


ORIGINAL ARTICLE

Open Access



# Geodata in labor market research: trends, potentials and perspectives

Kerstin Ostermann<sup>1,2\*</sup> , Johann Eppelsheimer<sup>1</sup>, Nina Gläser<sup>3</sup>, Peter Haller<sup>1</sup> and Martina Oertel<sup>1</sup>

## Abstract

This article shows the potentials of georeferenced data for labor market research. We review developments in the literature and highlight areas that can benefit from exploiting georeferenced data. Moreover, we share our experiences in geocoding administrative employment data including wage and socioeconomic information of almost the entire German workforce between 2000 and 2017. To make the data easily accessible for research, we create 1-square-kilometer grid cells aggregating a rich set of labor market characteristics and sociodemographics of unprecedented spatial precision. These unique data provide detailed insights into inner-city distributions for all German cities with more than 100,000 inhabitants. Accordingly, we provide an extensive series of maps in the Additional file 1 and describe Berlin and Munich in greater detail. The small-scale maps reveal substantial differences in various labor market aspects within and across cities.

**Keywords:** Georeferenced data, Microdata, Register-based data, Urban economics, Regional science, Labor economics, Neighborhood effects, Spatial economics, Segregation

**JEL classification:** J12, J31, R12, O18

## 1 Introduction

Today, individual geopositioning is ubiquitous. We use detailed georeferenced data (henceforth: geodata) to navigate driving routes, track after-work runs, and look up directions to a new restaurant. Companies profit from optimized logistics, agriculture and construction due to detailed information from orbital satellite systems. Whereas processing and utilizing detailed position data are common in many fields such as engineering and business administration, these skills have not been a primary subject in economics and sociology yet.

This article examines the potential of geodata in the social sciences. Moreover, the article presents multi-city evidence on how small-scale geodata can reveal inner-city developments and inequalities that have been hidden by administrative borders so far. The essential

characteristic of geodata is the assignment of each statistical identity to an exact location on the Earth's surface (Goodchild 2013). Currently, most spatial research in economics and sociology uses city district or county aggregates. However, spatially aggregated data face several limitations restricting the investigation of many research questions. In contrast, geodata allow to flexibly scale spatial information independently of administrative boundaries, resulting in three main advantages:

First, greater spatial depth enables the detailed investigation of topics such as segregation (Brakman et al. 2004; Eeckhout et al. 2014; Rosenthal and Strange 2008), neighborhood effects (Schönwälder and Söhn 2009) and mobility (Dauth and Haller 2018, 2020). Second, geodata can serve as a methodological tool. For instance, researchers can use geodata for the sampling of surveys or identifying neighborhood boundaries (Lee et al. 2008; Legewie and Schaeffer 2016), spatial shocks or family relations (Goldschmidt et al. 2017). Third, the potential of enriching existing data with geoinformation opens up possibilities for record linkage, e.g., with smartphone

\*Correspondence: kerstin.ostermann@iab.de

<sup>2</sup> Friedrich-Alexander University Erlangen-Nürnberg (FAU), Findelgasse 7-9, 90402 Nürnberg, Germany

Full list of author information is available at the end of the article

data (Bähr et al. 2018) as well as with genuine spatial data, such as satellite imagery (Henderson et al. 2012) and climate data (Rüttenauer 2018).

This is likely due to the lack of data and the complexity of processing them (Bayer et al. 2014; Vom Berge et al. 2014; Bügelmeier et al. 2015). However, increasing computational capacities and more suitable statistical tools facilitate research on and with geodata. As a result, the number of published studies using geo-data has been rapidly growing and will further increase given the variety of advantages geodata offers.

In this article, we highlight research potentials of geocoded labor market data with descriptive evidence from grid cell data as an example. Moreover, we share our experience in geocoding the employment biographies of almost the entire German workforce between 2000 and 2017. In addition to detailed daily information on employment and unemployment records, the data contain exact coordinates of workplaces and places of residence. This allows us to describe the German labor market with unprecedented spatial precision. Furthermore, this paper illustrates the potential of geodata by visualizing the labor market characteristics of all major German cities, of which two, Berlin and Munich, will be discussed in greater detail. We show that small-scale geodata can reveal substantial differences in fundamental labor market characteristics within and across cities.

This article is organized as follows: In Sect. 2, we review the recent literature, focusing on research that already uses or could benefit from using geodata. Next, in Sect. 3, we share our experiences in geocoding administrative labor market data. In Sect. 4, we provide small-scale descriptions of two large German cities, Berlin and Munich. In the final section, we conclude by identifying potential research areas and questions for the presented data set. Additionally, an extensive online appendix that contains fine-graded maps of labor market characteristics for all German cities with more than 100,000 inhabitants complements this article.

## 2 Potential research topics and trends in the relevant literature

In the following section, we provide a short overview of potential research fields, starting with questions covering larger regional areas and cities before moving towards research on neighborhoods and individual mobility. Although we present each topic separately, there are various dependencies across these research fields.

One of the most popular approaches to derive causal inference are “natural experiments” such as political reforms, mass layoffs or sudden economic or natural developments affecting entire regions (Ager et al. 2020; Ahlfeldt et al. 2015; Desmet and Henderson

2015; Gathmann et al. 2020). Natural experiments are of special interest for labor market research because they allow to rule out spatial sorting (Combes et al. 2008; Haller and Heuermann 2020). Geodata enable researchers to precisely evaluate the effect of regional shocks on individuals, subgroups, or entire local labor markets (Desmet and Henderson 2015; Oakes et al. 2015) with much higher precision than regional aggregates. One example for such an exogenous shock in Germany is the refugee inflow in 2015 and 2016. Using geodata, researchers can track refugee residences and workplaces within cities and can evaluate the integration process in a more detailed way than with regional aggregates. Moreover, flexible scaling enhances the selection of appropriate control regions for matching processes.

As a further large-scale topic, geodata contribute to insights for city and infrastructure planning which is connected to the locational choice for institutions, firms and workers (Duranton et al. 2015; Helsley 2004; Ottaviano and Thisse 2004). To capture metropolitan effects, Lucas and Rossi-Hansberg (2002) propose an equilibrium city model, which operates under the assumption that people live where they work. Using geodata, Dauth and Haller (2018) show that this assumption is—at least for Germany—only partially true. While US cities are mostly monocentric with clear districts for firms, workers and different employment groups, cities in other, e.g., European, countries might be structured differently, which makes it difficult to link them to existing theoretical and empirical models (Ahlfeldt et al. 2015; Dauth and Haller 2020; Duranton and Puga 2015). Tackling this issue, Ahlfeldt et al. (2015) use geodata in a quantitative theoretical model to estimate the dynamics of the internal city structure with heterogeneous centers. They build city “blocks” of 500 square meter grid cells (“grids”) to control for variation in the surroundings. In a second step, they combine their theoretical model with the natural experiment of the fall of the Berlin Wall and use inner-city variation across grids to provide causal evidence.

In addition to regional and city-related topics, geodata offer advantages on a smaller scale, enabling the detailed analysis of neighborhood effects. Although the concept of neighborhoods is quite diverse, research generally distinguishes between residential and workplace neighborhoods. Although research on workplace neighborhoods can considerably profit from the usage of geodata, we will focus on the research potentials for the literature on residential neighborhoods in this article. For the choice of residence, contextual factors such as the social context, quality of life, public goods, and housing costs play an important role (Dustmann et al. 2018; Kang et al. 2020;

Lee et al. 1994). Highlighting the relevance of social networks, Jahn and Neugart (2020) find significant job referral networks in German neighborhoods using geocoded data.

A prominent strand within neighborhood literature is the rise and development of segregation (Mossay and Picard 2019; Reardon and O'Sullivan 2004). Segregated subgroups can arise if characteristics are homogeneous within neighborhoods but heterogeneous between neighborhoods (Bayer et al. 2014; Cutler and Glaeser 1997; Graham 2018; Legewie and Schaeffer 2016; Schelling 1969). Small-scale geodata like grid cells provide a higher resolution for segregation patterns and their effects than county- or district-level data enabling not only a more fine-grained investigation on the base of grid cells but also comparisons between grid cells. For instance, vom Berge et al. (2014) use cross-sectional geocoded German employment data to visualize the distribution of low-income individuals for the German cities Berlin, Hamburg and Munich. Although providing only a snapshot for one year, vom Berge et al. (2014) already highlight that Munich and Berlin differ in their segregation patterns.

To investigate the rise of segregation, research cannot solely focus on a static definition of neighborhoods. Neighborhoods are dynamic environments that change and evolve over time due to exogenous events or selective individual mobility (Feijten and Van Ham 2009; Sharkey and Faber 2014). In general, similar individuals tend to choose neighborhoods with similar characteristics to their own (Durlauf 2004; Feijten and Van Ham 2009; Kremer 1997). Summing this selective residential choice up to a selective subgroup inflow on the aggregate level, neighborhoods might “tip”: The emerging subgroup drives minorities out of the neighborhood, causing endogenous mobility and segregation (Durlauf 2004; Schelling 1969, 1971). Such segregated neighborhoods can cause neighborhood conflicts, especially if neighborhood boundaries are contested (Legewie and Schaeffer 2016). For dynamic analyses of segregation developments, trend- or panel-data are necessary.<sup>1</sup> The investigation of dynamic compositional changes is especially relevant for high-density neighborhoods where housing alternatives are rare, particularly under the assumption that land and its users are heterogeneous (Card et al. 2008; Duranton and Puga 2015; Helsley 2004). As tight living conditions are most evident for larger cities, we focus on those in this article.

In addition to promoting descriptive research on segregation patterns and processes, geodata also offer new possibilities for the causal estimation of neighborhood

effects. As indicated in the beginning of this section, exploiting exogenous events is a popular strategy to account for endogenous neighborhood change (Chetty and Hendren 2018; Rossi-Hansberg et al. 2010). However, such events are rare and often identify local average treatment effects only. The geographically small scale of grid or point data enables other causal estimation techniques based on border distances or grid-cell variation. Exemplifying the potential of small-scale data, Bayer et al. (2008) use block-level variation within a wider neighborhood to estimate the causal effect of neighborhood referrals. Geocoded grid-cell data can easily improve their administrative block approach. Another example is the paper of Breidenbach et al. (2021), who use Berlin grid-cell data to estimate the causal effect of flight noise and proximity to the airport on housing rental prices. In exploiting the unexpected delays of the airport closure of Berlin-Tegel and inner-city variation in the exposure to flight noise, they show that flight noise reduces rental prices of treated neighborhoods by 2 to 5%.

