NBER WORKING PAPER SERIES

NEW DATA AND INSIGHTS IN REGIONAL AND URBAN ECONOMICS

Ran Abramitzky
Leah Platt Boustan
Adam Storeygard

Working Paper 33561
http://www.nber.org/papers/w33561

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2025

We thank Ethan Bergmann, Alison Campion, Frankie Fan, and Noah Simon for excellent research assistance, and Victor Couture, Stephan Heblich, Jackelyn Hwang, Isabela Manelici, Arianna Salazar Miranda, editors Dave Donaldson and Steve Redding, and participants at the Princeton Handbook conference for comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w33561

New Data and Insights in Regional and Urban Economics
Ran Abramitzky, Leah Platt Boustan, and Adam Storeygard
NBER Working Paper No. 33561
March 2025
JEL No. R0

ABSTRACT

This chapter surveys new data sources employed in urban and regional economics in the past decade and the insights they have enabled. We first provide a primer on the data sources, including advantages, disadvantages and use cases. Historical data sources include linked census records as well as digitized maps and directories. Contemporary data come from satellites, mobile phones, social media and wikis, posted prices and listings, e-commerce and payment card transactions, newly available administrative sources, and text, among others. We then discuss the advances these data have enabled in substantive areas throughout urban and regional economics, with historical and contemporary examples in each area. We conclude with some predictions and warnings

Ran Abramitzky
Department of Economics
Stanford University
579 Jane Stanford Way
Stanford, CA 94305
and NBER
ranabr@stanford.edu

Adam Storeygard
Department of Economics
Tufts University
177 College Avenue
Medford, MA 02115
and NBER
Adam.storeygard@tufts.edu

Leah Platt Boustan
Princeton University
Industrial Relations Section
Louis A. Simpson International Bldg.
Princeton, NJ 08544
and NBER
lboustan@princeton.edu

New Data and Insights in Regional and Urban Economics

Ran Abramitzky, Leah Boustan and Adam Storeygard[1]

Prepared for the *Handbook of Regional and Urban Economics*, Volume 6

February 2025

Abstract
This chapter surveys new data sources employed in urban and regional economics in the past decade and the insights they have enabled. We first provide a primer on the data sources, including advantages, disadvantages and use cases. Historical data sources include linked census records as well as digitized maps and directories. Contemporary data come from satellites, mobile phones, social media and wikis, posted prices and listings, e-commerce and payment card transactions, newly available administrative sources, and text, among others. We then discuss the advances these data have enabled in substantive areas throughout urban and regional economics, with historical and contemporary examples in each area. We conclude with some predictions and warnings.

Keywords: urban, regional, new data
JEL Codes: B4, N9, R0

I. Introduction

Urban economics has been enriched by a wealth of new data sources developed over the past two decades, covering both historical economies and the present. Some of these new sources offer greater spatial resolution, permitting analysis of phenomena such as visits to individual establishments that are only revealed at a fine scale. Others offer high-frequency observations on transactions and locations in space. Some sources provide panel data, linking individuals and locations across long periods for which only cross-sectional information had been available before. Others cover whole populations, or at least samples considerably larger than those found in typical surveys, enabling a more systematic study of activities such as migration and patenting, as well as the analysis of heterogeneity across subsamples. Some sources provide information on dyads, such as friendships, parent-child relationships, meetings between workers, and transactions linking consumers and firms, that were previously unavailable or present only in small samples. New methods allow researchers to create new quantitative datasets by imposing structure on qualitative data.

This chapter reviews these innovative data sources and illustrates how their use allows economists to provide new answers to a broad range of research questions. The expansion of historical data allows us to better understand the past and to provide additional evidence for historical persistence. We can now trace individuals from childhood to adulthood through record

[1] We thank Ethan Bergmann, Alison Campion, Frankie Fan, and Noah Simon for excellent research assistance, and Victor Couture, Stephan Heblich, Jackelyn Hwang, Isabela Manelici, Arianna Salazar Miranda, editors Dave Donaldson and Steve Redding, and participants at the Princeton Handbook conference for comments.

linkage to study the long-run effects of economic shocks and environmental conditions, and use newly-digitized historical maps to pinpoint exposure to local context, events, or transportation systems. Expanding our study of the past also opens up a wider set of natural experiments that can shed light on broader economic phenomena.

Modern data sources allow for the study of trade, commuting and social interactions at a much finer spatial scale. Satellite data offer measures of urban activity and environmental conditions, which can be particularly valuable in areas where other sources of information are sparse. Credit card transactions record trade at the local level, rather than from larger aggregates, allowing us to understand regional and urban linkages. Cell phone data provide a granular look at individuals' locations throughout the day, augmenting commuting data with information on consumption trips and trip chains. Cell phone call records and data culled from social media allows us to reconstruct networks for workers and friends. New text methods promise the creation of data on government regulations, social activity, transactions and prices relevant to urban economies.

This chapter covers a lot of ground and reviews studies of the past and the present, from rich and poor countries, and across many other research boundaries. Because of this broad scope, we need to make choices about what counts as new data versus more traditional sources, and what topics fall within urban and regional economics. Here we focus on data sources that were not or were barely used in economics 10-15 years ago.

We note several recent surveys that cover some of the areas we consider here. Hanlon and Heblich (2022) summarize historical work in urban economics. Kirchberger (2021) reviews data used to measure internal migration, including mobile phones, Donaldson and Storeygard (2016) survey the uses of remote sensing in economics, and Śpiewanowski, Talavera and Vi (2022) discuss scraped data. Most similar to our chapter are reviews focusing on new data in urban economics, which include Glaeser et al. (2018) and Saiz and Salazar-Miranda (2023). Glaeser et al. (2018) is primarily prospective, mostly considering potential rather than actual uses of new data. Relative to Saiz and Salazar-Miranda (2023), this chapter expands coverage of historical data and provides substantial discussion of how new data has expanded the areas of research inquiry – and thus facilitated new discoveries – in urban economics (see Section 3).

Many of the new sources of data rely on advances in machine learning (Mullainathan and Spiess, 2017). Like in other fields in economics, urban economists have adopted machine learning methods in recent years for prediction tasks with structured datasets (Chernozhukov, et al. 2018; Athey and Imbens, 2019; see, e.g., Derenoncourt 2022; Mookerjee and Slichter, 2023). However, where machine learning has the most distinct promise for urban economics is in its use for converting unstructured data – such as text, but especially maps and images – into usable variables. Machine learning approaches for using text as data are surveyed in Hirschberg and Manning (2015) and Gentzkow, Kelly, and Taddy (2019). For broader surveys on neural networks and deep learning, see LeCun, Bengio and Hinton (2015) and specific guides for economists written by Gorin (2021) and Dell (2024). Applications of these techniques for digitizing historical documents are described in Combes, Gobillon, and Zylberberg (2022).

We organize the chapter as follows. Section 2 provides a detailed list of the new data sources, both historical and modern. These include sources like linked historical data and digitized historical maps for the past, and satellite imagery and data from mobile phones, vehicle trackers and social media for the present. In each case, we review the core papers that introduce and describe the methodology needed to convert these sources into data to study urban topics. Section 3 reviews roughly a decade of studies in urban economics that have provided new insights by making use of these data sources. We subdivide these areas into studies of individual location and migration, either across or within cities and neighborhoods; studies of transportation networks and trade; and studies of the location of firms and transactions. We emphasize throughout Section 3 how the new data used in these studies allows for new substantive conclusions that advance our understanding of urban and regional economics. We conclude in Section 4 with suggestions for future work using these new data sources and for prospective new data sources that may emerge in future years.

II. New data

A. Historical data sources

i. Historical Census linked data

The US Census is far from a "new" data source, having provided the backbone of empirical research in urban economics and other applied fields for decades. Yet advances in record linkage have allowed researchers to convert (historical) census data into large panel datasets that follow individuals over time. This longitudinal data opens up a set of new research questions on spatial topics, including the determinants of geographic mobility, the long-run effect of childhood exposure to environmental conditions or economic shocks, and the causes and consequences of neighborhood change within cities.

Complete census records, including an individual's name and detailed location information, becomes available to the public 72 years after the Census is taken; the 1950 Census was just released in 2022. This complete-count data provides the opportunity to match individuals across census waves by attributes like first and last name, birth year and place of birth. The resulting large panel datasets, coupled with detailed census geography, opens new avenues for research in areas including immigrant assimilation, intergenerational mobility, and the long-term effects of environmental conditions in childhood.

Historical census data does not contain unique individual identifiers, like social security numbers, that could be used to link individuals over time. Instead, researchers have developed various linking algorithms that use personal attributes like first and last name, birth year and place of birth, that are either fixed or unlikely to change over time (Abramitzky et al., 2021). Yet individuals can be difficult to track in historical records due to common names, transcription and enumeration errors, age misreporting, mortality, under-enumeration, and international migration.

Abramitzky, Boustan and Eriksson (2012, 2014, 2019) introduced rule-based algorithms using a limited set of personal attributes to create matches between census pairs. Samples based on this method are posted on the Census Linking Project website. Extensions of this method are introduced in Abramitzky et al. (2019), which selects the most likely match using a probabilistic approach that considers differences in name and age, and Abramitzky et al. (2025), which adds other linking attributes including county of residence and names of parents and spouse to "break ties" among otherwise similarly likely multiple possible matches. The match rates of these rule-based approaches range between 15% and 45%.

Another set of linking algorithms, first developed by Feigenbaum (2016), use hand-linked data to train a machine learning algorithm, thereby generating links that mirror the choices of human coders. Price et al. (2021) and Buckles, et al. (2023) leverage much larger training datasets, based on millions of genealogical links made on FamilySearch, a large, public, wiki-style family tree. Their samples, which are posted on the Census Tree website, are able to match 62%-65% of the population, a statistic that includes both the crowd-sourced hand links and machine learning extensions. The Minnesota Population Center disseminates similar machine-learning based samples called the Multi-generational Longitudinal Panel (MLP), described by Goeken, et al. (2011) and Helgertz et al. (2022).

Following women over time, particularly from childhood into adulthood, is particularly challenging because linking algorithms rely on names and many women change their last name upon marriage. Craig, Eriksson and Niemesh (2022) create links for women using marriage certificates from Massachusetts between 1850 and 1910, which included women's maiden names. They match husbands and wives to adult census records using their shared married name and then match women to their childhood households using their maiden names. Althoff et al. (2024) instead leverage the first generation of social security application forms in the 1930s, for which women wrote down both their maiden and married names. Hand-linked data by genealogists or family members can link women before and after marriage if they have access to additional private information on maiden names and childhood household members (Buckles et al., 2023). Abramitzky et al. (2025) extends these rule-based algorithms to add links for never-married and ever-married women who are not likely to change their last name.

Abramitzky et al. (2019), Bailey et. al (2020b), and Abramitzky et. al (2021) highlight tradeoffs associated with these methods. When deciding between algorithms, machine learning offers a higher match rate but is less representative than rules-based algorithms, particularly for smaller racial and ethnic groups (Abramitzky et al., 2025). Zimran (2019) and Bailey et. al (2020b) encourage researchers to reweight linked data on observed attributes to match the population. Given no available linking method is perfect and each has advantages and disadvantages , we recommend using multiple approaches in empirical studies for robustness.

ii. Digitized historical maps and directories

Digital spatial data in Geographic Information Systems (GIS) is indispensable for a variety of modern urban applications but, until recently, historical maps were not compatible with this tool. In recent years, economic historians and other social scientists have digitized a wide range of historical maps, including census geography, and environmental and land management maps. These efforts have opened up study of historical neighborhoods and the effects of proximity to relevant geographic features like administrative boundaries, industrial sites, religious and cultural institutions and the epicenters of natural disasters.

Earlier work in this area was in large projects digitizing and disseminating basic census unit boundaries (e.g., census tracts in the US) from paper maps (see Fitch and Ruggles 2003 for the US, St. Hilaire et al. 2007 for Canada, Southall 2011, 2012, 2014 for the UK, Kunz and Boehler 2005 for Germany and Gay 2021 for France).

More recently, scholars have been expanding historical census digital mapping both into smaller subunits within cities (e.g., enumeration districts or census blocks) and into smaller places like towns and minor civil divisions. The Urban Transition Historical GIS Project provides digital maps of enumeration districts and geocodes (i.e., provides latitudes and longitudes based on street addresses) households in 39 US cities in the 1880 census (Logan et al., 2011). Shertzer et al. (2016) generate a public-use dataset of enumeration districts for ten northern cities from 1900 to 1930. They also aggregate them into slightly larger hexagons that address boundary changes and allow for easier comparisons across years

The Census Place Project provides public crosswalks linking census records to consistently-defined place names and points (latitude-longitude coordinates) as opposed to the polygons described above, from 1790-1940 based on a new method to geocode and standardize town names (Berkes et al. 2023). Previous work could only geocode 23% of households and individuals from 1790 to 1940; this method increased the match rate to 83%.

One drawback of census data is that it is only available at decadal intervals. In the intermediate years, details on city residents are available in historical city directories but hard to use at scale. Albers and Kappner (2023) develop an algorithm that converts scans of city directory pages into a geo-referenced household-level dataset, providing researchers with a customizable tool for specific cities and years of interest. Using the 1880 Berlin directory as a case study, they demonstrate that their semi-automatic approach yields results similar to more labor-intensive manual efforts.

Beyond census data, other historical maps produced by land surveyors, insurance companies and financial institutions, and maps drawn in earlier historical eras, can provide useful information about the location of economic activity. However, these maps are often more complicated to digitize because they use complicated color schemes to denote different land uses, overlay administrative boundaries with environmental features like rivers and roads, or incorporate textual labels that are hard to extract from the underlying map.

Recently, geographers and urban economists have turned to machine learning techniques to digitize such maps; these approaches are summarized in Combes, Gabillon and Zylberberg (2022). Random Forests and other classification tools can be used to digitize zones on a map delineated with different colors or shading. Digitizing boundaries and lines, like transportation networks, instead require deep learning tools like neural networks. These approaches are now being used by research teams to digitize land-use maps in France and the UK (Heblich, et al. 2024; Gorin, et al. 2024; Combes, et al. 2025). Related techniques to extract building footprints are described in Litvine et al. (2024), which combines historical maps and modern remote sensing data. Lin et al. (2023) present a similar method to extract building-level information from Sanborn insurance maps in the United States and illustrate the method in a case study of Columbus, Ohio. These methods are promising and can be applied to other maps and regions, offering insights into long-term urban development.

Tax collection records, notarial documents and cadastral maps (property boundaries) can be used to create detailed geocoded datasets for historical settings before the establishment of comprehensive census data. Carrion et al. (2016) use tax records for the Italian principality of Taranto to map historical places and trace their fiscal transactions in the 1450s. Zaragozi et al. (2019) extract geographic information from notarized documents and map the geography of economic activity in 17th and 18th century Catalonia and Valencia. DeBats (2008) uses tax records to geolocate households in two cities in the 19th century US, and Hedefalk et al. (2015) illustrate how property records allow the creation of longitudinal data following households over time in historical Sweden. Thus far, these methods have mostly been used in relatively small areas over short time periods. With new techniques, there is promise that these case studies could be expanded to more areas or over longer periods of time.

iii. Reconstructed historical transportation and communication routes

Beyond mapping the location of households or firms, GIS is also useful for reconstructing historical transportation infrastructure via waterways, roads or railroads.

Early efforts at digitizing historical navigable waterways and rail lines in the US are summarized in Atack (2013), which describes the process of building a historical GIS transportation database. Atack's approach works backwards from validated state-level digitized maps, using shapefiles to trace the rail network back through the 19th century. Atack (2018) also gathers information on canals and navigable rivers from various sources, with the precise locations of each canal determined using US Geological Survey topographic maps, satellite imagery and the historical accounts of individual canal projects. Bogart (2024) reviews the contributions of new cliometric work on transportation networks to understanding market integration, trade, urbanization, and aggregate income.

Scholars are now digitizing the historical road and rail systems of other countries beyond the US. Perret, Gribaudi, and Barthelemy (2015) use collaborative geocoding to collect and fill in missing information on the Cassini map of roads in 18th century France. Thévenin et al. (2016) study the impact of railway accessibility on population in France between 1860 and 1930. To measure railway accessibility, the researchers use the historical boundaries of 36,000 French communes and maps of railway lines and stations from the Department of Maps and Plans. Ciccarelli and Groote (2018) construct a digital database of railway lines in Italy at the provincial level, covering the period from 1839 to 1913. The route locations are reconstructed from a list of opening dates and the names of route segments, alongside modern shapefiles. Researchers have also digitized road and rail maps in many developing countries and regions, both for the present and into the recent past, including China (Faber 2014; Baum-Snow et al. 2017), India (Donaldson 2018; Allen and Atkin 2022, Baragwanath et al. 2024), and all of sub-Saharan Africa (Jedwab and Storeygard 2022; Burgess et al. 2015).

Reconstructing historical communication networks has also been a subject of recent research. Wang (2023) compiled information on telegraph lines in the US from a series of historical sources and digitized the network's growth from 1844 to 1852. Le (2024) creates maps of the leased telegraph wires used for transmitting news items in the early- and mid-20th century. Russell and Winkler (2023) create and post an animated map of the historical telegraph system from 1844 to 1862, allowing users to visualize the system at any point in time.

B. Modern data sources

i. Imagery

Remotely sensed data are widely available, and much, especially for environmental measurements of interest, and night lights, come from satellites operated by governments who provide them for free. Some of the highest resolution data are still commercial products that researchers purchase or access through cooperative agreements with individual providers or brokers.

Broadly speaking, most work using satellite data to measure income either uses relatively coarse resolution (~1 kilometer), univariate night lights data and simple relationships between log lights and log income estimated via OLS or similar methods, or much higher resolution (0.5-40 meter), multivariate (i.e. multiple ranges of the electromagnetic spectrum, known as channels or bands) daytime imagery from which a signal is extracted via machine learning techniques.[2] This work aggregates across the channels, and often across many pixels up to larger units of analysis.

a. ***Lights***

Night lights, in the form of the Defense Meteorological Satellite Program's Operational Linescan System (DMSP-OLS) data distributed by the US National Oceanic and Atmospheric Administration (NOAA; Elvidge et al. 1997, 2001, Baugh et al. 2010), were introduced to the

[2] Following the literature, we report approximate resolutions as round numbers of linear units. In practice, some datasets that are distributed as a single grid for large areas use spherical coordinates (latitude and longitude) and thus have effective resolutions that vary by direction and location, typically with the cosine of latitude. However, this variation is typically just an artefact of resampling.

economics literature by Henderson et al. (2011, 2012) and Chen and Nordhaus (2011).[3] Henderson et al. (2012) document a panel and long-difference cross-country lights-GDP elasticity of about 0.3. Other work has found similar elasticities for subnational regions globally (e.g. Hodler and Raschky 2014).

The DMSP main series is available in digital form for 1992 to 2013. Analogous data for 1972-1991 is available only in a physical film archive and has not, to our knowledge, been systematically analyzed. Since 2013, the orbits of remaining DMSP satellites have decayed. In practice, this means that their only night time overpass is around 3 am. The resulting data are distributed in a separate product called DMSP-E. Brimble et al. (2025) show that the DMSP-E-income relationship appears to be weaker than the DMSP-income one. They argue that this could be because the time series is not long enough yet, or because of the DMSP-E's 3 am overpass time, when lights are likely to be less sensitive to changes in economic activity. Work including Zheng et al. (2023) has also argued that recent trends may be different due to more rapid changes in lighting technology to light emitting diodes (LED), which emits light in wavelengths that are not as well detected by these sensors.

Recent methodological work has made advances in two broad directions. One has relaxed the assumptions used by Henderson et al. (2012) to optimally combine aggregate night lights measures with traditional GDP data to form better estimates of true economic activity (Pinkovskiy and Sala-i-Martin 2016; Hu and Yao 2022; Civelli, Gaduh and Sadek Yousef, forthcoming). The other has developed methods to address shortcomings in the pixel level data. Specifically, Bluhm and Krause (2022) have an adjustment for topcoding and Abrahams et al. (2018) address spatial blurring, also known as overglow. The spatial blurring means that work on scales under 3-5 km is rarely advisable. Topcoding is less prevalent in poor countries, although somewhat present in the very largest cities.

While macro applications use light as an alternative to unreliable national income data, subnational applications often consider cities in poor countries (Storeygard 2016; Campante and Yanagizawa-Drott 2017; Henderson, Storeygard and Deichmann 2017; Bluhm and Krause 2022; Bluhm, Lessman and Schaudt 2023). By their nature as spatially distinct regions denser than their neighbors, cities as a unit of analysis are less affected by spatial blurring, but more affected by topcoding, at least in middle- and high-income countries.

Since 2012, night lights data are available from an alternative source, the Day-Night Band (DNB) of the Visible Infrared Imaging Radiometer Suite (VIIRS) sensor on the Suomi National Polar-Orbiting Partnership (SNPP) satellite. As noted by Elvidge et al. (2013), and emphasized for economic applications by Gibson (2021) and Gibson et al. (2021), these data are technically superior to the DMSP-OLS, with a true resolution under 1 kilometer, higher sensitivity to low lights, less noise, and effectively no spatial blurring or topcoding. However, they have a 2:00AM overpass time, and Brimble et al. (2025) show that, like DMSP-E, VIIRS shows a weaker

[3] DMSP is a series of satellites and the OLS is a sensor on them. Correspondingly, VIIRS is a sensor on the SNPP satellite. However, in practice most literature refers to DMSP and VIIRS.

relationship with GDP, in changes over time, than DMSP-OLS does for all but the smallest geographies. Again, this could be due to a shorter time series, a later overpass time, and/or the switch to LEDs. Urban applications include Gandhi et al. (2022), who investigate flooding in a global panel of cities.

Three further points are worth noting. First, night lights-income elasticities from the literature may be less relevant in cases where electricity infrastructure is rapidly expanding, or when it is directly impacted by a natural disaster. Second, because lights are a noisy measure, they are most effective for considering long term changes. This is especially important for per capita uses, as most countries, especially poor countries, have reliable population data at most once per decade. Third, combining the VIIRS and DMSP data is, to our mind, an open question. While some work has attempted this (e.g. Li et al. 2020, Chiovelli et al. 2023), it relies on the cross-sectional relationship between the two sources, not the relationship in changes over time, which is considerably different and the basis for most (though not all) work in economics.

b. Multichannel satellites

A recent literature has used higher resolution satellites with multiple channels to predict income, consumption or wealth and related outcomes like poverty rates. Jean et al. (2016) apply "transfer learning" to a cross-section of three-channel (Red-Green-Blue) 2.5-meter daytime imagery from Google Maps. Specifically, Jean et al. (2016) use a convolutional neural network (CNN) trained on generic image data to predict night lights levels from the finer Google Maps data. This relationship is then used to predict average consumption and asset ownership for clusters of 2-45 households in five African countries using satellite data for a 100 km2 area surrounding the cluster. With nearly 50 million data points to predict each outcome, such data reduction methods are critical. Yeh et al. (2020) refine this method, using coarser (30-meter), freely available Landsat data and avoiding the intermediate step of predicting nightlights, to predict asset wealth in a larger set of African countries.[4]

Unlike the night lights work discussed above, much of this work is focused on cross-sectional relationships. An important exception is Khachiyan et al. (2022), who use a CNN to predict local economic growth in urbanized areas of the United States using 30-meter Landsat data. In the absence of baseline census data, it can explain more than a third of variation in log income changes 2000 to 2010 across 2.4-kilometer cells, a level at which night lights are too coarse to be very informative. Unfortunately, the model performs poorly out of sample, with zero explanatory power for changes between 2007 and 2017. While this is a stronger test than applied elsewhere to work on changes, it is a sign that more work is needed to use these data to measure changes.

