, provide essential infrastructure (including roads, parking, and rail or truck access), and establish regulations to control building size, lot coverage, and overall design harmony. Buildings in these parks were typically functional, reflecting the aesthetics of industrial efficiency. At the same time, when present, administrative offices were attached to or located near manufacturing facilities, maintaining a consistent utilitarian appearance (Mozingo 2011).

In the San Francisco Bay Area, one of the early centers for suburban manufacturing, *research/science park* became the dominant model for postwar corporate office locations, particularly on the San Francisco Peninsula, where white-collar staff worked in streamlined, industrially inspired environments, which prioritized efficiency, experimentation, and practicality, placing less emphasis on prestige, and more on function and innovation, in contrast to the landscaped corporate campuses common in other US regions, which prioritized status, stability, and long-term corporate image and innovation came more from internal R&D labs than from ecosystem interaction. The San Francisco Bay Area's industrial parks became highly successful because they combined flexibility with proximity to Stanford University, which provided talent pipelines (engineering graduates) and knowledge spillovers through faculty and research labs. This research park model attracted high-tech startups and established firms, creating dense clusters of innovation. Unlike traditional *corporate/business campuses* (isolated and landscaped), these industrial/research parks encouraged interaction, collaboration, and rapid prototyping.

Then, the model of *innovation zones* in the Bay Area, including Silicon Valley, the

South Bay, and North Bay tech corridors, can be seen as direct descendants of these early industrial parks. Proximity to universities and research institutions ensured a steady flow of talent and facilitated technology transfer. These early industrial parks laid the physical and social foundations that enabled the Bay Area to evolve into a globally recognized innovation ecosystem. Overall, this postwar industrial park model developed in the Bay Area, with its functional layout, integration of office and light-industrial spaces, and close ties to universities, continues to be replicated worldwide.

Drawing on the well-established case studies and typologies widely recognized as successful examples of innovation ecosystems discussed in this section, the review builds a conceptual foundation that bridges existing typologies and frameworks with the empirical dimension of this study. Rather than treating these cases as endpoints, they serve as a departure point, grounding the analysis and orienting it towards new analytical methods capable of producing a more comprehensive and nuanced understanding of places with high innovation performance.

2.2 Suburban Industrial, Science or Research (Institutional) Parks

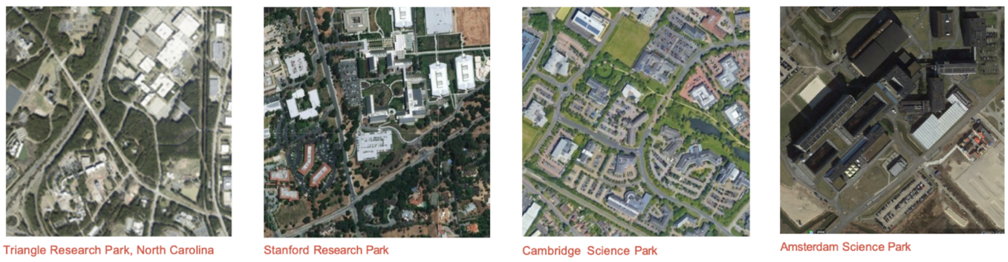


Figure 1: Aerial Views of Well-established Suburban Science or Research (Institutional) Parks, source: Google Earth

As described, Menlo Park and Stanford University pioneered new suburban zoning and development models that shaped the postwar office landscape of the San Francisco Peninsula. In 1948, Menlo Park introduced a new zoning district, among the first in the nation, designed specifically for “Administrative, Executive, Professional, and Research Institution” uses. Key features included large lots (minimum of 2 acres), strict limits: one-story buildings covering no more than 40% of the site, generous landscaping and setbacks, giving office properties the appearance of spacious suburban homes, high parking requirements, anticipating heavy car use, off-street parking only, unlike downtown districts of the time.

In 1951, it launched the Stanford Industrial Park (now the Stanford Research Park), a project that borrowed from the industrial park model but tailored it for office and research uses. Located adjacent to affluent suburban neighborhoods, not industrial areas, dependent on automobile access (no rail or trucking facilities), ample landscaping, green space, and low-rise buildings, custom buildings for tenants under long-term leases, the park expanded rapidly (Mozingo 2011). By 1953, Menlo Park expanded the district to over 70 acres as the initial sites were quickly developed. Later office corridors, such as Sand Hill Road, extended this model into the 1970s–80s and became a 'trademark' of the region’s suburban office identity. Stanford University faced restrictions that prevented it from selling its land. Therefore, it sought to monetize it by leasing parcels for development. By 1960, Stanford Industrial Park had become a national model in the US.

Architectural Forum praised it for transforming Palo Alto into a “satellite” city while maintaining its garden-like suburban character. Together, Menlo Park’s zoning experiment and Stanford’s development strategy defined a new suburban office landscape, which was low-rise, automobile-oriented, landscaped, and “clean”. Hundreds of regions worldwide have spent billions of dollars to build their versions of Silicon Valley, beginning by replicating the typology of this *science park*. The “recipe” was always the same; however, in most cases, the magic never happened. Several cities have imported a set of ideas from California about how the university can commercialize its research and profit as a result. The urban form of this suburban research park typology has been replicated to address the needs of commercial research across regions with very diverse socioeconomic contexts.

Castells' and Hall's 'Technopoles of the World' (Castells, Hall 1994) provides the most comprehensive study on the urban impact of technology clusters from Silicon Valley to other European countries. In their book, they evaluate the success in relation to the objectives pursued in the development of each park, while stating that it is industrial competitiveness, rather than scientific quality, that remains the fundamental goal of any technology park project. (Castells, Hall 1994). In other words, universities consider science parks to be just another form of property investment - a growing sector with great potential rather than a continuation of their role as research institutions. However, in Silicon Valley, these objectives are very narrow.

The model of the science park is commonly found in suburban areas worldwide, where traditionally isolated, sprawling areas are urbanizing through increased density and mixed-use activities (including retail and restaurants) rather than through separated uses. Originating from the Stanford Research Park model and exemplified by the Research Triangle Park in North Carolina and the Cambridge Science Park, among others (Figure 1), these parks are closely linked to universities and aim to commercialize research through entrepreneurial ecosystems.

While they symbolize technological progress and economic synergy, they also raise concerns around environmental impact, urban integration, and exclusivity. Even if, from a real estate perspective, this typology might have been successful, it has been insufficient to foster innovation ecosystems without creating environmental issues. At the same time, in a science park model, there is a paradox: the more governance structures, spatial layouts, and partnership frameworks are created to promote openness and collaboration, the more they can unintentionally impose boundaries. This way, they risk reducing spontaneity, narrowing the range of actors involved, and limiting the park's capacity to evolve organically (Sennett 2006).

2.3 Innovation Districts

Figure 2: Aerial Views of Well-established Innovation Districts, source: Google Earth

During the last decades, the model of innovation district has become the main expression of innovation activities in urban areas, established through public-private partnerships or top-down initiatives following other famous examples of districts, such as Kendall Square in Boston and similar districts. (Figure 2). This model appears to be an alternative to traditional innovation spaces such as science parks that often seek to foster 'serendipity 'in the suburban settings accessible by private vehicles, and with institutions working in effective silos. In contrast, the innovation district is a new, complementary urban model type (Katz, Wagner 2014), zones in cities where public and private actors work to attract entrepreneurs, startups, and business incubators, generally to revitalize depressed downtown areas. Especially in response to growing concerns about environmentally unsustainable development, the 'Innovation District' model presents an opportunity to become an arena where environmental practices are implemented to reduce carbon emissions and create denser residential and employment patterns that provide a better human experience in walkable environments.

Key characteristics of innovation districts include a diverse range of uses that attract a broad workforce and foster collaboration among governments, private firms, start-ups, entrepreneurs, business incubators, accelerators, and leading-edge anchor institutions (Puentes, Roberto 2008). These districts become vibrant, physically compact, transit-accessible, and technologically wired, offering mixed-use housing, offices, and retail. This approach appears effective in fostering prosperous business and innovation ecosystems, with major urban hubs successfully attracting skilled and young professionals from various regions of the globe. This recent type of ecosystem is primarily found in the downtowns and midtowns of central cities, where large-scale mixed-use development is centered or near historic waterfronts and where underutilized industrial buildings or warehouse districts are undergoing physical and economic transformation. Unlike traditional industrial clusters, which often develop organically, innovation districts are deliberately planned and supported by policies that integrate research institutions, startups, leading R&D organizations, incubators, and accelerators within mixed-use urban environments (Loures et al. 2007).

The Innovation District model has provided what Castells, Hall (1994) described as a "generation of synergy", as the "technological innovation depends on the process of learning by doing, rather than off-the-shelf operation manuals" and the success of various technopoles is evaluated in terms of the "milieux of innovation" (Castells, Hall 1994, p. 9). Jacobs (1969) argues that the primary source of knowledge spillovers lies outside the industry in which a firm operates. This implies that cities are significant hubs of innovation, given the extensive diversity of external knowledge sources they contain. Since Jacobs's writing and the return to a fondness for traditional street life, the notion that 'serendipity' in urban centers encourages innovation has been a cornerstone of urban theory.

2.4 Tech/Corporate/Business (Urban & Suburban) Parks

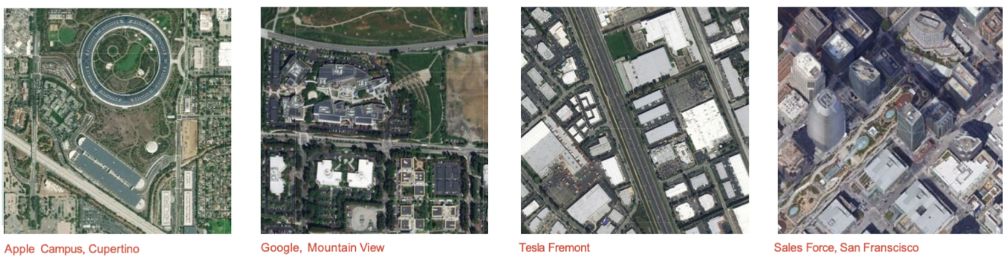


Figure 3: Well-established Tech/Corporate/Business Parks, source: Google Earth

However, that was not always the mainstay. By the end of the twentieth century, America's suburbs contained more office space than its central cities, while large corporate developments spread across suburban areas, becoming the primary feature of the spatial organization of industrial production (Mozingo 2011). Where Castells and Hall celebrate 'innovative milieux', an atmosphere of creative collaboration, which in their eyes is essential for the development of a technopole, Mozingo sees the value of natural seclusion in the suburbs for innovative companies; she sees the 'ideal of the pastoral' as the ideal of independence from society at large.