Moreover, geodata measure the effects of geographical distances more precisely than aggregates at higher administrative levels, thereby enhancing the analysis of individual mobility. Although the focus of this article is grid-cell data and inner-city distributions, individual mobility is a field of research with a high potential in the usage of geo-data. A broad body of research literature seeks to explain individual (non-)mobility (Arntz 2005; Chetty and Hendren 2018; Kennan and Walker 2011; Lee et al. 1994; Reichelt and Abraham 2017; Sorenson and Dahl 2016) and *commuting* (Dauth and Haller 2020). However, most of these analyses measure regional mobility as moving from one county or region to another, resulting in a bias for individuals living close to a border or moving within a district (Lee et al. 1994). Using geodata, mobility is now a continuous variable instead of a binary indicator that facilitates advanced estimation methods in mobility research (Dauth and Haller 2020). Currie et al. (2010), e.g., show that the distance to fast food restaurants in miles correlates with the individual's weight gain. Card (1993) uses college proximity as an instrument when examining the returns to schooling among young males in the US. Additionally, geodata researchers can either consider the initial position within an administrative unit explicitly or can neglect it completely.

Taken together, the review demonstrates that geodata improve a wide range of possible research topics and methods. First, geodata enable a more precise measurement of regional shocks and their effects. Second, geodata supersede the reliance on simplified city or neighborhood models without relying on assumptions about the distribution of productivity, income and

<sup>1</sup> Even though we are going to present trend data in this article, the underlying grid cell data provide also information on neighborhood in- and outflows.

socioeconomic characteristics within districts. Third, geodata enhance mobility research opening up a new scope of social science research.

### 3 A case study of geocoding

Even though some studies already use grid cell data to investigate city developments, neighborhood composition or individual mobility (Ahlfeldt et al. 2015; Jahn and Neugart 2020; Vom Berge et al. 2014), there is no available data set containing longitudinal and comprehensive labor market information on grid-cell level for a whole country as Germany. To provide such a data set, we geocoded administrative labor market data from Germany. In the following, we will shortly describe the characteristics of the Integrated Employment Biographies (IEB), the base of the data set used. Moreover, we give insight into the process of geocoding these particular data.

#### 3.1 Introduction to German administrative labor market data

The IEB contain register-based information about individuals who are employed (data available since 1975) or receive benefits according to the German Social Code (SGB). The IEB further include data of individuals searching for a job or receiving vocational guidance (data available since 2000) as clients of the German Federal Employment Agency (BA) or the local job centers. The IEB also contain information on individuals participating in programs of active labor market policies (data available since 2000).<sup>2</sup>

The spatial information in the base IEB was limited to separate units of municipalities and areas referring to administrative offices (“Arbeitsagenturen”) or local job centers. These units are not constant and underlie continuous changes due to fusions of political units or new layouts of local labor markets. Since the late 1990s, the IEB include not only the workplace or the agency that delivers benefits but also the residence of the individuals or the benefit units (“Bedarfsgemeinschaften”). Since 2000, this information has been based on mailing addresses. Time stamps are exact to the day when a new address is registered.

#### 3.2 Geocoding

In the following, we describe the process used to transform mail-exact address data from the IEB into geodata. The characteristic feature of geodata is the efficient storage of address information in points, lines or polygons. Each point contains two dimensions: the longitude on the x-axis and the latitude on the y-axis. Various points

result in lines, and multiple lines lead to a geometric object called a polygon. The latter can be an administrative unit on which data are spatially aggregated. However, independence from these administrative units is the most striking asset of geodata. Therefore, the final geocoded IEB store point data.

In previous years, the Institute for Employment Research (IAB) gained some experience with geocoding data sets: The first attempt was a sample of three due dates in 2007 to 2009 (Scholz et al. 2012), followed by the processing of the address histories of establishments, employees, and clients of job centers for the years 2000–2014 (Dauth and Haller 2018). The last reviewed version from 2019 contains the years 2000 to 2017 and all available address histories, called IEB GEO. This data set is a supplement to the IEB as well as to all other IAB data sets and samples that are connected to the register data, such as the IAB Establishment Panel (EP)<sup>3</sup>, the IAB Job Vacancy Survey (JVS)<sup>4</sup>, the Panel Study “Labour Market and Social Security” (PASS)<sup>5</sup> and the IAB-BAMF-SOEP Survey of Refugees<sup>6</sup>.

The IAB met several challenges to improve the future quality of references and shorten production time before the addresses of the IEB can be transformed to geocodes: One main challenge is that some addresses change over time because of new postcodes and new names of municipalities or streets. The used geocoding tool from infas360<sup>7</sup> refers to one single timestamp, in this case, to the end of 2017. Therefore, some historical information do not match the new notation, leading to inexact georeferences. In this case, we use technical links provided by the statistical Datawarehouse of the IAB. Usually, the Datawarehouse processes addresses into an identifier of a spatial unit, which is the common area of the postcode, community, Federal Employment Agency, and job center (statistical place identifier)<sup>8</sup>. If the units or unit names change, the linking document changes from an address to another statistical place or official name over time. Using this database, we add the new address notations for postcodes and names of municipalities to the pool of all

<sup>3</sup> <https://www.iab.de/en/erhebungen/iab-betriebspanel.aspx>.

<sup>4</sup> <https://www.iab.de/en/befragungen/stellenangebot.aspx>.

<sup>5</sup> [https://fdz.iab.de/en/FDZ\\_Individual\\_Data/PASS.aspx](https://fdz.iab.de/en/FDZ_Individual_Data/PASS.aspx).

<sup>6</sup> [https://fdz.iab.de/en/FDZ\\_Individual\\_Data/iab-bamf-soep/IAB-BAMF-SOEP-SUF1617v1.aspx](https://fdz.iab.de/en/FDZ_Individual_Data/iab-bamf-soep/IAB-BAMF-SOEP-SUF1617v1.aspx).

<sup>7</sup> <https://www.infas360.de/geokodierung/>.

<sup>8</sup> The statistical department of the Federal Employment Agency provides an overview of the different regional classifications <https://statistik.arbeitsagentur.de/DE/Navigation/Grundlagen/Klassifikationen/RegionaleGliederungen/RegionaleGliederungen-Nav.html>; see especially the combinations [https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Regionale-Gliederungen/Generische-Publikationen/Zusammenhang-Gebietsgliederungen.xlsx?\\_\\_blob=publicationFile&v=4](https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Regionale-Gliederungen/Generische-Publikationen/Zusammenhang-Gebietsgliederungen.xlsx?__blob=publicationFile&v=4).

<sup>2</sup> For more detailed information, see Jacobebbinghaus and Seth (2007).



historical addresses. For streets, no links were available until now, so gaps in the exact geocodes remain in this case.

Another issue is the implementation of address histories at different times with different standards. To solve this issue, we create a unique format that conforms with the geocoder tool and separates the house number from the street name. The geocoding tool is less successful in the case of several house numbers for one address (which is quite common for addresses of establishments), prompting the use of only the first number (e.g., instead of “Hauptstraße 100–104”, we refer to “Hauptstraße, 100”). Therefore, the coding quality for these addresses is less exact but without any missing house number information. Especially in the first years of the address histories, the address notation is poor due to shortening, typing or transmission errors. Therefore, we replace common or known notations with new standards. We also detect anonymous addresses such as lock boxes or refuges for battered women and set them to “missing” to protect secure personal information.

To georeference the addresses, we use the commercial tool of infas360. Unfortunately, the matching algorithms are business secrets and are therefore not available for scientific documentation or for developing another data preparation process. However, we derive some major principles and adjusted the processing accordingly. For example, the geocode quality is worse in some cases if postcode and municipality name do not match. Therefore, we geocode cases with minor results a second time without the postcode and include the geocode with the best quality. When the tool returns two codes belonging to different municipalities, we exclude these cases from further processing.

### 3.3 IEB GEO

In total, the address histories used include 420 million data rows with approximately 80 million different address notations. We pool these data as 43 million standardized notations with the geocoder tool returning 19 million geocodes. To keep the processing time manageable, we used two georeferencing processes in parallel. One geocoding passage ultimately lasted three days. The different measures of standardization therefore not only improved the data quality but also shortened the workflow. The quality of georeferences differs among the sources and increases over time. On average, approximately 95% of the geocodes are exact mailing addresses, making a strong base for further analyses.

As a variable of register data, the exact workplace or residence is highly sensitive information in terms of the German General Data Protection Regulation (GDPR). Due to the high sensitivity of the data, the IEB GEO is

not publicly available. Address information in connection with any social security information is highly secured and only available to the geocoding team. The juridical department of the IAB grants restricted access to IAB staff after a detailed description of the project. The IAB follows strict data protection measures as a matter of course.