Users of this data should keep these further points in mind. First, a recent literature has shown that measurement error in certain kinds of satellite data is likely to be non-classical, and that this

[4] Other work uses satellite data to predict related measures including the Human Development Index (Sherman et al. 2023) and poverty rates (Engstrom et al. 2022). Rolf et al. (2021) develops a method for predicting a wide range of socioeconomic indicators from satellite data "off the shelf" (i.e. with easily available data and limited customization).

is important for inference (e.g., Gibson et al. 2021). Methods have now been developed for addressing this in the binary (Alix-García and Millimet 2023) and continuous (Proctor et al. 2023) cases. Second, as with machine learning more generally, algorithms with good predictive power can be differentially biased along dimensions of interest including rural versus urban (Aiken et al. 2023) areas. Third, more positively, platforms such as Google Earth Engine allow users to combine several data sources and process them in the cloud.

c. Other remote sensing technologies and images to measure other attributes

Satellite data have been used to measure other physical attributes of interest in cities. Aerosol Optical Depth (AOD), a proxy for airborne particulate matter, has been measured for 3 kilometer grid squares at a daily frequency by the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor on NASA's Terra and Aqua satellites since 2000. Some work in urban and regional economics uses relatively raw AOD data for cloud-free pixel-days (Gendron-Carrier et al. 2022, Zou 2021), while other work (e.g. Yang et al. 2024) relies on more modeled measurements produced by atmospheric scientists (van Donkelaar et al. 2015, 2021). Nitrogen dioxide (NO_2) has been measured at 13 by 25 km resolution by the Ozone Monitoring Instrument (OMI) on NASA's Aura satellite since 2005 (Ezran et al. 2023). Other recently launched satellites, NASA's Orbiting Carbon Observatory-2 (OCO-2) launched in 2014 and ESA's Sentinel 5P (2017), measure the greenhouse gases CO2 and Methane in addition to NO2 at higher resolution. Beirle et al. (2019) show that NO2 emissions can be tracked to individual power plants with the TROPOMI sensor on Sentinel 5P. South Korea's Geostationary Environment Monitoring Spectrometer launched in 2020, NASA's TEMPO satellite, launched in 2023, and ESA's Sentinel 4 instruments, scheduled to be launched in 2024, each provide data on air pollutants on an hourly frequency over a more limited geographic range, unlike previous satellites which typically measure a given location only once per day. Together, they cover much of the northern hemisphere.[5] Balboni and Shapiro (2025, this volume) argue that new remote sensing data and machine learning algorithms for processing represent key opportunities for future work in spatial environmental economics.

Satellite-enhanced measures of precipitation are valuable in regions where terrestrial weather stations are relatively sparse, such as sub-Saharan Africa (Henderson et al. 2017). Marx et al. (2019) infer time-varying roof material from satellite imagery as a proxy for house investment in a Kenyan slum, and Gechter and Tsivanidis (2023) locate and demarcate slums using satellite data.

Building footprints and heights are an active area of data development. As noted below, some European countries have centralized administrative data, but for other regions, and for developing global datasets, remote sensing is critical. Building heights for individual cities, such as Nairobi (Henderson, Regan, and Venables, 2021), often comes from lidar (Light Detection and

[5] By their nature, only geostationary satellites can, on their own, provide continuous temporal coverage of a specific location, at the cost of limited spatial coverage and low resolution relative to the same sensor on a satellite in low Earth orbit.

Ranging) sensors, typically mounted on planes, not satellites, to achieve sufficient resolution. Lidar is an “active” sensor, in the sense that it sends light at a target and measures what comes back and when, much like radar does with radio waves and sonar with sound. Satellite data has been used to extract building footprints more broadly. As the technology improves, satellite-based lidar may make building heights cheaper to measure at scale. In 2023, Microsoft released a dataset of 1.2 billion building footprints, with height estimates based on imagery alone (i.e. not lidar) for 174 million in Western Europe, the US and Australia. Heris et al. (2020) report accuracy estimates for an earlier 2018 release of US-only data. Google has also released an Open Buildings dataset of building footprints for selected regions including Africa, Latin America and South and Southeast Asia using sub-meter resolution satellite data (Sirko et al. 2021), and more recently an annual raster product of building heights, 2016-2023, for the same regions based on 10-meter Sentinel-2 data (Sirko et al. 2023). We are not aware of uses in economics yet. Esch et al. (2024) calibrate Synthetic Aperture Radar (SAR) data from Esch et al. (2022) on building heights with a more traditional dataset of tall buildings to produce estimates of building volume and height on a global 90-meter grid.

Burchfield et al. (2006) report changes in urban land use between circa 1976 and 1992 in the United States, using aerial photography for 1976 and satellite data for 1992. An ongoing project by Druckenmiller, Hsiang, Madestam, and Tompsett records land use change in 60 former British colonies between the 1940s and 1990s based on digitization of 1.6 million aerial photos. They have committed to releasing these data in the next few years. Comparable data may exist in American, Russian and French archives for territories these countries controlled in the 20th century. The U.S. National Ocean Survey took aerial photos of several areas of the US coastline between the early 1920s and the 1960s. Currently only 10% have been digitized. Following a long tradition in traffic engineering dating back to Johnson (1928), contemporary aerial photos have just begun to be used by economists to count cars, in order to estimate a congestion elasticity with respect to traffic density (Mangrum and Molnar 2020).

Various kinds of satellite data have also been used to delineate urban areas, including Harari (2020) and several in a 2021 Special Issue of the *Journal of Urban Economics* (Baragwanath et al; Ch et al.; Dijkstra et al.; Dingel et al.; Duranton and Rosenthal). As other new data sources facilitate work comparing cities across countries, such delineation methods are becoming more important.[6]

Google Streetview is a terrestrial source of imagery that has been used to study cities, and specifically physical attributes of building and streetscapes, both programmatically (Naik et al. 2017; Galdo et al. 2021) and via human coders at scale (Harari and Wong 2024), following earlier work in sociology (Hwang and Sampson 2014). Seiferling et al. (2017) extract street trees algorithmically, potentially allowing careful efforts such as Han et al. (2024) to value such trees in a single city to scale dramatically. Salazar-Miranda et al. (2023) extract people in Streetview

[6] Delineating urban areas is a large literature in remote sensing. For a recent review focusing on comparisons across multiple cities, see Chen et al. (2024).

images to study street life. Saiz et al. (2018) show that counts of geotagged pictures of buildings on Google Maps and photo site Flickr predict survey ratings of their aesthetics.

Finally, we note one use of georeferenced data with extremely high temporal resolution: video. Salazar-Miranda et al. (2024) record movements and interactions of pedestrians in four public spaces of three cities, using movies recorded by Whyte (1980) and analogous ones from c. 2010. Relative to the earlier sociological literature, this work automatically codes behavior using a combination of algorithms: a CNN to identify people, a Kalman filter-based procedure to track them across frames, and a clustering algorithm to determine who is interacting.

ii. Mobile phones

a. Call data records (CDRs) from network providers

Mobile phones provide information about the location of the people who use them, and sometimes the vehicles they drive. Broadly, there are two kinds of cell phone data. Call data records (CDRs), provided by network operators, report the location of the phone at the time a call was made or received, as triangulated from the network of cell towers. In some cases, the counterparty to the call can also be identified. Several economics papers have made use of such data from a variety of network operators in many countries: Blumenstock (2012) and Blumenstock et al. (2023) in Rwanda, Büchel and Ehrlich (2020) and Büchel et. al (2020) in Switzerland, Barwick et al. (2023) and Li et al. (2023) in China, and Kreindler and Miyauchi (2023) in Colombo and Dhaka. At times, network operators have been open to sharing such data widely. Orange provided them for Cote d'Ivoire in 2011-12 and Senegal in 2013-14 as part of a "Data for Development Challenge". But most such work appears to be the result of ad hoc individual relationships between providers and individual researchers. As privacy concerns have grown around the world, such partnerships appear to be decreasing.

b. Smartphone aggregators

For research purposes, CDRs have been mostly replaced by data from smartphones, whose apps collect more accurate GPS-based locations at all times (regardless of whether a call is placed). A potential downside is that smartphone ownership is less representative in some contexts than all mobile phone users. Milusheva, Bjorkegren and Viotti (2021) discuss this issue and the resulting potential bias. It is less clear how opting out of location tracking by specific apps affects representativeness. A location data industry that aggregates the resulting location information across apps has developed in recent years. Keegan and Ng (2021) list 47 companies that were selling these data as of 2021. Urban and spatial economics researchers have used data from Safegraph (Athey et al. 2018 in San Francisco; Atkin et al. 2022 in Silicon Valley; Allcott, et al. 2020, Chen and Pope 2020, Monte, Porcher and Rossi-Hansberg, 2023, Cook 2025, and Abbiasov et al. 2024 in the US more broadly; and Blanchard et al. 2023 in Kenya, Nigeria and Tanzania), VenPath (Gupta et al. 2022a, 2022b), Replica (Cook, Currier and Glaeser 2024; Cook

2025) and Precisely PlaceIQ (Couture et al. 2022) in the US, Veraset (Almagro et al. in Chicago. 2024; Banick et al. 2021 in India), and Docomo Chizu NAVI in Japan (Miyauchi, Nakajima and Redding 2022). Other work has used analogous data from anonymous companies in the USA (Athey et al. 2021; Abraham et al. 2024) and China (Barwick et al. 2023).

While such data, like CDRs, typically only provide researchers with a very limited set of covariates, like the model and operating system of the mobile phone, researchers often impute other characteristics ecologically, based on data about home locations from the census and real estate and marketing databases. Home locations can typically be inferred from the location where the phone most often spends the night, in some cases at the level of the individual parcel (Cook 2025) or building (Couture et al. 2025).

Privacy concerns about these data have been particularly acute (Keegan and Ng 2021). Companies typically do not disclose the apps from which they draw location data. In 2022, Safegraph stopped selling its popular "Patterns" product, which was used in much of this research, after civil liberties groups raised concerns.[7] It now emphasizes products that estimate aggregate traffic over time to specific "points of interest" (POIs) such as stores and parks, without the information about travel between location pairs that is critical for some past work.[8] However, it appears that similar, perhaps exactly analogous, data are available from location data provider Advan.[9]

c. *Individual apps with individual location data*

Researchers have used location data from individual apps with which they have developed relationships. Most prominently, Uber has provided data on its trips to several groups of researchers, including some working on urban issues (Currier et al. 2023, Gonzalez-Navarro et al. 2022); other data on Uber trips have been the result of public records inquiries to local regulators (Gorback 2024) and direct queries (Rosaia 2024).[10] A benefit of Uber data is that driver and passenger locations validate each other, ensuring that they are real, discrete trips. A limitation, for some applications, is that for-hire Uber trips are unlikely to be representative of trips overall. Hausman et al. (2023) use data on trips from Waze in the US, aggregated by origin and destination ZIP code, capturing trips that are not for hire. Borker (2021) uses ratings on the safety of public locations in Delhi from SafetiPin, which crowdsources these ratings. Tang (2024) uses data on driver movements from an anonymous Chinese food delivery app.

Researchers have also designed their own apps to gather location information. Akbar et al. (2023a) use location data, collected every few seconds, divided into trips based on drivers throughout India reporting in the app the start and end of each trip they take. Kreindler (2024)

[7] https://www.vice.com/en/article/safegraph-to-close-shop-abortion-clinic-location-data/ , Accessed 20240816.
[8] Google's COVID-19 Community Mobility Reports provide related daily data for broad classes of establishments aggregated by administrative unit (e.g. counties in the US) between February 2020 and October 2022.
[9] https://www.deweydata.io/blog/advan-patterns-now-available, Accessed 20240816.
[10] Between 2017 and 2023, Uber publicly provided cell-to-cell speed estimates for small cells in several cities using data from real trips. However, the estimates combined data on actual trips between these cells with data from partial trips, making them hard to interpret.

goes further in this regard by using the app to provide drivers in Bangalore with financial incentives to avoid certain areas at certain times, a form of congestion pricing. Bjorkegren et al. (2024) design an app that analogously incentivizes bus riders to wait longer at a specific bus stop, using simple mobile phones and an on-site enumerator with a tablet.

iii. ***Derived speeds using vehicle location trackers and transit swipes***

Researchers have also used aggregate traffic speed estimates derived from vehicle locations by apps. Hanna, Kreindler and Olken (2017) in Jakarta, Akbar and Duranton (2017) in Bogota, and Akbar et al. (2023a,b) in cities throughout India and around the world all use data on the speed of individual trips queried from Google Maps using real-time data. These queried data are less likely to be reliable for extremely short trips (under 1 km), and one cannot make them representative without auxiliary data. However, importantly they do allow researchers to determine the speed of counterfactual trips that are not undertaken but in the choice set of real agents. Gu et al. (2021) work directly with Baidu Maps to generate speeds for road segments averaging 50 meters in length for cities throughout China. Alder, Song and Zhu (2023) use similar data for five Chinese cities from ridesharing firm DiDi.

Perhaps due to the highly regulated nature of the taxicab market in most cities, data on taxi trips are often readily available. While a much longer literature dating back to Camerer et al. (1997) uses data on individual taxi trips from individual taxi companies to study driver labor supply, a more recent literature uses more detailed data, obtained from city regulators on the universe of trips in a specific time period, to study more urban and spatial questions. These data typically include the distance and time of the trip in addition to the origin and destination. Work using these data includes Buchholz (2022), Gorback (2024), and Mangrum and Molnar (2020) in New York, and Almagro et al. (2024) in Chicago.

Freight vehicles are of particular interest for some questions. Alder, Song and Zhu (2023) use instantaneous locations of truck-based GPS transponders for over half a million trucks from an anonymous Chinese logistics firm, and Barnwal et al. (2024) use analogous data for India. Ducruet et al. (2024) use historical data on the set of merchant ships in each port around the world as reported by a trade publication between 1950 and 1990 to study the role of ports in city growth. A more recent technology, the Automatic Identification System, records ship locations transmitted continuously to satellites and/or ground-based receiving stations. It can be purchased on the open market and has primarily been used in the trade and industrial organization literature, though a recent contribution by Brancaccio, Kalouptsidi, and Papageorgiou (2024) has a somewhat more urban flavor.

Loop detectors, which count vehicles and record their speed anonymously using sensors embedded in the roadway, have been used since the 1960s by road agencies, but new data systems have made them more easily available in real time as well as historically for entire national highway systems, including those of England (Alder, Song and Zhu 2023) and Finland.

Electronic tolling systems that identify vehicles' locations at specific points on a highway using transponders or license plate readers are sometimes available from state and local departments of transportation and have also been used in several contexts, including New York City (Mangrum and Molnar 2020), Los Angeles (Bento et al. 2024), Minneapolis (Mattia 2023), and Seattle (Cook and Li 2023). Unlike loop detectors, depending on restrictions imposed by data providers, transponders can be used to create a panel dataset of vehicles.

Finally, nearly all urban transit systems now have electronic payment systems that can record the location and time of entries (and in some cases corresponding exits). Since pricing and practicality often incentivize re-use of cards, in principle these data can match trips across users, although privacy concerns can restrict such use. Such data have been accessed from transit authorities and used in several contexts, including London (Larcom et al. 2017), New York (Sun 2021; Sun and Mikhed 2023), Jakarta (Kreindler et al. 2023), Singapore (Lee and Tan 2024), Chicago (Almagro et al. 2024), and Lagos (Bjorkegren et al. 2024).

iv. Social media and wikis

Social media data allow researchers to link people to each other and to firms, with varying degrees of spatial precision. User profiles in social media sites including LinkedIn, Facebook, and Twitter allow users to report what city or region they live or work in, and the companies often have considerably more precise information from internet service providers and/or phone location tracking through their apps. Following Gee, Jones and Burke (2017), certain groups of economics researchers have developed relationships allowing them to perform analyses at scale with the largest social network, Facebook. In a series of papers, Bailey et al. (2018a, 2018b, 2020a) Kuchler et al. (2022) and Chetty et al. (2022a,b) have used such a spatially embedded social network from Facebook to study social capital and social connectedness across space. Bailey et al. (2018a) includes an explicit request for proposals from researchers interested in using the datasets they have compiled from raw Facebook data. Caetano and Maheshri (2019) use proprietary data from Foursquare to study gender segregation. This work has overwhelmingly been focused in rich countries and especially the US, suggesting opportunities for work elsewhere, especially in data-poor contexts.

In addition to being a listings site, Yelp is social media in the sense that users have partially public profiles and write public reviews that can be linked across reviewed establishments. Davis et al. (2019) harness this feature of the Yelp data via scraping.

OpenStreetMap (OSM), a user-contributed (wiki) map of the world, has been used by work including Currier et al. (2023), Akbar et al. (2023a, 2023b), Seidel (2023) and Brandily and Rauch (2024) to provide information about road types, as well as measures of the road network generated by OSMnx software (Boeing 2017), and OSM's routing engine (e.g. Herzog 2024).

v. E-commerce and payment card transactions

Two forms of transaction data have recently been used by researchers to study spatial questions: transactions from individual payment card companies or networks, and e-commerce transactions from individual large sellers or platforms. Data on individual credit card transactions from several sources have been used to study urban questions: JPMorganChase (Relihan 2024), Visa (Dolfen et al. 2023), Facteus (Allcott et al 2020), and an anonymous financial institution (Agarwal, Jensen and Monte 2020) in the United States, CaixaBank in Spain (Allen et al. 2020), Mastercard in Germany (Alipour et al. 2022), and UnionPay (Barwick et al. 2024) in China. Like Facebook data, these data require close collaboration with the data providers developed through individual relationships, though in some cases access appears to be somewhat more institutionalized, for example through the JPMorganChase Institute Academic Fellows program. Data can be based on a specific set of cardholders, as in Relihan (2024), or a set of points of sale, as in Allen et al. (2020), who can combine data on foreign and local cardholders using the same point-of-sale network. Fan et al. (2018) use confidential data on transactions by seller city, buyer city and product category from Alibaba in China.

vi. Posted prices/listings

Posted listings and prices in several product categories provide a wealth of information. Real estate listings and transaction data from Zillow have been scraped but also accessed through a formal program for academic research, the Zillow Transaction and Assessment Dataset (ZTRAX), now closed as of 2023 (Baum-Snow and Han 2024, Box-Couillard and Christensen 2024, and Special Issue with Introduction by Zabel, Phaneuf and Krause 2023). Bellet (2024) uses data scraped from Zillow. Researchers have scraped analogous data for Jakarta from Brickz Indonesia (Harari and Wong 2024) and 99.co (Hsiao 2023), Nairobi from Property24 (Henderson et al. 2021), New York from StreetEasy (Gupta et al. 2022b), and San Francisco from Craigslist (Pennington 2021).[11] Several scholars have studied short-term housing rental markets using AirBnB. These include Horn and Merante (2017) in Boston, Garcia-Lopez et al. (2020) in Barcelona, Edelman and Luca (2014) and Calder-Wang (2021) in New York, Koster et al. (2021) in Los Angeles, and Edelman, Luca and Svirsky (2017), Barron et al. (2021) and Farronato and Fradkin (2022) across many US cities. While most used scraped data, Farronato and Fradkin 2022 worked with the company directly. Outside of housing/accommodation markets, urban researchers have used Grubhub (Schiff, Cosman and Dai 2023) and Yelp (Davis et al. 2019) to study restaurants, and Yelp (Glaeser et al. 2017) and Google Places (Duprey et al. 2023) to study business openings and closings.

Analogous labor market listings data have also been used widely, though less for urban and spatial applications. The most common source is Lightcast (formerly Burning Glass), with which researchers have worked directly. Some urban and spatial applications include Papageorgiou (2022), Hazell et al. (2024), and Kleinman (2024).

vii. Administrative data

The use of administrative data in urban economics is far from new. However, certain newly used sources stand out in terms of their size and comprehensiveness. Most notably, Chetty et al.

[11] See also Besbris et al. (2021) who use Craigslist data for 50 cities.

(2014) and Chetty and Hendren (2018a,b) use annual US Internal Revenue Service (IRS) data on individual taxpayers and their dependents linked across generations and over time to study how intergenerational mobility varies across commuting zones, counties and neighborhoods within cities, respectively. One cannot learn about income mobility on such fine spatial scales with traditional panel datasets like the Panel Survey of Income Dynamics (PSID), because they are multiple orders of magnitude smaller. Wealth registers have been used to study housing markets in Denmark (Gruber et al. 2021, Andersen et al. 2022), and residential register data to study urban neighborhood microstructure in Norway (Krause and Seidel 2024).

The US Census Bureau has recently distributed two new datasets providing fine-grained detail for the whole country. The Master Address File (Sullivan and Genadek 2024) has been used to understand migration, and particularly residential vacancy chains (Mast 2023; French and Gilbert 2024). See also Bratu, Harjunen, and Saarimaa (2023) for Finland. Business Formation Statistics provide evidence on business starts that could be used to study entrepreneurship hyperlocally, in principle at the level of the Census block (Dinlersoz et al. 2023).

Several countries have begun to release data on individual transactions or aggregated supplier relationships between firms, collected as part of Value added tax (VAT) administration. Researchers have mined the spatial information embedded in these records to learn about spatial production networks in several countries including Chile (Arkolakis et al. 2023), China (Egger et al. 2024), India (Panigrahi 2022), Kenya (Wiedemann et al. 2024), and Turkey (Cosar et al. 2022 Demir et al. 2024).

National-level cadastres, with spatial data on building footprints and heights, among other attributes, are available for some European countries, include Spain (Arribas-Bel et al. 2021) and France (de Bellefon et al. 2021) and have been used to delineate urban areas. New administrative data from city governments include requests for services through 311 systems, which have been used in sociology (e.g. Levine and Gershenson 2014).

viii. Text as data

The use of text as data has been well-surveyed elsewhere (e.g. Gentzkow, Kelly and Taddy 2019). We note here three important text corpora that have recently been used to study urban issues, with clear opportunities for wider use. These applications include both historical and modern settings.

Posted prices and property transactions are embedded in newspaper text and in county records. The Historical Housing Prices project (Lyons et al. 2024) collected purchase and rental prices from 30 American cities from 1890 to 2006 using newspaper real estate listings. Shen and Ross (2021) use natural language processing to analyze attributes of housing units encoded in text data from the Multiple Listing Service in metropolitan Atlanta. These projects, already impressive in scope, could potentially be expanded to other cities and time periods. Property deeds have been recorded and stored in paper form in county records offices. Some of these documents have recently been scanned and posted online. Sood and Ehrman-Solberg (2023) investigate racially restrictive covenants extracted from deeds in Minneapolis. Work digitizing such information is underway in twelve other cities throughout the United States according to the National Covenants Research Coalition, and other attributes of a sale (beyond restrictive covenants) could presumably also be recovered.[12]

New techniques for natural language processing are being applied to recover local municipal regulations, including land use and zoning codes, that – to date – have been impossible to aggregate and harmonize. Bartik, Gupta and Milo (2024) use a large language model (LLM) to extract a set of attributes of local zoning codes such as setbacks and minimum lot sizes from the text of these codes for over four thousand US local governments. Related work is underway in adjacent disciplines (Axelrod, Lo and Bronin, 2023; Mleczko and Desmond 2023, Lawrimore, et al. 2024). Similar methods could be useful for studying other unstructured text corpora, including other local regulations, newspaper articles about urban events, or help-wanted ads for urban labor markets (see, e.g., Fang, Stewart and Tyndall, 2023 and Mitre-Becerril and MacDonald 2024). Dell et al. (2023) published a large collection of historical American newspapers concentrated in the late 19th century and other similar collections may be available soon.

