City Directories as a Source of Historical Microdata: Progress Report

Mathias Bühler
Davide Cantoni
Matthias Weigand*

March 2024

Abstract

This note represents a progress report on the digitization of city directories in Munich, 1845–1914. With information on names, occupations, and residential locations, directories provide a valuable source of microdata especially in a context where individual-level census records did not survive. We construct linked, complete-count data of over one million household-years based on 15 directories. Additionally, we precisely geolocate all historical addresses, drawing on a wide range of supplementary data. We discuss the construction of this dataset and present a novel approach to classify occupational standing based on the noun components of occupation titles. Finally, we show a series of descriptive findings on city growth, spatial inequality and social mobility that shed light on city development in one of the largest and fastest-growing, industrializing urban centers of Central Europe during the 19th and early 20th centuries. We view our approach as a pilot study towards the use of city directories as a source of rich individual-level microdata to study the economic and social history of cities.

*Bühler: Ludwig-Maximilians-Universität Munich. Email: mathias.buehler@econ.lmu.de. Cantoni: Ludwig-Maximilians-Universität Munich, CEPR, and CESifo. Email: davide.cantoni@econ.lmu.de. Weigand: Harvard University. Email: mweigand@g.harvard.edu. We acknowledge financial support from the LMUexcellent Investment Fund and from the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme (grant agreement n. 101044546, "CityRising"). We thank Kalle Kappner for thoughtful feedback. Lara Andermann, Alexander Basov, Beyza Gülmezoğlu, Fewo Nett, Fritz Stern, Alexander Weid, Maximiliane Winter, Elvira Yang, Christoph Ziegler, and Yannick Zurl provided excellent research assistance.

1 Introduction

Individual-level microdata from historical censuses, especially in their "full count" version, have amply proven their value for research in economics.[1] Unfortunately, for many continental European countries such individual-level or household-level historical census data are not available.[2] This is particularly true for the case of Germany, the largest European country as of 1900. Even though both the unified German Empire (after 1871) and the pre-unitary states such as Prussia, Bavaria, or Württemberg (before 1871) held regular censuses collecting increasingly detailed information about its residents, no individual-level records were kept. Enumerators' cards were destroyed after tabulation and publication of printed, aggregate statistics by the statistical offices.[3]

City directories promise to bridge this gap (Albers and Kappner, 2023). Directories have been published regularly, often yearly, for many cities, from small towns to the largest urban centers, in Europe and North America from the 19th century until the mid-20th century. While they do not obtain the comprehensiveness of a national census, they represent a large number of the population: in 1900, 44% of the population in Germany lived in cities with 5,000 inhabitants or more.[4]

Until the arrival of telephone directories, city directories were the primary source to locate and contact a person: to identify residents, city directories generally contained the names of the heads of households, their address, and their occupation. Figure 1 shows two famous examples of residents of Munich at the turn of the 20th century. If you traveled to Munich to meet Hermann Einstein (Albert's father), you would have looked up his address in the city directory. If you were an avid reader of Thomas Mann's novels and wanted to write him a letter, you would have consulted Munich's city directory in your own town's public library or post office.

In this report, we describe the potential of using city directories as a source for quantitative historical social science research. Funded by a grant of the European Research Council, we have digitized a number of directories of the city of Munich between 1845 and 1914, and are exploring

[1] Some recent examples, without claiming completeness, are the works by Abramitzky et al. (2020); Bazzi et al. (2020); Bleakley and Ferrie (2016); Collins and Wanamaker (2014); Derenoncourt (2022); Feigenbaum (2018); Fouka (2020); and Pérez (2019). See also bibliography here: `https://usa.ipums.org/usa/full_count_bibliography.shtml`, last accessed 2023-10-02.

[2] For Britain, historical census microdata for 1851–1911 are provided through the I-CeM project of the University of Essex, `http://icem.data-archive.ac.uk`. For Norway, see Ruggles et al. (2011). However, for most other continental European countries no such complete microdata for scientific use are available. In the case of France, for example, census records are usually kept in the *archives départementales* of each department. Some digitized lists are provided by `filae.com`, a commercial website affiliated to `myheritage.com`.

[3] See Gehrmann (2009) and Michel (1985) for a history of German censuses in the 19th century. Major exceptions are individual-level records of the 1925 census in Württemberg, which were donated to the state archive in Ludwigsburg in 1933 (Wietog, 2001), and the censuses of Schleswig-Holstein under Danish rule, which have been transcribed by the *Arbeits-Gemeinschaft Genealogie Schleswig-Holstein e.V.* and are kept in the Dansk Data Arkiv (`https://www.ddd.dda.dk/ddd-tysk/kiplink1.htm`, last accessed 2023-10-02).

[4] Arguably the largest downside of city directories, relative to census data, is the fact that they typically only report only the head of the household. They thus do not provide information on women (unless they were heads of households, e.g. as widows or if living as singles), children, servants, subtenants and people with no permanent address such as vagrants or day laborers. Albers and Kappner (2023) discuss the potential, but also the downsides and limitations, of city directories as a data source using the example of 1880 Berlin.

the potential applications of this data source. Various studies in economics and economic history have used single city directories as a source of information.[5] In this project, we digitize a large number of directories from the same city, carefully processing occupation and residence information, and highlight possible applications of this data.[6]

2 Munich's City Directories

2.1 Overview

We digitize 15 city directories for Munich before World War I: 1845–1910 in five-year intervals, and 1914. We additionally digitize the years of 1886, 1887, and 1893 to explore matching of households across short time periods (see Section 5).

A number of factors make the city of Munich uniquely suitable for an analysis of its historical directories. The first is its role as a large and rapidly growing, industrializing urban center during the 19th and early 20th centuries. Throughout this time period, Munich was the third-largest city in the area of today's Germany, and its growth rate outpaced that of the higher-ranked Hamburg and Berlin, growing sevenfold from around 90,000 inhabitants in 1845 to 640,000 in 1914.

A second factor is the accuracy and completeness of records: the Munich city directory was published by the police administration of the Bavarian state until 1914.[7] Building on a cooperation with the local city archive, we can supplement this information with rich details on the history of Munich that allows us to precisely geolocate streets that were renamed or changed course, as well as houses that have been torn down, replaced, or filled in.