Although the corporate campus succeeds in retaining talent, there is consensus that it creates a typology that resembles an enclave of like-minded people and that its detachment from the city prevents it from fostering the informal connections vital to innovation. That model of spatial organization of labor in a post-industrial economy had a pattern and an explanation; companies looked for mono-functional and isolated areas to create a productive workspace, leaving behind stressful urban centers in the mid-20th century to build a perfect paradise of controlled working environments, in which hierarchy and specialization could be adequately managed. Today, this typology of suburban corporate/technology/business park has been replicated globally by corporations that have developed their own park in the suburbs, as a model to attract business activities, while isolating these creative and most knowledge-intensive activities from the possibility of taking the best from cities: diversity, complexity, and cultural mix.

Mozingo (2011) provides the background behind this suburban model, describing this detachment of the corporate world from the city as 'the separatist geography of business,' giving three main reasons for its ubiquity: a) the 'ideal' suburban living, b) its ability to 'capture' a desirable workforce, and c) its accessibility by car. She suggests that 'the campus milieux, particularly its ambient landscape space, was part and parcel of technological discovery.' A university-like campus signaled that this was not simply a business, but an organization working towards a higher purpose. A campus also suggested a sense of community. With recreational facilities and social spaces, the low-rise campus aimed, symbolically and practically, to promote interaction and collaboration within the businesses.

In recent years, we observe a similar approach by corporations seeking to establish technology business parks in urbanized central business districts. There, the tower typology is used to pursue higher-density urban footprints, leasing multiple floors in skyscrapers to maximize space efficiency and minimize land use. This vertical concentration contrasts sharply with suburban campus models, supporting more sustainable urban growth patterns and enhancing access to transit and city amenities, in line with the new generation's preferred mode of travel: walking. For example, Salesforce Park in San Francisco, a publicly constructed space, represents a branded urban common where public infrastructure and corporate influence intersect to create a city amenity, an urbanized version of "corporate technology park", which is more open to the city and spontaneous interactions (Figure 3). If we reassess the park through the lens of the Quadruple Helix model (introduced by Carayannis, Campbell 2009, 2010), we see that it sits at the nexus of government planning, private-sector branding, and civil-society access. However, the critique is that such attempts still feel less public (Wiener 2019) and more like a "tech bubble", disconnected from the surrounding socioeconomic reality of the downtown area.

3 Methods to Assess Urban-Suburban Divides in Innovation Development

As discussed, the study contributes to the innovation literature by incorporating a local innovation research and policy perspective, building upon recent influential studies that have examined the geographies of innovation (Arzaghi, Henderson 2008, Delgado et al. 2010, Shearmur 2012, Carlino, Kerr 2015, Kerr, Kominers 2015, Jang et al. 2017, Rammer et al. 2019, Balland et al. 2020, Matsiuk et al. 2024). This approach emphasizes that a locality's innovation and invention capacities arise from the interplay of territorially connected factors, such as human capital, institutional presence, collaborative networks, and even urban morphology, rather than from a single enabling factor. This hybrid methodological framework will help policymakers and researchers gain insights into conditions at lower spatial scales and spur more inclusive economic transitions by focusing on places' actual strengths and areas for improvement.

Patents are widely used in empirical studies as proxies for innovation outcomes, as they provide a quantifiable and systematic measure of inventive output across time and space (Griliches 1990), and have been employed to trace knowledge spillovers and the geographic diffusion of innovation (Jaffe, Trajtenberg 2002, Audretsch, Feldman 1996). Despite acknowledged limitations regarding varying propensity to patent across industries and the non-patentability of all innovations (Feldman, Audretsch 1999, Holl et al. 2024), as well as the fact that innovation does not necessarily rely solely on technological invention (Castaldi 2024), patents remain the most consistently available indicator to capture innovation activity at regional and urban scales.

In addition to the traditional data and metrics, the study identifies opportunities to use alternative data sources to capture synergies at lower spatial scales and provide place-sensitive insights with greater granularity. Urban, API-mined data can contribute to a broader understanding of the divides and capacities of urban and suburban neighborhoods. In the context of this analysis, ZIP codes serve as the geographic unit of analysis, with the acknowledgment that in most cases in New York State and Massachusetts, one ZIP code corresponds to only part of a neighborhood, while a neighborhood typically spans 1–3 ZIP codes. In other cases, a ZIP code covers more than one neighborhood.

The study employs exploratory and statistical approaches to examine how specific territorial factors are associated with variation in patent-based innovation performance at the ZIP code level. A comparison of the means of specific observable neighborhood assets between Massachusetts and New York State is used as the primary statistical method to illustrate a neighborhood's underlying synergies. The study examines differences in means of neighborhood variables across distinct analytical samples divided by future performance outcomes observed 10 years later (patents, patent decade change) and population density. The goal is to examine how such earlier neighborhood conditions and behaviors, as captured in census, business and open-source POI data, relate to the subsequent presence of formal invention performance in the ZIP code (as captured in granted patent records).

The study assumes that a 10-year delay is appropriate, as patents typically emerge several years after the underlying research and development, and the "true" lag of innovation is distributed over many years rather than a single fixed value. (Pakes, Griliches 1984, Griliches 1990, Hall et al. 2005). This method enables comparisons between groups based on later invention performance (e.g., high vs. typical performers) and population density (low vs. high-density neighborhoods). Although the approach followed provides insights into potential associations, it does not establish causality, as the study is observational and does not manipulate variables to control for all possible confounding factors.

The analysis is applied to 15 neighborhood characteristics, ranging from commuting patterns to the presence of numerous formal and informal places of interaction (parks and cafes), which could be considered innovation-enabling factors, bridging the gap between traditional and digital methods for tracking innovation and providing insights to inform local innovation policies. The feature selection was based on a previous study which used a larger set of variables and sought to identify which variables were more strongly associated with patent-based innovation outcomes (Oikonomaki, Komninos 2026). The following section provides a detailed overview of the neighborhood assets selected for this study.

4 Data & Empirical Strategy

4.1 Mapping Traditional Innovation Outcomes Across Urban & Suburban Neighborhoods

The study quantifies invention outcomes by the number of granted patents based on the assignee's location using USPTO data, but it doesn't distinguish whether these records represent corporations, institutions, or individuals, as the main goal is to identify the presence of granted patents in specific ZIP codes. Since the research focuses on neighborhood innovation capacity, using assignee data is selected as the most appropriate location over inventor addresses, given that local innovation capacity is about the local ecosystem: firms, research centers, and institutions that can develop, scale, and commercialize innovations. Inventor addresses alone can overestimate activity in a neighborhood if inventors live there but work for firms elsewhere. Instead, using assignee locations ensures measuring the actual organizational presence that contributes to local economic and innovation capacity. Assignees are typically companies or organizations with official, publicly available addresses recorded in USPTO data, including full fields such as ZIP code, city, and country. This analysis focuses on two U.S. states, New York and Massachusetts, and assesses the geographic concentration of granted patents at the ZIP code level. Looking at the top 4 ZIP codes in these two US states with the highest concentrations, we observe that granted patents are distributed equally across urban and suburban ZIP codes (Appendix Figures A.1 and A.2).

Patent records are among the few widely accepted, conventional proxies for innovation that can be measured at the local level. While patent data offer valuable insights, they are subject to temporal limitations, with processing delays of months or years creating a considerable lag between invention and patent grant (Hall et al. 2005). This temporal disconnect treats patents as retrospective rather than real-time indicators of invention.

This narrow focus overlooks contributions from industries where patents are less relevant, such as software, services, and grassroots innovation (Bessen, Hunt 2007).

Furthermore, the patent system's scope is inherently limited, capturing only formally protected inventions while overlooking diverse forms of innovation, including process optimizations, open-source contributions, and internal organizational improvements that firms choose not to patent or cannot legally protect (Arundel, Kabla 1998). For instance, open-source software development, which plays a crucial role in technological advancement, is often underrepresented in patent data despite its significant impact on digital innovation (von Hippel 2005). Similarly, service-sector innovations, which usually rely on business model changes rather than technical inventions, remain largely invisible in patent-based analyses (Gallouj, Weinstein 1997).

Even the EU innovation scoreboard, which compared innovation metrics across all EU countries in 2024, showed that EU performance improved in all dimensions except intellectual assets (European Commission 2024). This might be because a meaningful portion of innovation does not follow a patenting path, due to a) the rapid pace of innovation (ICT evolves rapidly, often rendering technologies obsolete within a few years, making the effort and cost of applying for a patent less attractive). Patents are usually the least emphasized method of protecting intellectual property among firms in the majority of manufacturing industries (Cohen et al. 2000), b) the open standards for compatibility and interoperability across different platforms and devices that ICT industry relied on (applying patents might hinder collaborative development) c) high costs and complex process: filing, maintaining, and defending patents is costly and time-consuming; for smaller tech firms or startups, these costs can be prohibitive.

These limitations underscore why traditional innovation indicators often lack the coverage, granularity, and timeliness needed for effective STI policy making (Kinne, Lenz 2021). Despite these limitations, the study, as a second step, uses patents as a well-established indicator to assess neighborhood characteristics as potential enabling factors of innovation activity.