New text analysis methods have also been applied to large traditional datasets, including administrative data, for example in assigning individuals to social groups using their names (Asher et al. 2024, Gechter and Tsivanidis 2024, Box-Couillard and Christensen 2024).[13]

III. New forms of data have enabled many advances in urban economics

This section reviews recent papers that use the new data sources outlined above – both historical and modern – to generate novel insights in urban economics. We divide these papers into three literatures: household mobility and location choice, transportation networks that facilitate trade and commuting, and the location of firms, economic activity and transactions.

A. Household mobility and location choice

i. Regional location choice and migration chains

Linked historical census data enable researchers to study the migration of households and individuals across regions by providing information on place of residence at multiple points in time. Recent papers explore the determinants of location choice in the United States, as well as the implications of migration on family structures and outcomes.

Internal migration has been an important channel of upward mobility, particularly for minority subpopulations (Ward 2022). Collins and Wanamaker (2014) and Boustan (2016) follow pairs of brothers, one of whom remained in the US South and one of whom left the region, during the

[12] https://www.nationalcovenantsresearchcoalition.com/, accessed 20240830.
[13] In particular, Box-Couillard and Christensen use the Ethnicolr algorithm of Sood and Laohaprapanon (2018), while the other two papers develop their own methods.

Great Black Migration, and document that Black migrants more than doubled their earnings by leaving the South. Baran et al. (2023) show that leaving the South also benefited the children of migrants, who completed more years of schooling, although Derenoncourt (2022) outlines the limits of migration as a pathway to mobility for the second and third generation. The use of linked historical data to study internal migration and other factors that contributed to intergenerational mobility is surveyed in Abramitzky, Boustan and Matiashvili (2025).

The decision to move from one region to another is often influenced by family connections, migrant networks and local push factors. Eriksson (2020) uses trans-Atlantic linked data to trace Norwegian migrants in the US back to their home municipality, thereby documenting the strength of migrant networks to determine geographic location. Zimran (2022) compares the within-US migration experience of European immigrants to the US and the US-born. Green (2025) digitizes detailed records of men serving on US Naval ships during World War II and finds that networks established during military service encouraged migration across regions.

Exposure to local violence or natural disasters also encourages out-migration. Gabriel et al. (2023) show that Black families exposed to a lynching event have a greater probability of out-migration. Regions more affected by the San Francisco earthquake of 1906 experienced declines in population and economic activity, in part due to out-migration (Ager et al., 2020).

Compared to census linking, which can follow members of a household observed living together at a point in time, or residents of the same community, newer data sources can trace out more subtle connections identified by online social networks and phone calls. Researchers use these to study how social connections affect migration decisions. Büchel et al. (2020) demonstrate that social connections identified from call data records predict moves. People who live near few connections are more likely to move, and people tend to move to neighborhoods where they know people. Blumenstock et al. (2023) further disentangle this effect with phone call CDR data from Rwanda, showing that people are more likely to move to places with denser interconnections in their network, as opposed to places where their second and third order connections are more extensive. They interpret this pattern as being consistent with networks mattering more for social and economic support via cooperation, rather than information.

Researchers have traced out networks from other data sources. Tang (2024) uses confidential data on the behavior of drivers working for a delivery app to investigate the costs and benefits of migrants' tendency to cluster in the same destination industries and locations within cities as their peers from the same origin community. Bailey et al. (2022) study the assimilation of Syrian refugees in Germany using Facebook data, documenting not only friendship links between migrants and natives, but also migrants' use of the German language and membership in local groups. Egger et al. (2024) use Chinese VAT data to show that reduced *hukou* migration restrictions for a given county pair increases trade and capital flows within the pair as well.

Mobile phone data also allow researchers to understand human movement on time scales that have previously been largely unobserved. Commuting and travel surveys tell us about daily movements, and censuses tell us about moves over the course of years or decades. Blanchard et al. (2023) show that movements on the timescales of weeks and months also reveal important economic patterns. Within their sample of smartphone users in Nigeria, Kenya and Tanzania, which they are careful to acknowledge are not representative of national populations, approximately a day a week on average is spent at least 10 km from home, on average 35-50 km away, and concentrated in the largest cities with the greatest variety of amenities.

ii. Neighborhood formation and segregation within cities

The spatial distribution of where people live and work within cities has historically been influenced by racial and ethnic discrimination. New techniques using complete-count historical census data and historical maps have provided granular detail on residential location and patterns of segregation.

Traditional segregation measures rely on pre-established census geography, including census wards from 1890-1940 and census tracts from 1940 onward. These measures are limited to cities with available tract/ward boundaries and cannot capture segregation *within* wards or tracts (e.g., at the block level). Boustan (2012, 2013) digitizes block-level data on housing prices and racial composition for around 100 city-suburban border areas to study the demand for suburban residence from 1950-1980, during a period of Black in-migration to central cities. Logan and Parman (2017) introduce a new method of measuring historical residential segregation based on next-door neighbors listed above and below a particular household on a census manuscript page (see also Grigoryeva and Ruef, 2015). The new measure reveals that Black-white segregation rose substantially from 1880 to 1940, both in urban and rural areas. Eriksson and Ward (2019) adopt this measure to re-examine the residential segregation of immigrants in the US between 1850 and 1940.

Digitizing urban maps created by mortgage lenders and by municipalities also provide new insights into the historical process behind rising segregation in the early and mid-20th century. These maps include neighborhood credit scores assigned by the Home Owner's Loan Corporation (HOLC) in the 1930s, georeferenced data on covenants in property deeds, and local zoning regulations.

Fishback et al. (2023) combine HOLC maps newly digitized by the Mapping Inequality project with census data on individual households at the address level in 1930 and 1940. The HOLC gave letter grades (A-D) and color codes to neighborhoods by risk category. Neighborhoods with the lowest ranking (D) were colored red (hence the name 'redlining'), and their residents were much less likely to obtain mortgages. Fishback and co-authors argue that HOLC maps were *not* causally responsible for segregation, but rather reinscribed existing differences in true lending risk. Neighborhoods on the D side of C-D borders were already poorer in 1930, and Black neighborhoods in D areas were poorer than white neighborhoods (not the pattern one would expect if better-off Black neighborhoods were assigned a "D" grade for racial reasons). We discuss other work on HOLC maps below.

Racial restrictive covenants encoded in property deeds forbade the sale of certain parcels to members of racial minorities. These covenants were invalidated by the Supreme Court in 1948 but may have laid the groundwork for the built environment still in place today. Sood and Ehrman-Solberg (2023) combine detailed parcel-level data on racial covenants in Minneapolis (Hennepin County), collected by hand by the Mapping Prejudice Project, with geocoded modern records on sales transactions. Comparing units built before and after the 1948 court ruling, they find that formerly covenanted units were still 4-5% more expensive in 2020, suggesting that covenants were effective in persistently changing the composition of local neighborhoods.

Non-white households continue to be disadvantaged in the housing market. Building on smaller studies of one or a few cities, Box-Couillard and Christensen (2024) study racial price differentials using data from ZTRAX on 40 million repeat sales in 34 states over a 21-year period, allowing them to test for heterogeneity in several dimensions.They find that non-white buyers pay higher prices when buying from a seller outside their own racial group, and that this premium is larger in neighborhoods with more non-white residents, and, for black buyers, in more segregated neighborhoods..

Zoning laws create designated areas for single- and multi-family housing. While creation of a nation-wide dataset on local zoning ordinances is underway with new text methods (Bartik, Gupta and Milo, 2024), current work provides a proof of concept from case studies of specific cities or states (see Shanks, 2021). Shertzer et al. (2022) combine results from digitized zoning maps in Chicago and Seattle to demonstrate that permissive zoning rules (e.g., allowance for commercial and high-density housing) were targeted toward Black and immigrant neighborhoods (for earlier work, see also McMillen and McDonald 2002 and Shertzer et al. 2016a on Chicago and Twinam 2018 on Seattle). Expanding similar zoning databases to other cities and into suburban areas is a top priority for future research.

Local disamenities, including pollutants associated with industrial activity and water sources that contributed to the spread of communicable disease, influenced neighborhood quality in the past with persistent effects to today. Heblich et al. (2021) use geolocated industrial chimneys from Ordnance Survey maps in late 19th century England and model the dispersion of pollution from these locations based on wind direction. Combining this data with geolocated census data from 1881, they find that rising pollution triggered persistent residential sorting. Ambrus, Field and Gonzalez (2020) map local water pumps that contributed to major cholera outbreaks in mid-19th century London, along with the total number of deaths by house from the 1855 Cholera Inquiry Committee map. They likewise find long-run effects on neighborhood quality a century later. Persistence can depend on the presence or absence of commemoration that keeps memory of historical events alive. Bauer and Hnilo (2024) use newly-collected data on the placement of Stolpersteine, plaques that commemorate the residences of victims of Nazi terror in Berlin, on requests for on-line privacy today. Living near a Stolpersteine may increase the salience of the threat of authoritarianism, thereby increasing the demand for privacy.

Municipalities made investments to clean up the air and water pollution that emerge as a byproduct of manufacturing activity and population density. Coury et al. (2024) estimate the impact of piped water and sewers on residential land values in Chicago during the late 19th century. Leveraging newly collected data from newspaper land parcel transaction records and detailed maps of the piped water and sewer networks, they find that water and sewer access more than doubled property values in Chicago, with benefits surpassing costs by a factor of 60.

Certain local public goods that vary widely across neighborhoods can be hard to study even in the modern context because of their highly distributed nature. Currier et al. (2023) use Uber data to study road smoothness, which can vary dramatically across short distances, and is the target of much local infrastructure spending. Their data on driving behavior, and specifically avoidance of rough roads, allows them to estimate a cost of roughness, in addition to documenting its fine-grained features, including higher prevalence in poor neighborhoods within towns and cities. Asher et al. (2024) consider differential access to local public goods between Hindus and Muslims in India, which they show is only apparent when comparing neighborhoods within cities with highly geographically-precise data. Seidel (2023) shows that, on a global scale, more ethnically diverse regions report fewer local amenities in crowdsourced OSM data. Borker (2021) uses detailed crowdsourced measures of safety in public places that are location- and travel mode-specific. She finds that women face a stark tradeoff between safety in transit and university quality that is not (or barely) faced by men.

New data give us insight into more subtle forms of segregation beyond where people live, in where they socialize and consume.[14] Using Yelp reviews, Davis et al. (2019) show that restaurants are less segregated than residences, in a discrete choice consumption model. Using smartphone locations, a more reduced form method and indirect measure of race, Athey et al. (2021) document "experienced segregation" of racial groups across a much broader set of destination establishments, showing that it is also substantially lower than residential segregation, but correlated with residential segregation across cities.[15] Cook (2025) uses Replica smartphone data to consider consumption segregation across income groups in a discrete choice framework. He finds that while preferences for specific establishments differ across income groups, preferences for neighborhoods as consumption venues vary considerably less. He applies his framework to model the effects of gentrification, finding that amenities tailored to high-income residents have a far smaller negative effect on poorer incumbents than displacement from the neighborhood due to rising rents. Couture et al. (2025) also rely on a combination of smartphone locations and reviews, estimating preferences for interacting with people from different race and income groups in a discrete choice framework by measuring trips to different establishments within chains that have different customer mixes, controlling for distance to them.

[14] For a review of the longstanding sociology literature on urban activity space, including the use of new data, see Cagney et al. (2020).
[15] See also Nilforoshan et al. (2023) who look at segregation by income, as inferred from implied rents of home locations, and show that it is higher in larger US cities

They find that people in all racial groups prefer high income co-patrons, and that racial homophily is similar at all income levels.[16]

Abbiasov et al. (2024) studies the idea of the 15-minute city, promoted in some policy and planning circles. In a 15-minute city, people can access most local services within a 15-minute walk from home. Using Safegraph data, Abbiasov et al. show that the median US resident makes only 14 percent of consumption trips within this neighborhood. They further show that low-income neighborhoods with more local consumption tend to experience more income-segregated consumption than low-income neighborhoods with less local consumption. This positive relationship between local and income-segregated consumption is reversed for richer neighborhoods.

Chen and Pope (2020) document variation in physical mobility using Safegraph data. They find that smartphone users in the lowest-income quartile block groups within a commuting zone travel 12% less distance and to 13% fewer unique locations than users in the highest-income quartile block groups Cook et al. (2024) use Replica smartphone data to consider the experienced isolation of teens in school, who visit fewer locations and establishments than adults, especially in low income families. They show that conditional on family income, teens from richer and more suburban neighborhoods are more mobile, in the sense of traveling further, visiting more places and spending more time away from home/school/work.

iii. The effect of neighborhoods on individual outcomes

Living in a segregated neighborhood may influence economic outcomes for adults in the labor market and for children acquiring human capital. Recent work with historical linked datasets, digitized historical maps, or both has made progress on this question.

Recent papers examine the long-run effects of segregation induced by the HOLC redlining maps (described above) on Black-white inequality.[17] Aaronson et al. (2021a) use a block-level regression discontinuity across zone boundaries to assess the causal effect on a neighborhood of being assigned a lower grade by the HOLC. They find that receiving a lower neighborhood grade is associated with long-term reduction in home ownership rates, home valuations, and racial segregation. Faber (2020) documents a similar pattern in comparing cities of similar sizes that were or were not appraised by the HOLC. Salazar-Miranda et al. (2024) extend this analysis to the study of environmental outcomes, such as tree cover and drainage, and the associated effects on climate risk. Aaronson et al. (2021b) and Aaronson et al. (2023) extend these methods to study children who grew up in low-rated neighborhoods; these children were likelier to live in a high-poverty neighborhood as an adult, enjoyed less upward income mobility, and had less education and lower credit scores.

---

[16] See also Caetano and Maheshri (2019), who document gender segregation among users of the Foursquare social media app.

[17] However, note the critique of Fishback et al. (2023) described above that HOLC maps themselves may not have had a causal effect on lending behavior.

Segregation also contributes to differential exposure to local pollutants. Banzhaf, Mathews and Walsh (2024) document the evolution of racial disparities in exposure to air pollution in Pittsburgh using newly-digitized maps of "sootfall" in discrete locations combined with complete-count Census data on residential locations. Race was highly predictive of exposure to air pollution as early as 1910 and grew in importance by 1940, beyond controls for socio-economic status.

Geography matters for the outcomes of immigrants and their children too. Historically, immigrants to the US were more likely to settle in counties that exhibited high rates of upward mobility, generating a mobility advantage for the children of immigrants over the children of US-born parents who were raised in similar households (Abramitzky, et al. 2021). Abramitzky, Boustan and co-authors explore how living in an immigrant enclave affected upward mobility in two distinct settings: Jewish households that participated in the Industrial Removal Office dispersal program from New York City, and Polish households in neighborhoods with a newly-constructed Polish Catholic church. In both cases, using linked Census data, they find that children raised in enclave neighborhoods work in lower-wage, manual occupations, although enclaves provide other cultural benefits (Abramitzky, Boustan and Connor 2024; Abramitzky, Boustan and Giuntella 2024).

Neighborhoods remain important for individual outcomes in the 21st century. A series of papers by Chetty, Hendren and co-authors has used IRS income data linked across generations to study how neighborhoods affect intergenerational mobility. Chetty, Hendren and Katz (2016) evaluate the long term effects of the Moving To Opportunity experiment, which provided public housing residents in five US cities vouchers for moving to a low-poverty neighborhood in the mid-1990s. While shorter term evaluations found mixed results on adolescents, Chetty, Hendren and Katz (2016) show that children who were young at the time of the move saw large gains in their incomes as young adults. The IRS data, linked to children's social security numbers collected at the time of the experiment, is what allowed them to link the young adults' incomes to their experimental group. Tracking down participants up to 18 years after the experiment would have been essentially impossible otherwise.

Chetty et al. (2014), Chetty and Hendren (2018a,b) and Chetty et al. (forthcoming) show related patterns using the same IRS data on a national scale. Chetty et al. (2014) calculate intergenerational earnings elasticities, in percentile ranks, for each of the 741 commuting zones in the US. They show that high income mobility is correlated with low inequality, low segregation, good primary schools, high social capital and family stability.

Chetty and Hendren (2018a,b) examine the causal effect of childhood residence location on later life outcomes, including income and college attendance, available directly from the IRS or indirectly from the Social Security Administration. In the absence of experimental variation, they rely primarily on a mover's design. They show that spending more time in better neighborhoods, defined as places where children of permanent residents are more likely to move up in the national income distribution, causes those who move into them to earn more as adults. Given the richness of their data, they even include family fixed effects, to exclusively compare younger siblings, who spend more time in the new location (after a move) with older siblings, who spend more time in the old location (before the move).

Chetty et al. (forthcoming) reports adult outcomes, including income and incarceration, by tract of individuals' childhood residence, race, gender, and parental income. Examining a wider set of outcomes requires linking the IRS data to American Community Survey (ACS) and decennial Census data at the individual level. They find large variation in outcomes across childhood locations conditional on parent income, most of which is within-county, and best explained (statistically) by the mean attributes of adults in the children's neighborhood: their employment, income, and marital status. Identification strategies including a mover design imply that 60% of this variation is the causal effect of neighborhoods.

iv. The effects of local economic shocks on migration and population

Linked historical census data has encouraged exploration of the effects of local economic shocks both on internal migration and on children who were exposed to these events early in life. Shocks include negative events like natural disasters, crop failures or economic downturns, and positive events like investments in public health or the provision of welfare and relief funds.

The Dust Bowl, a set of severe dust storms in the Great Plains region in the 1930s, was one of the largest environmental disasters in US history. Hornbeck (2012) pioneered the digitization of historical maps to study the effect of Dust Bowl erosion. Highly-eroded counties suffered sharp declines in agricultural land values, revenue, and population. With access to linked census data, Arthi (2018) follows children exposed to this event, Long and Siu (2018) compare rates of in- and out-migration, and Hornbeck (2023) follows out-migrants. Children exposed to the Dust Bowl at young ages completed less schooling later in life. Migrants leaving the most eroded areas were negatively selected and earned less in their new destinations than migrants from other areas.

Population also responds to shocks to agricultural production due to weather or pests, and to trade shocks that affect the agricultural sector. Baker et al. (2020) studies the boll weevil's devastation of cotton production in the US South in the early 20th century. Linking children to their adult outcomes in the 1940 census, they find that children who grew up during a weevil infestation had higher educational attainment as adults because they spent critical years in school rather than working in the cotton fields. Heblich, Redding and Zylberberg (2024) documents that the import of New World grain into Britain following the repeal of the Corn Laws affected population in areas suitable to growing cereals; linked data shows that these residents were more likely to move across areas or switch occupations.

Childhood exposure to economic downturns, relief programs, and public health programs also affect the trajectory of outcomes later in life. Duque and Schmitz (2023) document that growing up in areas harder hit by the Great Depression reduces later-life employment, health, and economic well-being. In related work, Modrek et al. (2022) show that childhood exposure to New Deal emergency work-relief activity during the Depression moderates these “scarring” effects. Aizer et al. (2024) find that young men who participated in the Civilian Conservation Corps, a job training program during the New Deal, improved longevity and lifetime earnings. Beach et al. (2016) find that children who grew up after a city invested in clean water technology had higher educational attainment and earnings later in life.

Current scholarship suggests that workers are unlikely to move in response to recent economic shocks, giving rise to a geographic concentration of disadvantage (e.g., Autor, Dorn and Hanson, 2013). Linked census data allows scholars to assess whether this pattern was also true in the past or has changed over time. Choi (2023) finds that, like today, workers were unlikely to leave New England towns experiencing early deindustrialization of textile mills in the 1920s. Yet Abramitzky et al. (2023) show that internal migrants *did* flow to cities that experienced falling in-migration from Europe after the 1924 border closure, and Boone and Wilse-Samson (2023) documents that workers left cities for subsistence farming in rural areas during the Great Depression.

In work on more recent periods, researchers have used new data sources to investigate how population responds to the shocks of climate change and war. Henderson et al. (2017) document the effects of changes in aridity on population over the course of the late 20th and early 21st century. Because they can isolate individual cities and small regions with satellite data, they can show important heterogeneity in the relationship: aridity only drives urbanization in cities likely to have export industries, not those most likely to be dependent on demand from the local agricultural sector. Satellite data also allows for precise and systematic measures of what areas are exposed to floods. Patel (2024) exploits this to relate a systematic measure of daily flood exposure in Bangladesh to several outcomes including the destruction of individual buildings and outmigration. Kocornik-Mina et al. (2020) relate exposure to larger floods to city level outcomes in a global panel of cities, finding that, although low-elevation areas within cities flood more often, they recover no slower than areas uphill. Gandhi et al. (2022) focus more on adaptation, showing that more flood-prone cities grow more slowly, and that richer cities see fewer deaths and less loss of night lights. Hsiao (2023) develops and estimates a spatial model of sea level rise in an urban environment, using data from Jakarta. He shows that ex post governmental protection of threatened properties creates moral hazard, keeping people in coastal areas longer than they otherwise would remain. Lin et al. (2024) document that recent urban land development on the US East and Gulf coasts has been disproportionately concentrated in places at high risk of sea level rise, and highlight the tradeoff between risk and accessibility in a monocentric model of a growing city. Balboni and Shapiro (2025, this volume) further discuss natural disasters and climate change in cities.

Bombing campaigns can reshape the population of targeted areas. Riaño and Valencia Caicedo (2024) find substantial effects of bombing by the United States between 1964 and 1973 on population concentration two to five decades later in Laos, and argue that the fine resolution of their outcome data (night lights) is a likely source of difference from related work on Vietnam

(Miguel and Roland 2011). Redding and Sturm (2024) trace out the long-run effects of the bombing of London during World War II using newly-digitized maps of bomb damage, and find persistent declines in housing prices and shifts in local population composition, both in targeted neighborhoods and in areas nearby (see also Dericks and Koster (2021), which instead uses a mapping of where bombs landed). Chiovelli, Michalopoulos and Papaiaonnou (2024) document sizable local effects of landmine clearance on night lights in Mozambique, a novel shock in a data-scarce and poorer context.

B. Transportation and communication networks

i. Between cities/region: Railroads, highways, waterways

The expansion of the railroad network is widely regarded as a key driver of industrialization in the 19th-century United States and elsewhere. The impact of railroad extends beyond basic transportation of freight to urbanization, economic growth, and economic development. The digitization of historical railroad networks has ushered in a decade of work in economic history, trade and urban economics. Atack et al. (2010) were the first to use this new data to examine how local access to the railroad affected economic outcomes, finding modest effects on population density and larger effects on urbanization in the American Midwest. Bogart et al. (2022) document that the expansion of railroad connections in the UK in the mid-19th century accelerated the transition out of agriculture and the concentration of population in urban areas. Many of these rail lines were then dismantled by a series of mid-20th century route closures known colloquially as “the Beeching Axe.” Gibbons, Heblich and Pinchbeck (2024) show that these closures resulted in population loss and a decline in skilled workers in affected areas that were partially offset by expansions in the road network.