We obtain scanned copies of the directories from the Bavarian State Library and transcribe the entries with the help of a data entry company. Manual cross-checking of individual directory pages confirms that the transcription error is less than 0.1%.

A typical volume of the city directories contains a series of lists: a list of public offices and important addresses (such as churches, theaters, associations, post offices etc.), a list of businesses, and the list of all residents. We focus on the residents' list, which is sorted alphabetically by last name.[8]

Residents' lists indicate the first and last name of the primary tenant (owner or renter of an apartment), their occupation, and an exact address, which includes the floor within the building.

[5] Examples include Caesmann et al. (2021), Kappner (2021), and Siodla (2021).

[6] A similar project for the city of New York is the NYC Space/Time Directory by the New York Public Library (`http://spacetime.nypl.org/`, last accessed 2023-10-02)

[7] Munich is one of few cities in Germany with a directory published directly from official sources. In most other cases, directories were compiled by private publishers (Zwahr, 1968; see also Albers and Kappner, 2023, for a discussion of the incentives of private publishers). Following World War I, the Munich Chamber of Commerce and Industry took over the publication. It was replaced by a private directory publisher — *Adressbuchverlag Ruf* — after World War II.

[8] From 1875 onwards, directories also contain a second list of residents, sorted by address. When available, we rely on the latter list, since it minimizes transcription errors in addresses.

An excerpt of the 1914 directory, showing the first houses on Frühlingstraße, alongside its digitized table analogue, can be seen in Figure 2. To prepare the entries for analysis, we proceed in three steps: *(i)* we clean names, *(ii)* we standardize and geolocate addresses, and *(iii)* we process occupations by standardizing descriptions and matching them to status levels. We review these steps in more detail below.

2.2 Names

For each address, the entry first lists the owner of the building.[9] The following entries under each address are usually businesses located on the lower floors, such as cafés, shops, or artisan workshops. We remove entries of house owners (unless they also reside at the same address) and business entries, leaving us with a list of the resident households in the building. This results in 1,163,465 household entries across all directories.

We standardize first names of the head of household. First names were often abbreviated: "Jos." instead of Joseph or Josef, "Joh." instead of Johann or Johannes, "Frz." instead of Franz. We design a first name dictionary that maps abbreviations of first names to their complete version. Where an abbreviation is ambiguous, we include all possible options, ranked by likelihood (as defined by the relative frequency of the fully spelled-out first name in the data). We also separate out titles (such as "Dr.", "Prof.", or noble ranks).

2.3 Addresses

We precisely geolocate all 49,732 unique addresses (defined as a combination of a street name and house number) found across the 15 volumes of city directories. To this end, we exploit multiple features of the directory data and additional data sources. We proceed in four automated steps. Each step leaves a handful of observations we cannot handle in an automated fashion; we identify these and resolve them manually.

First, we construct a transition matrix of street names in Munich that records all renaming events. We project all street names to their contemporary name, thus creating a standardized, time-constant street name. We geolocate the minority of streets that do not exist today by hand.[10]

In a second step, we create standardized house numbers, again with contemporary house numbers as benchmarks. House numbers hardly changed over time, even in response to large changes in the cityscape: If a block was torn down, these house numbers were not re-assigned — otherwise, house numbers on the entire street would have been reshuffled. We identify the house numbers

[9]Note that until the 1951 *Wohnungseigentumsgesetz* (Building Ownership Act), fractional ownership of a building (condominium) was not legally possible in Germany. Tenants of apartments were necessarily renters. In most cases, the building owner resided elsewhere.

[10]The primary source for re-naming is `https://stadtgeschichte-muenchen.de/strassen`. To hand-code nonexistent streets or changes in street course, we consult both house-level historical maps and information on the location of streets, as provided in the directories and from the Munich city archive.

that did change, and manually crosswalk them to today's house numbers.[11] Figure 3 shows how Frühlingstraße was renamed to Eduard-Schmid-Straße, but did not change course over time, with house numbers staying the same.

The above approach results in 25,317 unique "standardized" addresses, which we match to locations in a third step. We match 13,588 addresses directly to a contemporary location and its exact coordinates. Another 7,611 addresses without a contemporary counterpart reflect minor changes in the cityscape: individual buildings that were torn down or enlarged. We interpolate these missing addresses based on the historical evidence on the continuity of house numbers.[12] This leaves around 4,118 addresses that are unmatched because of major city restructuring. These are manually geolocated using historical maps.

2.4 Occupations

Across all city directories, we have 166,870 unique occupation titles. We standardize (abbreviated) occupation titles and correct typos, thereby reducing the number of titles to 17,340.

To analyze the socioeconomic composition of the city across space and time in a quantitative way, it is essential to assign a (cardinal or ordinal) rank to each occupation. Our reference data comes from the 1925 census of occupations (*Berufszählung*) of the German Empire. It provides a complete count of 6,573 specific occupation titles, along with a status category (on a scale from 1 to 9), as assessed by the census officials.[13]

We hence rely on an "occupational standing" method: we assign to each occupation, in all years, the occupational status of workers with that occupation in the reference year.[14] The method implicitly assumes that the relative status of different occupations is constant over time, and that there is no heterogeneity across workers in the same occupation. Since occupation titles in the directories are very specific, we consider within-occupation heterogeneity to be minimal; additionally, we confirm that the relative occupation status did not change for the subset of occupations listed in earlier, less comprehensive occupation censuses going back to 1882. We hence consider this approach a useful approximation for quantitative analysis.

[11] To identify house number changes, we exploit that the directories list the exact house numbers at street intersections. Since street intersections did typically not change location, we can identify whether a numbering change occurred. In that case, we use historical maps to characterize the house number changed.

[12] See Appendix A.2 for more details.

[13] The 1925 occupation census set out to record the "status within the occupation". It classifies blue-collar workers (status 1-3), white-collar workers (status 4-6), and managers as well as capital owners (status 7-9). Within the groups, the census distinguishes by substitutability, ranking, for example, technicians (status 2) above unskilled workers (status 1). The census also records information on industry and sector classification at three levels of granularity, as well as an across-industry grouping of similar skill-set occupations. To the best of our knowledge, there are no comparably detailed and comprehensive records of wages in the time period and area of interest. Appendix Figure A.1 shows the exact score groupings.