4.2 Mapping Innovation Enabling Factors through POI Data

Today, digital tools are revolutionizing the ways innovation enabling factors are tracked by providing real-time, large-scale, and more inclusive insights compared to traditional methods. Digital footprints, such as company websites, job postings, and research collaborations, offer valuable indicators of innovation activity, capturing emerging trends that may not be reflected in patent filings alone (Younge, Kuhn 2016). For instance, platforms like LinkedIn job postings can reveal shifts in workforce demand for emerging technologies (Cockburn et al. 2018). Additionally, firm websites and press releases provide timely insights into product launches and strategic shifts, complementing conventional innovation metrics (Brynjolfsson, McElheran 2016). These digital sources not only expand the scope of innovation measurement but also democratize access to data, allowing for a more dynamic and granular understanding of technological change.

This way, we gain access to a wealth of real-time information, enabling us to discern trends and insights that may not be immediately apparent in conventional datasets. While traditional methods of data collection and analysis lag behind the pace of neighborhood development, analyses using alternative data sources can illuminate the broader economic implications of these movements. As part of a new exploration of innovation capacity, innovation spaces were geocoded using the Google Maps Geocoding API based on their entity names and keywords such as accelerator, co-working space, incubator, innovation center, innovation hub, innovation park, start-up, tech hub, and technology park. The geocoded locations of these spaces were retrieved in Python through the Google Maps API and then filtered to include only ZIP codes within the geographical boundaries of the two US states, New York and Massachusetts. For the purposes of this study, all these spaces were aggregated into a single feature.

To effectively address the research questions, it is crucial to assess the distribution of this newly constructed variable across the two states and to examine aerial views of the top four ZIP codes. In the case of the two US states, we observe that innovation infrastructure, such as tech hubs, incubators, accelerators, and innovation centers or start-ups are primarily located in urban areas, particularly near downtown areas, which can provide the needed vibrancy; this pattern is encountered in both states, as all top eight ZIP codes examined are situated in an urban context (Appendix Figures A.3 and A.4). Overall, this method can serve as an example for other metropolitan and regional scientists, economic geographers, and innovation analysts to track changes in the concentration of innovation spaces.

4.3 Mapping Innovation Capacities through Socioeconomic and Business Characteristics

For the ZIP code-level estimates of population density, the 2012 ACS 5-Year Estimates are retrieved from Census Bureau. The 2012 ACS 5-Year Estimates are also used to measure the median age and the percentage of population aged 25-34. As discussed, places with large concentrations of creative-class workers, who are disproportionately young, highly educated adults, rank highest as centers of innovation and high-tech industry (Florida 2002). The percentage of the population with a graduate degree serves as a proxy for the workforce capable of generating new technologies. For population density, land-area data are obtained from the Tiger Line shapefile, and total population data are from the 2012 ACS 5-Year Estimates.

ACS 5-Year Estimates are also used to examine commuting patterns to work, either a) walking, b) biking, c) using the car or van, d) using public transportation, or e) working from home, and the median income, which serves as a proxy for residents' economic capacity. Given that a substantial portion of the skilled workforce in these tech/innovation clusters originates from overseas, this analysis captures the number of H-1B visa applications filed using data from the OFLC website. The proportion of the highly skilled foreign-born workforce is a proxy that reflects the demand for a high level of education and, simultaneously, shows cultural diversity in a neighborhood. Other traditional indicators of innovation activity, such as the concentration of business registrations and patterns of firm-level R&D expenditure obtained from the Wharton Data Portal, were used across all 2-digit NAICS sectors.

4.4 Mapping Livability Spatial Features through OSM Data

For this analysis, which assesses how the urban environment is associated with patent generation, the concentration of significant infrastructure and urban spaces previously associated with vibrancy and livability is captured using open-source data. Livability indices typically include factors such as green spaces, mobility options, public amenities, and overall environmental quality, and are closely linked to a locality's innovation capacity. High urban quality attracts and retains high-skill workers who drive innovation, supporting Florida's argument that talent, amenities, and creative urban environments are fundamental to regional economic performance (Florida 2002, 2014).

In this context, OpenStreetMap (OSM) data offers significant advantages, particularly due to its free accessibility and daily updates (Zhou 2018). OSM data has emerged as a valuable resource for assessing enabling factors of innovation, particularly in the context of urban development, infrastructure, and land use. As a crowd-sourced geographic information system (GIS), OSM provides high-resolution spatial data on transportation networks, land use, points of interest, and urban structures. Incorporating spatial aspects, using tags such as cafés, squares, parks, and bus stops, captures potential factors that enable innovation. Such spaces are collected through the OpenStreetMap API and using ZIP codes of the two US states. For our analysis, these variables are combined with other socio-economic and business features to evaluate how the spatial distribution of amenities, urban infrastructure and spatial connectivity influence innovation dynamics at finer geographical resolution.

Despite these benefits, the use of OSM data presents some challenges: its collection is uncoordinated and not designed for academic purposes, and it relies on international volunteer activity, whereby volunteers map features of physical space by recording their geometries, coordinates, and feature tags (e.g., amenity = cafe). Previous studies have used these data to measure spaces of encounter in neighborhoods, which are crucial for

Table 1: Clusters Summary of LISA Cluster Types Corresponding to Massachusetts ZIP Codes

Cluster Type	ZIP Code Count	% of Total	Interpretation
High-High, Innovation Hotspots	8	3.5	Innovation hotspots
High-Low, Isolated Innovation Nodes	3	1.3	Isolated innovation nodes
Low-High, Spillover / At-Risk Zones	5	2.2	Spillover / at-risk zones
Low-Low, Underperforming Areas	29	12.6	Underperforming areas
Not Significant	186	80.5	No meaningful spatial structure

understanding urban change (Zhang, Pfoser 2019). OSM has been previously used to construct databases (Boeing, Waddell 2017) for comparative analyses across geographies (US cities). Similarly, OSM data can enable a geospatial approach by facilitating the assessment of innovation-enabling factors and helping policymakers and businesses identify neighborhoods with high patent-generation potential. For instance, by mapping clusters of innovative firms and assessing their relation with access to transportation, educational institutions, and amenities, decision-makers could develop more targeted policies to foster innovation ecosystems.

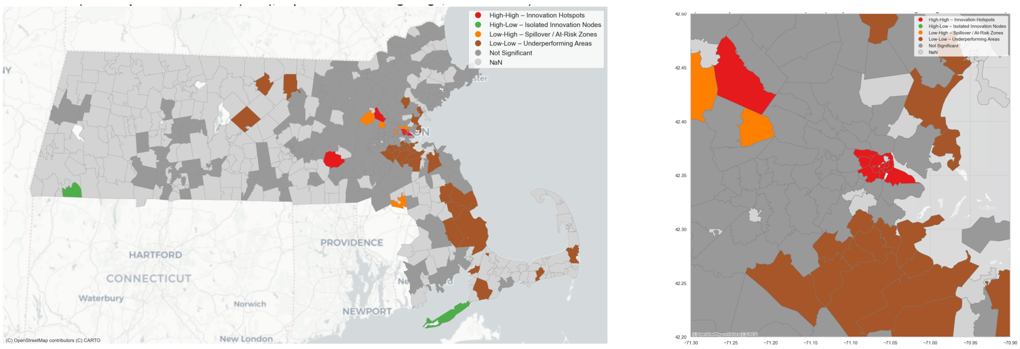
Additionally, spatial machine learning techniques applied to OSM data can reveal patterns, highlighting how factors such as urban density, public services, and access to key infrastructure can enable direct innovative endeavors (Kerr et al. 2014, Fehder et al. 2014). OSM often allows access to finer-grained information at lower spatial scales, thereby enhancing the potential to analyze more neighborhood characteristics. Despite their widespread use in other domains, API-mined urban informatics have not been extensively used to assess the innovation landscape. For this reason, such data are collected as part of this study.

5 Exploratory Analysis and Summary of Findings

5.1 Mapping the Spatial Structure of Innovation: Local Moran's I Cluster and Outlier Analysis in MA and NY

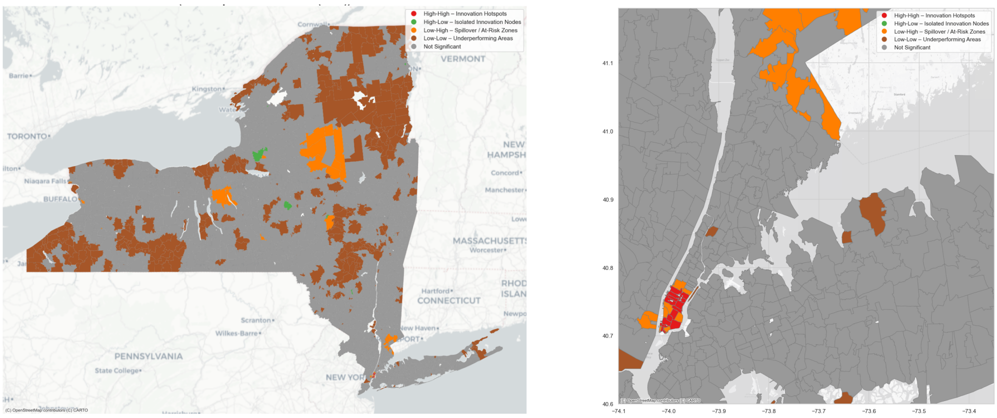
To assess the spatial structure of innovation development across urban and suburban contexts, the study draws on a suite of spatial analytical methods, including Local Moran's I cluster analysis for hot-spot detection, to reveal not only where innovation activity concentrates but also how the conditions sustaining it change as one moves from dense urban cores to suburban neighborhoods. The Local Moran's I maps indicate that statistically significant spatial clustering of patent-based innovation is relatively limited across Massachusetts (Table 1). High–High innovation hot-spots account for only 3.5% of all spatial units (8 ZIP codes), suggesting that strong, self-reinforcing innovation ecosystems are rare and highly localized. High–Low clusters, representing isolated innovation nodes (1.3%, 3 ZIP codes), highlight places that perform well in terms of patent intensity despite being surrounded by weaker innovation contexts, pointing to localized institutional or sectoral advantages that do not readily spill over to neighboring areas. Low–High spillover or at-risk zones (2.2%, 5 ZIP codes) appear infrequently, indicating that proximity to innovation hot-spots does not automatically translate into higher local innovation performance. By contrast, Low–Low underperforming areas comprise 12.6% of the region (29 ZIP codes), revealing pockets of persistent innovation disadvantage where low patenting rates are reinforced by similarly weak surrounding contexts. The most striking result is the dominance of statistically non-significant areas, which represent 80.5% of all units (186 ZIP codes). This suggests that, for the majority of the region, patent intensity does not exhibit strong spatial autocorrelation and is instead shaped by highly localized or non-spatial factors. Taken together, these findings point to a fragmented innovation landscape, where only a small subset of places forms coherent spatial clusters, while most areas operate outside clearly defined spatial innovation dynamics.