Donaldson and Hornbeck (2016) add insights to the study of railroad infrastructure from the trade literature, operationalizing a concept of local market access. Using digitized data on railroads and waterways, they calculate the least-cost county-to-county freight routes by decade. A county’s market access improves when it becomes cheaper to trade with another county, especially if that other county has a larger population and a higher trade cost with other countries. They estimate that the expansion of the railroad network generated substantial increases in agricultural activity, as proxied by agricultural land values. Hornbeck and Rotemberg (2024) collect new data from the Census of Manufactures and document railroad-induced gains in manufacturing productivity. Donaldson (2018) documents similar improvements in economic activity using newly-digitized railroad data for colonial India.

Waterways as a means of trade expanded with new infrastructure like locks, bridges, and flood control. Bleakley and Lin (2012) study portage sites at river rapids along the East Coast. Historically, traders would bring their boats ashore and walk with them across these portage site. Using population and night light data, Bleakley and Lin show that cities and towns sprung up at these locations and remain there today even after the locational advantage disappeared over time. Tompsett (2024) leverages variation in the timing and location of bridge construction over major rivers using data from the National Bridge Inventory. Improved connectivity increased economic activity over large spatial scales, but had some negative externalities in local areas as residents were able to buy products from farther away.

New shipping technology also helped leverage global trade by waterway. Pascali (2017) documents how shipping times declined with steamship technology, using data on bilateral trade for 16,000 country pairs. Shipping times fell the most for routes that did not have favorable wind patterns. Heblich, Redding and Voth (2023) use data on 36,000 slave voyages with detailed information on ownership, dates, and ports of origin and destination to assess the effect of slave wealth on British industrialization. Ducruet, Cuyala and El Hosni (2018) compile 120 years of data from the Lloyd's List on vessel movements between ports and trace how the association between city size and trade volume has changed over time. Ducruet et al. (2024) show that containerization increased shipping flows to cities with ports well-suited to the associated larger ships, but less so for cities with geographic constraints on development, and that population was not affected, suggesting that market access benefits were offset by local congestion costs. Counterfactuals estimated in their model of port development and trade quantify the benefit of containerization on global welfare.

Moving (substantially) further back in time, Barjamovic et al. (2019) estimate a structural gravity model of long-distance trade in the Bronze Age and examine the role of geographical features and trade routes in the growth of cities. By transcribing and digitizing ancient commercial records of Assyrian merchants from the 19th century BCE, the authors extract a subset of records that explicitly mention trade between cities. The data provide a remarkable level of detail on city construction and trade practices in an ancient context. Using their model, the authors are able to locate lost ancient cities, which at times align with findings from historians who used different methods. Bakker, et al. (2021) extend the interest in ancient history into the Iron Age, studying the value of coastal access to trading partners on economic activity, as proxied by archeological sites circa 750 BCE. Using newly-digitized maps and least cost paths, Flückiger et al. (2022) demonstrate that Roman transport networks throughout Europe continue to influence economic integration between regions today, while earlier pre-Roman trade does not. Michaels and Rauch (2018) add that persistence of the Roman network traps some cities in sub-optimal locations today. Hornbeck, Michaels and Rauch (2024) study the "agglomeration shadows" around some ancient port cities, the presence of which discouraged other nearby population centers from forming.

New data have facilitated much recent work on transportation in poor and middle-income countries. Storeygard (2016) and Jedwab and Storeygard (2022) document the importance of reduced transport costs on the growth of African cities. In Storeygard (2016), higher transport costs, as proxied by the interaction between distance and fuel costs, reduced night lights in the short run. Using newly digitized maps, Jedwab and Storeygard (2022) show that increases in market access due to road building over the course of the late 20th century increases the population of cities. Burgess et al. (2015) further document the role of politics in the construction of the road system over this period in Kenya.

Alder (2016) and Baragwanath et al. (2024) study Indian highway construction in the late 1990s and 2000s, using different trade models. They both find, using night lights, that the Golden Quadrilateral increased Indian GDP. Baragwanath et al. (2024) show that in contrast, other highways built in the same period had a small negative net impact after accounting for construction costs. They also show that gains are highly concentrated, such that even district-level data masks a great deal of heterogeneity. Besides building roads, in 2017 India ended the ability of individual states to charge taxes on imports from other states. Barnwal et al. (2024) document that this increase in market access increased household expenditures. Cosar et al. (2021) uses VAT data to show that road upgrades in Turkey increased trade, manufacturing employment, and wages. Faber (2014) finds more negative effects of China's more extensive National Trunk highways on peripheral counties in China.

Using digitized maps, Allen and Atkin (2022) consider Indian road building over a longer period covering most of the late 20th century. They document that new roads lowered trade costs and induced farmers to specialize more and to grow less risky crops. Carnap et al. (2024) use satellite data to show how market access drives related producer choices in modern Mozambique.

Gertler et al. (2024) document the effect of road maintenance on a variety of local outcomes, including increases in local manufacturing jobs and incomes. Although the type of road roughness data that they use has been used by engineers since the 1980s, we are unaware of previous use in the economics literature.

Finally, Campante and Yanagizawa-Drott (2018) show the importance of air transport in the distribution of economic activity across cities globally. They again rely on night lights as an outcome because of their comparability and availability at the city level across many countries. They find that differential increases in flights under 6,000 miles, a regulatory limit, helped cities with more counterparts under this threshold as relatively long flights became more common over time.

Beyond the effects on trade and economic activity, the construction of new transportation infrastructure also affects households living near newly opened stations or interchanges. Historically, the expansion of transport networks wrought the largest changes on market access in rural areas. Chan (2024) finds that the expansion in market access due to railroad expansion in the US encouraged local households to remain in agriculture, holding back the economic advancement of their children. With linked historical census data, Chan shows that boys exposed to market access shocks had lower educational attainment and income in adulthood compared to their counterparts. Black et al. (2015) instead find that southern Black children who grew up near a railroad station were more likely to participate in the Great Black Migration to northern industrial cities. Using detailed Medicare data and death records, the authors show that, in the long run, these migrants experienced deteriorations in their health relative to children who grew up nearby (but further from a station) who remained in the South.

ii. Within-city transport

Newly digitized maps on transportation networks within cities – including commuter rail, streetcars and highways in the past, and subways and ridesharing today – have opened up a new set of questions about commuting and the location of economic activity. One of the first papers to exploit these new maps for historical cities was Heblich et al. (2020), which created new digitized maps for commuter rail, subway and tram in late 19th and early 20th century London. Alongside recovered data on bilateral commuting flows, they show that the invention of the steam railroad led to the first large-scale separation of workplace and residence, contributing to urban agglomeration.

The use of historical transportation maps to study commercial development and spatial sorting has been expanding. You (2021) combines newly digitized data on street car routes in Boston with digitized data on firm location from city directories to study the effect of urban transit on business size and location. Boston streetcars were electrified in a short seven-year period, increasing the speed at which people could move around the city. You finds a corresponding decline in small grocers, replaced by larger and more modern markets. Lee (2022) digitized the location and construction dates of the New York City subway and commuter railroads. She finds that central neighborhoods that receive subway connections lost population and ended up with residents that were poorer and more likely to be African American. New zoning laws reinforced these trends. Brooks and Lutz (2019) show that historical streetcar routes continue to influence current land use patterns in Los Angeles, primarily through the development of zoning codes that froze historical patterns in place.

Baum-Snow (2007, 2020) studies the relationship between highway construction and the suburbanization of population and employment within cities from 1950 to 1990. These studies recover historical highway construction by working backwards, matching a digital map of modern interstate highways with a database of segment construction from the Federal Highway Administration. Baum-Snow (2007) instruments for highway construction with rays included in the original 1947 construction plan and finds that each new highway passing through a central city reduces its population by 18 percent. Michaels (2008) considers the effect of these highways on long-distance trade. Morten and Oliveira (2024) apply similar methods to study the effect of highways on trade and migration in Brazil.

While transportation improvements have facilitated faster travel, urban residents still typically spend a substantial portion of their day travelling between work and home and elsewhere in the city. Traditional data on these movements in the form of representative surveys are mostly limited to rich countries.[18] Because they are expensive to collect, even these are typically repeated only at intervals of 5-10 years or more, and with sample sizes that limit the study of variation within cities. New data have allowed researchers to study travel at finer temporal and spatial scales, though often at the cost of lacking information about individual travelers. In rich countries, a typical workaround is to assign travelers attributes of their home location (e.g. census tract, block group, or even building). Kreindler and Miyauchi (2023) show that call

[18] Exceptions include 11 surveys conducted by the Japanese International Cooperation Agency (JICA) in 11 cities in poor and middle-income countries, and Bogota (Akbar and Duranton 2017).

records can be used to estimate commuting flows with sufficient precision to derive estimates of income by skill group in two middle-income countries, without training data.

Social media and vehicle tracker data is also useful for studying mobility within a city. Bailey et al. (2020a) document that social connectedness between people in New York, as defined by Facebook connections, is more sensitive to travel time between locations rather than geographic distance, and that residents in areas with greater access to transit have geographically broader social networks. Using speed estimates from Google Maps, Akbar et al. (2023a) document a high variance in urban travel speeds across Indian cities and show that it is mostly driven not by congestion, but rather by differences in uncongested speed. Akbar et al. (2023b) show that this remains true in 1200 cities in 152 countries, and that national income per capita explains most of the variation, consistent with a model where higher incomes result in lower density and more and better roads. Currier et al. (2023), noted above, use Uber data to show that poor cities and neighborhoods in the US have substantially worse roads, which increase their transport costs.

Papers using new data have analyzed the effect of ridesharing on cities.[19] Gonzalez-Navarro et al. (2022) use proprietary trip-level Uber data from areas surrounding new urban rail stations to document that Uber and urban rail are complements. This result relies on the fine-grained nature of individual trip data. Effects are largest within 100 meters of a new station and disappear beyond 300 meters. Gorback (2024) shows that ridesharing increases amenities and house prices in urban neighborhoods relatively inaccessible by transit.

Modern expansions of transport infrastructure have had a wide variety of impacts in urban areas. Gu et al. (2021) show that new subway lines in China reduce traffic congestion, while Gendron-Carrier et al. (2022) find that new subway stations globally reduce satellite-measured particulate pollution, presumably from cars. Gupta et al. (2022b) document the capitalization of the Second Avenue subway in New York into local real estate prices. In each case, spatial specificity of the data on location (or pollution and traffic) within cities is critical. Barwick et al. (2024) study the avoidance of pollution and extreme climate events that access to intercity transport infrastructure facilitates using credit card transaction data. They find that the Chinese high speed rail system indeed facilitated such avoidance behavior. Lee and Tan (2024) use farecard data from Singapore to estimate a quantitative spatial model with multiple types. They find that benefits from a new transit line in Singapore accrued almost entirely to high-income workers.

iii. Optimal transport networks and optimal use of existing networks

Granular data has allowed several groups of researchers to study optimality in transport, from the perspective of both individual travelers and city policymakers. Larcom et al. (2017) use transit card data to show that commuters who were forced to change their route due to closure of some stations during a transit strike did not revert to their old route after the strike ended, implying that their old route was suboptimal. Almagro et al. (2024) and Kreindler et al. (2023) develop models

[19] See also Rosaia (2024) discussed below.

designed to consider the optimal design of transport systems, in Chicago and Jakarta, respectively. Almagro et al. (2024) focus on pricing and the choice between public and private modes conditional on the network, accounting for congestion. Kreindler et al. (2023) emphasize the design of the public network. In both cases, it is critical that they have access to both transit data and mobile phone-based location data that measures a larger set of flows between location pairs across multiple travel modes. Buchholz (2022) and Rosaia (2024) characterize how policies could improve the efficiency of taxi and ride-hailing markets, respectively, in New York City.

iv. Congestion and its pricing

Congestion is a particular focus of the economic analysis of transportation because it is an externality, and while this does not imply that its optimal amount is zero, in principle policies intended to reduce it can increase social surplus. Several recent papers evaluate such policies. Because traffic conditions often vary substantially across both space and time, even within small windows, fine-grained data are crucial. Kreindler (2024) experimentally implements congestion pricing within the dense urban street network of Bangalore using a custom app, varying price by time and location across a sample of drivers. He finds that because delays are linear in traffic density in this context with a modest slope, short-run welfare gains from plausible congestion charges would be small even if time-varying and implemented optimally. Herzog's (2024) complementary ex post evaluation of the London Congestion Charge, combining administrative data and OSM routing in a quantitative spatial framework, is somewhat more optimistic, finding that its benefits exceed costs. Mattia (2023) and Cook and Li (2023) use evidence from existing toll lanes with variable pricing on American highways. Interpreting evidence from two different contexts through the lens of two different structural models, they obtain strikingly different results. Mattia finds that express toll lanes in Minneapolis-St. Paul induce enough additional congestion on the non-toll lanes to make everyone worse off on average, especially drivers with a lower value of time. Cook and Li (2023) find benefits from tolling in metro Seattle that accrue disproportionately to drivers with low value of time, primarily due to the option value the lanes provide for when drivers are in a particular hurry. Consistent with this, Bento et al. (2024) estimate that willingness to pay for an express lane is driven almost entirely by a discrete disutility of arriving late. Hanna, Kreindler and Olken (2017) find substantial effects of a carpooling policy in Jakarta on congestion using Google Maps travel times. Mangrum and Molnar (2020) evaluate the congestion effects of additional vehicles induced onto the streets of New York City by new technologies (rideshare) and policies (taxi expansion), and Gu et al. (2021) consider the congestion impact of 45 new subway lines in China.

Without evaluating a specific policy, similar to Kreindler (2024), Akbar and Duranton (2017) estimate a relatively small deadweight loss from congestion in Bogota, supplementing traditional data on actual trips with data on the speed of counterfactual trips from Google Maps to address endogeneity concerns. Akbar et al. (2023a, 2023b) use Google Maps trip speeds to show that variation in speeds across cities, in India and globally, respectively, are driven primarily by variation in uncongested speed, not congestion. Akbar et al. (2023b) show that much of the variation in uncongested speed, can be explained by national income, and that much of this income effect, in turn, can be explained by more major roads and lower density, consistent with a simple model where richer cities invest more in roads and spread out over larger land areas.

v. Communication/information networks

A series of recent papers recover and digitize data on historical communication and information networks.

In the early modern period, information would often flow through merchant networks that sent emissaries to ports around the world and would return with both goods and information. Koudijs (2015, 2016) collects detailed data on boats sailing from London to Amsterdam in the 18th century. Weather delays often prevented information from getting from one market to another and Koudijs finds that asset prices respond strongly to delays in trade information. Raster (2024) compiles data on more than 1 million voyages into Baltic trading ports from 1500-1860, tracing out networks of market-relevant information.

The development of the telegraph in the mid-19th century sped up the pace of information exchange. Steinwender (2018) combines data on the establishment of the trans-Atlantic telegraph with detailed information on cotton prices in New York and Liverpool to show that enhanced information led to price convergence and efficiency gains. Juhasz and Steinwender (2018) extend this work to a global setting and show that telegraph connections particularly helped foster trade in "codifiable" products like yarn that could be described in words but not in "non-codifiable" products that may require physical observation, like finished cloth. Wang (2023) builds data on the telegraph network within the US and shows that residents of areas connected with telegraph to Washington D.C. were more likely to vote in presidential elections, presumably because of the value of receiving political information.

The cost of sending letters by mail also declined over time due to the development of reliable roads and government policy designed to facilitate communication. Hanlon et al. (2022) geolocate the network of post offices and post roads in the UK in the 1830s and 1840s and study the effect of a standardization of postal prices on communication between scientists, as measured by citations of more distant patents in new patent applications.

Büchel and Ehrlich (2020) show that a more modern technology, mobile phones, are complementary to population density in promoting social interaction. Bailey et al. (2018a) construct dyadic indices of social connections between US counties based on Facebook friendship links, and document several facts about these connections, including quantifying their correlations with inverse distance and flows of trade, migration and innovation. Bailey et al. (2018b) then document that information transmitted through social networks affect housing market expectations and purchase behavior. Using Turkish VAT data, Demir et al. (2024) show that when a region receives fast internet access, its firms are more likely to become suppliers, shifting but also diversifying supply chains for buyers.

C. Location of firms, economic activity and transactions

i. Long-run determinants: Geography and institutions

A large literature studies the long-term effects of historical conditions on the contemporary location of economic activity. Earlier work in this area relied on hand-collection of historical sources that could then be matched to contemporary datasets (e.g., Dell, 2010; Voigtlander and Voth, 2012; Michalopoulos and Papaioannou, 2013). New techniques for the georeferencing of archeological and agricultural data allow questions about historical persistence to be pushed further back in time.

Suitability for agriculture can explain a large portion of the variation across space in economic activity today, particularly in today's rich countries, which urbanized early when transport costs were higher (Henderson et al. 2018). Lehner and Philippe (2025) build a dataset with time-varying measures of agricultural suitability in Europe over 500 years, which offers promising new insights for this literature.

Agricultural development affected not only population density but also social and governmental structures. Mayshar, Moav and Pascali (2022) argue that areas suitable for growing cereals rather than root vegetables later developed centralized hierarchies because of the storability of grains that made them easier to tax. Allen, Bertazzini and Heldring (2023) add that state formation emerged as a response to natural disasters that required collective action. They explore this relationship in the context of flooding and river shifts in the Tigris and Euphrates river plains. Bazzi et al. (2020) find that today's residents in areas that spent more decades exposed to the American frontier – that is, periods of low population density, in part due to low agricultural suitability – exhibit more individualism and opposition to government activity. These papers are all based on new mapping techniques.

The introduction of crops from the New World had profound effects on the path of development in Europe and elsewhere. The FAO Global Agro-Ecological Zones Database (GAEZ) has facilitated the study of this and other questions. Nunn and Qian (2011) show that Old World areas suitable for growing potatoes experienced rapid population growth and urbanization after 1700. Under some conditions, this heightened agricultural productivity gave rise to political conflict (Iyigun, Nunn and Qian 2017; Dincecco, Fenske and Menon 2024). Fiszbein (2022) combines the GAEZ with historical agricultural data for the US to show that agricultural diversity led to industrialization and long-run regional development.

ii. Agglomeration

Digitizing detailed historical maps has allowed scholars to document how concentrations of economic activity within cities have evolved over time. Research has leveraged historical events like the 1872 Boston Fire, the 1906 San Francisco Earthquake and the construction of the Berlin Wall.

The division of Berlin after the Second World War and the reunification of the city after the fall of the Iron Curtain created a series of natural experiments in commuting times between locations. Ahlfeldt et al. (2015) digitize data on land prices, population and location of work for the residents of thousands of city blocks in Berlin in three periods: 1936, 1986 and 2006. They use this data to estimate a quantitative model of internal city structure that incorporates agglomeration forces and document highly localized production and residential externalities.

Because current buildings and other durable capital pose a friction to redevelopment, disaster events can actually improve long-run economic conditions in a city. Hornbeck and Keniston (2017) use tax records to create parcel-level snapshots of the built environment in Boston before and after the Great Fire of 1872. They find that the reconstruction of burned buildings initiated a virtuous cycle, leading to upgrades in nearby buildings as well. Siodla (2015) documents a similar pattern after the 1906 San Francisco Earthquake by combining newly digitized plot-level records from Sanborn insurance maps with modern land use records from the Planning Department. Takeda and Yamagichi (2024) assemble detailed historical population data for the city of Hiroshima and chronicle a remarkable bounce-back of the central city after the atomic bomb blast of 1945. Brooks et al. (2024) provide a counterpoint to these optimistic case studies, showing that blocks destroyed during the 1968 riot in Washington DC took decades to be rebuilt. This pattern is consistent with predictions about how disasters can positively affect otherwise growing areas but negatively affect declining areas.

Agglomeration continues to be a central force in determining the concentration of economic activity across space today. Using mobile phone data, Atkin et al. (2022) unpack a key mechanism of agglomeration – the role of in-person meetings in information flows. Because mobile phones record highly precise spatial and temporal information, the authors are able to identify likely meetings between workers from different firms, and find that these meetings predict subsequent patent citations across these firm-pairs.

E-commerce has supplanted certain transactions that had been in-person, and indirectly affected others, with important implications for the spatial location of economic activity. Fan et al. (2018) use a model in which e-commerce reduces firms' fixed cost of entering a new market to show that e-commerce benefits accrue to small and remote cities in China. However, Couture et al. (2021) use a randomized controlled trial to show that *rural* access to e-commerce has little benefit for producers in China, and mostly benefits younger and wealthier consumers, in contrast to the hopes of programs supporting such access for rural development. In somewhat of a contrast to Fan et al. (2018), Dolfen et al. (2023) show that in the United States, consumers in densely populated counties benefited most from the introduction of e-commerce.[20]

iii. Land use, housing, and real estate

Posted prices and rents for housing units have long been available in newspaper advertisements and scholars are just beginning to plumb this data for academic use. Gray and Bowman (2021) collect data for Manhattan rental prices from 1880-1910 and geocode sample units. Lyons et al. (2024) expand this type of analysis to both rentals and home sales in 30 cities over a full century

[20] See also Relihan (2024) described below.

(1890-2006).[21] The authors uncover several new facts about urban home prices in the United States, including a steady increase in urban rents over the century and a sharper increase in housing sale prices than previously documented from 1965-1995.

Baum-Snow and Han (2024) leverage the national coverage of ZTRAX to estimate housing supply elasticities at the level of individual US Census tracts. Identification relies on labor demand shocks aggregated to residential locations using a model of commuting. Although policy debates about housing prices often occur at the level of metro areas, they find greater variation within metro areas than across them.

One phenomenon that is widely believed to have limited housing supply and increased housing prices in recent years is restrictive zoning. Previous systematic work has been limited in geographic or substantive scope due to the sheer complexity of such regulations. Using a large language model, Bartik et al. (2024) systematically categorize regulations from over 4,000 US municipalities. Among other facts, they document widespread restrictions against multifamily housing. They also show that while affordable housing mandates follow a monocentric pattern in cities throughout the country, minimum lot requirements only vary with distance to the city center in the Northeast. Pennington (2021) shows that exogenous increases in the market-rate housing stock in San Francisco lower prices nearby, while income-restricted (affordable) housing has no such spillover effect.

Asquith, Mast and Reed (2023) and Mast (2023) show how large new apartment buildings decrease rents in two dimensions. The relative rarity of such buildings requires comprehensive datasets covering many cities. Asquith, Mast and Reed (2023) document effects on nearby existing rental units. Mast (2023) instead focuses on the vacancy chains induced by moves into such new buildings. He finds that a large share of such chains lead to low-income neighborhoods, drawing renters out of them and reducing their market rents.

A large literature, cited above in Section 2, has studied the effects of short-term rentals on the broader housing market in many cities. While most provide reduced form estimates based on the rollout of a particular rental platform or regulations affecting it, Calder-Wang (2021) develops a structural model of the New York housing market in which Airbnb increases short-term supply and links the short- and long-term markets, deriving estimates of small average losses to renters concentrated among wealthier residents of central neighborhoods.

New data allow researchers to study social influences on the housing market. Bellet (2024) uses Zillow data to show that when their neighbors build larger homes, homeowners' feel reduced satisfaction with their own homes, causing them to take on debt to expand them. Bailey et al. (2018b) show that non-local social networks also influence the housing market. People whose Facebook friends saw house price appreciation in their (the friends') local markets are more likely to buy, to buy more (larger homes), and to pay more.

[21] Some cities have other sources of information on housing prices. For example, Nicholas and Scherbina (2013) digitized data on individual home sales for New York City from the *Real Estate Record and Builders' Guide*. However, these sources are not available for most locations.