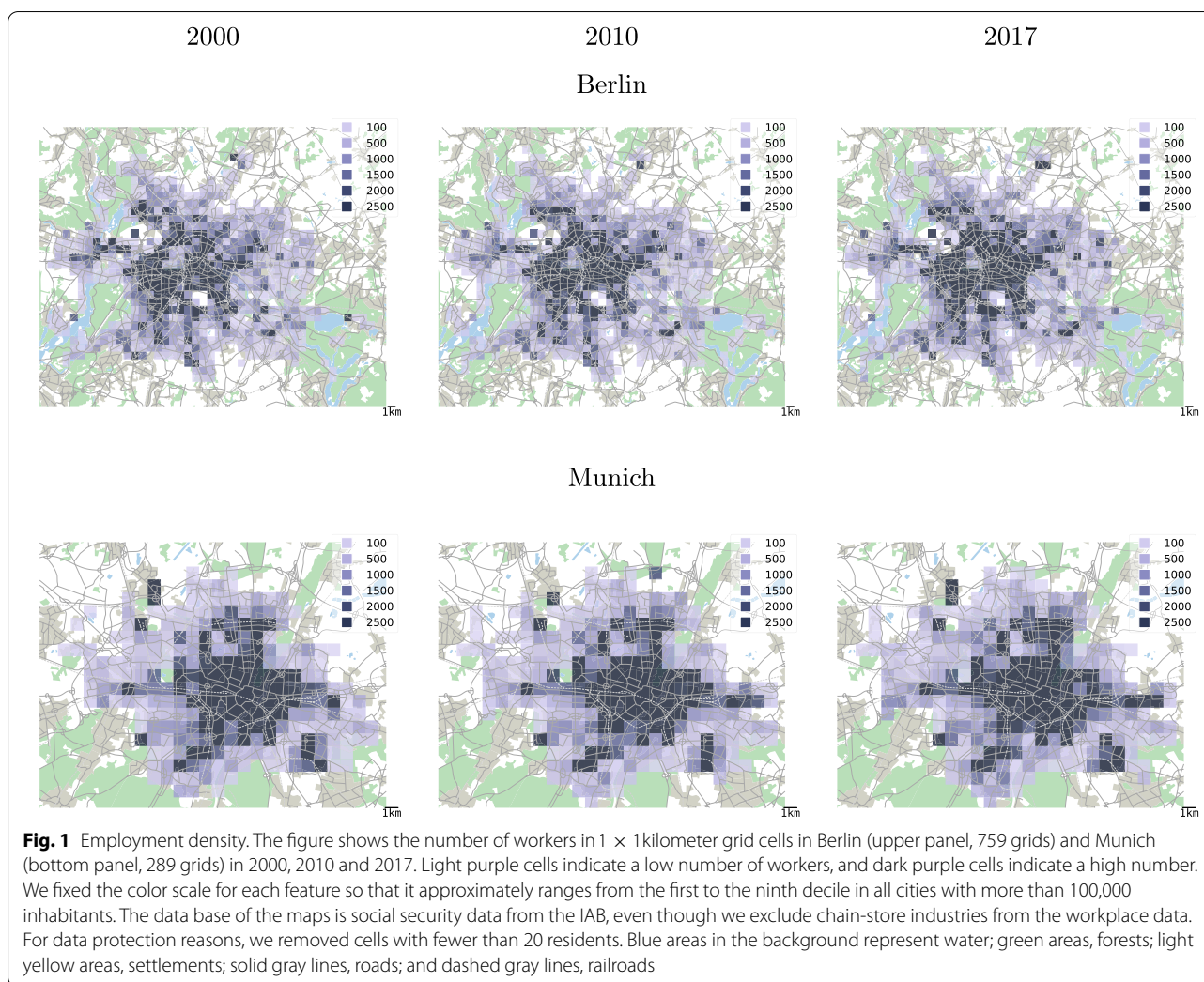
To meet the data protection guidelines, we designed the IEB GEO as a system of several data sets with different sensitivity and access modes: The five histories<sup>9</sup> contain only an anonymous Geo-ID along with anonymized identifiers of persons, establishments or SGB-II-benefit units, begin-/enddate with some variables describing the quality and two markers of moves between addresses. A second data set contains information on the relation between the point-ID and six available anonymous grid-cell-IDs 100 m<sup>2</sup> (100 m<sup>2</sup>, 500 m<sup>2</sup>, and 1000 m<sup>2</sup>-grids in Lambert projection (LAEA) and Universal Transversal Mercator-Projection Zone 32 (UTM32)). Seven separated data sets contain the official codes and two additional projection systems (Gauß-Krüger-Projection and World Geodetic System 1984), and the last data set links the identifiers of the IEB to those of the IEB GEO.

To comply with the GDPR, the design of the IEB GEO is available at different levels of anonymization according to the scientific purpose. For some analyses, anonymous geogrid identifier are sufficient. In other cases, users can compute distances with remote data access. If necessary, users have to apply for geocodes or grid codes in different granularities to combine the IEB GEO with other geodata or points of interest or, as in the example below, to produce maps of labor market characteristics in 1 × 1 kilometer grid cells illustrating the labor market structure of cities.

## 4 Results: labor market characteristics of selected cities

Having explained our experiences with geocoding social security data, the following section shows labor market insights and developments on a fine scale enabling analyses within and irrespective to administrative boundaries. We illustrate the potential of such data by investigating various inner-city labor market characteristics. Based on a series of maps, we describe the spatial distribution of workplaces, residencies, wages, employment types, and skills. All maps are based on the full

<sup>9</sup> Referring to (a) the place of establishments, place of residence of (b) employees, (c) clients of the Federal Employment Agency and d) job center-clients of authorized municipalities that deliver data via the transmission standard XSozial-BA-SGB II, and the place of residence of e) benefit units following §7 SGB II.

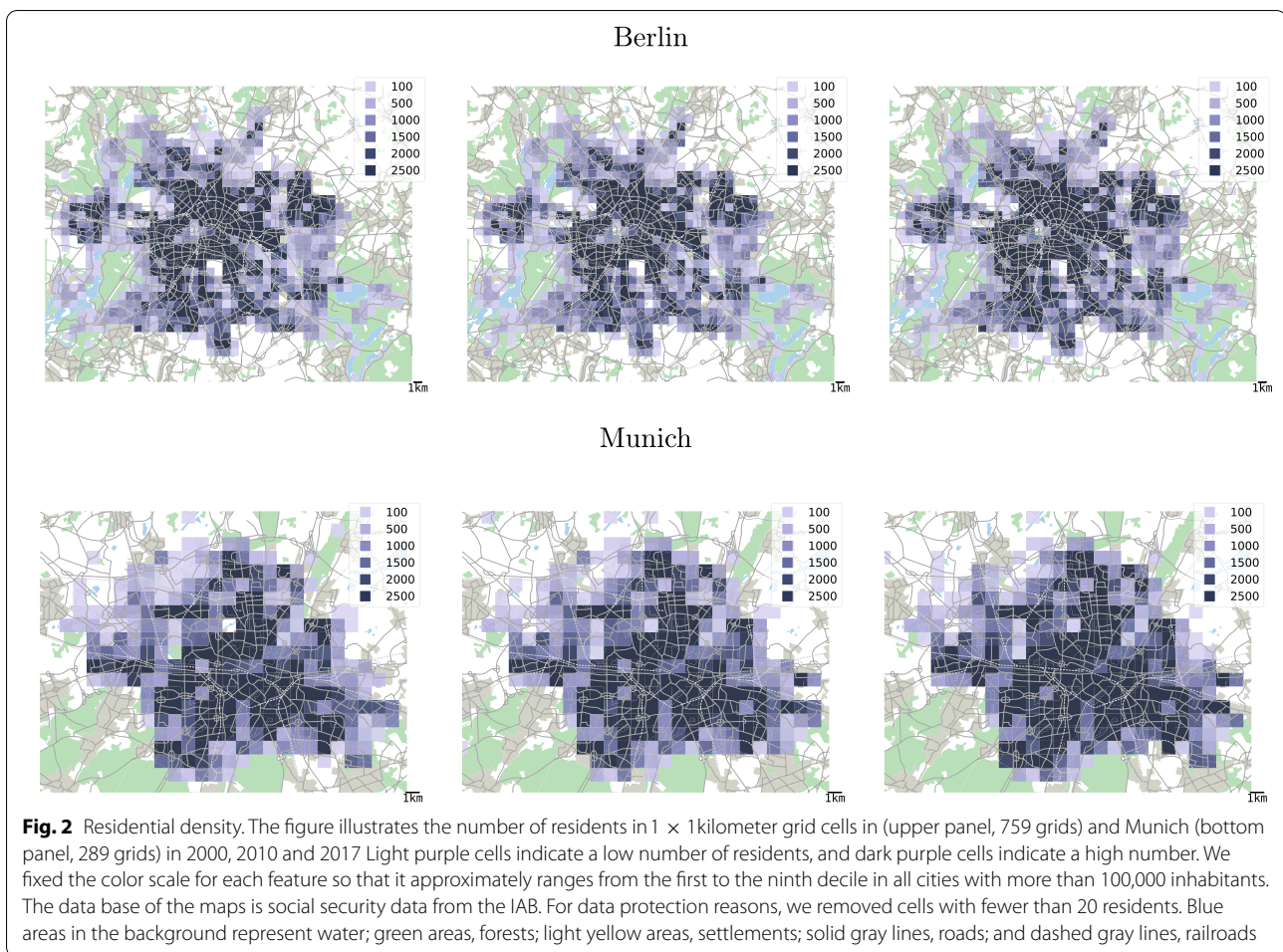


IEB GEO and visualize the distribution of labor market characteristics in  $1 \times 1$  kilometer grid cells.

For data protection reasons, we censored cells with fewer than 20 residents or, in case of the employment density, with fewer than four establishments. We refer readers to the extensive online supplement, which contains more than 2000 maps for all German cities with over 100,000 inhabitants. These maps show that many German cities differ substantially in their shape from a monocentric city structure. The general shape of Düsseldorf, for instance, (pp. 53–55), follows the form of a left-faced arc, whereas the shape of Bremen (pp. 29–31) follows the large river Weser from east to west. However, this study focuses on two of the largest cities in Germany: Berlin and Munich. These cities are interesting subjects because they exhibit diametrically different histories and infrastructure.

#### 4.1 Employment and residential density

Figures 1 and 2 illustrate the employment and residential density in Berlin and Munich. To measure employment density, we count all workers in their workplace grid cell. German firms have to register at least one of their establishments per municipality and industry by law, which makes workplace information highly reliable in general. However, firms that operate several establishments in a municipality within the same industry are only obliged to register one of them. In such cases, it cannot be guaranteed that individuals work in the grid they are registered. To prevent errors, we follow Dauth and Haller (2020) and exclude the following chain-store industries from the workplace data: construction, financial intermediation, public service, retail trade, temporary agency work and transportation. The exclusion of chain store industries leads to slightly underestimated employment densities.



The map for Berlin (Fig. 1, upper panel) indicates a loose employment agglomeration towards the city center in 2017. However, some extensions reach out towards the peripheries highlighting the importance of alternative agglomeration models like the model of Ahlfeldt et al. (2015). Employment density has grown over the years in Berlin and shifted from a slight tendency to the west towards the city center.

In the bottom panel of Fig. 1, the employment density in Munich shows an increasing agglomeration towards the city center. The few extensions in certain regions around the city might be caused by plants of large firms around the belt of Munich.