[14] This is similar to classifications developed by sociologists, such as the Duncan Socioeconomic Index based on the 1950 U.S. census (for a review of these indices, see Hauser and Warren, 1997). A similar approach more geared to historical occupational titles is the HISCO/HISCLASS classification (Van Leeuwen and Maas, 2011).

Predicting occupation scores. Owing to the detailed nature of the 1925 occupation census, we match 69% of inhabitants to an exact occupation. To accurately predict the occupation status of unmatched occupations, we leverage that occupation titles listed in the directories are highly specific. Reflecting word structures of the German language, most are compound nouns, with information embedded in each component. Consider again the example of Frühlingstraße in Figure 2. At house number 1, Anna Königsberger is listed as *Oberstationsmeisterswitwe*, literally translated as "head-station-master-widow", denoting the widow of the head manager of a railroad station. Intuitively, the component "head" should increase the occupational standing relative to the occupation of a station master. Similarly, J. Heigenmooser, a *Magistratsassistent* or "magistrate-assistant", should be of lower rank than a magistrate.

We operationalize these intuitions as follows to match our occupation titles to the 1925 census of occupations. First, we decompose occupation titles from the city directories into component nouns, using a string splitter library to obtain a set of over 200 possible splits for each occupation, and subsequently predict the most likely correct split.[15] Next, we predict the status of an occupation based on its components. We match the (split) occupation titles from the city directories to the components of occupation titles the 1925 census (which we split manually). Based on the set of matches between the two datasets, we then predict a score for each occupation in the directories.[16]

Revisiting the example of the "head-station master-widow", the 1925 occupation census lists station masters with a score of 6 (out of 9). The matching algorithm removes the title of widow, which does not indicate an occupation, and it adjusts for the "head" prefix, so that Anna Königsberger has a status of 7.35. The "magistrate-assistant" J. Heigenmooser has a score of 5.71, lower than magistrates, who receive a score of 7.[17]

3 Descriptive Findings

3.1 Population Growth

Figure 4 shows the growth of Munich over time. The left panel shows shows population counts from official statistics, whereas the right panel shows household entries in the directories. In 1914, 640,000 inhabitants lived in 212,747 households in the city, a considerable increase over the 90,000 inhabitants in 1845.

[15] We rely on the Python package *compound-split*: `https://pypi.org/project/compound-split/`

[16] For more details, refer to Appendix A.3.

[17] Depending on the specific goal of the analysis, we can account for widower or retiree status by adjusting the occupation score according to pension rates at the time. In our main analyses, which focus on occupational outcomes, we do not include these adjustments.

3.2 Occupation Structure

In Figure 5, we examine occupations of households: while the diversity of (cleaned) occupation titles increased with the number of inhabitants (left panel), the distribution of occupation scores exihibits no discernible trends over time (right panel).

3.3 Geographical Growth

We next turn to the spatial extent of the city. Figure 6 illustrates the intensive and extensive margins of city growth: within the area of the city in 1845, density increased over time, with nearly 300 households per hexagonal grid cell of 200 meters width in 1914 (left panel). Additionally, the city grew in size: in the right panel, we plot the area covered by grid cells that contain at least one household over time.

In Figure 7, we illustrate both facts on a map that shows the household density of (populated) 200 meters grid cells in 1845, 1865, 1895, and 1914: the city becomes more densely settled and grows in size.

4 Spatial Inequality

Our data furthermore allows us to inspect the extent and evolution of spatial inequality in Munich. We zoom in on hexagonal grid cells of 50 meters width, and classify a grid cell as "blue collar" if more than 50% of the households in the cell have a below-median predicted occupation score. Figure 8 shows the distribution of blue collar workers across the city. As the city grows, spatial segregation between blue- and white collar grid cells increases.

In Figure 9, we examine the dissimilarity index (Massey and Denton, 1988) as a measure of occupation segregation for each year in our data. Reflecting visual patterns in the maps, measured segregation between blue and white collar households increases over time.[18]

5 Linking individuals across directories

Directories offer only a cross-section of the residents of the city at a point in time. As opposed to census records, they do not contain information that allows to link individuals across generations. But how feasible is it to follow an individual through different years just by observing the entries in the directories, ideally observing their spatial and social mobility?

To illustrate the potential of linking households across directories, we additionally digitize the directories of 1886 and 1887 and trace individuals across the time period 1885–1887. We restrict our analysis to those names (first and last names) that occur only once in the directory, and to exact

[18] Appendix Figure A.2 shows that the pattern is similar across alternative segregation indices: information theory (Theil, 1967) and neighbor-based segregation (Logan and Parman, 2017).

name matches. For example, the 1885 directory lists 45,618 households with unique names, which are 71% of all households in that year. Out of those, we observe 31,647 households in each year 1885–1887.

In this linked sample, 27.7% of households changed residential location within the city between 1885 and 1887. Those that changed location were 8.9 percentage points more likely to upgrade occupation status than those who stayed, compared to a baseline likelihood of 4.3%. They also moved closer to the city center by 107 meters on average, roughly 10% of the mean distance of households to the city center in 1885.

References

Abramitzky, Ran, Leah Boustan, and Katherine Eriksson, "Do Immigrants Assimilate More Slowly Today Than in the Past?," *American Economic Review: Insights*, March 2020, *2* (1), 125–41.

Albers, Thilo N.H. and Kalle Kappner, "Perks and Pitfalls of City Directories as a Micro-Geographic Data Source," *Explorations in Economic History*, 2023, *87*, 101476.

Bazzi, Samuel, Martin Fiszbein, and Mesay Gebresilasse, "Frontier Culture: The Roots and Persistence of "Rugged Individualism" in the United States," *Econometrica*, 2020, *88* (6), 2329–2368.

Bleakley, Hoyt and Joseph Ferrie, "Shocking Behavior: Random Wealth in Antebellum Georgia and Human Capital Across Generations," *Quarterly Journal of Economics*, 2016, *131* (3), 1455–1495.