An emerging literature on cities in low-income countries has considered informal settlements or slums, which by their nature are often difficult to study with administrative sources. Henderson et al. (2021) build a dynamic structural model of building development with two technologies, formal and slum, and a cost of converting between them. They back out this cost for Nairobi based on observed redevelopment in initially formal versus slum areas, using a rich combination of survey, lidar and scraped real estate data. Harari and Wong (2024) provide a long-term evaluation of a slum upgrading program in Jakarta using an overlapping collection of data including street photos, administrative data, scraped real estate data, and historical maps. They find that targeted sites are more likely than their neighbors to remain informal, which they attribute to the higher density and formalization costs that the program induced. Gechter and Tsivanidis (2024) evaluate the effect of plausibly exogenous high-rise construction on nearby slums in Mumbai using a combination of satellite and administrative data. They find that such construction accelerates slum redevelopment, with substantial costs to slum dwellers that could, in principle, easily be compensated from the overall surplus generated by redevelopment. Michaels et al. (2021) evaluate a "Sites and Services" intended to circumvent informal settlement by providing surveyed plots, roads and water mains in selected areas of several Tanzanian cities prior to settlement, using a combination of historical maps, satellite data and surveys. Unlike the upgraded areas studied by Harari and Wong (2024), they find that compared to nearby areas that were more likely to be settled informally, the treated areas had better-quality housing decades later.

Harari (2020) studies the shape of city boundaries in India using satellite data. She finds that cities with compact shapes have grown faster than more elongated or concave cities, consistent with compactness reducing transportation costs for city residents. Heblich, Nagy, Trew and Zylberberg (2024) infer historical city boundaries by applying machine learning methods to historical maps. Like Harari (2020), they argue that physical constraints around a city's early footprint contributed to land fragmentation and stunted city growth. Late 19th century city size, as identified via this procedure, has a limited impact on 1970s skill levels.

Finally, Grupp et al. (2024) evaluate the effect of protecting land from urban development to protect biodiversity in the European Union, using satellite data on night lights and vegetation. They find that strong protections have no impact because they are implemented in places where such development was ex ante unlikely.

iv. Labor markets

New data have facilitated work on local labor markets. Although some traditional sources document the location of market work at fixed work sites, mobile phones can track how daytime location varies on a daily or hourly basis. High-frequency information on location has become especially important with the sharp rise of remote work since the COVID-19 pandemic. Abraham et al. (2024) document the tract-level correlates of on-site work (i.e., at a fixed, non-home location) by following the devices of 4.2 million initially on-site workers, and show that variation in on-site work across tracts even within counties exceeds variation across counties. Monte, Porcher and Rossi-Hansberg (2023) develop a model with multiple equilibria in commuting modalities and show that it is consistent with the fact that most cities almost entirely returned to their pre-pandemic commuting equilibrium, but that the reduction in commuting due to work from home has been longer-lasting in the largest cities. Li et al. (2023) document that reduced commuting during the COVID-19 pandemic in China's Guangdong Province reflected unemployment, not work from home, and varied widely across cities.

Mobile phone call records have facilitated the study of referrals in local labor markets. Barwick et al. (2023) document the role of referrals in job changes. They show that job switchers increase their calls to friends who already work at their new firm shortly before their job switch. This call volume provides novel evidence about the quantity of information flow in such referrals, as well as distinguishing information flow from a preference for working with friends, and the detailed home and work locations provided by the CDRs allow the authors to distinguish referrals from homophily and sorting.

Burning Glass/Lightcast job openings data for the US labor market have allowed researchers to uncover other new facts about local labor markets. Hazell et al. (2024) document patterns in wage-setting across cities within firms. They show that occupation-specific wages and their growth are strongly driven by firm-level factors, with a surprisingly small role for location-level factors, including other local prices. They note a tradeoff between the Burning Glass data and administrative data from the Longitudinal Employer-Household Dynamics (LEHD) database. While LEHD data are more representative, especially relative to the subset of Burning Glass data with posted wages, wages from LEHD are noisier because they must be imputed from limited information on hours worked. Papageorgiou (2022) uses the Burning Glass data to document the greater variety of occupations available in larger cities, which then motivates a model in which better labor market matches drive agglomeration. Kleinman (2024) also harnesses these data to investigate the role of large, multi-establishment employers in labor market trends including rising inequality across workers and locations.

v. Consumers

Continuous location data from mobile phones have allowed researchers to unpack consumer behavior, including local effects of the rise of e-commerce and work from home. Athey et al. (2018) use smartphone location data to estimate consumer preferences for restaurants. The data contain enough locational detail and enough visits per consumer to estimate heterogeneity across consumers in a highly flexible model. Hausman et al. (2023) use Waze data on trip volumes aggregated by location pair and time of day in 370 US cities to argue that accessibility of amenities is at least as important as accessibility of jobs in determining housing prices and rents. Miyauchi et al. (2021) use these data in Japan to identify trip chains, whereby people combine multiple trip purposes, including work, shopping or entertainment, and embed them in a quantitative spatial model of a city. In their context, trip chaining amplifies the effect of work-from-home on downtown locations.

Data about purchase locations from payment cards provides further insight into consumer activity. Relihan (2024) finds that access to online groceries decreases local purchases of goods but increases the use of local services. Duguid et al. (2023) trace the impact of work-from-home in the US, documenting that retail establishments leave large cities, particularly in residential neighborhoods as residents moved away to take advantage of work-from-home. Agarwal, Jensen and Monte (2020) use credit card data from 2003, before the widespread adoption of e-commerce, to document the extent of clustering of purchases near home for an earlier period.

Finally, as noted above in discussing neighborhood effects, recent work on experienced segregation has focused on segregation in the consumption of local services using granular smartphone-based individual location data linked to establishments (Athey et al. 2021, Cook 2025 and Couture et al. 2025).

vi. Local political economy

New sources of spatial data have illuminated the relationship between the allocation of resources across space and political economy considerations. Hodler and Raschky (2014) show that regions where the current head of state was born tend to have more night lights, perhaps because of heightened government transfers; this relationship is stronger in countries with weak political institutions. Burgess et al. (2015) find related patterns of favoritism in the distribution of roads, digitized from a panel of historical maps, to co-ethnic regions of the president in Kenya. Bai and Jia (2023) trace out shifts in provincial capitals over six regimes spanning a thousand years of Chinese history and find large but transitory effects of proximity to local power on population and economic activity. Bluhm, Lessmann and Schaudt (2024) study provincial capitals globally over a recent 30-year period, exploiting variation due to new provinces, common over this period of widespread decentralization. On this shorter time scale, they find that becoming a capital increases night lights, and more so for larger provinces.

At a finer spatial scale, Marx et al. (2019) show that residential investment on individual rented dwellings in a Nairobi slum, as proxied by newer roofs visible from satellites, increase, while rents decrease, when local politicians are ethnically aligned with their occupants, as opposed to their landlords. Zeng and Zhou (2024) show that promotion incentives induce Chinese mayors to inflate their reported local GDP, relative to night lights. Yang et al. (2024) show that similar incentives lead Chinese cities to focus pollution abatement on areas near ground-based sensors, leaving areas further from sensors relatively worse off, despite high pollution levels (detected from satellites).[22]

IV. Conclusion

The dramatic leap in new data sources about urban economies past and present is the outgrowth of advances in data capture, processing, and storage. New algorithms have facilitated the analysis of unstructured data such as text and images, and allow for linkage across millions of historical

[22] See also Zou (2021) who shows that *temporal* intermittency in ground-based pollution data collection affects pollution levels seen in satellite data.

records. Hardware like satellites and smartphones have become considerably cheaper and more widespread. These forces have also encouraged certain companies and governments to collect more data on their customers, citizens and lands, and in some cases to share this data with researchers.

We expect substantial improvements to current data sources in both the short- and medium-term, allowing for researchers to address an expanded set of the research questions with these methods. In the short run, for historical data, we note the recent efforts to follow women over time across historical datasets by using maiden names to link between childhood or adulthood, or by linking within marital status. We expect machine learning methods will improve the digitization of more "boutique" historical maps (e.g., as drafted by land surveyors or insurance companies) or hand-drawn maps from the early modern period, allowing for researchers to study exposure to environmental or built conditions. We encourage further efforts to digitize transport networks beyond the US and UK, and to add more communication networks like the telegraph, fiber optic cables and postal routes, as well as other forms of infrastructure like water, sewers, electricity and gas lines.

In the short run, for modern data, we anticipate continued improvements to the use of satellite imagery, including efforts to deal with measurement error, measurement of new attributes including carbon emissions, geographic coverage of high resolution information (for example, on climate and weather), and ever-longer time series. We expect the use of multivariate daytime imagery will expand beyond cross-sectional comparisons to changes within locations over time. Linking cell phone movements to home parcels or buildings to learn more about the phone's owner is still quite new. If this process becomes more feasible and widespread, it will allow for much richer analysis of urban mobility. New data sources with fine spatial resolution are being released at a rapid pace, including data on building footprints and heights from satellite imagery and government sources like the Master Address File from US Census or business-to-business transactions collected for valued added tax purposes in many countries, in some cases linked with employer-employee matched data

In the medium term, we anticipate a surge of new urban data from textual and photographic sources. Current efforts to create a national database on US zoning rules using large language models can be expanded to other local regulations. Urban newspapers can be mined for rental prices, vacancy rates in the labor market via help-wanted ads, and measures of public opinion. Pilot datasets based on historical maps, urban photography (both aerial and street view), as well as from a growing set of video sources, will likely be expanded and standardized. Researchers can also harvest their own data by building phone apps or fielding social media surveys to engage the public.

It is hazardous to speculate too far into the future about what tools might be invented and therefore what sorts of urban data might arise in the longer run. However, we believe that a vision of this future can already be imagined. Advances in machine learning will improve our ability to turn text, images and video into data. Corporate and governmental efforts to collect data about a growing number of our daily activities - from commuting to consumption - will expand. Many activities, past and present, that we can now only proxy for or describe with qualitative sources will soon be measurable at scale, using the techniques described in this chapter. We imagine that much of what we have reviewed in this chapter will look like "pilot" studies of much larger research efforts a decade hence.

Although these new data sources provide many advantages, their use also poses logistical and ethical challenges. First, some of these new sources are collected by corporations and subject to narrow terms of service agreements. Access to this proprietary data is often provided selectively to researchers who establish personal connections, rather than being made widely available to the research community. Even when access is more widespread, availability for ongoing projects can also be rescinded (e.g., Zillow's ZTRAX database and Uber Mobility).[23] Use of proprietary data can limit the replicability of published results, and carries the risk that data providers will shape research agendas by only authorizing projects on topics likely to reflect favorably on them (Zingales 2013). Access to certain kinds of new, often sensitive government data can often be relationship-based as well, posing related problems. Privacy restrictions will likely continue to limit the study of certain topics, such as geographic variation in health. Second, many of these new sources require subject matter expertise in new methods and measurement. Often, successful projects require assembling a research team and collaborating with specialists in geography, computer science or other adjacent disciplines. These efforts may be easier to achieve for senior scholars at large research universities or at institutions with larger research budgets to hire research assistants. At the same time, we have some hope that technological advances, including large language models like ChatGPT, can democratize access to sources that require scraping or processing data.

Some of these data sources also come along with intellectual challenges. First, the need to "know thy data" remains paramount. New datasets have limitations, many of which are well-known to data producers and other specialists. Although downstream users cannot be expected to know as much as the specialists, they must be careful to understand the data well enough to know where its limitations affect their own use. For example, some of these data sources are not representative of the population of interest, so attention must be paid to questions of external validity. Most satellite data of interest to social scientists are unavailable under clouds. Datasets based on use of specific apps, websites, credit cards or phone services are drawn from specific sets of consumers. As the composition of an app's users changes over time, this compositional shift can obscure changes in the underlying phenomenon to be measured such as mobility or overall purchasing activity. Second, the technical frontier for some of the approaches, like machine learning and text analysis, is changing so quickly that research that is state-of-the-art in one period may be out of date in the next. Researchers may choose to redo their analysis with

[23] An alternative to using proprietary data is to "scrape" such data from public websites. The 2023 American Economic Association (AEA) Data Legality policy notes that the legality of scraped data "is not settled under current US law", and strongly implies that if it becomes illegal under settled law, this would affect the willingness of AEA journals to accept such research.

new methods when they arise, which may be important for producing lasting knowledge but can also slow down publication.

Despite these challenges, we are confident that this wave of new data will continue to transform urban economics in the coming decade. We encourage urban economists, both young and old, to familiarize themselves with these data sources and to become conversant in some of the methods needed to build new data from textual corpora, digital traces, and images and video of the world around us, including large language models and deep learning more broadly. We emphasize the word "conversant" because we do not think that all of us need to become experts in these techniques. Rather, we anticipate and encourage interdisciplinary collaboration with scholars around the university in data science, computational linguistics, computer science, geography and the natural sciences who know these methods well and can thus complement the research focus and conceptual framework specific to urban economics.

References

Aaronson, D., Hartley, D., & Mazumder, B. (2021). The Effects of the 1930s HOLC "Redlining" Maps. *American Economic Journal: Economic Policy*, *13*(4), 355–392. https://doi.org/10.1257/pol.20190414

Aaronson, D., Hartley, D., Mazumder, B., & Stinson, M. (2023). The Long-Run Effects of the 1930s Redlining Maps on Children. *Journal of Economic Literature*, *61*(3), 846–862. https://doi.org/10.1257/jel.20221702

Abbiasov, T., Heine, C., Sabouri, S., Salazar-Miranda, A., Santi, P., Glaeser, E., & Ratti, C. (2024). The 15-minute City Quantified Using Human Mobility Data. *Nature Human Behavior*, *8*(March 2024), 445–455.

Abraham, K. G., Ashoori, M., Darzi, A., Gonzalez-Prieto, N., Haltiwanger, J. C., Kabiri, A., & Ozbay, E. Y. (2024). Local Variation in Onsite Work during the Pandemic and its Aftermath. *NBER Working Paper Series*.

Abrahams, A., Oram, C., & Lozano-Gracia, N. (2018). Deblurring DMSP nighttime lights: A new method using Gaussian filters and frequencies of illumination. *Remote Sensing of Environment*, *210*, 242–258. https://doi.org/10.1016/j.rse.2018.03.018

Abramitzky, R., Ager, P., Boustan, L., Cohen, E., & Hansen, C. W. (2023). The Effect of Immigration Restrictions on Local Labor Markets: Lessons from the 1920s Border Closure. *American Economic Journal: Applied Economics*, *15*(1), 164–191. https://doi.org/10.1257/app.20200807

Abramitzky, R., Boustan, L., Brookes Gray, H., Eriksson, K., Perez, S., Postel, H., Rashid, M., & Simon, N. (2025). Finding John Smith: Using Extra Information for Historical Record Linkage. *Working Paper*.

Abramitzky, R., Boustan, L., & Eriksson, K. (2019). To the New World and Back Again: Return Migrants in the Age of Mass Migration. *Industrial & Labor Relations Review*, *72*(2), 300–322. https://doi.org/10.1177/0019793917726981

Abramitzky, R., Boustan, L., Eriksson, K., Feigenbaum, J., & Pérez, S. (2021). Automated Linking of Historical Data. *Journal of Economic Literature*, *59*(3), 865–918. https://doi.org/10.1257/jel.20201599

Abramitzky, R., Boustan, L., Jácome, E., & Pérez, S. (2021). Intergenerational Mobility of Immigrants in the United States over Two Centuries. *American Economic Review*, *111*(2), 580–608. https://doi.org/10.1257/aer.20191586

Abramitzky, R., Boustan, L. P., & Connor, D. (2024). Leaving the Enclave: Historical Evidence on Immigrant Mobility from the Industrial Removal Office. *Journal of Economic History*, *84*(2).

Abramitzky, R., Boustan, L. P., & Eriksson, K. (2012). Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review*, *102*(5), 1832–1856. https://doi.org/10.1257/aer.102.5.1832

Abramitzky, R., Boustan, L. P., & Eriksson, K. (2014). A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration. *Journal of Political Economy*, *122*(3), 467–506. https://doi.org/10.1086/675805

Abramitzky, R., Boustan, L. P., & Giuntella, O. (2025). *Enclaves and Assimilation in the Age of Mass Migration: Evidence from Ethnic Catholic Churches* (NBER Working Paper 33362). NBER Working Paper. https://papers.ssrn.com/abstract=5103552

Abramitzky, R., Boustan, L. P., & Matiashvili, T. (2025). *Intergenerational Mobility Over Two Centuries* (SSRN Scholarly Paper 5083771). Social Science Research Network. https://papers.ssrn.com/abstract=5083771

Adukia, A., Asher, S., Jha, K., Novosad, P., & Tan, B. (2022). Residential Segregation and Unequal Access to Local Public Services in India: Evidence from 1.5m Neighborhoods. *Working Paper*. https://www.semanticscholar.org/paper/Residential-Segregation-and-Unequal-Access-to-Local-Adukia-Asher/5f18e15d3cd628346f32aeb2d4f74c21442add48

Agarwal, S., Jensen, J. B., & Monte, F. (2020). Consumer Mobility and the Local Structure of Consumption Industries. *Human Capital and Economic Opportunity Global Working Group*. https://doi.org/10.3386/w23616

Ager, P., Eriksson, K., Hansen, C. W., & Lønstrup, L. (2020). How the 1906 San Francisco earthquake shaped economic activity in the American West. *Explorations in Economic History*, *77*, 101342. https://doi.org/10.1016/j.eeh.2020.101342

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., & Wolf, N. (2015). The Economics of Density: Evidence From the Berlin Wall. *Econometrica*, *83*(6), 2127–2189. https://doi.org/10.3982/ECTA10876

Aiken, E., Rolf, E., & Blumenstock, J. (2023). Fairness and representation in satellite-based poverty maps: Evidence of urban-rural disparities and their impacts on downstream policy. *arXiv:2305.01783*. https://doi.org/10.48550/ARXIV.2305.01783

Aizer, A., Early, N., Eli, S., Imbens, G., Lee, K., Lleras-Muney, A., & Strand, A. (2024). The Lifetime Impacts of the New Deal's Youth Employment Program. *The Quarterly Journal of Economics*, *139*(4), 2579–2635. https://doi.org/10.1093/qje/qjae016

Akbar, P. A., & Duranton, G. (2017). Measuring the cost of congestion in highly congested city: Bogota. *CAF Working Paper*.

Akbar, P., Couture, V., Duranton, G., & Storeygard, A. (2023a). Mobility and Congestion in Urban India. *American Economic Review*, *113*(4), 1083–1111. https://doi.org/10.1257/aer.20181662

Akbar, P., Couture, V., Duranton, G., & Storeygard, A. (2023b). *The Fast, the Slow, and the Congested: Urban Transportation in Rich and Poor Countries* (w31642; p. w31642). National Bureau of Economic Research. https://doi.org/10.3386/w31642

Albers, T. N. H., & Kappner, K. (2023). Perks and pitfalls of city directories as a micro-geographic data source. *Explorations in Economic History*, *87*, 101476. https://doi.org/10.1016/j.eeh.2022.101476

Alder, S. (2016). Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2856050

Alder, S., Song, Z. (Michael), & Zhu, Z. (2023). On (Un)Congested Roads: A Quantitative Analysis of Infrastructure Investment Efficiency using Truck GPS Data. *Working Paper*.

Alipour, J.-V., Falck, O., Krause, S., Krolage, C., & Wichert, S. (2022). Working from Home and Consumption in Cities. *CESifo Working Paper*, *10000*.

Alix-García, J., & Millimet, D. L. (2023). Remotely Incorrect? Accounting for Nonclassical Measurement Error in Satellite Data on Deforestation. *Journal of the Association of Environmental and Resource Economists*, *10*(5), 1335–1367. https://doi.org/10.1086/723723

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., & Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of Public Economics*, *191*, 104254. https://doi.org/10.1016/j.jpubeco.2020.104254

Allen, R. C., Bertazzini, M. C., & Heldring, L. (2023). The Economic Origins of Government. *American Economic Review*, *113*(10), 2507–2545. https://doi.org/10.1257/aer.20201919

Allen, T., & Atkin, D. (2022). Volatility and the Gains From Trade. *Econometrica*, *90*(5), 2053–2092. https://doi.org/10.3982/ECTA14411

Allen, T., Fuchs, S., Ganapati, S., Graziano, A., Madera, R., & Montoriol-Garriga, J. (2020). Urban Welfare: Tourism in Barcelona. *Working Papoer*.

Almagro, M., Barbieri, F., Castillo, J. C., Hickok, N., & Salz, T. (2024). Optimal Urban Transportation Policy: Evidence from Chicago. *NBER Working Paper Series*.

Althoff, L., Brookes Gray, H., & Reichardt, H. (2024). America's Rise in Human Capital Mobility. *Working Paper*.

Ambrus, A., Field, E., & Gonzalez, R. (2020). Loss in the Time of Cholera: Long-Run Impact of a Disease Epidemic on the Urban Landscape. *American Economic Review*, *110*(2), 475–525. https://doi.org/10.1257/aer.20190759

Andersen, S., Badarinza, C., Liu, L., Marx, J., & Ramadorai, T. (2022). Reference Dependence in the Housing Market. *American Economic Review*, *112*(10), 3398–3440. https://doi.org/10.1257/aer.20191766

Arkolakis, C., Huneeus, F., & Miyauchi, Y. (2023). Spatial Production Networks. *NBER Working Paper Series*. https://doi.org/10.3386/w30954

Arribas-Bel, D., Garcia-López, M.-À., & Viladecans-Marsal, E. (2021). Building(s and) cities: Delineating urban areas with a machine learning algorithm. *Journal of Urban Economics*, *125*, 103217. https://doi.org/10.1016/j.jue.2019.103217

Arthi, V. (2018). "The Dust Was Long in Settling": Human Capital and the Lasting Impact of the American Dust Bowl. *The Journal of Economic History*, *78*(1), 196–230.

Asher, S., Jha, K., Novosad, P., Adukia, A., & Tan, B. J. (2024). Residential Segregation and Unequal Access to Local Public Services in India: Evidence from 1.5m Neighborhoods. *Working Paper*.

Asquith, B. J., Mast, E., & Reed, D. (2023). Local Effects of Large New Apartment Buildings in Low-Income Areas. *The Review of Economics and Statistics*, *105*(2), 359–375. https://doi.org/10.1162/rest_a_01055

Atack, J. (2013). On the Use of Geographic Information Systems in Economic History: The American Transportation Revolution Revisited. *The Journal of Economic History*, *73*(2), 313–338. https://doi.org/10.1017/S0022050713000284

Atack, J. (2018). Creating Historical Transportation Shapefiles of Navigable Rivers, Canals and Railroads for the United States Before World War I. In I. Gregory, D. Debats, & D. Lafreniere (Eds.), *The Routledge Handbook of Spatial History* (pp. 168–183). Routledge.

Atack, J., Bateman, F., Haines, M., & Margo, R. A. (2010). Did Railroads Induce or Follow Economic Growth?: Urbanization and Population Growth in the American Midwest, 1850–1860. *Social Science History*, *34*(2), 171–197. https://doi.org/10.1017/S0145553200011202

Athey, S., Blei, D., Donnelly, R., Ruiz, F., & Schmidt, T. (2018). Estimating Heterogeneous Consumer Preferences for Restaurants and Travel Time Using Mobile Location Data. *AEA Papers and Proceedings*, *108*, 64–67. https://doi.org/10.1257/pandp.20181031

Athey, S., Ferguson, B., Gentzkow, M., & Schmidt, T. (2021). Estimating experienced racial segregation in US cities using large-scale GPS data. *Proceedings of the National Academy of Sciences*, *118*(46), e2026160118. https://doi.org/10.1073/pnas.2026160118

Athey, S., & Imbens, G. W. (2019). Machine Learning Methods That Economists Should Know About. *Annual Review of Economics*, *11*(1), 685–725. https://doi.org/10.1146/annurev-economics-080217-053433

Atkin, D., Chen, K., & Popov, A. (2022). The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley. *NBER Working Paper Series*.