To measure the residential density, we counted all individuals in their grid of residence. Due to the origin of the data, the data only include individuals in the German social security system, such as employees, registered unemployed individuals, individuals in labor market programs, and recipients of unemployment benefits. Therefore, the data do not provide information

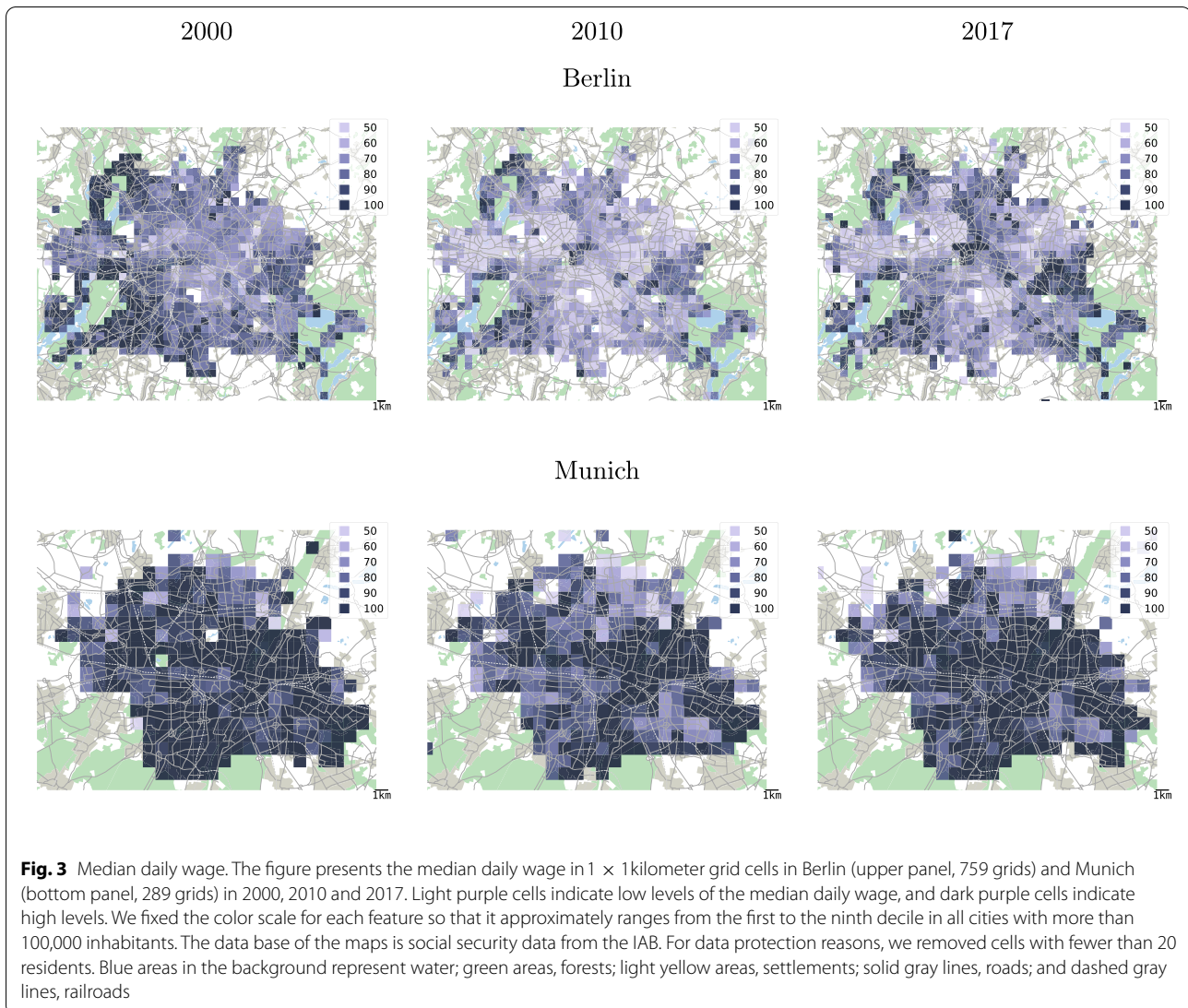
about self-employed individuals, civil servants, students, retirees, pure homemakers or children.

Figure 2 shows the residential density in the two cities. The distribution of residents is scattered over the different districts of Berlin, creating a multicentric cityscape. While still appearing slightly more concentrated in the west, the population density shifted, similar to the employment density, towards the geographical center of Berlin over time.

In Munich, the population density is slightly more concentrated in the southern part of the city. It shows steady growth, exceeding the threshold of 3000 inhabitants in most of the grids in 2017. This high density confirms previous findings, which show that Munich is the city with the highest population density in Germany (Statistisches Bundesamt 2019).

In both of the displayed cities, the employment density shows a radiating pattern that is likely to correlate with the main transportation routes of each city. The residential density seems to be more centered in Munich,





whereas Berlin is more multicentric, showing diversity in districts. Additionally, there seems to be an agglomeration trend over time in employment as well as residential density.<sup>10</sup>

#### 4.2 Wages

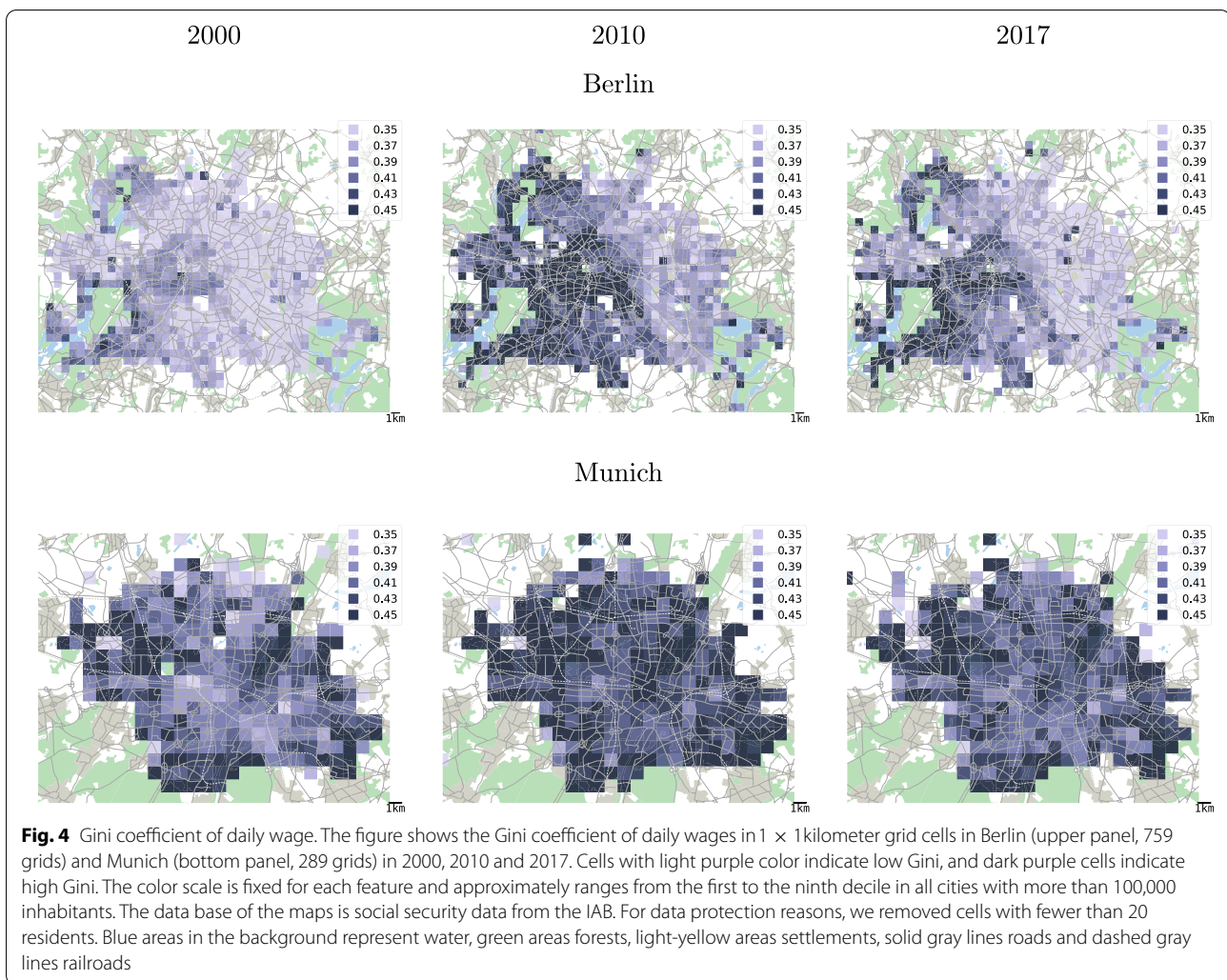
Figures 3 and 4 show the median daily wages of residents and the Gini coefficients in Berlin and Munich. We use both variables as measures for wage segregation and inequality in neighborhoods. The maps for the median daily

wage illustrate between-neighborhood inequality and the Gini coefficient visualizes within-neighborhood inequality. If all wages within a grid cell were equal, the Gini coefficient would be zero. If one inhabitant earns all, the Gini would be equal to 1. The wage information in the register data is highly reliable in general because employers are legally obliged to report wages. However, as typical for social security data, earnings are right-censored at the social security threshold, which affects approximately 10% of the German workforce. We impute top-coded wages using a two-stage procedure similar to Dustmann et al. (2009) and Card et al. (2013) before computing median wages and Gini coefficients.

The concentration of high wages in Berlin (Fig. 3, upper panel) is even more multicentric than the distribution of employment and residential density. In 2017, multiple

<sup>10</sup> Munich and Berlin are only examples of German cities. Other cities show different unusual patterns. For example, the density of residents in Dresden (maps on pp. 47–49 in the online appendix) is shaped as two diagonal lines across the River Elbe rather than a clear city center concentration, giving geographical conditions a decisive role.