Caesmann, Marcel, Bruno Caprettini, Hans-Joachim Voth, and David Yanagizawa-Drott, "Going Viral: Propaganda, Persuasion and Polarization in 1932 Hamburg," *CEPR Discussion Paper Series*, 2021, (DP16356).

Collins, William J. and Marianne H. Wanamaker, "Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data," *American Economic Journal: Applied Economics*, 2014, *6* (1), 220–252.

Derenoncourt, Ellora, "Can You Move to Opportunity? Evidence from the Great Migration," *American Economic Review*, February 2022, *112* (2), 369–408.

Feigenbaum, James J., "Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940," *Economic Journal*, 2018, *128* (612), F446–F481.

Fouka, Vasiliki, "Backlash: The Unintended Effects of Language Prohibition in U.S Schools after Word War 1," *Review of Economic Studies*, May 2020, *87* (1), 204–239.

Gehrmann, Rolf, "German Census-Taking Before 1871," *Max Planck Institute for Demographic Research Laboratory of Historical Demography Working Paper*, August 2009, *023.*

Hauser, Robert M. and John Robert Warren, "Socioeconomic Indexes for Occupations: A Review, Update, and Critique," *Sociological Methodology*, 1997, *27* (1), 177–298.

Kappner, Kalle, "Dense, Diverse and Healthy? Mixed-Income Housing and the Spread of Urban Epidemics," *Mimeo*, 2021.

Logan, Trevon D and John M Parman, "The National Rise in Residential Segregation," *Journal of Economic History*, 2017, *77* (1), 127–170.

Massey, Douglas S and Nancy A Denton, "The Dimensions of Residential Segregation," *Social forces*, 1988, *67* (2), 281–315.

Michel, Harald, "Volkszählungen in Deutschland. Die Erfassung des Bevölkerungsstandes von 1816 bis 1933," *Jahrbuch für Wirtschaftsgeschichte*, 1985, (2), 79–91.

Pérez, Santiago, "Intergenerational Occupational Mobility across Three Continents," *Journal of Economic History*, 2019, *79* (2), 383–416.

Ruggles, Steven, Evan Roberts, Sula Sarkar, and Matthew Sobek, "The North Atlantic Population Project: Progress and Prospects," *Historical methods*, 2011, *44* (1), 1–6.

Siodla, James, "Firms, fires, and firebreaks: The impact of the 1906 San Francisco disaster on business agglomeration," *Regional Science and Urban Economics*, 2021, *88*, 103659.

Theil, Henri, "Economics and Information Theory," *Studies in Mathematical and Managerial Economics*, 1967, *7*.

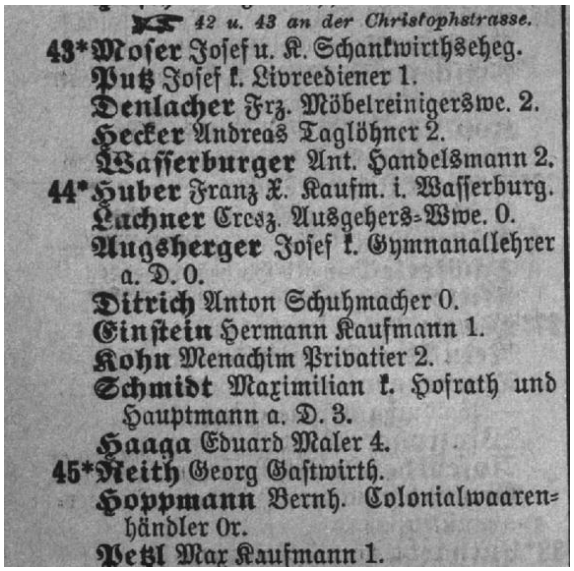
Van Leeuwen, Marco H. D. and Ineke Maas, *HISCLASS: A Historical International Social Class Scheme*, Leuven University Press, 2011.

Wietog, Jutta, *Volkszählungen unter dem Nationalsozialismus: Eine Dokumentation zur Bevölkerungsstatistik im Dritten Reich.*, Berlin: Duncker & Humblot, 2001.

Zwahr, Hartmut, "Das deutsche Stadtadreßbuch als orts- und sozialgeschichtliche Quelle," *Jahrbuch für Regionalgeschichte*, 1968, *3*, 204–229.

Figures and Tables

Hermann Einstein
(Kanalstraße 44, Floor 1)



42 u. 43 an der Christophstrasse.
43* Moser Josef u. K. Schankwirthsehefg.
Putz Josef k. Livreediener 1.
Denlacher Frz. Möbelreinigerswe. 2.
Hecker Andreas Taglöhner 2.
Wasserburger Ant. Handelsmann 2.
44* Huber Franz X. Kaufm. i. Wasserburg.
Lachner Cresz. Ausgehers-Wwe. 0.
Augsberger Josef k. Gymnanallehrer a. D. 0.
Ditrich Anton Schuhmacher 0.
Einstein Hermann Kaufmann 1.
Kohn Menachim Privatier 2.
Schmidt Maximilian k. Hofrath und Hauptmann a. D. 3.
Haaga Eduard Maler 4.
45* Reith Georg Gastwirth.
Hoppmann Bernh. Colonialwaarenhändler Or.
Petzl Max Kaufmann 1.

Thomas Mann
(Ainmillerstraße 31, Floor 3)

31* Trump Friedrich Baumeister
Popp Franz Architekt 0
Neumann Max Architekt 0
Ribbentrop Konstantin v. K. preuß. Leutnant d. L. 0
Reichert Hippolyt K. Major 1
Heine Rudolf K. pr. Zollrevisions-Inspektor a. D. 1
Holzschuher Hans Psychol. u. Schriftsteller 1
Teschner Elisab. Privatiere 1
Traut Heinrich Photograph 2
Elger Rudolf Architekt 2
Verstl Richard K. Leutnant 3
Huber Max Oberlehrer 3
Mann Thomas Schriftsteller 3
Biebel Anton Büreaudiener 4

Figure 1: Munich Residents at the Turn of the 20th Century

Note Figure shows the residential location famous Munich residents as listed in the Munich directories. Left panel shows Hermann Einstein, taken from p. 188 of the 1893 Munich directory. Right panel shows Thomas Mann, taken from p. 663 of the 1905 Munich directory.