Autor, D. H., Dorn, D., & Hanson, G. H. (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review*, *103*(6), 2121–2168. https://doi.org/10.1257/aer.103.6.2121

Axelrod, J., Lo, L., & Bronin, S. C. (2023). *Automating Zoning Data Collection: Results from a Pilot Effort to Automate National Zoning Atlas Methodologies*. Urban Institute. https://www.urban.org/research/publication/automating-zoning-data-collection

Bai, Y., & Jia, R. (2023). The Economic Consequences of Political Hierarchy: Evidence from Regime Changes in China, 1000-2000 C.E. *The Review of Economics and Statistics*, 1–45. https://doi.org/10.1162/rest_a_01091

Bailey, M., Cao, R., Kuchler, T., & Stroebel, J. (2018). The Economic Effects of Social Networks: Evidence from the Housing Market. *Journal of Political Economy*, *126*(6), 2224–2276. https://doi.org/10.1086/700073

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., & Wong, A. (2018). Social Connectedness: Measurement, Determinants, and Effects. *Journal of Economic Perspectives*, *32*(3), 259–280. https://doi.org/10.1257/jep.32.3.259

Bailey, M., Farrell, P., Kuchler, T., & Stroebel, J. (2020). Social connectedness in urban areas. *Journal of Urban Economics*, *118*, 103264. https://doi.org/10.1016/j.jue.2020.103264

Bailey, M. J., Cole, C., Henderson, M., & Massey, C. (2020). How Well Do Automated Linking Methods Perform? Lessons from US Historical Data. *Journal of Economic Literature*, *58*(4), 997–1044. https://doi.org/10.1257/jel.20191526

Bailey, M., Johnston, D., Koenen, M., Kuchler, T., Russel, D., & Stroebel, J. (2022). The Social Integration of International Migrants: Evidence from the Networks of Syrians in Germany. *NBER Working Paper Series*.

Baker, R. B., Blanchette, J., & Eriksson, K. (2020). Long-Run Impacts of Agricultural Shocks on Educational Attainment: Evidence from the Boll Weevil. *The Journal of Economic History*, *80*(1), 136–174. https://doi.org/10.1017/S0022050719000779

Bakker, J. D., Maurer, S., Pischke, J.-S., & Rauch, F. (2021). Of Mice and Merchants: Connectedness and the Location of Economic Activity in the Iron Age. *The Review of Economics and Statistics*, *103*(4), 652–665.

Balboni, C. A., & Shapiro, J. S. (2025). *Spatial Environmental Economics* (Working Paper 33377). National Bureau of Economic Research. https://doi.org/10.3386/w33377

Banzhaf, H. S., Mathews, W., & Walsh, R. (2024). Hell with the Lid Off: Racial Segregation and Environmental Equity in America's Most Polluted City. *NBER Working Paper Series*. https://doi.org/10.3386/w32950

Baragwanath, K., Goldblatt, R., Hanson, G., & Khandelwal, A. K. (2021). Detecting urban markets with satellite imagery: An application to India. *Journal of Urban Economics*, *125*, 103173. https://doi.org/10.1016/j.jue.2019.05.004

Baragwanath, K., Hanson, G. H., Khandelwal, A. K., Liu, C., & Park, H. (2024). Using Satellite Imagery to Detect the Impacts of New Highways: An Application to India. *NBER Working Paper Series*.

Baran, C., Chyn, E., & Stuart, B. A. (2022). *The Great Migration and Educational Opportunity*. Upjohn Research Working Papers.

Barjamovic, G., Chaney, T., Coşar, K., & Hortaçsu, A. (2019). Trade, Merchants, and the Lost Cities of the Bronze Age. *The Quarterly Journal of Economics*, *134*(3), 1455–1503. https://doi.org/10.1093/qje/qjz009

Barnwal, P., Dingel, J. I., Krishna, P., Iurchenko, D., & Van Leemput, E. (2024). *Internal Trade Barriers in India*.

Barron, K., Kung, E., & Proserpio, D. (2021). The Effect of Home-Sharing on House Prices and Rents: Evidence from Airbnb. *Marketing Science*, *40*(1), 23–47. https://doi.org/10.1287/mksc.2020.1227

Bartik, A., Gupta, A., & Milo, D. (2024). The Costs of Housing Regulation: Evidence From Generative Regulatory Measurement. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4627587

Barwick, P. J., Liu, Y., Patacchini, E., & Wu, Q. (2023). Information, Mobile Communication, and Referral Effects. *American Economic Review*, *113*(5), 1170–1207. https://doi.org/10.1257/aer.20200187

Barwick, P. J., Waxman, A., Li, S., Wu, J., & Xia, T. (2024). Efficiency and Equity Impacts of Urban Transportation Policies with Equilibrium Sorting. *American Economic Review (Forthcoming)*.

Bauer, S., & Hnilo, F. (2023). Scars of the Gestapo: Remembrance and Privacy Concerns. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4328676

Baugh, K., Elvidge, C. D., Ghosh, T., & Ziskin, D. (2010). Development of a 2009 Stable Lights Product using DMSP-OLS data. *Proceedings of the Asia-Pacific Advanced Network*, *30*(0), Article 0. https://doi.org/10.7125/apan.30.17

Baum-Snow, N. (2007). Did Highways Cause Suburbanization? *The Quarterly Journal of Economics*, *122*(2), 775–805. https://doi.org/10.1162/qjec.122.2.775

Baum-Snow, N. (2020). Urban Transport Expansions and Changes in the Spatial Structure of U.S. Cities: Implications for Productivity and Welfare. *The Review of Economics and Statistics*, *102*(5), 929–945. https://doi.org/10.1162/rest_a_00855

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., & Zhang, Q. (2017). Roads, Railroads, and Decentralization of Chinese Cities. *The Review of Economics and Statistics*, *99*(3), 435–448. https://doi.org/10.1162/REST_a_00660

Baum-Snow, N., & Han, L. (2024). The Microgeography of Housing Supply. *Journal of Political Economy*, *132*(6), 1897–1946. https://doi.org/10.1086/728110

Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q., & Brandt, L. (2020). Does investment in national highways help or hurt hinterland city growth? *Journal of Urban Economics*, *115*, 103124. https://doi.org/10.1016/j.jue.2018.05.001

Bazzi, S., Fiszbein, M., & Gebresilasse, M. (2020). Frontier Culture: The Roots and Persistence of "Rugged Individualism" in the United States. *Econometrica*, *88*(6), 2329–2368. https://doi.org/10.3982/ECTA16484

Beach, B., Ferrie, J., Saavedra, M., & Troesken, W. (2016). Typhoid Fever, Water Quality, and Human Capital Formation. *The Journal of Economic History*, *76*(1), 41–75. https://doi.org/10.1017/S0022050716000413

Beirle, S., Borger, C., Dörner, S., Li, A., Hu, Z., Liu, F., Wang, Y., & Wagner, T. (2019). Pinpointing nitrogen oxide emissions from space. *Science Advances*, *5*(11), eaax9800. https://doi.org/10.1126/sciadv.aax9800

Bellet, C. S. (2024). The McMansion effect: Positional externalities in U.S. suburbs. *Journal of Public Economics*, *238*, 105174. https://doi.org/10.1016/j.jpubeco.2024.105174

Bento, A., Roth, K., & Waxman, A. (2024). The Value of Urgency: Evidence from Real-Time Congestion Pricing. *Journal of Political Economy Microeconomics*, *2*(4), 786–851. https://doi.org/10.1086/731832

Berkes, E., Karger, E., & Nencka, P. (2023). The census place project: A method for geolocating unstructured place names. *Explorations in Economic History*, *87*, 101477. https://doi.org/10.1016/j.eeh.2022.101477

Besbris, M., Schachter, A., & Kuk, J. (2021). The Unequal Availability of Rental Housing Information Across Neighborhoods. *Demography*, *58*(4), 1197–1221. https://doi.org/10.1215/00703370-9357518

Bjorkegren, D., Duhaut, A., & Tsivanidis, N. (2024). Public and Private Transit: Evidence from Lagos. *Working Paper*.

Black, D. A., Sanders, S. G., Taylor, E. J., & Taylor, L. J. (2015). The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South. *American Economic Review*, *105*(2), 477–503. https://doi.org/10.1257/aer.20120642

Blanchard, P., Gollin, D., & Kirchberger, M. (2023). *High-Frequency Human Mobility in Three African Countries*.

Bleakley, H., & Lin, J. (2012). Portage and Path Dependence. *The Quarterly Journal of Economics*, *127*(2), 587–644. https://doi.org/10.1093/qje/qjs011

Bluhm, R., & Krause, M. (2022). Top lights: Bright cities and their contribution to economic development. *Journal of Development Economics*, *157*, 102880. https://doi.org/10.1016/j.jdeveco.2022.102880

Bluhm, R., Lessmann, C., & Schaudt, P. (2021). *The Political Geography of Cities* (SSRN Scholarly Paper 3953012). Social Science Research Network. https://doi.org/10.2139/ssrn.3953012

Blumenstock, J. E. (2012). Inferring patterns of internal migration from mobile phone call records: Evidence from Rwanda. *Information Technology for Development*, *18*(2), 107–125. https://doi.org/10.1080/02681102.2011.643209

Blumenstock, J. E., Chi, G., & Tan, X. (2023). Migration and the Value of Social Networks. *Review of Economic Studies*, rdad113. https://doi.org/10.1093/restud/rdad113

Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, *65*, 126–139. https://doi.org/10.1016/j.compenvurbsys.2017.05.004

Bogart, D. (2024). A survey of economic history research on transport. *Working Paper*.

Bogart, D., You, X., Alvarez-Palau, E. J., Satchell, M., & Shaw-Taylor, L. (2022). Railways, divergence, and structural change in 19th century England and Wales. *Journal of Urban Economics*, *128*, 103390. https://doi.org/10.1016/j.jue.2021.103390

Boone, C. D. A., & Wilse-Samson, L. (2023). Structural Change and Internal Labor Migration: Evidence from the Great Depression. *Review of Economics and Statistics*, *105*(4), 962–981. https://doi.org/10.1162/rest_a_01116

Borker, G. (2021). *Safety First: Perceived Risk of Street Harassment and Educational Choices of Women*. The World Bank. https://doi.org/10.1596/1813-9450-9731

Boustan, L. P. (2012). School Desegregation and Urban Change: Evidence from City Boundaries. *American Economic Journal: Applied Economics*, *4*(1), 85–108. https://doi.org/10.1257/app.4.1.85

Boustan, L. P. (2013). Local public goods and the demand for high-income municipalities. *Journal of Urban Economics*, *76*, 71–82. https://doi.org/10.1016/j.jue.2013.02.003

Boustan, L. P. (2016). *Competition in the Promised Land: Black Migrants in Northern Cities and Labor Markets*. Princeton University Press.

Box-Couillard, S., & Christensen, P. (2024). Racial Housing Price Differentials and Neighborhood Segregation. *NBER Working Paper Series*. https://doi.org/10.3386/w32815

Brancaccio, G., Kalouptsidi, M., & Papageorgiou, T. (2024). Investment in Infrastructure and Trade: The Case of Ports. *NBER Working Paper Series*. https://doi.org/10.3386/w32503

Brandily, P., & Rauch, F. (2024). Within‐city roads and urban growth. *Journal of Regional Science*, *64*(4), 1236–1264. https://doi.org/10.1111/jors.12699

Bratu, C., Harjunen, O., & Saarimaa, T. (2023). JUE Insight: City-wide effects of new housing supply: Evidence from moving chains. *Journal of Urban Economics*, *133*, 103528. https://doi.org/10.1016/j.jue.2022.103528

Brimble, P., Larrinaga, A. E., & Motohashi, K. (2025). VIIRS and the future of night lights data in economics. *Working PAper*.

Brooks, L., & Lutz, B. (2019). Vestiges of Transit: Urban Persistence at a Microscale. *The Review of Economics and Statistics*, *101*(3), 385–399. https://doi.org/10.1162/rest_a_00817

Brooks, L., Rose, J., & Veuger, S. (2024). Destruction, Policy, and the Evolving Consequences of Washington, DC's 1968 Civil Disturbance. *The Review of Economics and Statistics*, 1–45. https://doi.org/10.1162/rest_a_01455

Büchel, K., & Ehrlich, M. V. (2020). Cities and the structure of social interactions: Evidence from mobile phone data. *Journal of Urban Economics*, *119*, 103276. https://doi.org/10.1016/j.jue.2020.103276

Büchel, K., Ehrlich, M. V., Puga, D., & Viladecans-Marsal, E. (2020). Calling from the outside: The role of networks in residential mobility. *Journal of Urban Economics*, *119*, 103277. https://doi.org/10.1016/j.jue.2020.103277

Buchholz, N. (2022). Spatial Equilibrium, Search Frictions, and Dynamic Efficiency in the Taxi Industry. *The Review of Economic Studies*, *89*(2), 556–591. https://doi.org/10.1093/restud/rdab050

Buckles, K., Haws, A., Price, J., & Wilbert, H. E. B. (2023). Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project. *NBER Working Paper Series*.

Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., & Padró i Miquel, G. (2015). The Value of Democracy: Evidence from Road Building in Kenya. *American Economic Review*, *105*(6), 1817–1851. https://doi.org/10.1257/aer.20131031

Caetano, G., & Maheshri, V. (2019). Gender segregation within neighborhoods. *Regional Science and Urban Economics*, *77*, 253–263. https://doi.org/10.1016/j.regsciurbeco.2019.05.004

Cagney, K. A., Cornwell, E. Y., Goldman, A. W., & Cai, L. (2020). Urban Mobility and Activity Space. *Annual Review of Sociology*, *46*(Volume 46, 2020), 623–648. https://doi.org/10.1146/annurev-soc-121919-054848

Calder-Wang, S. (2021). The Distributional Impact of the Sharing Economy: Evidence from New York City. *Working Paper*.

Camerer, C., Babcock, L., Loewenstein, G., & Thaler, R. (1997). Labor Supply of New York City Cabdrivers: One Day at a Time*. *The Quarterly Journal of Economics*, *112*(2), 407–441. https://doi.org/10.1162/003355397555244

Campante, F., & Yanagizawa-Drott, D. (2018). Long-Range Growth: Economic Development in the Global Network of Air Links. *The Quarterly Journal of Economics*, *133*(3), 1395–1458. https://doi.org/10.1093/qje/qjx050

Carrion, D., Migliaccio, F., Minini, G., & Zambrano, C. (2016). From historical documents to GIS: A spatial database for medieval fiscal data in Southern Italy. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *49*(1), 1–10. https://doi.org/10.1080/01615440.2015.1023877

Ch, R., Martin, D. A., & Vargas, J. F. (2021). Measuring the size and growth of cities using nighttime light. *Journal of Urban Economics*, *125*, 103254. https://doi.org/10.1016/j.jue.2020.103254

Chacha, P. W., Kirui, B. K., & Wiedemann, V. (2024). Supply Chains in Times of Crisis: Evidence from Kenya's Production Network. *World Development*, *173*, 106363. https://doi.org/10.1016/j.worlddev.2023.106363

Chan, J. (2024). The long-run effects of childhood exposure to market access shocks: Evidence from the US railroad network expansion. *Explorations in Economic History*, *91*, 101503. https://doi.org/10.1016/j.eeh.2022.101503

Chen, M. K., & Pope, D. G. (2020). Geographic Mobility in America: Evidence from Cell Phone Data. *NBER Working Paper Series*.

Chen, X., & Nordhaus, W. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, *108*(21), 8589–8594.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, *21*(1), C1–C68. https://doi.org/10.1111/ectj.12097

Chetty, R., Friedman, J. N., Hendren, N., Jones, M. R., & Porter, S. R. (2025). The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility. *American Economic Review*, *forthcoming*. https://doi.org/10.3386/w25147

Chetty, R., & Hendren, N. (2018). The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates*. *The Quarterly Journal of Economics*, *133*(3), 1163–1228. https://doi.org/10.1093/qje/qjy006

Chetty, R., Hendren, N., & Katz, L. F. (2016). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review*, *106*(4), 855–902. https://doi.org/10.1257/aer.20150572

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *. *The Quarterly Journal of Economics*, *129*(4), 1553–1623. https://doi.org/10.1093/qje/qju022

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., Johnston, D., Koenen, M., Laguna-Muggenburg, E., Mudekereza, F., Rutter, T., Thor, N., Townsend, W., Zhang, R., Bailey, M., … Wernerfelt, N. (2022a). Social capital I: Measurement and associations with economic mobility. *Nature*, *608*(7921), 108–121. https://doi.org/10.1038/s41586-022-04996-4

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., Johnston, D., Koenen, M., Laguna-Muggenburg, E., Mudekereza, F., Rutter, T., Thor, N., Townsend, W., Zhang, R., Bailey, M., … Wernerfelt, N. (2022b). Social capital II: Determinants of economic connectedness. *Nature*, *608*(7921), 122–134. https://doi.org/10.1038/s41586-022-04997-3

Chiovelli, G., Michalopoulos, S., Papaioannou, E., & Sequeira, S. (2021). *Forced Displacement and Human Capital: Evidence from Separated Siblings* (SSRN Scholarly Paper 3994268). Social Science Research Network. https://papers.ssrn.com/abstract=3994268

Choi, J. (2023). *The Effect of Deindustrialization on Local Economies: Evidence from New England Textile Towns* (SSRN Scholarly Paper 4772455). Social Science Research Network. https://doi.org/10.2139/ssrn.4772455

Ciccarelli, C., & Groote, P. (2018). The Spread of Railroads in Italian Provinces: A GIS Approach. *Scienze Regionali*, *2*, 189–224. https://doi.org/10.14650/90221

Civelli, A., Gaduh, A., & Yousuf, A. S. (2024). Measuring Economic Growth with a Fully Identified Three-Signal Model. *The Review of Economics and Statistics*, 1–30. https://doi.org/10.1162/rest_a_01489

Collins, W. J., & Wanamaker, M. H. (2014). Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data. *American Economic Journal: Applied Economics*, *6*(1), 220–252. https://doi.org/10.1257/app.6.1.220

Combes, P.-P., Duranton, G., Gobillon, L., & Gorin, C. (2025). *Measuring land use changes by (machine) learning from historical maps* [Working Paper].

Combes, P.-P., Gobillon, L., & Zylberberg, Y. (2022). Urban economics in a historical perspective: Recovering data with machine learning. *Regional Science and Urban Economics*, *94*, 103711. https://doi.org/10.1016/j.regsciurbeco.2021.103711

Cook, C. (2025). Heterogeneous Preferences for Neighborhood Amenities: Evidence from GPS Data. *The Review of Economics and Statistics*, 1–43. https://doi.org/10.1162/rest_a_01551

Cook, C., Currier, L., & Glaeser, E. (2024). Urban mobility and the experienced isolation of students. *Nature Cities*, *1*(1), 73–82. https://doi.org/10.1038/s44284-023-00007-3

Cook, C., & Li, P. Z. (2023). Value Pricing or Lexus Lanes? The Distributional Effects of Dynamic Tolling. *Working Paper*.

Coşar, A. K., Demir, B., Ghose, D., & Young, N. (2022). Road capacity, domestic trade and regional outcomes. *Journal of Economic Geography*, *22*(5), 901–929. https://doi.org/10.1093/jeg/lbab045

Coury, M., Kitagawa, T., Shertzer, A., & Turner, M. A. (2024). The Value of Piped Water and Sewers: Evidence from 19th Century Chicago. *The Review of Economics and Statistics*, 1–47. https://doi.org/10.1162/rest_a_01483

Couture, V., Dingel, J. I., Green, A., & Handbury, J. (2025). Demographic Preferences and Income Segregation. *Working Paper*.

Couture, V., Dingel, J. I., Green, A., Handbury, J., & Williams, K. R. (2022). JUE Insight: Measuring movement and social contact with smartphone data: a real-time application to COVID-19. *Journal of Urban Economics*, *127*, 103328. https://doi.org/10.1016/j.jue.2021.103328

Couture, V., Faber, B., Gu, Y., & Liu, L. (2021). Connecting the Countryside via E-Commerce: Evidence from China. *American Economic Review: Insights*, *3*(1), 35–50. https://doi.org/10.1257/aeri.20190382

Currier, L., Glaeser, E. L., & Kreindler, G. E. (2023). Infrastructure Inequality: Who Pays the Cost of Road Roughness? *NBER Working Paper Series*.

Davis, D. R., Dingel, J. I., Monras, J., & Morales, E. (2019). How Segregated Is Urban Consumption? *Journal of Political Economy*, *127*(4).

de Bellefon, M.-P., Combes, P.-P., Duranton, G., Gobillon, L., & Gorin, C. (2021). Delineating urban areas using building density. *Journal of Urban Economics*, *125*, 103226. https://doi.org/10.1016/j.jue.2019.103226

DeBats, D. A. (2009). Using GIS and Individual-Level Data for Whole Communities: A Path Toward the Reconciliation of Political and Social History. *Social Science Computer Review*, *27*(3), 313–330. https://doi.org/10.1177/0894439308329759

Dell, M. (2010). The Persistent Effects of Peru's Mining Mita. *Econometrica*, *78*(6), 1863–1903. https://doi.org/10.3982/ECTA8121

Dell, M. (2024). *Deep Learning for Economists* (Working Paper 32768). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w32768/w32768.pdf

Dell, M., Carlson, J., Bryan, T., Silcock, E., Arora, A., Shen, Z., D'Amico-Wong, L., Le, Q., Querubin, P., & Heldring, L. (2023). American Stories: A Large-Scale Structured Text Dataset of Historical U.S. Newspapers. *Working Paper*.