The same analysis for New York reveals a markedly different spatial structure of innovation (Table 2). High–High innovation hot-spots are exceedingly rare, accounting for just 0.9% of spatial units (17 ZIP codes), indicating that strongly clustered centers of



(a) Massachusetts (b) Boston Metro Area

Figure 4: Local Moran's I Map of Patent Intensity for Massachusetts (Granted Patents per 1,000 Residents, 2022), author's elaboration



(a) New York State (b) New York Metro Area

Figure 5: Local Moran's I Map of Patent Intensity for New York State (Granted Patents per 1,000 Residents in 2022), author's elaboration

patent activity are highly exceptional. In contrast to Massachusetts neighborhoods (ZIP codes), however, Low–Low clusters, representing underperforming areas, are far more prevalent, comprising 21.4% of all areas (397 ZIP codes). This pattern suggests that innovation disadvantage in New York State is more often spatially reinforced, with large contiguous areas characterized by persistently low patent intensity. High–Low clusters, interpreted as isolated innovation nodes, are relatively uncommon, representing only 0.5% of spatial units (10 ZIP codes). This indicates that instances in which individual locations substantially outperform their surrounding context are rare, pointing to a limited role for highly localized, site-specific innovation advantages that do not diffuse spatially. Low–High spillover or at-risk zones account for 2.1% of areas (39 ZIP codes), suggesting modest potential for innovation diffusion from nearby high-performing areas, but also highlighting locations that may be failing to capitalize on their innovative surroundings.

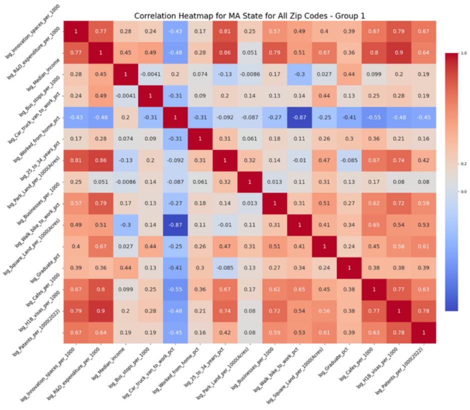
The majority of New York ZIP codes are statistically non-significant (75.1%, 1,393 ZIP codes), underscoring the overall weakness of spatial autocorrelation in patent intensity across the state. Taken together, these results point to an innovation geography in New York that is neither strongly clustered nor dominated by isolated high-performing nodes but instead characterized by extensive areas of spatially persistent underperformance embedded within a largely unstructured spatial landscape. In comparison, Massachusetts is distinguished by a small number of coherent innovation hot-spots alongside more clearly defined spatial patterns of advantage and disadvantage, whereas New York exhibits a more diffuse and uneven innovation structure with limited evidence of spatial spillovers.

Table 2: Clusters Summary of LISA Cluster Types Corresponding to New York State's ZIP codes

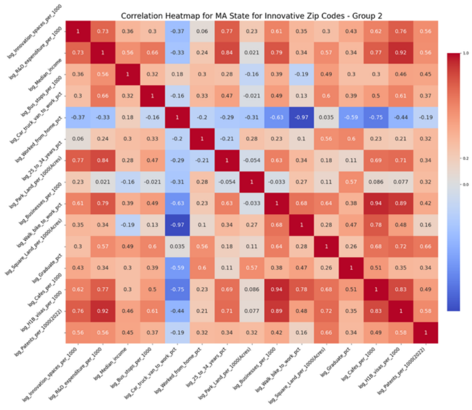
Cluster Type	ZIP Code Count	% of Total	Interpretation
High-High, Innovation Hotspots	17	0.9	Innovation hotspots
High-Low, Isolated Innovation Nodes	10	0.5	Isolated innovation nodes
Low-High, Spillover / At-Risk Zones	39	2.1	Spillover / at-risk zones
Low-Low, Underperforming Areas	397	21.4	Underperforming areas
Not Significant	1393	75.1	No meaningful spatial structure

5.2 A Comprehensive Overview of Analytical Samples and a Comparative Analysis between Massachusetts and New York State

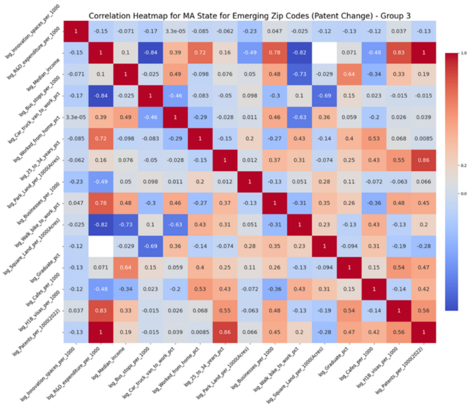
Revealing specific neighborhood patterns in educational attainment, commutes, or urban morphology helps pinpoint neighborhoods with distinct needs or unique characteristics. This section presents a comparison of mean differences for selected variables across four analytical groups. More specifically, the four samples were created for both Massachusetts and New York State according to the following criteria: The first sample includes all ZIP codes with census data in the state, while the second sample consists of ZIP codes that are within the 10th percentile for patents granted in 2022. Analyzing ZIP codes with varying levels of granted patents enables a better understanding of how different factors associate with patent-based innovation performance. The third group includes ZIP codes that are in the top 10th percentile for change in the number of patent records during the decade 2012-2022.



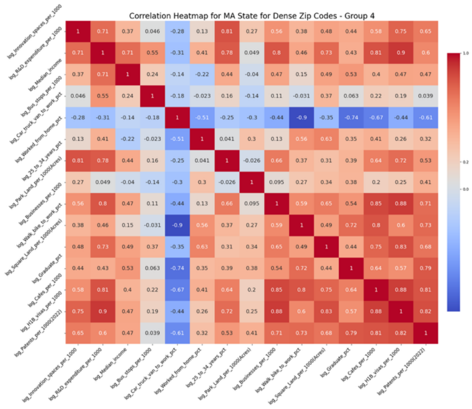
(a) All ZIP codes



(b) Innovative ZIP codes

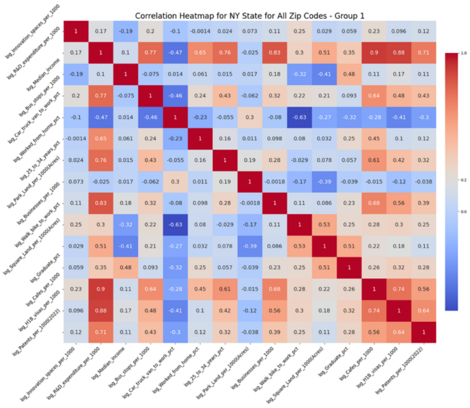


(c) Emerging ZIP codes

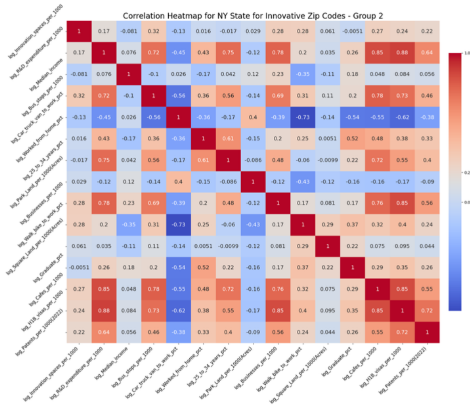


(d) Dense ZIP codes

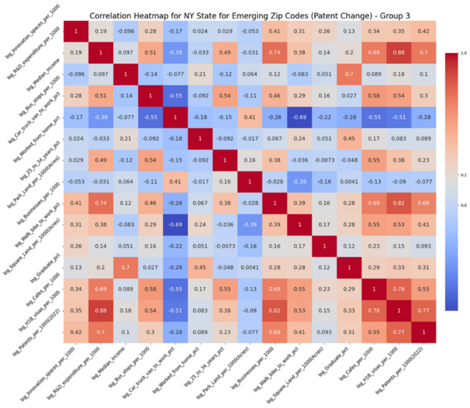
Figure 6: Pearson's correlation matrix of the most significant log-transformed variables for the state of Massachusetts, authors' elaboration



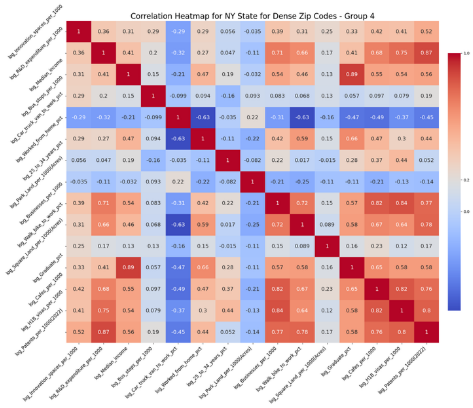
(a) All ZIP codes



(b) Innovative ZIP codes



(c) Emerging ZIP codes



(d) Dense ZIP codes

Figure 7: Pearson's correlation matrix of the most significant log-transformed variables for the state of New York, authors' elaboration

The fourth is a group of urban ZIP codes that ranked in the 10^{th} percentile for population density in 2022. Urban areas differ significantly in population density, which can affect economic, social, and technological dynamics; high-density districts may exhibit different levels of invention than low-density areas. This last group helps examine how population density is associated with innovation outcomes. Overall, the selection of the criteria helps capture a range of urban contexts, from highly innovative, densely populated districts to less innovative, sparsely populated ones, providing a more comprehensive analysis. By assessing neighborhoods across these two dimensions, we can better understand how urban density and innovation are interrelated and how they influence economic outcomes.