* **Frühlingstraße.** SO

1—25 **7** 26—32 **9**

P.A. III — 16. Bez. — St.A. II

K.Pf. Au, 26—32 Giesing, Pr.Pf.St.Joh. K.Sch. 1 m. 25 Mariahilfplatz, 26 m. 32 Columbusstr., Pr.Sch. 1 m. 25 Herrnstr., 26 m. 32 Herzog Wilhelmstr.

(Ihrer herrlichen landschaftlichen Lage nach mit dem Frühlinge vergleichbar.)

Zieht am rechten Isarufer flußaufwärts von der Zeppelinstraße zur Corneliusbrücke, an der Mariahilfstraße, am Bereiteranger, der Ohlmüller-, Boos-, Albani- und Schlotthauerstraße vorüber, zur Sommerstraße.

☛ *5 an der Mariahilfstrasse.*

1***Kreisgemeinde von Oberbayern**
Kreislehrerinnen-Bildungsanstalt für Oberbayern mit Seminar-übungsschule, Präparandinnenschule und Seminar
Dietsch Georg Diener 0
Königsberger Anna Oberstations-meisterswe. 0
Heigenmooser J. Seminardirektor 1
— Magistratsassistent 1
3***Dosch** Gg. Baumeister
Café-Restaurant Cornelius 0
Knappich Alois Restaurateur 0
Böglmüller Gangulf Hausmeister 0
Boswau & Knauer G. m. b. H. Hoch- und Tiefbaugeschäft 1
Stellwag Robert Bankbeamter 1
Langermaier Johanna Kaufm.-We. 1
Diamand Markus Kaufmann 3
— Esther Kleidergeschäftsinhaberin 3
Neuhütl Art. Bankob.-Beamt.a.D. 3
Gaßmann Arthur Kaufmann 4
Lengauer Adelh. Schulverweserin 4

Harrer Anna Museumsdienerswe. 1
Reim Betty Sekretärswe. 1
— Franz Magistratsfunktionär 1
Meyer Oswald Rechtsprakt. 1
Scherm Joh. pens. Bez.-Amtsdien. 2
Wachter Frieda Hilfslehrerin 2
Färber Karl Kaufmann 3
Schreiber Rudolf Buchhalter 3

Rückgebäude.

Heßdörfer Joseph Schreiner 0
Bachmeier Georg Schreiner 1
Hauser Peter Taglöhner 1
Gschwendtner Ther. Privat.-We. 1
8***Beyer** Anna Kaufmannswe.
Frank Babette Kaufmannswe. 0
Kremer Gustav Werkmeister 0
Hitzelberger Wendelin Lohndiener 0
Vodvarsky Elise Goldarbeiterswe. 1
Mainer Karl Rechtsprakt. 1
Tafelmeier Ludw. Photograph 1
Divora Joseph Sekretär 2
Huber Therese Wagenwärterswe. 2
— Xaver Mechaniker 2
Rothmeier Georg Schauspieler 2
Winterstein Georg Schriftsetzer 3
Steinmüller Frdr. Wagenbauer 3
Pischl Michael Buchdrucker 3
9***Lang** Quirin Gastwirt 0
***Lang** Ther. dessen Gattin 0
Gasthaus zur Anlage 0
Kühlbeck Georg Maschinist 0
Fottner Sebastian Handelsmann 0
— Sebastian Eisendreher 0
Berkmann Hugo Stukkateur 1
Hasch Georg Wildbrethändler 1
Gaißer Karl Oberexpeditor a. D. 2
Pongratz Ludw. Notariatsbuchhalt. 2
Huber Ludwig Schneidermeister 3

street	number	house_owner	floor	last_name	first_name	occupation
Frühlingstraße	1	1		Kreisgemeinde von ...		
Frühlingstraße	1	0	0	Dietsch	Georg	Diener
Frühlingstraße	1	0	0	Königsberger	Anna	Oberstationsmeisterswe.
Frühlingstraße	1	0	1	Heigenmooser	J.	Seminardirektor
Frühlingstraße	1	0	1	Heigenmooser		Magistratsassistent
Frühlingstraße	3	1		Dosch	Gg.	Baumeister
Frühlingstraße	3	0	0	Café-Restaurant		
Frühlingstraße	3	0	0	Knappich	Alois	Restaurateur
Frühlingstraße	3	0	0	Böglmüller	Gangulf	Hausmeister
Frühlingstraße	3	0	1	Boswau & Knauer ...		
Frühlingstraße	3	0	1	Stellwag	Robert	Bankbeamter
Frühlingstraße	3	0	1	Langermaier	Johanna	Kaufm. We.
Frühlingstraße	3	0	3	Diamand	Markus	Kaufmann
Frühlingstraße	3	0	3	Diamand	Esther	Kleidergeschäftsinhaberin
Frühlingstraße	3	0	3	Neuhütl	Art.	Bankob.-Beamt.a.D.
Frühlingstraße	3	0	4	Gaßmann	Arthur	Kaufmann

Figure 2: Raw and Digitized Directory Excerpt

Note Frühlingstraße on p. 1034 of the 1914 Munich directory. Note that we abbreviate long business names for display purposes.

Straßenbenennungen

Straße	von	Grund	bis	Grund
Frühlingstraße	3.1874	Erstnennung		Aufhebung
Eduard-Schmid-Straße	1946	Umbenennung		

Figure 3: Frühlingstraße (Today: Eduard-Schmid-Straße) in 1909 and Today

Note Top panel shows data entry on the renaming of Frühlingstraße to Eduard-Schmid-Straße in 1946. Middle panel shows Frühlingstraße in 1909 and today. Bottom Panel shows continuity of house numbers by zooming in on a street segment. Data sources: Stadtgeschichte, Stadtatlas, and OpenStreetMaps.