Demir, B., Fieler, A. C., Xu, D. Y., & Yang, K. K. (2024). O-Ring Production Networks. *Journal of Political Economy*, *132*(1), 200–247. https://doi.org/10.1086/725703

Demir, B., Javorcik, B., & Panigrahi, P. (2024). *Breaking Invisible Barriers: Does Fast Internet Improve Access to Input Markets?*

Derenoncourt, E. (2022). Can You Move to Opportunity? Evidence from the Great Migration. *American Economic Review*, *112*(2), 369–408. https://doi.org/10.1257/aer.20200002

Dericks, G. H., & Koster, H. R. A. (2021). The billion pound drop: The Blitz and agglomeration economies in London. *Journal of Economic Geography*, *21*(6), 869–897. https://doi.org/10.1093/jeg/lbaa022

Dijkstra, L., Florczyk, A. J., Freire, S., Kemper, T., Melchiorri, M., Pesaresi, M., & Schiavina, M. (2021). Applying the Degree of Urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. *Journal of Urban Economics*, *125*, 103312. https://doi.org/10.1016/j.jue.2020.103312

Dincecco, M., Fenske, J., Gupta, B., & Menon, A. (2024). *Conflict and Gender Norms* (SSRN Scholarly Paper 4751560). Social Science Research Network. https://doi.org/10.2139/ssrn.4751560

Dingel, J. I., Miscio, A., & Davis, D. R. (2021). Cities, lights, and skills in developing economies. *Journal of Urban Economics*, *125*, 103174. https://doi.org/10.1016/j.jue.2019.05.005

Dinlersoz, E., Dunne, T., Haltiwanger, J., & Penciakova, V. (2023). The Local Origins of Business Formation. *Working Papers*, Article 23–34. https://ideas.repec.org//p/cen/wpaper/23-34.html

Dolfen, P., Einav, L., Klenow, P. J., Klopack, B., Levin, J. D., Levin, L., & Best, W. (2023). Assessing the Gains from E-Commerce. *American Economic Journal: Macroeconomics*, *15*(1), 342–370. https://doi.org/10.1257/mac.20210049

Donaldson, D. (2018). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*, *108*(4–5), 899–934. https://doi.org/10.1257/aer.20101199

Donaldson, D., & Hornbeck, R. (2016). Railroads and American Economic Growth: A "Market Access" Approach. *The Quarterly Journal of Economics*, *131*(2), 799–858. https://doi.org/10.1093/qje/qjw002

Donaldson, D., & Storeygard, A. (2016). The View from Above: Applications of Satellite Data in Economics. *Journal of Economic Perspectives*, *30*(4), 171–198. https://doi.org/10.1257/jep.30.4.171

Ducruet, C., Cuyala, S., & El Hosni, A. (2018). Maritime networks as systems of cities: The long-term interdependencies between global shipping flows and urban development

(1890–2010). *Journal of Transport Geography*, *66*, 340–355. https://doi.org/10.1016/j.jtrangeo.2017.10.019

Ducruet, C., Juhász, R., Nagy, D. K., & Steinwender, C. (2024). All aboard: The effects of port development. *Journal of International Economics*, *151*, 103963. https://doi.org/10.1016/j.jinteco.2024.103963

Duguid, J., Kim, B., Relihan, L., & Wheat, C. (2023). The Impact of Work-from-Home on Brick-and-Mortar Retail Establishments: Evidence from Card Transactions. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4466607

Duprey, T., Rigobon, D. E., Kotlicki, A., & Schnattinger, P. (2023). Timely Business Dynamics Using Google Places. *AEA Papers and Proceedings*, *113*, 135–139. https://doi.org/10.1257/pandp.20231050

Duque, V., & Schmitz, L. L. (2023). Early-life Exposure to the Great Depression and Long-term Health and Economic Outcomes. *Journal of Human Resources*, 0421-11584R1. https://doi.org/10.3368/jhr.0421-11584R1

Duranton, G., & Rosenthal, S. S. (2021). Special issue on delineation of urban areas. *Journal of Urban Economics*, *125*, 103352. https://doi.org/10.1016/j.jue.2021.103352

Edelman, B., & Luca, M. (2014). Digital Discrimination: The Case of Airbnb.com. *Harvard Business School Working Papers*, Article 14–054. https://ideas.repec.org//p/hbs/wpaper/14-054.html

Edelman, B., Luca, M., & Svirsky, D. (2017). Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment. *American Economic Journal: Applied Economics*, *9*(2), 1–22. https://doi.org/10.1257/app.20160213

Egger, D., Faber, B., Li, M., & Lin, W. (2024). Rural-Urban Migration and Market Integration. *Working Paper*.

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., & Davis, E. R. (1997). Mapping of city lights using DMSP Operational Linescan System data. *Photogrammetric Engineering and Remote Sensing*, *63*(6). https://trid.trb.org/View/483519

Elvidge, C. D., Baugh, K. E., Zhizhin, M., & Hsu, F.-C. (2013). Why VIIRS data are superior to DMSP for mapping nighttime lights. *Proceedings of the Asia-Pacific Advanced Network*, *35*(0), 62. https://doi.org/10.7125/APAN.35.7

Elvidge, C. D., Imhoff, M. L., Baugh, K. E., Hobson, V. R., Nelson, I., Safran, J., Dietz, J. B., & Tuttle, B. T. (2001). Night-time lights of the world: 1994–1995. *ISPRS Journal of Photogrammetry and Remote Sensing*, *56*(2), 81–99. https://doi.org/10.1016/S0924-2716(01)00040-5

Engstrom, R., Hersh, J., & Newhouse, D. (2022). Poverty from Space: Using High Resolution Satellite Imagery for Estimating Economic Well-being. *The World Bank Economic Review*, *36*(2), 382–412. https://doi.org/10.1093/wber/lhab015

Eriksson, K. (2020). Ethnic enclaves and immigrant outcomes: Norwegian immigrants during the Age of Mass Migration. *European Review of Economic History*, *24*(3), 427–446. https://doi.org/10.1093/ereh/hez013

Eriksson, K., & Ward, Z. (2019). The Residential Segregation of Immigrants in the United States from 1850 to 1940. *The Journal of Economic History*, *79*(4), 989–1026. https://doi.org/10.1017/S0022050719000536

Esch, T., Brzoska, E., Dech, S., Leutner, B., Palacios-Lopez, D., Metz-Marconcini, A., Marconcini, M., Roth, A., & Zeidler, J. (2022). World Settlement Footprint 3D - A first three-dimensional survey of the global building stock. *Remote Sensing of Environment*, *270*, 112877. https://doi.org/10.1016/j.rse.2021.112877

Esch, T., Deininger, K., Jedwab, R., & Palacios-Lopez, D. (2024). *Outward and Upward Construction: A 3D Analysis of the Global Building Stock*. Washington, DC: World Bank. https://doi.org/10.1596/1813-9450-10866

Faber, B. (2014). Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *The Review of Economic Studies*, *81*(3), 1046–1070. https://doi.org/10.1093/restud/rdu010

Faber, J. (2020). We Built This: Consequences of New Deal Era Intervention in America's Racial Geography. *American Sociological Review*, *85*(5).

Fan, J., Tang, L., Zhu, W., & Zou, B. (2018). The Alibaba effect: Spatial consumption inequality and the welfare gains from e-commerce. *Journal of International Economics*, *114*, 203–220. https://doi.org/10.1016/j.jinteco.2018.07.002

Fang, L., Stewart, N., & Tyndall, J. (2023). Homeowner politics and housing supply. *Journal of Urban Economics*, *138*(C). https://ideas.repec.org//a/eee/juecon/v138y2023ics0094119023000785.html

Farronato, C., & Fradkin, A. (2022). The Welfare Effects of Peer Entry: The Case of Airbnb and the Accommodation Industry. *American Economic Review*, *112*(6), 1782–1817. https://doi.org/10.1257/aer.20180260

Feigenbaum, J. J. (2016). A Machine Learning Approach to Census Record Linking. *Working Paper*.

Fishback, P. V., LaVoice, J., Shertzer, A., & Walsh, R. P. (2023). The HOLC Maps: How Race and Poverty Influenced Real Estate Professionals' Evaluation of Lending Risk in the 1930s. *The Journal of Economic History*, *83*(4), 1019–1056. https://doi.org/10.1017/S0022050723000475

Fiszbein, M. (2022). Agricultural Diversity, Structural Change, and Long-Run Development: Evidence from the United States. *American Economic Journal: Macroeconomics*, *14*(2), 1–43. https://doi.org/10.1257/mac.20190285

Fitch, C. A., & Ruggles, S. (2003). Building the National Historical Geographic Information System. *Historical Methods*, *36*(1), 41–51. https://doi.org/10.1080/01615440309601214

Flückiger, M., Hornung, E., Larch, M., Ludwig, M., & Mees, A. (2022). Roman Transport Network Connectivity and Economic Integration. *The Review of Economic Studies*, *89*(2), 774–810. https://doi.org/10.1093/restud/rdab036

French, R., & Gilbert, V. (2024). Suburban Housing and Urban Affordability: Evidence from Residential Vacancy Chains. *Joint Center for Housing Studies Meyer Fellowship Paper*.

Gabriel, R., Haws, A., Bailey, A. K., & Price, J. (2023). The Migration of Lynch Victims' Families, 1880–1930. *Demography*, *60*(4), 1235–1256. https://doi.org/10.1215/00703370-10881293

Galdo, V., Li, Y., & Rama, M. (2021). Identifying urban areas by combining human judgment and machine learning: An application to India. *Journal of Urban Economics*, *125*, 103229. https://doi.org/10.1016/j.jue.2019.103229

Gandhi, S., Kahn, M. E., Kochhar, R., Lall, S., & Tandel, V. (2022). Adapting to Flood Risk: Evidence from a Panel of Global Cities. *Working Paper*. https://papers.ssrn.com/abstract=4134933

Garcia-López, M.-À., Jofre-Monseny, J., Martínez-Mazza, R., & Segú, M. (2020). Do short-term rental platforms affect housing markets? Evidence from Airbnb in Barcelona. *Journal of Urban Economics*, *119*, 103278. https://doi.org/10.1016/j.jue.2020.103278

Gay, V. (2020). Mapping the Third Republic: A Geographic Information System of France (1870–1940). *TSE Working Papers*, Article 20–1151. https://ideas.repec.org//p/tse/wpaper/124767.html

Gechter, M., & Tsivanidis, N. (2023). Spatial Spillovers from High-Rise Developments: Evidence from the Mumbai Mills. *Working Paper*.

Gee, L. K., Jones, J., & Burke, M. (2017). Social Networks and Labor Markets: How Strong Ties Relate to Job Finding on Facebook's Social Network. *Journal of Labor Economics*, *35*(2), 485–518. https://doi.org/10.1086/686225

Gendron-Carrier, N., Gonzalez-Navarro, M., Polloni, S., & Turner, M. A. (2022). Subways and Urban Air Pollution. *American Economic Journal: Applied Economics*, *14*(1), 164–196. https://doi.org/10.1257/app.20180168

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as Data. *Journal of Economic Literature*, *57*(3), 535–574. https://doi.org/10.1257/jel.20181020

Gertler, P. J., Gonzalez-Navarro, M., Gračner, T., & Rothenberg, A. D. (2024). Road maintenance and local economic development: Evidence from Indonesia's highways. *Journal of Urban Economics*, *143*, 103687. https://doi.org/10.1016/j.jue.2024.103687

Gibbons, S., Heblich, S., & Pinchbeck, E. (2024). *The Spatial Impacts of a Massive Rail Disinvestment Program: The Beeching Axe* (w32800; p. w32800). National Bureau of Economic Research. https://doi.org/10.3386/w32800

Gibson, J. (2021). Better Night Lights Data, For Longer. *Oxford Bulletin for Economics and Statistics*, *3*, 770–791.

Gibson, J., Olivia, S., Boe-Gibson, G., & Li, C. (2021). Which night lights data should we use in economics, and where? *Journal of Development Economics*, *149*, 102602. https://doi.org/10.1016/j.jdeveco.2020.102602

Glaeser, E., Kim, H., & Luca, M. (2017). *Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity* (w24010; p. w24010). National Bureau of Economic Research. https://doi.org/10.3386/w24010

Glaeser, E. L., Kominers, S. D., Luca, M., & Naik, N. (2018). Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life. *Economic Inquiry*, *56*(1), 114–137. https://doi.org/10.1111/ecin.12364

Goeken, R., Huynh, L., Lynch, T. A., & Vick, R. (2011). New Methods of Census Record Linking. *Historical Methods*, *44*(1), 7–14.

Gonzalez-Navarro, M., Hall, J. D., Williams, R., Wheeler, H., & Williams, R. (2022). Uber Versus Trains? Worldwide Evidence from Transit Expansions. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3959760

Gorback, C. (2024). Ridesharing and the Redistribution of Economic Activity. *Working Paper*. https://doi.org/10.2139/ssrn.4956559

Gorin, C. (2021). Neural Networks for Economists: A Primer. *Working Paper*.

Gorin, C., Combes, P.-P., Gobillon, L., & Robert-Nicoud, F. (2024). *Urbanization and urban divergence: France 1760-2020* [Working Paper].

Gray, R., & Bowman, R. (2021). Locating the Manhattan housing market: GIS evidence for 1880-1910. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *54*(3), 151–171. https://doi.org/10.1080/01615440.2020.1832007

Green, A. (2025). Networks and Geographic Mobility: Evidence from World War II Navy Ships. *Working Paper*.

Gregory, I., DeBats, D. A., & Lafreniere, D. (2018). *The Routledge companion to spatial history*. Routledge.

Grigoryeva, A., & Ruef, M. (2015). The Historical Demography of Racial Segregation. *American Sociological Review*, *80*(4), 814–842.

Gruber, J., Jensen, A., & Kleven, H. (2021). Do People Respond to the Mortgage Interest Deduction? Quasi-Experimental Evidence from Denmark. *American Economic Journal: Economic Policy*, *13*(2), 273–303. https://doi.org/10.1257/pol.20170366

Grupp, T. E., Mishra, P., Reynaert, M., & Van Benthem, A. (2024). An Evaluation of Protected Area Policies in the European Union. *Working Paper*.

Gu, Y., Jiang, C., Zhang, J., & Zou, B. (2021). Subways and Road Congestion. *American Economic Journal: Applied Economics*, *13*(2), 83–115. https://doi.org/10.1257/app.20190024

Gupta, A., Mittal, V., Peeters, J., & Van Nieuwerburgh, S. (2022). Flattening the curve: Pandemic-Induced revaluation of urban real estate. *Journal of Financial Economics*, *146*(2), 594–636. https://doi.org/10.1016/j.jfineco.2021.10.008

Gupta, A., Van Nieuwerburgh, S., & Kontokosta, C. (2022). Take the Q train: Value capture of public infrastructure projects. *Journal of Urban Economics*, *129*, 103422. https://doi.org/10.1016/j.jue.2021.103422

Han, L., Heblich, S., Timmins, C., & Zylberberg, Y. (2024). Cool cities: The value of urban trees. *NBER Working Paper Series*. https://doi.org/10.3386/w32063

Hanlon, W. W., & Heblich, S. (2022). History and urban economics. *Regional Science and Urban Economics*, *94*, 103751. https://doi.org/10.1016/j.regsciurbeco.2021.103751

Hanlon, W. W., Heblich, S., Monte, F., & Schmitz, M. (2022). A Penny for Your Thoughts. *NBER Working Paper Series*. https://doi.org/10.3386/w30076

Hanna, R., Kreindler, G., & Olken, B. A. (2017). Citywide effects of high-occupancy vehicle restrictions: Evidence from "three-in-one" in Jakarta. *Science (New York, N.Y.)*, *357*(6346), 89–93. https://doi.org/10.1126/science.aan2747

Harari, M. (2020). Cities in Bad Shape: Urban Geometry in India. *American Economic Review*, *110*(8), 2377–2421. https://doi.org/10.1257/aer.20171673

Harari, M., & Wong, M. (2024). Slum Upgrading and Long-run Urban Development: Evidence from Indonesia. *Working Paper*.

Hausman, N., Samuels, P. R., Cohen, M. C., & Sasson, R. (2023). *Urban Pull: The Roles of Amenities and Employment* (SSRN Scholarly Paper 3670974). Social Science Research Network. https://doi.org/10.2139/ssrn.3670974

Hazell, J., Patterson, C., Sarsons, H., & Taska, B. (2024). National Wage Setting. *NBER Working Paper Series*.

Heblich, S., Nagy, D. K., Trew, A., & Zylberberg, Y. (2023). The death and life of great British cities. *Working Paper*. https://ideas.repec.org//p/upf/upfgen/1867.html

Heblich, S., Redding, S. J., & Sturm, D. M. (2020). The Making of the Modern Metropolis: Evidence from London. *The Quarterly Journal of Economics*, *135*(4), 2059–2133. https://doi.org/10.1093/qje/qjaa014

Heblich, S., Redding, S. J., & Zylberberg, Y. (2024). *The Distributional Consequences of Trade: Evidence from the Grain Invasion*.

Heblich, S., Seror, M., Xu, H., & Zylberberg, Y. (2022). *Industrial Clusters in the Long Run: Evidence from Million-Rouble Plants in China* (w30744; p. w30744). National Bureau of Economic Research. https://doi.org/10.3386/w30744

Heblich, S., Trew, A., & Zylberberg, Y. (2021). East-Side Story: Historical Pollution and Persistent Neighborhood Sorting. *Journal of Political Economy*, *129*(5).

Hedefalk, F., Harrie, L., & Svensson, P. (2015). Methods to Create a Longitudinal Integrated Demographic and Geographic Database on the Micro-Level: A Case Study of Five Swedish Rural Parishes, 1813–1914. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *48*(3), 153–173. https://doi.org/10.1080/01615440.2015.1016645

Helgertz, J., Price, J., Wellington, J., Thompson, K. J., Ruggles, S., & Fitch, C. A. (2022). A new strategy for linking U.S. historical censuses: A case study for the IPUMS multigenerational longitudinal panel. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *55*(1), 12–29. https://doi.org/10.1080/01615440.2021.1985027

Henderson, J. V., Regan, T., & Venables, A. J. (2021). Building the City: From Slums to a Modern Metropolis. *The Review of Economic Studies*, *88*(3), 1157–1192. https://doi.org/10.1093/restud/rdaa042

Henderson, J. V., Squires, T., Storeygard, A., & Weil, D. (2018). The Global Distribution of Economic Activity: Nature, History, and the Role of Trade1. *The Quarterly Journal of Economics*, *133*(1), 357–406. https://doi.org/10.1093/qje/qjx030

Henderson, J. V., Storeygard, A., & Deichmann, U. (2017). Has climate change driven urbanization in Africa? *Journal of Development Economics*, *124*, 60–82. https://doi.org/10.1016/j.jdeveco.2016.09.001

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring Economic Growth from Outer Space. *American Economic Review*, *102*(2), 994–1028. https://doi.org/10.1257/aer.102.2.994

Henderson, V., Storeygard, A., & Weil, D. N. (2011). A Bright Idea for Measuring Economic Growth. *American Economic Review*, *101*(3), 194–199. https://doi.org/10.1257/aer.101.3.194

Heris, M. P., Foks, N. L., Bagstad, K. J., Troy, A., & Ancona, Z. H. (2020). A rasterized building footprint dataset for the United States. *Scientific Data*, *7*(1), 207. https://doi.org/10.1038/s41597-020-0542-3

Herzog, I. (2024). The city-wide effects of tolling downtown drivers: Evidence from London's congestion charge. *Journal of Urban Economics*, *144*, 103714. https://doi.org/10.1016/j.jue.2024.103714

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, *349*(6245), 261–266. https://doi.org/10.1126/science.aaa8685

Hodler, R., & Raschky, P. A. (2014). Regional Favoritism. *The Quarterly Journal of Economics*, *129*(2), 995–1033. https://doi.org/10.1093/qje/qju004

Horn, K., & Merante, M. (2017). Is home sharing driving up rents? Evidence from Airbnb in Boston. *Journal of Housing Economics*, *38*, 14–24. https://doi.org/10.1016/j.jhe.2017.08.002

Hornbeck, R. (2012). The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe. *American Economic Review*, *102*(4), 1477–1507. https://doi.org/10.1257/aer.102.4.1477

Hornbeck, R. (2023). Dust Bowl Migrants: Environmental Refugees and Economic Adaptation. *The Journal of Economic History*, *83*(3), 645–675.

Hornbeck, R., & Keniston, D. (2017). Creative Destruction: Barriers to Urban Growth and the Great Boston Fire of 1872. *American Economic Review*, *107*(6), 1365–1398. https://doi.org/10.1257/aer.20141707

Hornbeck, R., Michaels, G., & Rauch, F. (2024). *Identifying Agglomeration Shadows: Long-run Evidence from Ancient Ports* (w32634; p. w32634). National Bureau of Economic Research. https://doi.org/10.3386/w32634

Hornbeck, R., & Rotemberg, M. (2024). Growth Off the Rails: Aggregate Productivity Growth in Distorted Economies. *Journal of Political Economy*, *132*(2).

Hsiao, A. (2023). Sea Level Rise and Urban Adaptation. *Working Paper*.

Hu, Y., & Yao, J. (2022). Illuminating economic growth. *Journal of Econometrics*, *228*(2), 359–378.

Hwang, J., & Sampson, R. J. (2014). Divergent Pathways of Gentrification: Racial Inequality and the Social Order of Renewal in Chicago Neighborhoods. *American Sociological Review*, *79*(4), 726–751. https://doi.org/10.1177/0003122414535774

Iyigun, M., Nunn, N., & Qian, N. (2017). *The Long-Run Effects of Agricultural Productivity on Conflict, 1400-1900* (SSRN Scholarly Paper 3088933). Social Science Research Network. https://doi.org/10.2139/ssrn.3088933

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790–794. https://doi.org/10.1126/science.aaf7894

Jedwab, R., & Storeygard, A. (2022). The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa 1960–2010. *Journal of the European Economic Association*, *20*(1), 1–38. https://doi.org/10.1093/jeea/jvab027

Johnson, A. N. (1928). Maryland Aerial Survey of Highway Traffic Between Baltimore and Washington. *Highway Research Board Proceedings*, *8*, 106–115.

Juhász, R., & Steinwender, C. (2018). Spinning the Web: Codifiability, Information Frictions and Trade. *Working Paper*.

Keegan, J., & Ng, A. (2021). *There's a Multibillion-Dollar Market for Your Phone's Location Data – The Markup*. The Markup. https://themarkup.org/privacy/2021/09/30/theres-a-multibillion-dollar-market-for-your-phones-location-data

Khachiyan, A., Thomas, A., Zhou, H., Hanson, G., Cloninger, A., Rosing, T., & Khandelwal, A. K. (2022). Using Neural Networks to Predict Microspatial Economic Growth. *American Economic Review: Insights*, *4*(4), 491–506. https://doi.org/10.1257/aeri.20210422

Kirchberger, M. (2021). Measuring internal migration. *Regional Science and Urban Economics*, *91*, 103714. https://doi.org/10.1016/j.regsciurbeco.2021.103714

Kleinman, B. (2024). Wage Inequality and the Spatial Expansion of Firms. *Working Paper*.

Kocornik-Mina, A., McDermott, T. K. J., Michaels, G., & Rauch, F. (2020). Flooded Cities. *American Economic Journal: Applied Economics*, *12*(2), 35–66. https://doi.org/10.1257/app.20170066

Koster, H. R. A., van Ommeren, J., & Volkhausen, N. (2021). Short-term rentals and the housing market: Quasi-experimental evidence from Airbnb in Los Angeles. *Journal of Urban Economics*, *124*, 103356. https://doi.org/10.1016/j.jue.2021.103356

Koudijs, P. (2015). Those Who Know Most: Insider Trading in Eighteenth-Century Amsterdam. *Journal of Political Economy*, *123*(6), 1356–1409. https://doi.org/10.1086/683839

Koudijs, P. (2016). The Boats That Did Not Sail: Asset Price Volatility in a Natural Experiment. *The Journal of Finance*, *71*(3), 1185–1226. https://doi.org/10.1111/jofi.12312

Krause, M., & Seidel, A. (2024). Unlocking neighborhood density. *Journal of Urban Economics*, *144*, 103708. https://doi.org/10.1016/j.jue.2024.103708

Kreindler, G. (2024). Peak‐Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications. *Econometrica*, *92*(4), 1233–1268. https://doi.org/10.3982/ECTA18422

Kreindler, G. E., & Miyauchi, Y. (2023). Measuring Commuting and Economic Activity inside Cities with Cell Phone Records. *Review of Economics and Statistics*, *105*(4), 899–909. https://doi.org/10.1162/rest_a_01085

Kreindler, G., Gaduh, A., Graff, T., Hanna, R., & Olken, B. A. (2023). Optimal Public Transportation Networks: Evidence from the World's Largest Bus Rapid Transit System in Jakarta. *NBER Working Paper Series*.