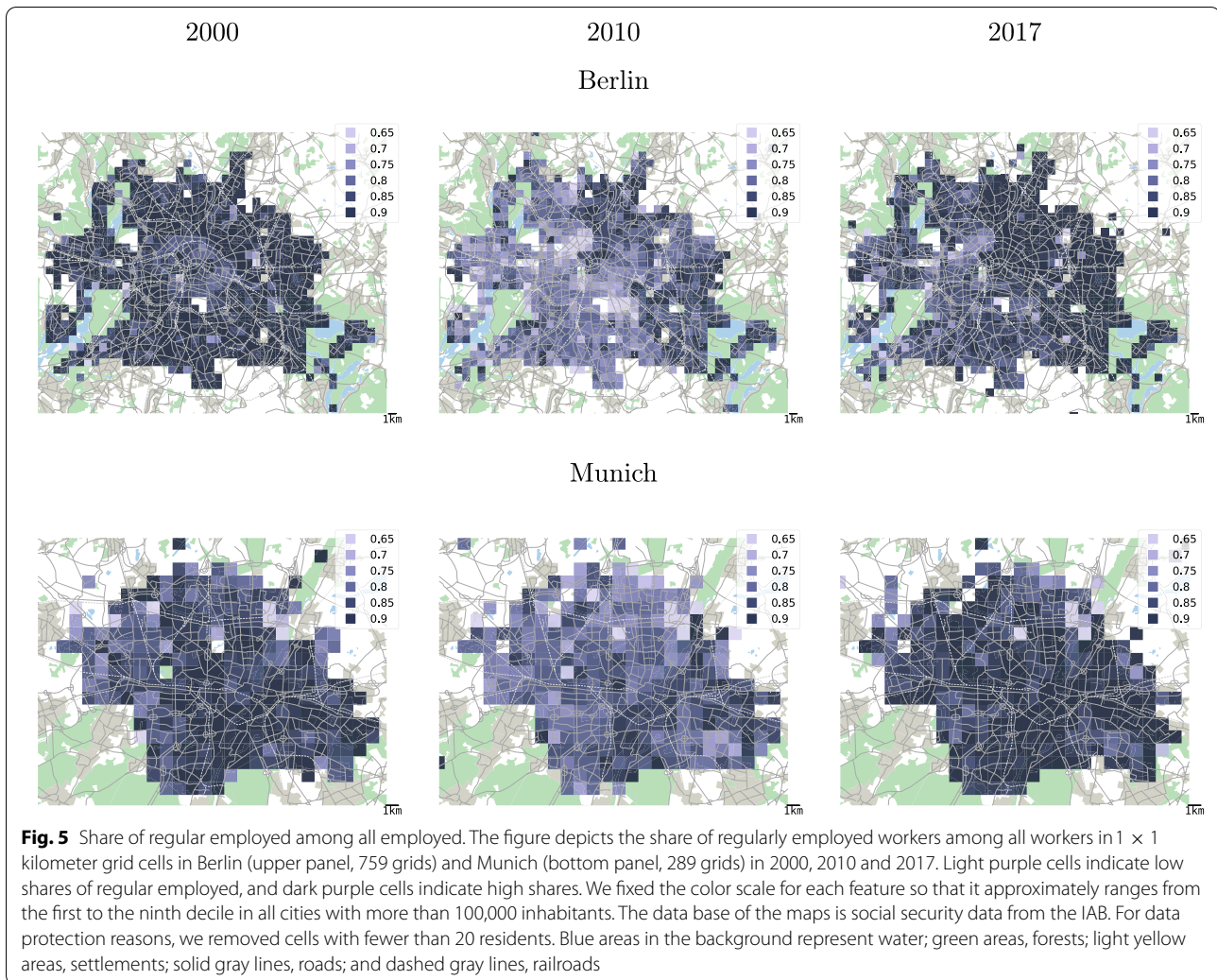


high-wage centers spread across the north, southwest, southeast and the center of Berlin. The median wage is the highest and most equally spread in 2000 before declining and agglomerating over time with no clear visually detectable pattern. Adding a dynamic perspective to the cross-sectional findings of vom Berge et al. (2014), we do see an increasing income segregation within larger neighborhood clusters across the city since 2010.

Munich (Fig. 3, bottom panel) has a persistently high level in the median wages. Slightly smaller median wages are only temporarily evident for 2010. However, the only small percentage of lower median income grids on the periphery in 2017 indicates that the city had recovered from this situation.

The Gini coefficient draws a completely different picture (Fig. 4). In the maps of Berlin (upper panel), the city is clearly divided along the former border of the West and the German Democratic Republic (GDR), with the western part showing noticeably higher inequality

within neighborhoods in 2010 and 2017, with a Gini of over 0.45. This was not always the case: the prevalent segregation occurred sometime between 2000 and 2010, with a sharp incline in inequality between the former West and the former GDR. This pattern and development can have several reasons, ranging from political (the major social security reform in 2005) or economic reasons (global finance crisis in 2008) to segregation processes and private infrastructure investments. As the maps on low-paid workers of vom Berge et al. (2014) do not show such a sharp division along the former border in 2009, the inner-city distribution of low-paid workers does not solely drive this pattern. In fact, the relation of low-paid workers to high-paid workers seems to differ systematically between the former West and the former GDR. A comparison with other German cities of the former GDR indicates that the low Gini coefficient in East Berlin in 2017 might be a feature of East German cities: Although, e.g., Chemnitz (p. 36 in the Additional file 1),



Dresden (p.48 in the Additional file 1), Leipzig (p. 132 in the Additional file 1) and Magdeburg (p.144 in the Additional file 1) show a slightly higher Gini coefficients than East Berlin in 2010, the inequality within neighborhoods is remarkably low in all of those cities in 2017. As we are only providing visual and non-systematic evidence, future research should examine the potential reasons of this specific pattern in East German cities more precisely by using appropriate statistical models and a full observation period of 18 years instead of 3-year snapshots.

Wage inequality in Munich follows the pattern of the median wages, with increasing inequality from 2000 to 2010 and a slight recovery as of 2017 (Fig. 4, bottom panel). However, inequality within neighborhoods is, in contrast to the median wage distribution, higher in certain parts of the city belt.

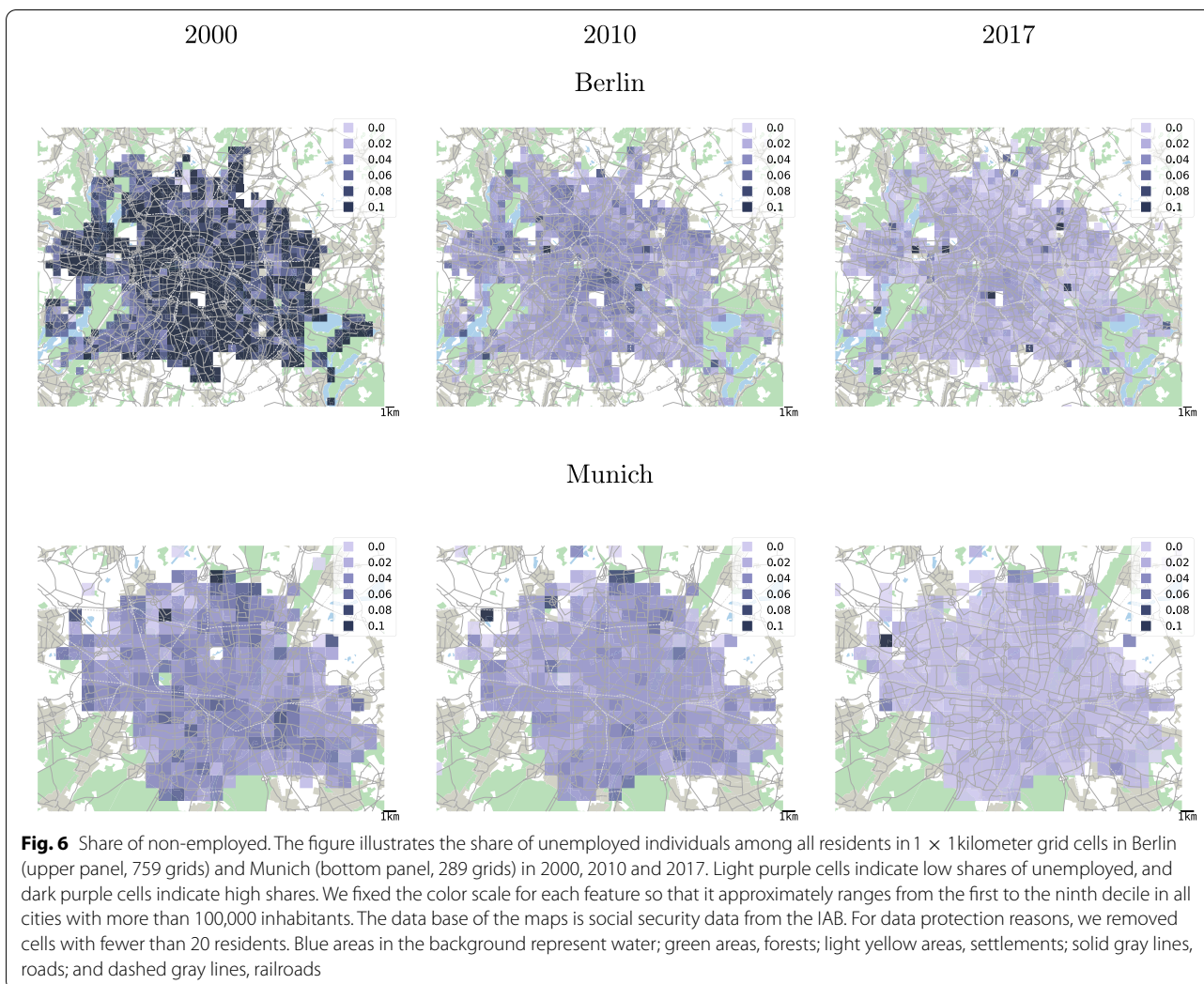
Although the wage inequality for both cities seems to be highest in 2010 indicating a non-linear trend, the inner-city distribution of the wage inequality differs

strongly between the two cities. Berlin has little inequality within neighborhoods in a large part of the city and high inequality in the southwestern part, dividing the city into two parts. In contrast, Munich has a high inequality across large parts of the city. Additionally, median wages are steadily high in Munich indicating low inequality between neighborhoods. Conversely, wages in Berlin are distributed heterogeneously across the city, again creating a multicentric picture of segregated neighborhood clusters. The comparison of the two cities stresses that inequality within and between neighborhoods can differ substantially from each other highlighting the importance of different measures and levels of segregation.

### 4.3 Employment types

This subsection sheds further light on employment and non-employment using the residential information of the IEB GEO. Figure 5 depicts the share of regularly employed individuals who are subject to social insurance





among all employed individuals in Berlin and Munich. Figure 6 displays the share of non-working individuals (henceforth unemployed individuals) among all individuals in our data. We define unemployed individuals as individuals who are registered unemployed, recipients of social security benefits, or those who participate in labor market measurements and do not have a parallel employment spell.