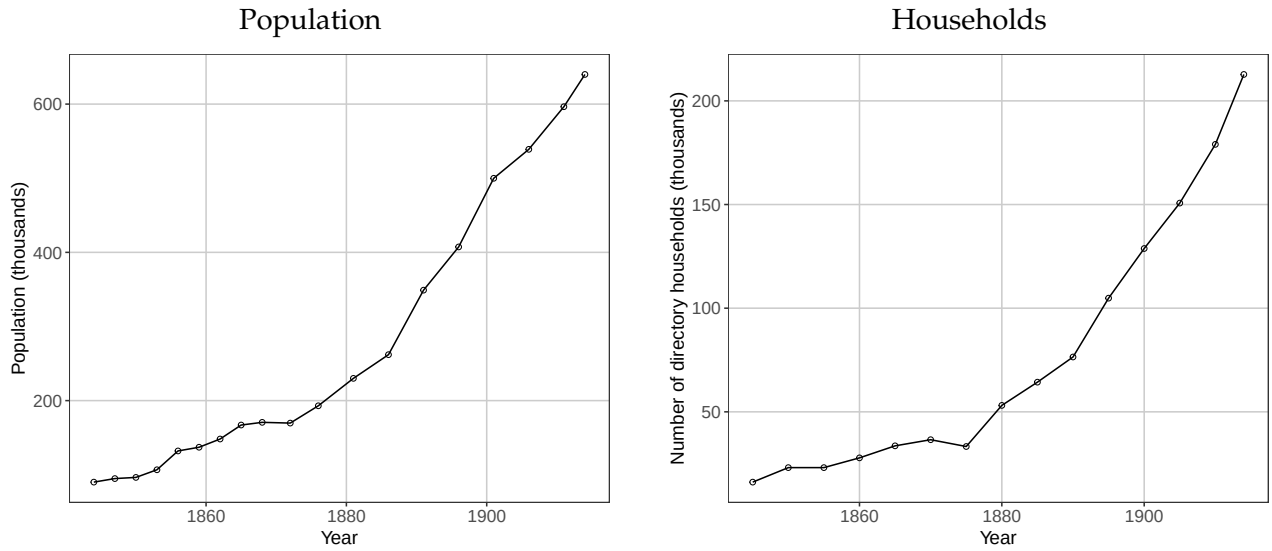


Figure 4: Population

Note Figure shows the number of Munich inhabitants from official statistics (left panel), and the number of households listed in the Munich directories (right panel) 1845 – 1914.

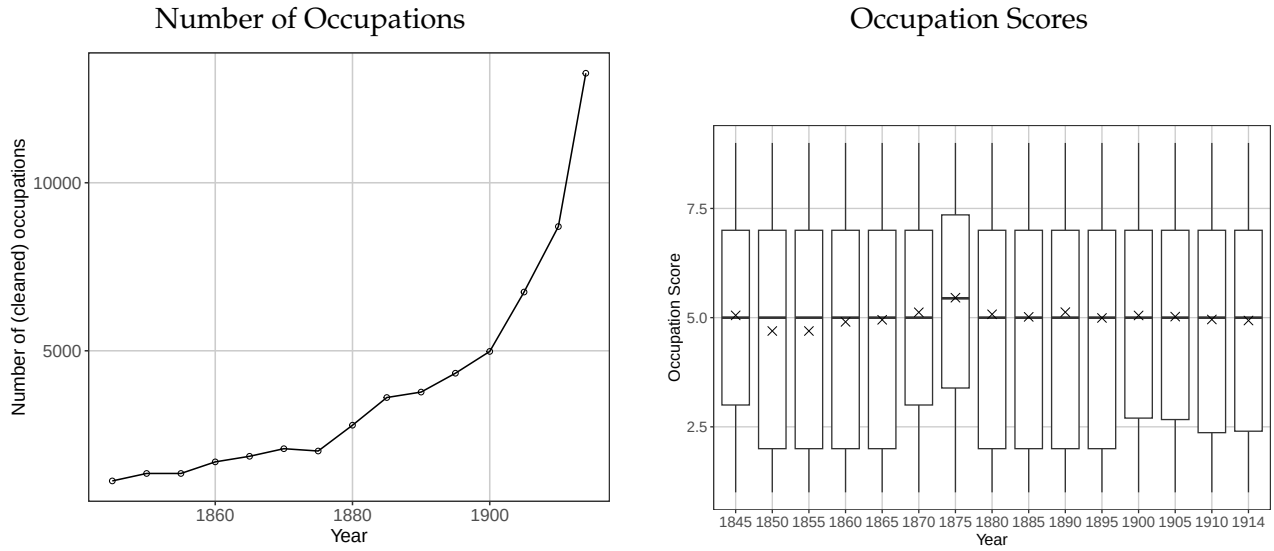


Figure 5: Occupations

Note Figure shows the number of unique occupations in the directories (left panel) and the distribution of occupation scores for households (right panel) in Munich 1845 – 1914. The box plot depicts median (line) and mean (crosses), as well as 10th, 25th, 75th, and 90th percentile of the data, separately for each year.