Kuchler, T., Russel, D., & Stroebel, J. (2022). JUE Insight: The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook. *Journal of Urban Economics*, *127*, 103314. https://doi.org/10.1016/j.jue.2020.103314

Kunz, A., & Boehler, W. (2005). HGIS Germany: An information system on German states and territories from 1820 to 1914. *Historical Geography*, *33*, 145–147.

Laohaprapanon, S., & Sood, G. (2018). Predicting Race and Ethnicity From the Sequence of Characters in a Name. *arXiv*. https://doi.org/10.48550/arXiv.1805.02109

Larcom, S., Rauch, F., & Willems, T. (2017). The Benefits of Forced Experimentation: Striking Evidence from the London Underground Network. *The Quarterly Journal of Economics*, *132*(4), 2019–2055. https://doi.org/10.1093/qje/qjx020

Lawrimore, M. A., Sanchez, G. M., Cothron, C., Tulbure, M. G., BenDor, T. K., & Meentemeyer, R. K. (2024). Creating spatially complete zoning maps using machine learning. *Computers, Environment and Urban Systems*, *112*, 102157. https://doi.org/10.1016/j.compenvurbsys.2024.102157

Le, Q. (2024). Network Competition and Exclusive Contracts: Evidence from News Agencies. *Working Paper*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, K. H., & Tan, B. J. (2024a). Urban Transit Infrastructure and Inequality. *The Review of Economics and Statistics*, 1–46. https://doi.org/10.1162/rest_a_01442

Lee, K. H., & Tan, B. J. (2024b). Urban Transit Infrastructure and Inequality. *The Review of Economics and Statistics*, 1–46. https://doi.org/10.1162/rest_a_01442

Lee, S. K. (2022). *When Cities Grow: Urban Planning and Segregation in the Prewar US* [Working Paper].

Lehner, A., & Philippe, D. (2025). A time-varying index for agricultural suitability across Europe from 1500–2000. *Scientific Data*, *12*(1), 101. https://doi.org/10.1038/s41597-024-04194-z

Levine, J. R., & Gershenson, C. (2014). From political to material inequality: Race, immigration, and requests for public goods. *Sociological Forum*, *29*(3), 607–627. https://doi.org/10.1111/socf.12106

Li, T., Barwick, P. J., Deng, Y., Huang, X., & Li, S. (2023). The COVID-19 pandemic and unemployment: Evidence from mobile phone data from China. *Journal of Urban Economics*, *135*, 103543. https://doi.org/10.1016/j.jue.2023.103543

Li, X., Zhou, Y., Zhao, M., & Zhao, X. (2020). A harmonized global nighttime light dataset 1992–2018. *Scientific Data*, *7*(1), 168. https://doi.org/10.1038/s41597-020-0510-y

Lin, Y., Li, J., Porr, A., Logan, G., Xiao, N., & Miller, H. J. (2023). Creating building-level, three-dimensional digital models of historic urban neighborhoods from Sanborn Fire Insurance maps using machine learning. *PLOS ONE*, *18*(6), e0286340. https://doi.org/10.1371/journal.pone.0286340

Lin, Y., McDermott, T. K. J., & Michaels, G. (2024). Cities and the sea level. *Journal of Urban Economics*, *143*, 103685. https://doi.org/10.1016/j.jue.2024.103685

Litvine, A. D., Starzec, A., Younis, R., Faula, Y., Coustaty, M., Shaw-Taylor, L., & Églin, V. (2024). Built-up areas of nineteenth-century Britain. An integrated methodology for extracting high-resolution urban footprints from historical maps. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *57*(1), 1–19. https://doi.org/10.1080/01615440.2024.2324163

Logan, J. R., Jindrich, J., Shin, H., & Zhang, W. (2011). Mapping America in 1880: The Urban Transition Historical GIS Project. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *44*(1), 49–60. https://doi.org/10.1080/01615440.2010.517509

Logan, T. D., & Parman, J. M. (2017a). Segregation and Homeownership in the Early Twentieth Century. *American Economic Review*, *107*(5), 410–414. https://doi.org/10.1257/aer.p20171081

Logan, T. D., & Parman, J. M. (2017b). The National Rise in Residential Segregation. *The Journal of Economic History*, *77*(1), 127–170. https://doi.org/10.1017/S0022050717000079

Long, J., & Siu, H. (2018). Refugees from Dust and Shrinking Land: Tracking the Dust Bowl Migrants. *The Journal of Economic History*, *78*(4), 1001–1033. https://doi.org/10.1017/S0022050718000591

Lyons, R. C., Shertzer, A., Gray, R., & Agorastos, D. N. (2024). The Price of Housing in the United States, 1890-2006. *NBER Working Paper Series*.

Mangrum, D., & Molnar, A. (2020). The marginal congestion of a taxi in New York City. *Working Paper*.

Marx, B., Stoker, T. M., & Suri, T. (2019). There Is No Free House: Ethnic Patronage in a Kenyan Slum. *American Economic Journal: Applied Economics*, *11*(4), 36–70. https://doi.org/10.1257/app.20160484

Mast, E. (2023). JUE Insight: The effect of new market-rate housing construction on the low-income housing market. *Journal of Urban Economics*, *133*, 103383. https://doi.org/10.1016/j.jue.2021.103383

Mattia, A. (2023). The Distribution of Value of Time: An Analysis from Traffic Congestion and Express Lanes. *Working Paper*.

Mayshar, J., Moav, O., & Pascali, L. (2022). The Origin of the State: Land Productivity or Appropriability? *Journal of Political Economy*, *130*(4), 1091–1144. https://doi.org/10.1086/718372

McMillen, D., & McDonald, J. (2002). Land Values In A Newly Zoned City. *The Review of Economics and Statistics*, *84*(1), 62–72.

Michaels, G. (2008). The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System. *The Review of Economics and Statistics*, *90*(4), 683–701.

Michaels, G., Nigmatulina, D., Rauch, F., Regan, T., Baruah, N., & Dahlstrand, A. (2021). Planning Ahead for Better Neighborhoods: Long-Run Evidence from Tanzania. *Journal of Political Economy*, *129*(7), 2112–2156. https://doi.org/10.1086/714119

Michaels, G., & Rauch, F. (2018). Resetting the Urban Network: 117-2012. *The Economic Journal*, *128*(608), 378–412.

Michalopoulos, S., & Papaioannou, E. (2013). Pre-Colonial Ethnic Institutions and Contemporary African Development. *Econometrica*, *81*(1), 113–152.

Milusheva, S., Bjorkegren, D., & Viotti, L. (2021). *Assessing Bias in Smartphone Mobility Estimates in Low Income Countries*. 364–378. https://doi.org/10.1145/3460112.3471968

Mitre-Becerril, D., & MacDonald, J. M. (2024). Does urban development influence crime? Evidence from Philadelphia's new zoning regulations. *Journal of Urban Economics*, *142*(C). https://ideas.repec.org//a/eee/juecon/v142y2024ics0094119024000378.html

Miyauchi, Y., Nakajima, K., & Redding, S. (2021a). The Economics of Spatial Mobility: Theory and Evidence Using Smartphone Data. *NBER Working Paper Series*. https://doi.org/10.3386/w28497

Miyauchi, Y., Nakajima, K., & Redding, S. J. (2021b). Consumption Access and Agglomeration: Evidence from Smartphone Data. *CEPR Discussion Paper*, *DP15839*.

Mleczko, M., & Desmond, M. (2023). Using natural language processing to construct a National Zoning and Land Use Database. *Urban Studies*, *60*(13), 2564–2584. https://doi.org/10.1177/00420980231156352

Modrek, S., Roberts, E., Warren, J. R., & Rehkopf, D. (2022). Long-Term Effects of Local-Area New Deal Work Relief in Childhood on Educational, Economic, and Health Outcomes Over the Life Course: Evidence From the Wisconsin Longitudinal Study. *Demography*, *59*(4), 1489–1516. https://doi.org/10.1215/00703370-10111856

Monte, F., Porcher, C., & Rossi-Hansberg, E. (2023). Remote Work and City Structure. *NBER Working Papers*, Article 2023–016. https://ideas.repec.org//p/hka/wpaper/2023-016.html

Mookerjee, S., & Slichter, D. (2023). Test scores, schools, and the geography of economic opportunity. *Journal of Urban Economics*, *137*, 103589. https://doi.org/10.1016/j.jue.2023.103589

Morten, M., & Oliveira, J. (2024). The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City. *American Economic Journal: Applied Economics*, *16*(2), 389–421. https://doi.org/10.1257/app.20180487

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, *31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

Naik, N., Kominers, S. D., Raskar, R., Glaeser, E. L., & Hidalgo, C. A. (2017). Computer vision uncovers predictors of physical urban change. *Proceedings of the National Academy of Sciences*, *114*(29), 7571–7576. https://doi.org/10.1073/pnas.1619003114

Naik, N., Raskar, R., & Hidalgo, C. A. (2016). Cities Are Physical Too: Using Computer Vision to Measure the Quality and Impact of Urban Appearance. *American Economic Review*, *106*(5), 128–132. https://doi.org/10.1257/aer.p20161030

Nunn, N., & Qian, N. (2011). The Potato's Contribution to Population and Urbanization: Evidence From A Historical Experiment. *The Quarterly Journal of Economics*, *126*(2), 593–650. https://doi.org/10.1093/qje/qjr009

Panigrahi, P. (2022). *Endogenous Spatial Production Networks: Quantitative Implications for Trade and Productivity* [Working Paper]. https://www.cesifo.org/en/publications/2021/working-paper/endogenous-spatial-production-networks-quantitative-implications

Papageorgiou, T. (2022). Occupational Matching and Cities. *American Economic Journal: Macroeconomics*, *14*(3), 82–132. https://doi.org/10.1257/mac.20180122

Pascali, L. (2017). The Wind of Change: Maritime Technology, Trade, and Economic Development. *American Economic Review*, *107*(9), 2821–2854. https://doi.org/10.1257/aer.20140832

Patel, D. (2024). Floods. *Working Paper*.

Pennington, K. (2021). *Does Building New Housing Cause Displacement?: The Supply and Demand Effects of Construction in San Francisco* (SSRN Scholarly Paper 3867764). Social Science Research Network. https://doi.org/10.2139/ssrn.3867764

Perret, J., Gribaudi, M., & Barthelemy, M. (2015). Roads and cities of 18th century France. *Scientific Data*, *2*(1), 150048. https://doi.org/10.1038/sdata.2015.48

Pinkovskiy, M., & Sala-i-Martin, X. (2016). Lights, Camera … Income! Illuminating the National Accounts-Household Surveys Debate. *The Quarterly Journal of Economics*, *131*(2), 579–631. https://doi.org/10.1093/qje/qjw003

Price, J., Buckles, K., Van Leeuwen, J., & Riley, I. (2021). Combining family history and machine learning to link historical records: The Census Tree data set. *Explorations in Economic History*, *80*, 101391. https://doi.org/10.1016/j.eeh.2021.101391

Proctor, J., Carleton, T., & Sum, S. (2023). *Parameter Recovery Using Remotely Sensed Variables* (w30861; p. w30861). National Bureau of Economic Research. https://doi.org/10.3386/w30861

Raster, T. (2024). Breaking the ice: The persistent effects of pioneers on trade relationships. *Working Paper*.

Redding, S., & Sturm, D. (2024). *Neighborhood Effects: Evidence from Wartime Destruction in London* (w32333; p. w32333). National Bureau of Economic Research. https://doi.org/10.3386/w32333

Regan, T., Chiovelli, G., Michalopoulos, S., & Papaioannou, E. (2023). Illuminating Africa? *Working Papers*, Article 2023–11. https://ideas.repec.org//p/gwi/wpaper/2023-11.html

Relihan, L. E. (2024). *Clicks and Bricks: The Complementarity of Online Retail and Urban Services*. https://ssrn.com/abstract=4065950

Riaño, J. F., & Valencia Caicedo, F. (2024). Collateral Damage: The Legacy of the Secret War in Laos. *The Economic Journal*, *134*(661), 2101–2140. https://doi.org/10.1093/ej/ueae004

Rolf, E., Proctor, J., Carleton, T., Bolliger, I., Shankar, V., Ishihara, M., Recht, B., & Hsiang, S. (2021). A generalizable and accessible approach to machine learning with global satellite imagery. *Nature Communications*, *12*(1), 4392. https://doi.org/10.1038/s41467-021-24638-z

Rosaia, N. (2024). *Competing Platforms and Transport Equilibrium*.

Russell, E., & Winkler, L. (2023). Animating History: Digitally Mapping the United States Telegraph System. *Technology and Culture*, *64*(4), 1211–1225. https://doi.org/10.1353/tech.2023.a911001

Saiz, A. (2010). The Geographic Determinants of Housing Supply. *The Quarterly Journal of Economics*, *125*(3), 1253–1296. https://doi.org/10.1162/qjec.2010.125.3.1253

Saiz, A., & Salazar-Miranda, A. (2023). *Understanding Urban Economies, Land Use, and Social Dynamics in the City: Big Data and Measurement*.

Salazar-Miranda, A., Fan, Z., Baick, M. B., Hampton, K. N., Duarte, F., Loo, B. P. Y., Glaeser, E. L., & Ratti, C. (2024). Shifting Patterns of Social Interaction: Exploring the Social Life of Urban Spaces Through A.I. *NBER Working Paper Series*. https://doi.org/10.3386/w33185

Salazar-Miranda, A., Zhang, F., Sun, M., Leoni, P., Duarte, F., & Ratti, C. (2023). Smart curbs: Measuring street activities in real-time using computer vision. *Landscape and Urban Planning*, *234*, 104715. https://doi.org/10.1016/j.landurbplan.2023.104715

Schiff, N., Cosman, J., & Dai, T. (2023). Delivery in the city: Differentiated products competition among New York restaurants. *Journal of Urban Economics*, *134*, 103509. https://doi.org/10.1016/j.jue.2022.103509

Seidel, A. (2019). *A Global Map of Amenities: Public Goods, Ethnic Divisions and Decentralization* (SSRN Scholarly Paper 3457141). Social Science Research Network. https://doi.org/10.2139/ssrn.3457141

Seiferling, I., Naik, N., Ratti, C., & Proulx, R. (2017). Green streets − Quantifying and mapping urban trees with street-level imagery and computer vision. *Landscape and Urban Planning*, *165*, 93–101. https://doi.org/10.1016/j.landurbplan.2017.05.010

Shanks, B. (2021). Land Use Regulations and Housing Development: Evidence from Tax Parcels and Zoning Bylaws in Massachusetts. *Working Paper*.

Shen, L., & Ross, S. (2021). Information value of property description: A Machine learning approach. *Journal of Urban Economics*, *121*, 103299. https://doi.org/10.1016/j.jue.2020.103299

Sherman, L., Proctor, J., Druckenmiller, H., Tapia, H., & Hsiang, S. (2023). *Global High-Resolution Estimates of the United Nations Human Development Index Using Satellite Imagery and Machine-learning* (w31044; p. w31044). National Bureau of Economic Research. https://doi.org/10.3386/w31044

Shertzer, A., Twinam, T., & Walsh, R. P. (2016). Race, Ethnicity, and Discriminatory Zoning. *American Economic Journal: Applied Economics*, *8*(3), 217–246. https://doi.org/10.1257/app.20140430

Shertzer, A., Twinam, T., & Walsh, R. P. (2022). Zoning and segregation in urban economic history. *Regional Science and Urban Economics*, *94*, 103652. https://doi.org/10.1016/j.regsciurbeco.2021.103652

Shertzer, A., Walsh, R. P., & Logan, J. R. (2016). Segregation and neighborhood change in northern cities: New historical GIS data from 1900–1930. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *49*(4), 187–197. https://doi.org/10.1080/01615440.2016.1151393

Siodla, J. (2015). Razing San Francisco: The 1906 disaster as a natural experiment in urban redevelopment. *Journal of Urban Economics*, *89*, 48–61. https://doi.org/10.1016/j.jue.2015.07.001

Sirko, W., Brempong, E. A., Marcos, J. T. C., Annkah, A., Korme, A., Hassen, M. A., Sapkota, K., Shekel, T., Diack, A., Nevo, S., Hickey, J., & Quinn, J. (2023). *High-Resolution Building and Road Detection from Sentinel-2* (arXiv:2310.11622). arXiv. https://doi.org/10.48550/arXiv.2310.11622

Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y. S. E., Dauphin, Y., Keysers, D., Neumann, M., Cisse, M., & Quinn, J. (2021). *Continental-Scale Building Detection from High Resolution Satellite Imagery* (arXiv:2107.12283). arXiv. https://doi.org/10.48550/arXiv.2107.12283

Sood, A., & Ehrman-Solberg, K. (2023). The Long Shadow of Housing Discrimination: Evidence from Racial Covenants. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4606234

Southall, H. (2011). Rebuilding the Great Britain Historical GIS, Part 1: Building an indefinitely scalable statistical database. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *44*(3), 149–159. https://doi.org/10.1080/01615440.2011.589774

Southall, H. (2012). Rebuilding the Great Britain Historical GIS, part 2: A geo-spatial ontology of administrative units. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *45*(3), 119–134. https://doi.org/10.1080/01615440.2012.664101

Southall, H. (2014). Rebuilding the Great Britain Historical GIS, Part 3: Integrating qualitative content for a sense of place. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *47*(1), 31–44. https://doi.org/10.1080/01615440.2013.847774

Śpiewanowski, P., Talavera, O., & Vi, L. (2022). Applications of Web Scraping in Economics and Finance. In P. Śpiewanowski, O. Talavera, & L. Vi, *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190625979.013.652

Steinwender, C. (2018). Real Effects of Information Frictions: When the States and the Kingdom Became United. *American Economic Review*, *108*(3), 657–696. https://doi.org/10.1257/aer.20150681

Storeygard, A. (2016). Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa. *The Review of Economic Studies*, *83*(3), 1263–1295. https://doi.org/10.1093/restud/rdw020

Sullivan, J., & Genadek, K. (2024). *Using the Census Bureau's Master Address File for Migration Research*.

Sun, M. (2016). The Puzzle of Mistaken Millions: The MTA Surcharge and the Surge of Money onto MetroCards. *Working Paper*.

Sun, M. (Aggie), & Mikhed, V. (2023). Do Transit Fares Hurt the Poor? Evidence from Repeated Weekly Pass Commuters in New York City. *Working Paper*.

Takeda, K., & Yamagishi, A. (2024). The economic dynamics of city structure: Evidence from Hiroshima's recovery. *LSE Research Online Documents on Economics*, Article 126823. https://ideas.repec.org//p/ehl/lserod/126823.html

Tang, Y. (2024). To Follow the Crowd? Benefits and Costs of Migrant Networks. *Working Paper*.

Thevenin, T., Mimeur, C., Schwartz, R., & Sapet, L. (2016). Measuring one century of railway accessibility and population change in France. A historical GIS approach. *Journal of Transport Geography*, *56*, 62–76. https://doi.org/10.1016/j.jtrangeo.2016.08.017

Tompsett, A. (2024). Bridges. *Working Paper*.

Twinam, T. (2018). The long-run impact of zoning: Institutional hysteresis and durable capital in Seattle, 1920–2015. *Regional Science and Urban Economics*, *73*, 155–169. https://doi.org/10.1016/j.regsciurbeco.2018.08.004

van Donkelaar, A., Hammer, M. S., Bindle, L., Brauer, M., Brook, J. R., Garay, M. J., Hsu, N. C., Kalashnikova, O. V., Kahn, R. A., Lee, C., Levy, R. C., Lyapustin, A., Sayer, A. M., & Martin, R. V. (2021). Monthly Global Estimates of Fine Particulate Matter and Their Uncertainty. *Environmental Science & Technology*, *55*(22), 15287–15300. https://doi.org/10.1021/acs.est.1c05309

van Donkelaar, A., Martin, R. V., Brauer, M., & Boys, B. L. (2015). Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. *Environmental Health Perspectives*, *123*(2), 135–143. https://doi.org/10.1289/ehp.1408646

Voigtländer, N., & Voth, H.-J. (2012). Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany*. *The Quarterly Journal of Economics*, *127*(3), 1339–1392. https://doi.org/10.1093/qje/qjs019

von Carnap, T., Christian, P., Tompsett, A., & Zwager, A. (2024). Market access and agricultural development: Evidence from remotely-sensed land use patterns at river crossings in Mozambique. *Working Paper*.

Wang, T. (2023). The Electric Telegraph, News Coverage and Political Participation. *NBER Working Paper Series*.

Ward, Z. (2022). Internal Migration, Education, and Intergenerational Mobility: Evidence from American History. *Journal of Human Resources*, *57*(6), 1981–2011. https://doi.org/10.3368/jhr.58.2.0619-10265R2

Whyte, W. H. (2001). *The Social Life of Small Urban Spaces* (8th ed. edition). Project for Public Spaces.

Yang, L., Lin, Y., Wang, J., & Peng, F. (2024). Achieving Air Pollution Control Targets with Technology-Aided Monitoring: Better Enforcement or Localized Efforts? *American Economic Journal: Economic Policy*, *16*(4), 280–315. https://doi.org/10.1257/pol.20220810

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., & Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications*, *11*(1), 2583. https://doi.org/10.1038/s41467-020-16185-w

You, W. (2021). The Economics of Speed: The Electrification of the Streetcar System and the Decline of Mom-and-Pop Stores in Boston, 1885–1905. *American Economic Journal: Applied Economics*, *13*(4), 285–324. https://doi.org/10.1257/app.20180795

Zabel, J., Phaneuf, D. J., & Krause, A. (2023). Introduction to the Special Issue Property Value Analysis using ZTRAX: Applications under the Approaching Sunset. *Journal of Housing Economics*, *62*(C). https://econpapers.repec.org/article/eeejhouse/v_3a62_3ay_3a2023_3ai_3ac_3as1051137723000566.htm

Zaragozí, B., Giménez-Font, P., Belda-Antolí, A., & Ramón-Morte, A. (2019). A graph-based analysis for generating geographical context from a historical cadastre in Spain (17th and 18th centuries). *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *52*(4), 228–243. https://doi.org/10.1080/01615440.2019.1590269

Zeng, J., & Zhou, Q. (2024). Mayors' promotion incentives and subnational-level GDP manipulation. *Journal of Urban Economics*, *143*, 103679. https://doi.org/10.1016/j.jue.2024.103679

Zheng, Q., Seto, K. C., Zhou, Y., You, S., & Weng, Q. (2023). Nighttime light remote sensing for urban applications: Progress, challenges, and prospects. *ISPRS Journal of Photogrammetry and Remote Sensing*, *202*, 125–141. https://doi.org/10.1016/j.isprsjprs.2023.05.028

Zimran, A. (2019). Sample-Selection Bias and Height Trends in the Nineteenth-Century United States. *The Journal of Economic History*, *79*(1), 99–138. https://doi.org/10.1017/S0022050718000694

Zimran, A. (2022). US immigrants' secondary migration and geographic assimilation during the Age of Mass Migration. *Explorations in Economic History*, *85*, 101457. https://doi.org/10.1016/j.eeh.2022.101457

Zingales, L. (2013). Preventing Economists' Capture. In D. Carpenter & D. Moss (Eds.), *Preventing Regulatory Capture: Special Interest Influence and How to Limit It* (pp. 124–151). Cambridge University Press.

Zou, E. Y. (2021). Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality. *American Economic Review*, *111*(7), 2101–2126. https://doi.org/10.1257/aer.20181346