In Berlin (Fig. 5, upper panel), the distribution of regularly employed individuals is relatively even in 2017. However, the division between East and West Berlin is clearly visible, as the eastern area has a higher share of regular employment. The segregation trend is also traceable in the employment status: the equally distributed share of regularly employed individuals in 2000 evolves into a more segregated inner-city distribution in 2010 and 2017.

In Munich (Fig. 5, bottom panel), regularly employed individuals are equally distributed with only a few

exceptions. This image has not changed substantially in recent decades other than a marginal decrease in 2010.

The distribution of unemployment draws a different picture (Fig. 6). Whereas the share of unemployed was generally high in 2000, it decreased in Berlin over the years. It is equally low across entire Berlin in 2017. The same decrease in unemployed individuals applies to Munich but at a different starting level. The share of unemployed individuals is overall low to nonexistent across the entire city and peripheries.

Employment development in both cities shows decreasing unemployment, which is in agreement with the nationally declining number of unemployed individuals in Germany, especially since the social assistance (SGB II) reforms in 2005 (Bundesagentur für Arbeit 2020). The share of unemployed individuals in Berlin is higher than that in Munich. In both cities, unemployment is almost equally distributed, with a few exceptions of high-unemployment grids. Whereas Berlin is more divided into two

areas, the distribution of regular employment relationships in Munich appears to be more equal.

#### 4.4 Skills

A final series of maps illustrates the distribution of high-, medium- and low-skilled residents in Berlin and Munich. In the definition of skill levels, we follow the common classification in labor economics: low-skilled residents are individuals without vocational training, medium-skilled residents are individuals who had completed vocational training, and high-skilled residents are individuals with a degree from a university or university of applied science. Figures 7 and 8 present the geographical distribution of these three groups in Berlin and Munich in 2000, 2010 and 2017.

Berlin (Fig. 7) shows a diverse distribution of skills at first sight. A closer look reveals an agglomeration of high-skilled workers around the center and the southwestern side of the city in 2017. In contrast, a lower share of high-skilled workers reside in the northwestern part where the flight corridor of Berlin-Tegel is located. The lower representation of high-skilled individuals in the northwestern part of the city indicates a correlation between airport noise and skill-level. Using our new grid data on labor market characteristics, researchers can estimate the causal effect of airport noise on labor market outcomes in exploiting the unexpected delays similar to the strategy of Breidenbach et al. (2021) for rental prices.

Strengthening this research potential, areas with a high share of high-skilled residents are the exact areas in which the share of medium-skilled workers is noticeably low. The share of low-skilled workers does not match this segregated picture but has a segregation of its own: It is clearly divided between the former East-West border, but with its highest share in the northwestern part of the city where the flight corridor of the Berlin-Tegel airport is located. While the share and trend of agglomeration of medium- and high-skilled workers increased over the years, the share of low-skilled workers decreased from 2000 to 2017, with lasting East-West segregation.

Munich (Fig. 8), in contrast, again shows less diversity. In 2017, the skill distribution of the entire city has an exhaustive share of at least 35% high-skilled workers. This number increased steadily in size and across the city from 2000 onward, forming the largest skill share in 2017. This trend to a higher share of high-skilled individuals might be driven by a German-wide trend of increasing shares of high-skilled workers over the years. Alternatively, a city-specific reason might be the high rent and cost of living in the city (Kholodilin and Mense 2012). The share of medium-skilled workers in Munich is contrarily small, especially in the city center, matching the findings of Eeckhout et al. (2014, p. 555) that

“large cities disproportionately attract both high- and low skilled workers, while average skills are constant across city size”. The share of low-skilled workers is slightly higher and almost evenly distributed over the city, with a slightly higher concentration on the northeastern side. The shares of medium- and low-skilled workers decline over the years and are substituted by the increasing share of high-skilled individuals.

What strikes attention is that in both cities, despite their distinct differences in structure and centers, high- and medium-skilled individuals are segregated. The residence choice of low-skilled individuals follows a different pattern. We find a similar pattern of residence segregation by skill level for, e.g., Cologne (German “Köln”, pp. 125–127 in Additional file 1) and Leipzig (pp. 131–133 in Additional file 1).

Overall, Munich and Berlin differ from each other in various labor market characteristics. Berlin has a rather multicentric structure, which might be driven by historical reasons or sheer size. Furthermore, many characteristics show a clear East-West division as the former separation of the city seems to still play a decisive role in the agglomeration of the workforce. Munich, alternatively, appears more centered and shows a less diverse picture of labor market characteristics. Having already detected several inner-city patterns in both cities, we also stress the necessity to explain and understand these patterns in using more years and additional data. In this aspect, future research should exploit the possibility of combining these labor market data with other geodata.

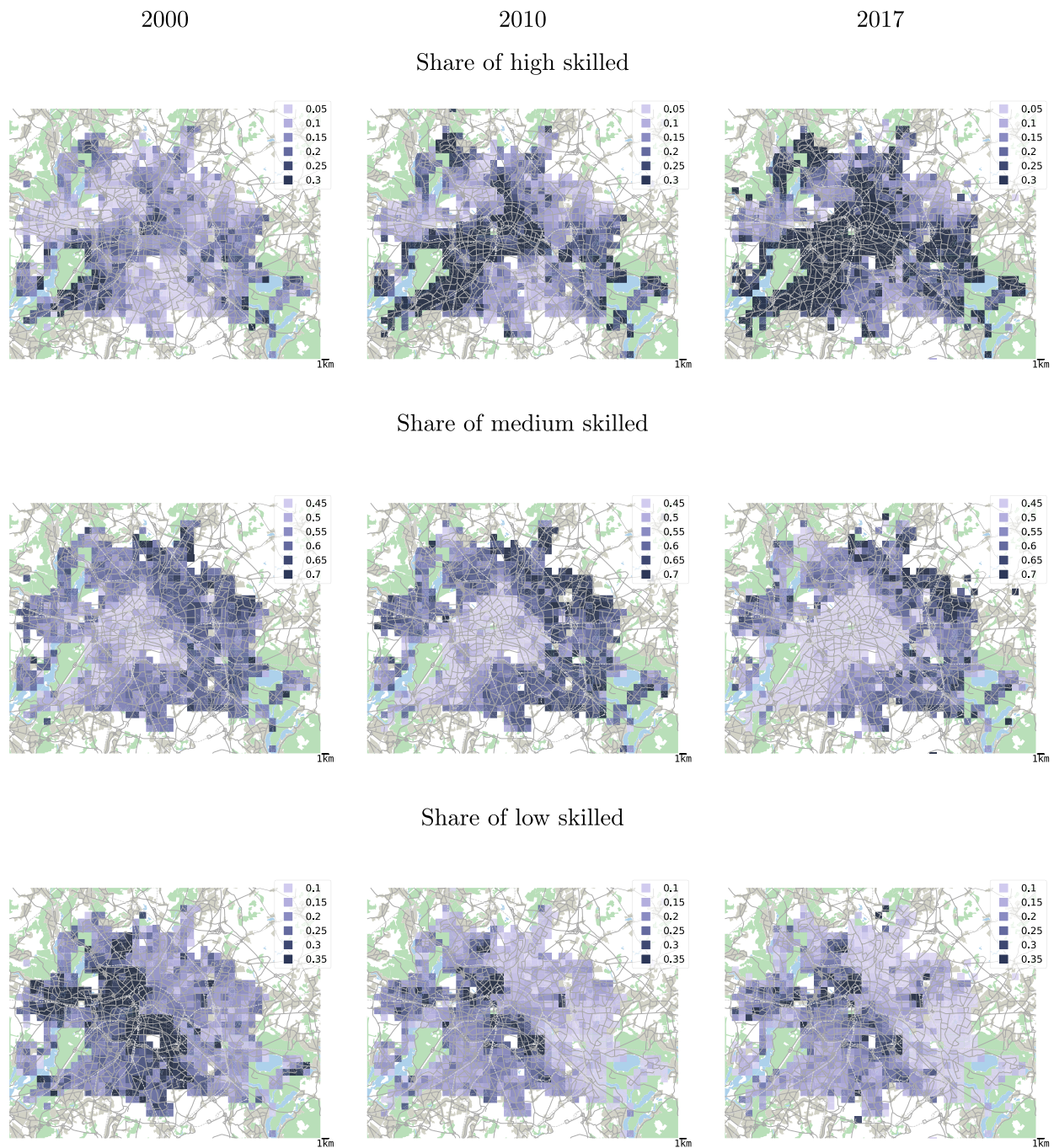
## 5 Discussion and conclusions

Geodata are one of the furthest-reaching developments for regional and urban economics. Nevertheless, the literature that uses geodata is still comparatively small. This article provides an overview of research areas that profit from and already use geocoded data. Geodata enrich analyses on the regional scale and further provide insight into spatial relationships on the city or individual scale.