The correlation matrices for the four groups of the two states also provide an additional assessment of the relationships between the neighborhood's asset variables captured in 2012 and the granted patents recorded in 2022, thereby identifying relationships among variables. To achieve this, two sets of transformed features were generated: one standardized and the other log-transformed. The log-transformed features were used to evaluate the relationships between the variables, as their natural logarithms improve linearity. In both states, our analysis reveals a significant correlation (0.78 for Group 1 and 0.82 for Group 4 of MA, and 0.64 for Group 1 and 0.80 in Group 4 of NY) between the number of H-1B applications filed in 2012 and the number of patent applications filed in that neighborhood 10 years later (2022) (Figures 6 and 7). Group 1, all ZIP codes of New York State, has more ZIP codes with exceptionally high numbers of H-1B applications, which may indicate ZIP codes with high concentrations of companies that frequently sponsor H-1B visas, while Group 1 of Massachusetts has fewer extreme outliers.

Additionally, in both states, a moderate to strong correlation is observed between the patent-based innovation outcomes and the R&D expenditure per 1000 residents, 0.56-0.64 for Groups 1,2,4 while we observe a very strong correlation for emerging neighborhoods (Group 3) in Massachusetts, and ranging from 0.64 in Group 2 to 0.87 in Group 4 for New York. In both states, a strong correlation (0.76 in NY and 0.81 in MA) is observed between the number of cafes and the number of granted patents in dense neighborhoods for 2022 (Group 4), with moderate correlations in Group 2 (ZIP codes with the highest patent activity in 2022). In Massachusetts, we find a high correlation between granted patents and the number of innovation spaces per 1000 residents (0.67), the square land (acres) per 1000 residents (0.68), the percentage of people with graduate degree (0.79), the business registrations per 1000 residents (0.71) and the percentage of people walking to work (0.73) within the dense ZIP codes (Group 4). In New York State, in the same group (4), we find a high correlation between granted patents and the business registrations per 1000 residents (0.77) and the percentage of people walking to work (0.78).

5.3 A Comparative Analysis of Differences in Means between Massachusetts and New York State

For the statistical analysis, Welch's ANOVA and post hoc multiple comparisons tests were conducted in SPSS to assess whether there were significant differences in the selected neighborhood variables among the four analytical samples. The mean was selected over the median as the selection criterion because it better captures variability and is more suitable for comparison. This assessment helps us better understand the relationship of the identified significant assets with neighborhoods' innovation performance, quantified by patents. Levene's test of homogeneity of variances was used to validate Welch's ANOVA assumptions. When significant differences were observed in ANOVA with the Welch correction, Games-Howell tests were used to identify which sample groups differed. The significance level was set at $\alpha = 0.05$.

First, we find notable differences in H-1B visa applications across groups, but New York State shows greater within-group variance in H-1B visa application counts. This suggests a higher concentration of H-1B visa holders in New York ZIP codes with higher patenting activity than in Massachusetts (Tables 3 and 4). In both states, Group 2 (innovators) with Group 4 (dense) neighborhoods exhibit the lowest percentage of people commuting to work by car/truck/van and the highest percentage walking or biking to work, showing greater walkability in these ZIP codes, especially when compared to Group 1, which includes both urban and suburban ZIP codes. We find that Group 2 has the highest mean number of bus stops per 1,000 residents in both Massachusetts (2.60) and New York (6.59), indicating more public transit facilities in these areas, with Group 1 and other groups far behind. Group 2 also has the highest mean percentage of residents with a graduate degree in Massachusetts (0.30) and New York (0.20), reflecting the highest educational attainment in these ZIP codes.

On the other hand, median income differences are significant between Groups 1 and 2 in both states, with Group 2 (innovators) and Group 3 (emerging) consistently showing higher median incomes. However, Massachusetts exhibits a smaller range of income variability across groups than New York. This indicates a more balanced economic distribution across groups in Massachusetts. New York shows significant differences in the parkland (acres) per 1000 residents between groups, while Group 2 in both states shows a substantial difference in the number of cafes, which is confirmed by the correlation matrices; in both states, the number of cafes shows a significant correlation with the number of granted patents in Group 4, the dense ZIP codes. Looking at the means of analytical groups in New York, we see that the density of cafes in ZIP codes with high patenting activity is approximately ten times higher than that of typical ZIP codes within the state, suggesting that these areas are probably situated in high-density urban neighborhoods, more often characterized by a high concentration of such amenities. New York generally exhibits greater inter-group variability, particularly in H-1B visas, income, and cafes. Massachusetts shows similar trends but with slightly less variability, indicating more evenly distributed characteristics across neighborhoods.

Table 3: Statistics & Mean Comparisons across Massachusetts' Analytical Samples

	M / SD G1. All ZIP Codes	M / SD G2. Innovators	M / SD G3. Emerging	M / SD G4. Dense	F	p	η_p^2
H-1B applications per 1000 residents	0.74 ± 6.40 [G2, G3]	6.11 ± 19.48 [G1]	5.63 ± 19.35 [G1]	3.92 ± 15.85	**2.87**	**0.04**	**0.03**
Years 25 to 34 pct	0.006 ± 0.03	0.004 ± 0.006	0.002 ± 0.003	0.002 ± 0.004	0.66	0.57	0.03
Car truck van to work (pct)	0.82 ± 0.17 [G2, G3, G4]	0.63 ± 0.28 [G1, G4]	0.66 ± 0.27 [G1, G4]	0.49 ± 0.21 [G1, G2, G3]	**43.62**	**<.001**	**0.20**
Walk bike to work (pct)	0.05 ± 0.10 [G2, G3, G4]	0.17 ± 0.19 [G1]	0.14 ± 0.18 [G1]	0.20 ± 0.18 [G1]	**86.49**	**<.001**	**0.16**
Work from home (pct)	0.05 ± 0.05	0.05 ± 0.03	0.05 ± 0.02	0.04 ± 0.04	1.09	>0.35	0.005
Graduate degree (pct)	0.17 ± 0.12 [G2, G3]	0.30 ± 0.13 [G1, G4]	0.29 ± 0.12 [G1, G4]	0.21 ± 0.15 [G2, G3]	**19.41**	**<.001**	**0.09**
Median age	42.05 ± 8.03 [G2, G3, G4]	38.74 ± 7.94 [G1, G4]	37.26 ± 6.81 [G1, G4]	32.70 ± 6.38 [G1, G2, G3]	**26.64**	**<.001**	**0.11**
R&D expenditure per 1000 residents	0.28 ± 3.55 [G2, G3]	2.67 ± 10.98 [G1]	2.40 ± 10.82 [G1]	1.10 ± 4.05	1.93	0 .12	0.02
Median income	34929.56 ± 11612.46 [G3]	41271.60 ± 17467.43 [G4]	41409.50 ± 15504.83 [G1, G4]	31534.94 ± 18526.00 [G2, G3]	**5.35**	**0.002**	**0.04**
Bus stops per 1000 residents	1.36 ± 2.61 [G2, G3, G4]	2.60 ± 3.76 [G1]	2.57 ± 3.67 [G1]	2.88 ± 2.19 [G1]	**8.81**	**<.001**	**0.04**
Park land (acres) per 1000 residents	27.06 ± 343.80 [G2, G3, G4]	20.81 ± 50.68 [G1, G3, G4]	13.09 ± 25.84 [G1]	10.40 ± 23.92 [G1]	0.075	0.97	0.00
Cafes per 1000 residents	0.22 ± 0.65 [G2]	0.86 ± 1.63 [G1]	0.74 ± 1.55	0.74 ± 1.55	**13.825**	**<.001**	**0.06**
Business registrations per 1000 residents	63.19 ± 127.77 [G2, G3, G4]	168.37 ± 365.20 [G1]	154.85 ± 362.24 [G1]	148.92 ± 362.30 [G1]	**3.23**	**0.02**	**0.04**
Square land (acres) per 1000 residents	0.03 ± 0.17	0.21 ± 0.48	0.18 ± 0.47	0.22 ± 0.45	**2.28**	**0.085**	**0.02**
Innovation spaces per 1000 residents	0.03 ± 0.19 [G2, G3, G4]	0.20 ± 0.58 [G1]	0.16 ± 0.54 [G1]	0.17 ± 0.55 [G1]	**3.44**	**0.02**	**0.04**

Note: Means (M) and standard deviations (SD) of measures by analytical sample. Values are in relative form (percentages, counts, and $ for median income). Welch's ANOVA results: F statistic, p-value (p), and partial η_p^2. Group 1: All Massachusetts' ZIP codes, Group 2: Innovative ZIP codes (patents in top 10 percentile), Group 3: Emerging ZIP codes (Patent Change in top 10 percentile), Group 4: Dense urban ZIP codes (population density in top 10 percentile). The group labels in parentheses (G1–G4) indicate which groups exhibit statistically significant differences in means based on post hoc tests.