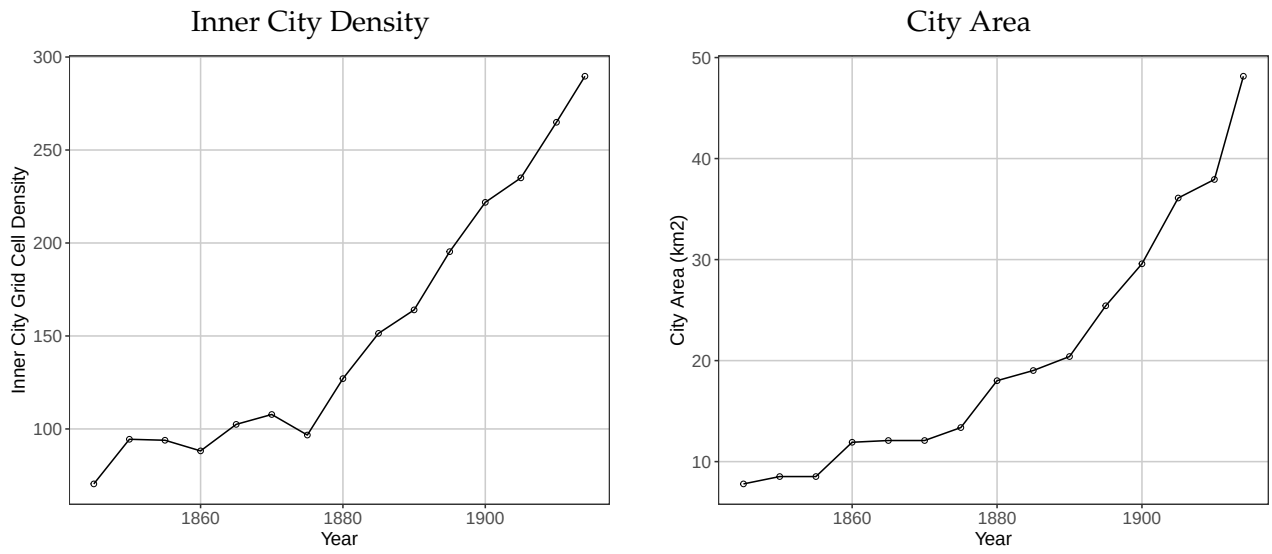


Figure 6: Household Density and City Area

Note Figure shows the number of households per hexagonal grid cell of 200 meters width in the inner city, and the city area in square kilometers in Munich 1845 – 1914. The inner city incorporates the districts of Altstadt-Lehel, Ludwigsvorstadt-Isarvorstadt, Maxvorstadt, Schwanthalerhöhe, and Au-Haidhausen. City area is calculated based on the area covered by 50m hexagonal grid cells that contain at least one household.

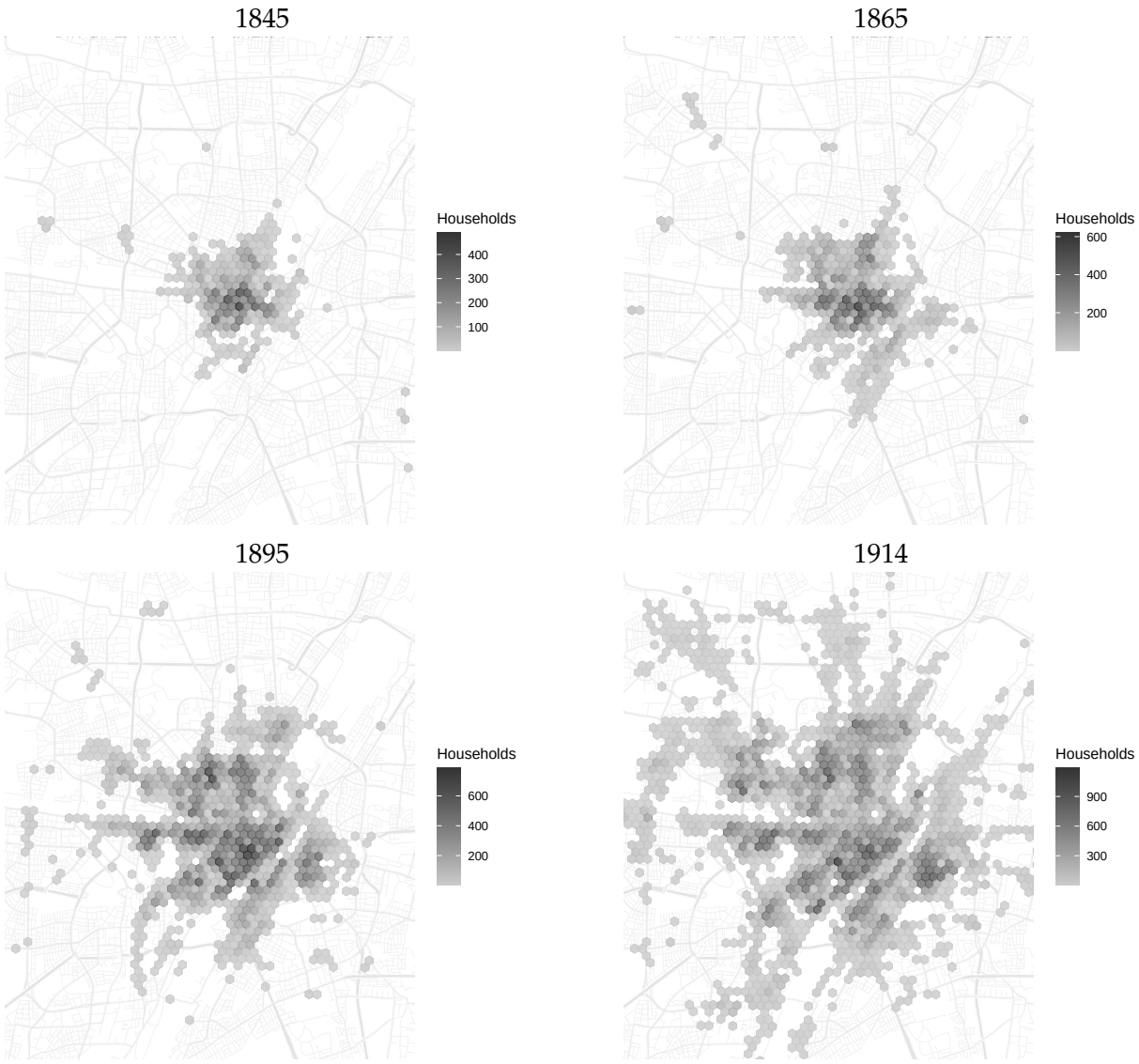


Figure 7: Household Density

Note Figure shows the number of households in hexagonal grid cells of 200 meters width in the years 1845, 1865, 1895, and 1914. Base map is the street grid of modern Munich.