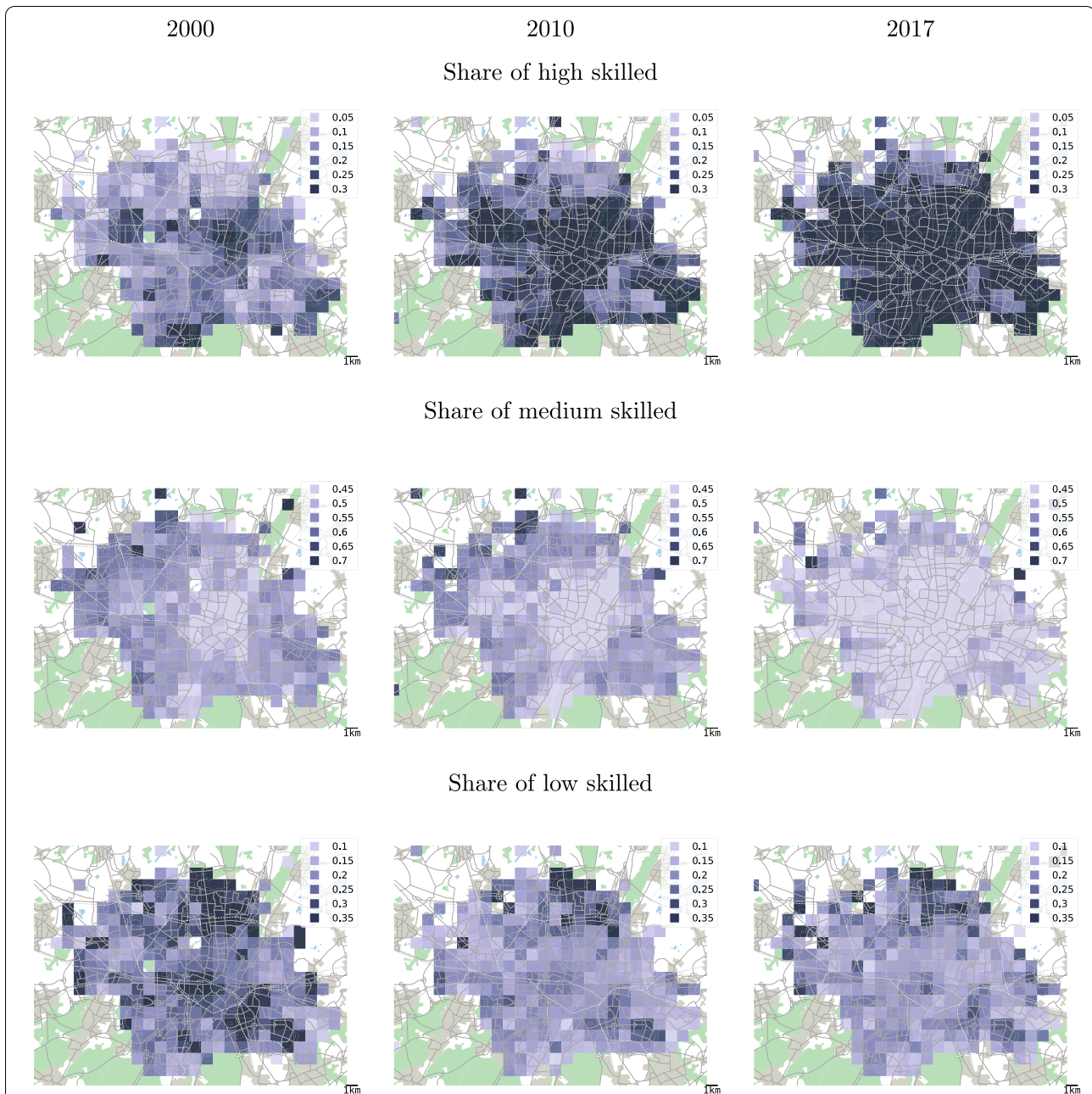
To foster the usage of geodata, we share our experiences in generating and preparing employment and labor market data at the IAB. The resulting data set IEB GEO contains georeferenced and register-based information on all individuals who were subject to the German social security system from 2000 to 2017. These linkable data provide 350 million consolidated episodes with 19 million different geocodes, of which 95% are on the level of exact mailing addresses. The small-scale, rich, and highly reliable information make the IEB GEO a worldwide unique and high-potential data set.

To illustrate the potential of the IEB GEO, the Additional file 1 provides maps of all German cities with more than 100,000 inhabitants. Every map displays the





**Fig. 7** Skills in Berlin. The figure shows the share of high-skilled (top layer), medium-skilled (middle layer) and low-skilled individuals (bottom layer) among all residents in our data in  $1 \times 1$  kilometer grid cells in Berlin (759 grids) in 2000, 2010 and 2017. Light purple cells indicate low shares, and dark purple cells indicate high shares. We fixed the color scale for each feature so that it approximately ranges from the first to the ninth decile in all cities with more than 100,000 inhabitants. The data base of the maps is social security data from the IAB. For data protection reasons, we removed cells with fewer than 20 residents. Blue areas in the background represent water; green areas, forests; light yellow areas, settlements; solid gray lines, roads; and dashed gray lines, railroads



**Fig. 8** Skills in Munich. The figure presents the share of high-skilled (top layer), medium-skilled (middle layer) and low-skilled individuals (bottom layer) among all residents in our data in  $1 \times 1$  kilometer grid cells in Munich (289 grids) in 2000, 2010 and 2017. Light purple cells indicate low shares, and dark purple cells indicate high shares. We fixed the color scale for each feature so that it approximately ranges from the first to the ninth decile in all cities with more than 100,000 inhabitants. The data base of the maps is social security data from the IAB. For data protection reasons, we removed cells with fewer than 20 residents. Blue areas in the background represent water; green areas, forests; light yellow areas, settlements; solid gray lines, roads; and dashed gray lines, railroads

inner-city distribution of one labor market indicator on a  $1 \times 1$  kilometer grid-cell level (e.g., wages, unemployment and skills). This article exemplarily describes the cities Berlin and Munich in greater detail. We observe large differences within and across these two cities in

the employment and resident density, the distribution of wages, employment status and skills. Whereas Berlin shows a multicentric pattern in the median daily wages, the former division of East and West Germany is visible in wage inequality as well as in the share of regularly

employed and low-skilled individuals. In contrast, Munich is more centered and shows a less diverse inner-city distribution. The descriptive results highlight the need for further research using geodata to identify determinants of inner-city developments.

From a broader perspective, many German cities have not developed monocentrically, as traditional city equilibrium models assume. Therefore, we emphasize the importance of alternative theoretical models such as that of Ahlfeldt et al. (2015). Our data at hand allows to identify the dynamics of agglomeration effects with higher temporal frequency. Hence, future research can determine spatial equilibrium models with more precision. In addition, our maps highlight the high prevalence of segregation in Germany. We often find visible patterns of increasing segregation between larger neighborhood clusters by median daily wage especially for cities in the eastern part of Germany like Dresden and Leipzig, or in the Ruhr-region like Bochum and Bottrop. However, we also find examples of decreasing (e.g., Hamburg and Cologne) or constant (e.g., Bonn or Mainz) segregation that underlines the necessity of investigating these different trends over time more comprehensively.

The approach used in this study has some limitations. We only reported exemplary and descriptive evidence for three separate years and two cities. Although we hint at reasons and developments, inference about (causal) relationships of the visualized distributions and their changes over time is beyond the scope of this study. However, the detected patterns and differences within and across the two cities Berlin and Munich provide high-potential starting points for relevant research topics using the full panel data of the IEB GEO.

A rather minor data limitation of the IEB GEO is that it relies on social security data only. Therefore, the IEB GEO provide no information about self-employed, civil servants, students, children or pure homemakers. Future research can partly solve this issue by spatially merging the IEB GEO to other geodata, which combination was previously restricted to the county level for analyses with the IEB.<sup>11</sup>

With data such as the IEB GEO, future research should analyze various topics of social sciences, as the examples in Sects. 2 and 4 have shown. By exploiting the advantages of geodata, research can provide more fine-scaled, causal evidence for the impact of regional shocks on neighborhood effects and individual distance thresholds. Overall, this study shows the potential and perspectives of the usage of geodata enriched by comprehensive

descriptive evidence for all large cities in Germany. By sharing experiences on the implementation and preparation of geodata as well as examples of visualization, we encourage the social sciences community to exploit the potential of these new data.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12651-022-00310-x>.

**Additional file 1.** Online appendix containing maps for all German cities with more than 100,000 inhabitants for the years 2000, 2010 and 2017. The maps visualize the inner-city distribution of the residential density, the employment density, the median wages, the gini-coefficient, the share of regular employed and unemployed as well as the share of low-, medium- and high-skilled residents.

## Acknowledgements

The authors thank two anonymous referees, Philipp Breidenbach, Wolfgang Dauth, Malte Reichelt and Sandra Schaffner for many helpful comments and suggestions. Moreover, we thank Sebastian Bähr and Konstantin Körner for their help in substantially revising the grid cell data. We thank two anonymous referees and the editors of the Journal for Labour Market Research for helpful comments. We also thank Elisabeth Roß, Haika Otholt, Petra Prietz and Barbara Wünsche for excellent legal advice on data privacy.

## Author contributions

All the authors have read and approved the final manuscript.

## Funding

We gratefully acknowledge financial support from the Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz e.V. Competition (K165/2018/Segregation and regional mobility). Kerstin Ostermann acknowledges financial support from the graduate program of the IAB and the Friedrich-Alexander University Erlangen-Nürnberg (GradAB). The funding did not influence the design of the study, analysis, and interpretation of data.

## Availability of data and materials

The datasets analysed during the current study are not publicly available as the authors use administrative data of the Institute for Employment Research. The data are social data with administrative origin which are processed and kept by Institute for Employment Research (IAB) according to Social Code III. There are certain legal restrictions due to the protection of data privacy. The data contain sensitive information and therefore are subject to the confidentiality regulations of the German Social Code (Book I, Section 35, Paragraph 1). The data are held by the IAB (email: [iab@iab.de](mailto:iab@iab.de), phone: +49 911 1790) and are on-site available on reasonable request. The code is available and archived at the Research Data Centre of the IAB; see <https://iab.de/en/daten/replikationen.aspx> for further information. The authors are willing to assist (Kerstin Ostermann, [kerstin.ostermann@iab.de](mailto:kerstin.ostermann@iab.de)).

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests

### Author details

<sup>1</sup>Institute for Employment Research (IAB), Regensburger Str. 100, 90478 Nürnberg, Germany. <sup>2</sup>Friedrich-Alexander University Erlangen-Nürnberg (FAU),

<sup>11</sup> For an overview on geodata for Germany, visit the website of the RWI, <https://www.rwi-essen.de/en/forschung-beratung/weitere/forschungsdaten-zentrum-ruhr/datenangebot>.



Findelgasse 7-9, 90402 Nürnberg, Germany. <sup>3</sup>German Institute for Economic Research (DIW), Mohrenstr. 58, 10117 Berlin, Germany.