Table 4: Statistics & Mean Comparisons across New York's Analytical Samples

	M/SD G1. All ZIP Codes	M/SD G2. Innovators	M/SD G3. Emerging	M/SD G4. Dense	F	p	η_p^2
H-1B applications per 1000 residents	1.51 ± 46.57 [G2]	13.09 ± 146.68 [G1]	1.42 ± 6.62	2.20 ± 9.78	0.68	0.56	0.03
Years 25 to 34 pct	0.02 ± 0.51	0.12 ± 1.62	0.002 ± 0.015	0.001 ± .011	1.37	0.25	0.02
Car truck van to work (pct)	0.80 ± 0.22 [G2, G3, G4]	0.65 ± 0.32 [G1, G4]	0.64 ± 0.31 [G1, G4]	0.31 ± 0.22 [G1, G2, G3]	**287.86**	**<.001**	0.23
Walk bike to work (pct)	0.04 ± 0.08 [G2, G3, G4]	0.09 ± 0.13 [G1]	0.07 ± 0.09 [G1, G4]	0.11 ± 0.09 [G1, G3]	**37.61**	**<.001**	0.05
Worked from home (pct)	0.04 ± 0.05 [G4]	0.05 ± 0.04 [G4]	0.04 ± 0.02 [G4]	0.03 ± 0.02 [G1, G2, G3]	**10.42**	**<.001**	0.05
Graduate degree (pct)	0.12 ± 0.10 [G2, G3]	0.20 ± 0.13 [G1, G3, G4]	0.17 ± 0.11 [G1, G2]	0.15 ± 0.12 [G2]	**34.86**	**<.001**	0.05
Median age	41.70 ± 8.12 [G2, G3, G4]	39.11 ± 7.51 [G1, G4]	38.99 ± 6.59 [G1, G4]	35.89 ± 5.44 [G1, G2, G3]	**63.82**	**<.001**	0.05
R&D expenditure per 1000 residents	0.21 ± 7.74	2.11 ± 24.44	0.18 ± 0.91	0.16 ± 0.74	**0.41**	**<.001**	0.04
Median income	30995.82 ± 12555.55 [G2, G3]	37862.36 ± 20098.96 [G1]	35383.25 ± 16781.82 [G1]	34028.17 ± 20335.41 [G1]	**12.67**	**<.001**	0.23
Bus stops per 1000 residents	0.87 ± 17.79 [G2]	6.59 ± 55.97 [G1]	2.34 ± 23.42	1.25 ± 0.82	1.02	0.38	0.04
Park land (acres) per 1000 residents	0.82 ± 2.04 [G4]	1.00 ± 1.62 [G4]	0.71 ± 1.12 [G4]	0.40 ± 1.01 [G1, G2, G3]	**22.37**	**<.001**	0.03
Cafes per 1000 residents	0.38 ± 8.08 [G2]	2.81 ± 25.43 [G1, G4]	0.71 ± 6.18	0.38 ± 0.78 [G2]	0.80	0.49	0.04
Business registrations per 1000 residents	105.76 ± 1873.08 [G2]	572.29 ± 5891.90 [G1]	92.99 ± 308.84	68.39 ± 160.02	0.99	0.39	0.03
Square land (acres) per 1000 residents	0.001 ± 0.02 [G2]	0.01 ± 0.07 [G1]	0.005 ± 0.03	0.006 ± 0.03	**3.02**	**.030**	0.05
Innovation spaces per 1000 residents	0.02 ± 0.31	0.04 ± 0.25	0.02 ± 0.06	0.01 ± 0.04	0.62	0.60	0.01

Note: Means (M) and standard deviations (SD) of measures by analytical sample. Values are presented in relative form (percentages, counts, and $ for median income). Welch's ANOVA results: test statistic value (F), p-value (p), and partial eta-squared (η_p^2). Group 1: All Massachusetts ZIP codes, Group 2: Innovative ZIP codes (patents in top 10 percentile), Group 3: Emerging ZIP codes (Patent Change in top 10 percentile), Group 4: Dense urban ZIP codes (population density in top 10 percentile). The group labels in parentheses (G1–G4) indicate which groups exhibit statistically significant differences in means based on post hoc tests.

6 Conclusions, Limitations, and Future Outlook

In this study, a comparative analysis of neighborhood characteristics across urban and suburban neighborhoods was conducted by merging data across diverse categories: a) the social characteristics, b) the economic, c) land use context, d) mobility behaviors, e) urban morphology, and f) infrastructure to understand associations with patenting activities and concentration of innovation spaces. First, correlation matrices are used to assess the association of each neighborhood feature with patent-based innovation outcomes. These findings indicate that several features are strongly associated with high levels of granted patents (H-1B visa applications, percentage of people walking to work, percentage of people with a graduate degree, median income, low- and high-tech innovation spaces, number of cafes, and parks). Secondly, differences in the means of selected variables helped us identify the main distinctions across neighborhoods with varying levels of patenting activity and population density. The differences in significant variables across four analytical samples from the two US states are examined to ensure a comprehensive assessment across a range of contexts.

This approach enables the identification of patterns and differences specific to each analytical sample (e.g., typical, innovators, emerging, or dense neighborhoods), offering deeper insights into how the significant variables identified behave within each sample. The analysis of mean differences across samples highlights neighborhoods with different levels of patent-generating activity and density in 2022, systematically showing distinct levels of these key neighborhood characteristics measured a decade earlier in 2012. If these differences are assessed by urban planners, policymakers, and stakeholders, they could tailor interventions and resource allocations to promote development across diverse neighborhoods accordingly.

The results from the difference-in-means comparison provide strong evidence that specific socioeconomic, infrastructure, and spatial variables are more strongly associated with local patent-based innovation performance. The analysis confirmed that specific neighborhood assets, such as the percentage of people walking or biking to work, the percentage of people with a graduate degree, median age, median income, H-1B visa applications, and bus stops per 1000 residents, exhibit similarities in mean differences across analytical samples of the neighborhoods of the two US states. Massachusetts exhibited trends similar to New York State with respect to the key variables, albeit with slightly less variability; few variables show significant mean differences across neighborhoods within a single state.

This exploratory geospatial data analysis emphasized the role of large, urban, and API-mined data collected via Google Maps and OpenStreetMap, enabling fine-grained decision-making at finer spatial scales across both urban and suburban ZIP codes. By tracking the growth of specific variables before patent activity occurs, policymakers and researchers can monitor neighborhood dynamics to inform local innovation policy. The rise of digital tools and API-mined data has transformed how innovation activities are tracked. These methods provide large-scale and granular insights into innovation activity, capturing trends from diverse sources, including open-source repositories. Unlike 'traditional' proprietary data and measures, open data tracking facilitates geospatial analysis of patent-based innovation-enabling factors across regions and industries, highlighting emerging hubs that might otherwise be overlooked. However, challenges such as data reliability, ethical concerns, and potential biases must be addressed to ensure accurate and fair representation.

Ultimately, policymakers can leverage urban informatics to complement traditional data sources and metrics in designing targeted policies that support underrepresented regions, emerging industries, and grassroots innovation. By integrating real-time insights, local governments can foster more adaptive and evidence-based innovation strategies. Future research should focus on enhancing data accuracy and validation methods. Additionally, expanding digital innovation tracking to developing regions and informal sectors will provide a more inclusive view of innovation trends. For this to happen, collaboration among academia, industry, and policymakers will be essential to refine and scale these methodologies.

References

Arundel A, Kabla I (1998) What percentage of innovations are patented? Empirical estimates for European firms. *Research Policy* 27: 127–141. CrossRef

Arzaghi M, Henderson J (2008) Networking off Madison Avenue. *Review of Economic Studies* 75: 1011–1038. CrossRef

Audretsch D, Feldman M (1996) R&D spillovers and the geography of innovation and production. *American Economic Review* 86: 630–640

Balland P, Boschma R, Crespo J, Rigby D (2020) Smart specialization policy in the European Union: Relatedness, knowledge complexity and regional diversification. *Regional Studies* 54: 1252–1268. CrossRef

Balland P, Rigby D (2017) The geography of complex knowledge. *Economic Geography* 93: 1–23. CrossRef

Berry C, Glaeser E (2005) The divergence of human capital levels across cities. *Papers in Regional Science* 84: 407–444. CrossRef

Bessen J, Hunt R (2007) An empirical look at software patents. *Journal of Economics & Management Strategy* 16: 157–189. CrossRef

Boeing G, Waddell P (2017) New insights into rental housing markets across the United States: Web scraping and analyzing craigslist rental listings. *Journal of Planning Education and Research* 37: 457–476. CrossRef

Boschma R (2005) Proximity and innovation: A critical assessment. *Regional Studies* 39: 61–74. CrossRef

Brynjolfsson E, McElheran K (2016) The rapid adoption of data-driven decision-making. *American Economic Review* 106: 133–139. CrossRef

Capdevila I (2015) Co-working spaces and the localized dynamics of innovation in Barcelona. *International Journal of Innovation Management* 19: 1540004. CrossRef

Carayannis E, Campbell D (2009) 'mode 3' and 'quadruple helix': toward a 21st century fractal innovation ecosystem. *International Journal of Technological Management* 46: 201. CrossRef

Carayannis E, Campbell D (2010) Triple helix, quadruple helix and quintuple helix and how do knowledge, innovation and environment relate to each other? *International Journal of Social Ecology and Sustainable Development* 1: 41–69

Carlino G, Kerr W (2015) Agglomeration and innovation. In: *Handbook of Regional and Urban Economics*, Volume 5. Elsevier, 349–404. CrossRef

Castaldi C (2024) The geography of urban innovation beyond patents only: New evidence on large and secondary cities in the United States. *Urban Studies* 61: 1248–1272. CrossRef

Castells M, Hall P (1994) *Technopoles of the world: The making of twenty-first-century industrial complexes*. Routledge, London

Cockburn I, Henderson R, Stern S (2018) The impact of artificial intelligence on innovation. NBER Working Paper No. 24449. CrossRef

Cohen W, Nelson R, Walsh J (2000) Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not). NBER Working Paper No. 7552. CrossRef

Cooke P, Parrilli MD, Curbelo JL (2012) *Innovation, Global Change and Territorial Resilience*. Edward Elgar Publishing. CrossRef

Cortright J, Mayer H (2002) Signs of life: The growth of biotechnology centers in the U.S. *Greater Philadelphia Regional Review*. https://www.researchgate.net/publication/242301940_Signs_of_Life_The_Growth_of_Biotechnology_Centers_in_the_US

Delgado M, Porter M, Stern S (2010) Clusters and entrepreneurship. *Journal of Economic Geography* 10: 495–518. CrossRef

Diemer A, Iammarino S, Rodríguez-Pose A, Storper M (2022) The regional development trap in Europe. *Economic Geography* 98: 487–509. CrossRef

D'Este P, Guy F, Iammarino S (2013) Shaping the formation of university–industry research collaborations: What type of proximity does really matter? *Journal of Economic Geography* 13: 537–558. CrossRef

European Commission – European Commission: Directorate-General for Research and Innovation (2024) European innovation score-board 2024. Publications Office of the European Union, https://data.europa.eu/doi/10.2777/779689

Ezcurra R, Rodríguez-Pose A (2013) Political decentralization, economic growth and regional disparities in the OECD. *Regional Studies* 47: 388–401. CrossRef

Fehder D, Hochberg Y, Barrot J, Cohen S, Hausman N, Murray F, Nanda R, Stern S (2014) Accelerators and the regional supply of venture capital investment. Available at SSRN: https://ssrn.com/abstract=2518668. CrossRef

Feldman M (2001) The entrepreneurial event revisited: Firm formation in a regional context. *Industrial and Corporate Change* 10: 861–891. CrossRef

Feldman MP, Audretsch DB (1999) Innovation in cities: Science-based diversity, specialization and localized competition. *European Economic Review* 43: 409–429. CrossRef

Florida R (2002) *The rise of the creative class: And how it's transforming work, leisure, community and everyday life*. Basic Books, New York

Florida R (2014) *The rise of the creative class—revisited*. Basic Books, New York

Gallouj F, Weinstein O (1997) Innovation in services. *Research Policy* 26: 537–556. CrossRef

Griliches Z (1990) Patent statistics as economic indicators: A survey. *Journal of Economic Literature* 28: 1661–1707

Hall B, Jaffe A, Trajtenberg M (2005) Market value and patent citations. *RAND Journal of Economics* 36: 16–38

Herrera L, Nieto M (2008) The national innovation policy effect according to firm location. *Technovation* 28: 540–550. CrossRef

Hidalgo C, Klinger B, Barabási A, Hausmann R (2007) The product space conditions the development of nations. *Science* 317: 482–487. CrossRef

Holl A, Martínez C, Casado C (2024) The changing geography of innovation: Comparing urban, suburban and rural areas. *Growth and Change* 55. CrossRef

Iammarino S (2018) *FDI, global value chains, and local development*. Palgrave Macmillan, London

Jacobs J (1969) *The death and life of great American cities*. Random House, New York

Jaffe AB, Trajtenberg M (2002) *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. The MIT Press, Cambridge MA. CrossRef

Jang S, Kim J, Zedtwitz M (2017) The importance of spatial agglomeration in product innovation: A microgeography perspective. *Journal of Business Research* 78: 143–154. CrossRef

Katz B, Wagner J (2014) The rise of innovation districts: A new geography of innovation in America. Brookings institution

Kerr W, Kominers S (2015) Agglomerative forces and cluster shapes. *Review of Economics and Statistics* 97: 877–899. CrossRef

Kerr W, Nanda R, Rhodes-Kropf M (2014) Entrepreneurship as experimentation. *Journal of Economic Perspectives* 28: 25–48. CrossRef

Kinne J, Lenz D (2021) Predicting innovative firms using web mining and deep learning. *PLOS ONE* 16: 0249071. CrossRef

Leyshon A, Thrift N (1997) *Money/space: Geographies of monetary transformation*. Routledge, London

Loures L, Santos R, Panagopoulos T (2007) Urban parks and sustainable city planning: The case of Portimão, Portugal. *WSEAS Transactions on Environment and Development* 3: 171–180

Mariotti I, Akhavan M, Matteo D (2021) The geography of coworking spaces and the effects on the urban context: Are pole areas gaining? In: Mariotti I, Akhavan M, Matteo D (eds), *New Workplaces—Location Patterns, Urban Effects and Development Trajectories*. Springer, Cham, 169–194. CrossRef

Matsiuk N, Doloreux D, Shearmur R (2024) Beyond buzz: Knowledge interactions, innovation, and neighborhood characteristics. *Journal of Economic Geography* lbae026. CrossRef

Mewes L, Broekel T (2022) Technological complexity and economic growth of regions. *Research Policy* 51: 104550. CrossRef

Mozingo L (2011) *Pastoral capitalism: A history of suburban corporate landscapes*. MIT Press, Cambridge, MA. CrossRef

OECD (2018) *Job creation and local economic development 2018: Preparing for the future of work*. OECD Publishing, Paris. CrossRef

Oikonomaki E, Komninos N (2026) Predicting local innovation through AI and geospatial models: Revisiting the innovation district theory. *Technovation* 156: 103544. CrossRef

Pakes A, Griliches Z (1984) Patents and r&d at the firm level: A first look. In: Griliches Z (ed), *R&D, Patents, and Productivity*. University of Chicago Press, Chicago IL, 55–72. CrossRef

Pike A, Rodríguez-Pose A, Tomaney J (2006) *Local and regional development*. Routledge, London. CrossRef

Pinheiro F, Balland P, Boschma R, Hartmann D (2022) The dark side of the geography of innovation: Relatedness, complexity and regional inequality in Europe. *Regional Studies* 59. CrossRef

Puentes R, Roberto E (2008) Commuting to opportunity: The working poor and commuting in the United States. Brookings institution

Rammer C, Kinne J, Blind K (2019) Knowledge proximity and firm innovation: A microgeographic analysis for Berlin. *Urban Studies* 57: 996–1014. CrossRef

Roche M, Oettl A, Catalini C (2024) Proximate (co-)working: Knowledge spillovers and social interactions. *Management Science* 70: 8245–8264. CrossRef

Rodríguez-Pose A, Dijkstra L, Poelman H (2024) The geography of EU discontent and the regional development trap. *Economic Geography* 100: 213–245. CrossRef

Rodríguez-Pose A, Gill N (2003) The global trend towards devolution and its implications. *Environment and Planning C: Government and Policy* 21: 333–351. CrossRef

Rodríguez-Pose A, Wilkie C (2017) Revamping local and regional development through place-based strategies. *Cityscape* 19: 151–170

Sennett R (2006) *The culture of the new capitalism*. Yale University Press, New Haven

Shearmur R (2012) The geography of intrametropolitan KIBS innovation: Distinguishing agglomeration economies from innovation dynamics. *Urban Studies* 49: 2331–2356. CrossRef

Storper M (1995) The resurgence of regional economies, ten years later: The region as a nexus of untraded interdependencies. *European Urban and Regional Studies* 2: 191–221. CrossRef

Vazquez-Barquero A (1999) Inward investment and endogenous development. *Entrepreneurship & Regional Development* 11: 79–93. CrossRef

von Hippel E (2005) *Democratizing innovation*. MIT Press, Cambridge, MA. CrossRef

Wiener A (2019) The floating utopia of Salesforce Park. *The New Yorker*

World Bank (2009) *World development report 2009: Reshaping economic geography*. World Bank, Washington, DC

Younge K, Kuhn J (2016) Patent-to-patent similarity: A vector space model. SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2709238

Zhang L, Pfoser D (2019) Using OpenStreetMap point-of-interest data to model urban change. *PLOS ONE* 14: 0212606. CrossRef

Zhou Q (2018) Exploring the relationship between density and completeness of urban building data in OpenStreetMap. *International Journal of Geographical Information Science* 32: 257–281. CrossRef

A Appendix:

A.1 Patent Records (Granted) in New York State

Figure A.1: Aerial views of the top 4 ZIP codes with the highest number of granted patent records in 2022 for New York State, author's elaboration

A.2 Patent Records (Granted) in Massachusetts

Figure A.2: Aerial views of the top 4 ZIP codes with the highest number of granted patent records in 2022 for Massachusetts, author's elaboration

A.3 Innovation Spaces in New York State

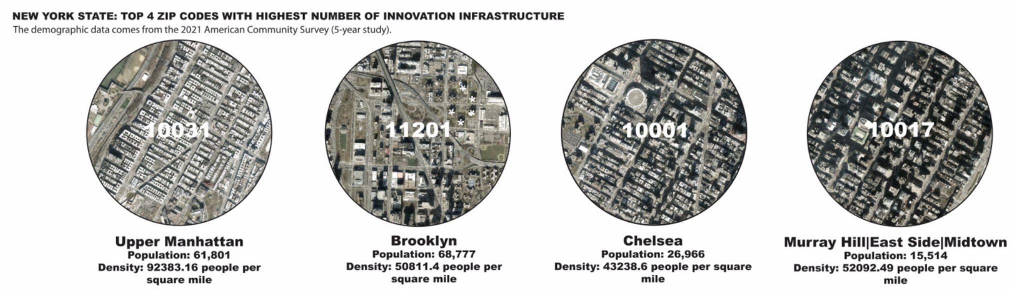


Figure A.3: Aerial views of the top 4 ZIP codes with the highest number of innovation spaces in 2022 for New York State, author's elaboration

A.4 Innovation Spaces in Massachusetts

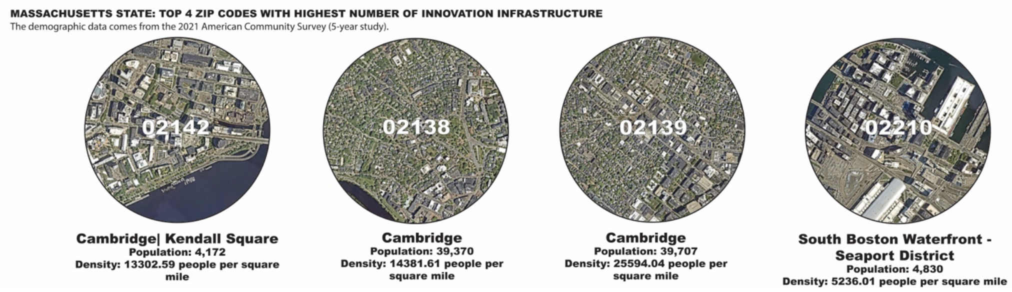


Figure A.4: Aerial views of the top 4 ZIP codes with the highest number of innovation spaces in 2022 for Massachusetts, author's elaboration