Received: 12 October 2021 Accepted: 13 April 2022

Published: 3 June 2022

## References

- Ager, P., Eriksson, K., Hansen, C.W., Lønstrup, L.: How the 1906 San Francisco earthquake shaped economic activity in the American West. *Explor. Econ. Hist.* **77**, 101342 (2020)
- Ahlfeldt, G.M., Redding, S.J., Sturm, D.M., Wolf, N.: The economics of density: evidence from the Berlin Wall. *Econometrica* **83**(6), 2127–2189 (2015)
- Arntz, M.: The geographical mobility of unemployed workers. ZEW-Centre for European Economic Research Discussion Paper 05-034 (2005)
- Bähr, S., Haas, G.-C., Keusch, F., Kreuter, F., Trappmann, M.: IAB-SMART-Studie: Mit dem Smartphone den Arbeitsmarkt erforschen. In IAB-Forum: Das neue Onlinemagazin des Instituts für Arbeitsmarkt- und Berufsforschung, pp. 09–01. IAB (2018)
- Bayer, P., Fang, H., McMillan, R.: Separate when equal? Racial inequality and residential segregation. *J. Urban Econ.* **82**, 32–48 (2014)
- Bayer, P., Ross, S.L., Topa, G.: Place of work and place of residence: informal hiring networks and labor market outcomes. *J. Polit. Econ.* **116**(6), 1150–1196 (2008)
- Brakman, S., Garretsen, H., Schramm, M.: The spatial distribution of wages: estimating the Helpman-Hanson model for Germany. *J. Reg. Sci.* **44**(3), 437–466 (2004)
- Breidenbach, P., Cohen, J., Schaffner, S.: Continuation of air services at Berlin-Tegel and its effects on apartment rental prices. Available at SSRN 3840560 (2021)
- Bügelmeyer, E., Schaffner, S., Schanne, N., Scholz, T.: Das DIW-IAB-RWI-Nachbarschaftspanel: Ein Scientific-Use-File mit lokalen Aggregatdaten und dessen Verknüpfung mit dem deutschen Sozio-ökonomischen Panel. RWI Materialien 97, RWI (2015)
- Bundesagentur für Arbeit: Blickpunkt Arbeitsmarkt: Monatsbericht zum Arbeits- und Ausbildungsmarkt. <https://www.arbeitsagentur.de/datei/ba146273.pdf> (2020)
- Card, D.: Using geographic variation in college proximity to estimate the return to schooling. NBER Working Paper 4483 (1993)
- Card, D., Heining, J., Kline, P.: Workplace heterogeneity and the rise of West German wage inequality. *Q. J. Econ.* **128**(3), 967–1015 (2013)
- Card, D., Mas, A., Rothstein, J.: Tipping and the dynamics of segregation. *Q. J. Econ.* **123**(1), 177–218 (2008)
- Chetty, R., Hendren, N.: The impacts of neighborhoods on intergenerational mobility I: childhood exposure effects. *Q. J. Econ.* **133**(3), 1107–1162 (2018)
- Combes, P.-P., Duranton, G., Gobillon, L.: Spatial wage disparities: sorting matters! *J. Urban Econ.* **63**(2), 723–742 (2008)
- Currie, J., DellaVigna, S., Moretti, E., Pathania, V.: The effect of fast food restaurants on obesity and weight gain. *Am. Econ. J. Econ. Policy* **2**(3), 32–63 (2010)
- Cutler, D.M., Glaeser, E.L.: Are ghettos good or bad? *Q. J. Econ.* **112**(3), 827–872 (1997)
- Dauth, W., Haller, P.: Berufliches Pendeln zwischen Wohn- und Arbeitsort: Klarer Trend zu längeren Pendeldistanzen. IAB-Kurzbericht 10/2018 (2018)
- Dauth, W., Haller, P.: Is there loss aversion in the trade-off between wages and commuting distances? *Reg. Sci. Urban Econ.* **83**, 103527 (2020)
- Desmet, K., Henderson, J.V.: The geography of development within countries. In: Duranton, G., Henderson, V., Strange, W. (eds.) *Handbook of Regional and Urban Economics*, vol. 5, pp. 1457–1517. North-Holland, Amsterdam (2015)
- Duranton, G., Henderson, V., Strange, W.: *Handbook of Regional and Urban Economics*, vol. 5A. North-Holland, Amsterdam (2015)
- Duranton, G., Puga, D.: Urban land use. In: Duranton, G., Henderson, V., Strange, W. (eds.) *Handbook of Regional and Urban Economics*, vol. 5, pp. 467–560. North-Holland, Amsterdam (2015)
- Durlauf, S. N.: Neighborhood effects. In J. V. Henderson and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics*, vol. 4, Chapter 50, pp. 2173–2242. Amsterdam: North-Holland (2004)
- Dustmann, C., Fitzenberger, B., Zimmermann, M.: Housing expenditures and income inequality. ZEW-Centre for European Economic Research Discussion Paper 18-048 (2018)
- Dustmann, C., Ludsteck, J., Schönberg, U.: Revisiting the German wage structure. *Q. J. Econ.* **124**(2), 843–881 (2009)
- Eeckhout, J., Pinheiro, R., Schmidheiny, K.: Spatial sorting. *J. Polit. Econ.* **122**(3), 554–620 (2014)
- Feijten, P., Van Ham, M.: Neighbourhood change... reason to leave? *Urban Stud.* **46**(10), 2103–2122 (2009)
- Gathmann, C., Helm, I., Schönberg, U.: Spillover effects of mass layoffs. *J. Eur. Econ. Assoc.* **18**(1), 427–468 (2020)
- Goldschmidt, D., Klosterhuber, W., Schmieder, J.F.: Identifying couples in administrative data. *J. Labour Market Res.* **50**(1), 29–43 (2017)
- Goodchild, M.F.: The quality of big (geo) data. *Dialogues Hum. Geogr.* **3**(3), 280–284 (2013)
- Graham, B.S.: Identifying and estimating neighborhood effects. *J. Econ. Lit.* **56**(2), 450–500 (2018)
- Haller, P., Heuermann, D.F.: Opportunities and competition in thick labor markets: evidence from plant closures. *J. Reg. Sci.* **60**(2), 273–295 (2020)
- Helsley, R.W.: Urban political economics. In: Henderson, J.V., Thisse, J.-F. (eds.) *Handbook of Regional and Urban Economics*, vol. 4, Chapter 54, pp. 2381–2421. Elsevier, Amsterdam (2004)
- Henderson, J.V., Storeygard, A., Weil, D.N.: Measuring economic growth from outer space. *Am. Econ. Rev.* **102**(2), 994–1028 (2012)
- Jacobebbinghaus, P., Seth, S.: The German integrated employment biographies sample IEBS. *Schmollers Jahrbuch* **127**(2), 335–342 (2007)
- Jahn, E., Neugart, M.: Do neighbors help finding a job? Social networks and labor market outcomes after plant closures. *Labour Econ.* **65**, 101825 (2020)
- Kang, Y., Zhang, F., Peng, W., Gao, S., Rao, J., Duarte, F., Ratti, C.: Understanding house price appreciation using multi-source big geo-data and machine learning. *Land Use Policy* (Online first), 104919 (2020)
- Kennan, J., Walker, J.R.: The effect of expected income on individual migration decisions. *Econometrica* **79**(1), 211–251 (2011)
- Kholodilin, K.A., Mense, A.: German cities to see further rises in housing prices and rents in 2013. *DIW Econ. Bull.* **2**(12), 16–26 (2012)
- Kremer, M.: How much does sorting increase inequality? *Q. J. Econ.* **112**(1), 115–139 (1997)
- Lee, B.A., Oropesa, R.S., Kanan, J.W.: Neighborhood context and residential mobility. *Demography* **31**(2), 249–270 (1994)
- Lee, B.A., Reardon, S.F., Firebaugh, G., Farrell, C.R., Matthews, S.A., O'Sullivan, D.: Beyond the census tract: patterns and determinants of racial segregation at multiple geographic scales. *Am. Sociol. Rev.* **73**(5), 766–791 (2008)
- Legewie, J., Schaeffer, M.: Contested boundaries: explaining where ethnic diversity provokes neighborhood conflict. *Am. J. Sociol.* **122**(1), 125–161 (2016)
- Lucas, R.E., Rossi-Hansberg, E.: On the internal structure of cities. *Econometrica* **70**(4), 1445–1476 (2002)
- Mossay, P., Picard, P.: Spatial segregation and urban structure. *J. Reg. Sci.* **59**(3), 480–507 (2019)
- Oakes, J.M., Andrade, K.E., Biyoow, I.M., Cowan, L.T.: Twenty years of neighborhood effect research: an assessment. *Curr. Epidemiol. Rep.* **2**(1), 80–87 (2015)
- Ottaviano, G., Thisse, J.-F.: Agglomeration and economic geography. In: Henderson, J.V., Thisse, J.-F. (eds.) *Handbook of Regional and Urban Economics*, vol. 4, Chapter 58, pp. 2563–2608. Elsevier, Amsterdam (2004)
- Reardon, S.F., O'Sullivan, D.: Measures of spatial segregation. *Sociol. Methodol.* **34**(1), 121–162 (2004)
- Reichelt, M., Abraham, M.: Occupational and regional mobility as substitutes: a new approach to understanding job changes and wage inequality. *Soc. Forces* **95**(4), 1399–1426 (2017)
- Rosenthal, S.S., Strange, W.C.: The attenuation of human capital spillovers. *J. Urban Econ.* **64**(2), 373–389 (2008)
- Rossi-Hansberg, E., Sarte, P.-D., Owens, R., Ill.: Housing externalities. *J. Polit. Econ.* **118**(3), 485–535 (2010)
- Rüttenauer, T.: Neighbours matter: a nation-wide small-area assessment of environmental inequality in Germany. *Soc. Sci. Res.* **70**, 198–211 (2018)
- Schelling, T.C.: Models of segregation. *Am. Econ. Rev.* **59**(2), 488–493 (1969)
- Schelling, T.C.: Dynamic models of segregation. *J. Math. Sociol.* **1**(2), 143–186 (1971)



- Scholz, T., Rauscher, C., Reiher, J., Bachteler, T.: Geocoding of German administrative data: the case of the Institute for Employment Research. FDZ-Methodenbericht 9 (2012)
- Schönwälder, K., Söhn, J.: Immigrant settlement structures in Germany: general patterns and urban levels of concentration of major groups. *Urban Stud* **46**(7), 1439–1460 (2009)
- Sharkey, P., Faber, J.W.: Where, when, why, and for whom do residential contexts matter? Moving away from the dichotomous understanding of neighborhood effects. *Annu. Rev. Sociol.* **40**, 559–579 (2014)
- Sorenson, O., Dahl, M.S.: Geography, joint choices, and the reproduction of gender inequality. *Am. Sociol. Rev.* **81**(5), 900–920 (2016)
- Statistisches Bundesamt. Alle politisch selbständigen Gemeinden mit ausgewählten Merkmalen am 30.09.2019 (3. Quartal 2019) (2019). <https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/Archiv/GVAuszugQ/AuszugGV3QAktuell.html>
- Vom Berge, P., Schanne, N., Schild, C.-J., Trübswetter, P., Wurdack, A., Petrovic, A.: Eine räumliche Analyse für Deutschland: Wie sich Menschen mit niedrigen Löhnen in Großstädten verteilen. IAB-Kurzbericht 12/2014 (2014)

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---