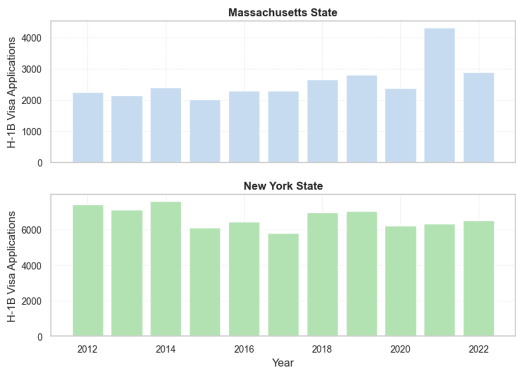
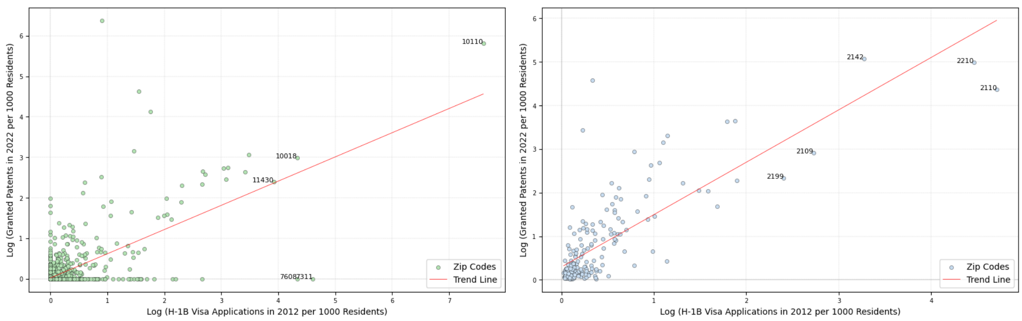


Figure A.5: H-1B visa applications in Massachusetts and New York, 2012-2022, author's elaboration



(a) NY State (b) MA State

Figure A.6: Relationship between log-transformed variables of the H-1B visa applications in 2012 and granted patents in 2022 per 1000 residents, author's elaboration

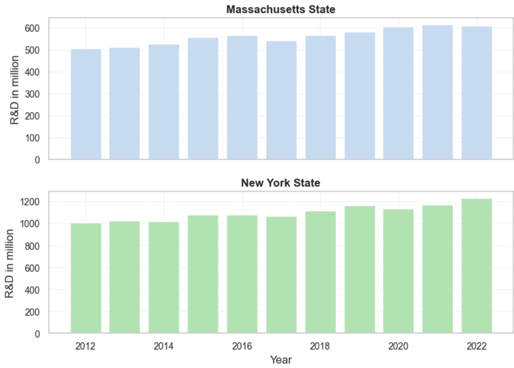
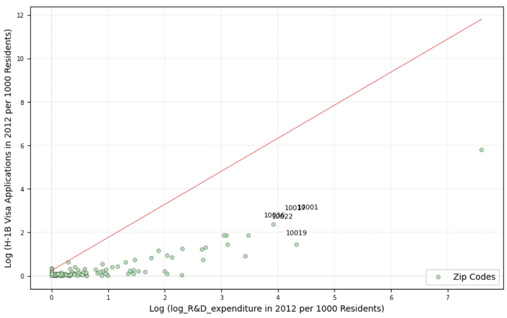
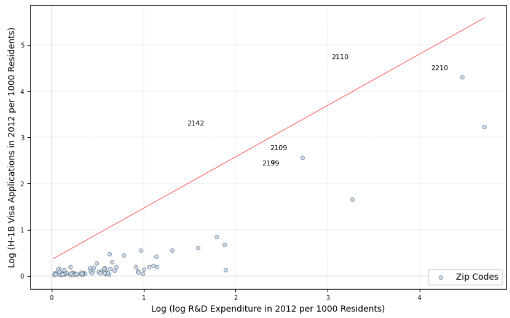


Figure A.7: R&D expenditures in million USD in Massachusetts and New York, 2012-2022, author's elaboration

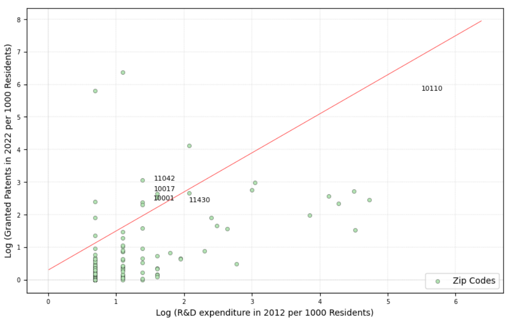


(a) NY State

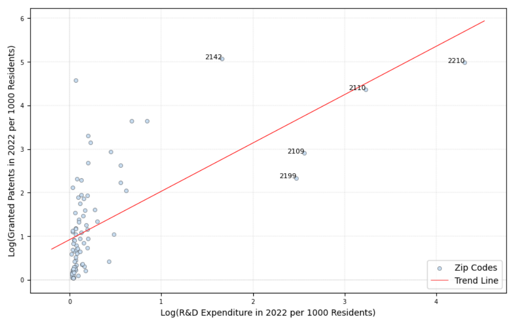


(b) MA State

Figure A.8: Relationship between log-transformed variables of the H-1B visa applications in 2012 and firm-level R&D expenditure in 2012 per 1000 residents, author's elaboration



(a) NY State



(b) MA State

Figure A.9: Relationship between log-transformed variables of the granted patents in 2022 and R&D expenditure in 2012 per 1000 residents, author's elaboration

Table A.1: Descriptive Statistics for the Selected Variables for New York State

Group	Variable	N	Median	Mean	SD	Year Collected	Source
Social	H-1B visa applications per 1000 residents	1678	0.0	1.671	48.982	2012	OFLC
Social	25 to 34 years pct	1856	0.002	0.023	0.517	2012	Census Bureau
Social	Graduate pct	1856	0.099	0.128	0.101	2012	Census Bureau
Social	Median age	1854	41.6	41.747	8.014	2012	Census Bureau
Economic	Median income	1845	28390.0	31180.6	12361.9	2012	Census Bureau
Economic	R&D expenditure per 1000 residents	159	0.078	2.525	26.427	2012	Warton Data Portal
Economic	Business registrations per 1000 residents	1847	36.982	106.277	1877.635	2012	Warton Data Portal
Infrastructure	Occupied housing units pct	1856	1444.0	4295.9	6452.7	2012	Census Bureau
Infrastructure	Cafes per 1000 residents	658	0.176	2.58	51.996	Approx. 2012	Open-StreeetMap
Mobility & Environmental	Park Land (Acres) per 1000 residents	1331	7.774	155.055	1004.22	Approx. 2012	Open-StreeetMap
Mobility & Environmental	Square Land (Acres) per 1000 residents	47	0.008	0.056	0.144	Approx. 2012	Open-StreeetMap
Mobility & Environmental	Car truck van to work pct	1844	0.886	0.807	0.214	2012	Census Bureau
Mobility & Environmental	Walk bike to work pct	1844	0.025	0.048	0.083	2012	Census Bureau
Mobility & Environmental	Worked from home pct	1844	0.036	0.046	0.051	2012	Census Bureau
Mobility & Environmental	Bus stops per 1000 residents	579	0.584	2.817	31.787	Approx. 2012	Open-StreeetMap
Innovation Outcomes	Granted Patents (2022) per 1000 residents	1856	0.0	0.879	15.977	2022	USPTO
Innovation Activities	Innovation spaces per 1000 residents	1856	0.0	0.024	0.319	2022	Google Maps

Table A.2: Descriptive Statistics for the Selected Variables for Massachusetts

Group	Variable	N	Median	Mean	SD	Year Collected	Source
Social	H-1B visa applications per 1000 residents	237	0.21	1.552	9.203	2012	OFLC
Social	25 to 34 years pct	495	0.149	0.18	0.123	2012	Census Bureau
Social	Graduate pct	495	42.2	42.054	8.037	2012	Census Bureau
Social	Median age	495	42.2	42.054	8.037	2012	Census Bureau
Economic	Median income	495	33892.0	34929.562	11612.469	2012	Census Bureau
Economic	R&D expenditure per 1000 residents	93	0.084	1.495	8.119	2012	Warton Data Portal
Economic	Business registrations per 1000 residents	493	42.816	63.448	127.975	2012	Warton Data Portal
Infrastructure	Occupied housing units pct	495	3390.0	4814.125	4749.455	2012	Census Bureau
Infrastructure	Cafes per 1000 residents	437	4.217	30.659	365.809	Approx. 2012	Open-StreeetMap
Mobility & Environmental	Park Land (Acres) per 1000 residents	437	4.217	30.659	365.809	Approx. 2012	Open-StreeetMap
Mobility & Environmental	Square Land (Acres) per 1000 residents	48	0.01	0.189	0.587	Approx. 2012	Open-StreeetMap
Mobility & Environmental	Car truck van to work pct	493	0.883	0.827	0.167	2012	Census Bureau
Mobility & Environmental	Walk bike to work pct	493	0.023	0.053	0.098	2012	Census Bureau
Mobility & Environmental	Worked from home pct	493	0.046	0.055	0.053	2012	Census Bureau
Mobility & Environmental	Bus stops per 1000 residents	283	1.337	2.387	3.09	Approx. 2012	Open-StreeetMap
Innovation Outcomes	Granted Patents (2022) per 1000 residents	234	0.331	3.97	16.612	2022	USPTO
Innovation Activities	Innovation spaces per 1000 residents	79	0.069	0.194	0.465	2022	Google Maps