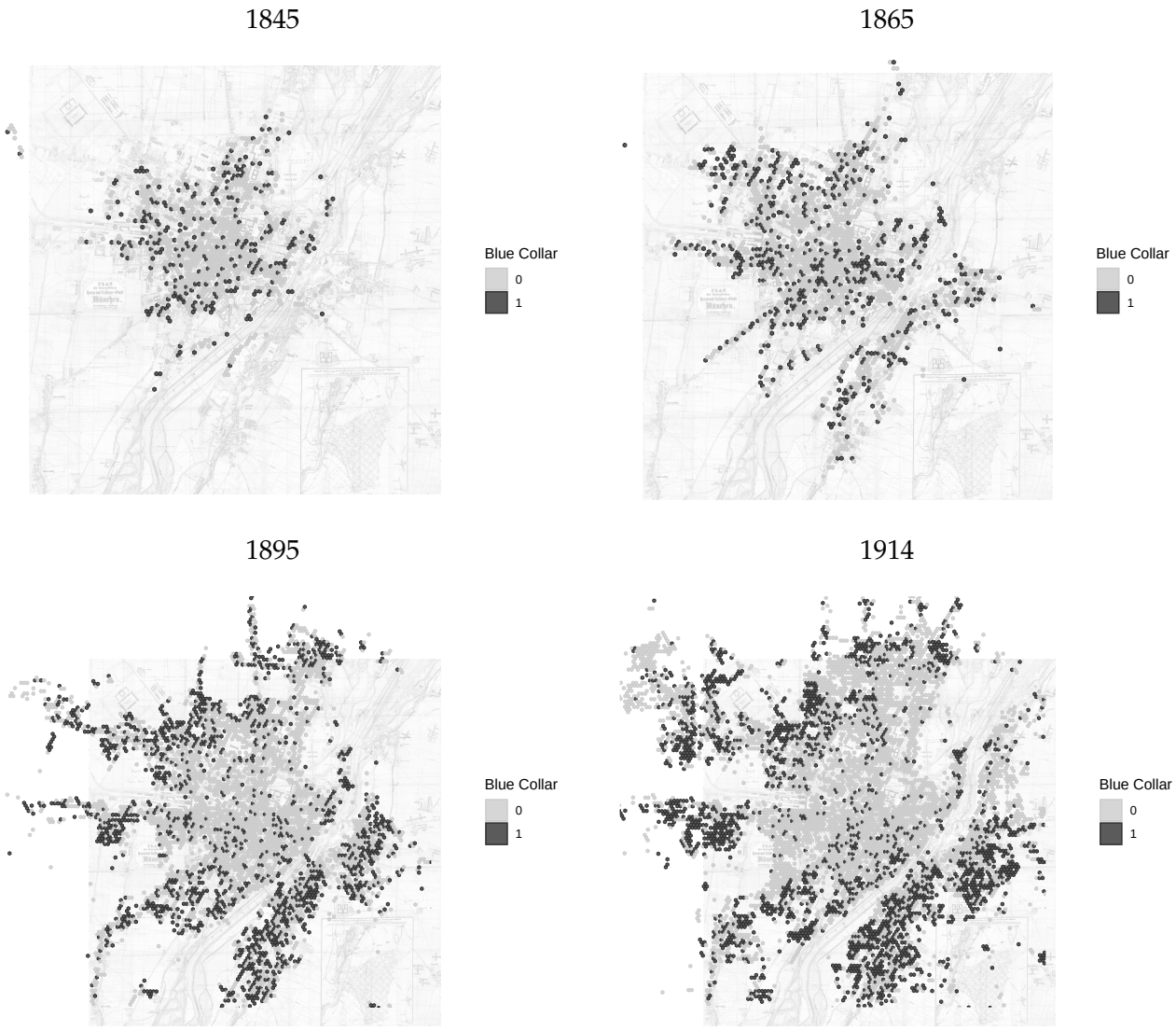


Figure 8: Blue Collar Workers

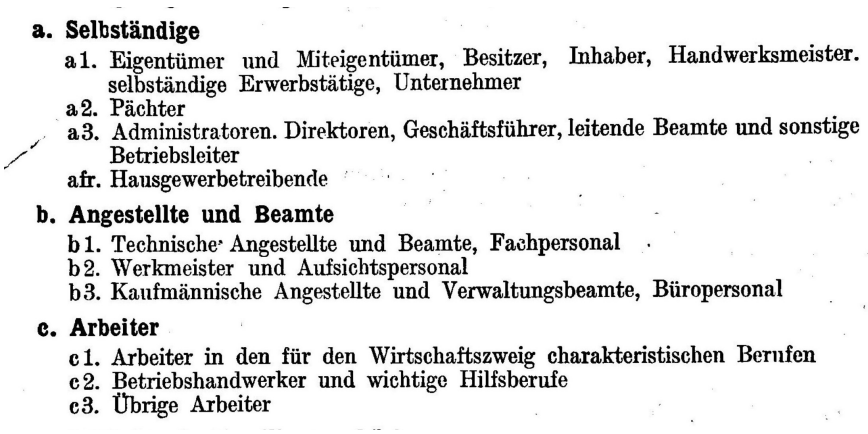
Note Figure shows grid cells of 50 meters width, classified by whether a cell comprises more than 50% blue collar workers. Plotted are the years 1845, 1865, 1895, and 1914. Base map is a historical map of 1865 Munich. A grid cell is classified as "blue collar" if more than 50% of the households in the cell have a below-median predicted occupation score.

Figure 9: Dissimilarity Index of Economic Segregation

Note Figure shows the evolution of the dissimilarity index of occupation-based spatial segregation over time. We classify a household as "blue collar" if it has a below-median predicted occupation score. Dissimilarity is then computed at the level of grid cells of 200 meters width, on the basis of the relative shares of blue collar and white collar workers in each cell.

A Supplementary Appendix

A.1 Figures and Tables



a. Selbständige

a1. Eigentümer und Miteigentümer, Besitzer, Inhaber, Handwerksmeister. selbständige Erwerbstätige, Unternehmer
a2. Pächter
a3. Administratoren. Direktoren, Geschäftsführer, leitende Beamte und sonstige Betriebsleiter
afr. Hausgewerbetreibende

b. Angestellte und Beamte

b1. Technische Angestellte und Beamte, Fachpersonal
b2. Werkmeister und Aufsichtspersonal
b3. Kaufmännische Angestellte und Verwaltungsbeamte, Büropersonal

c. Arbeiter

c1. Arbeiter in den für den Wirtschaftszweig charakteristischen Berufen
c2. Betriebshandwerker und wichtige Hilfsberufe
c3. Übrige Arbeiter

Figure A.1: Description of Occupation Scores

Note Figure is take from p. 11 of the 1925 occupation census. It describes the occupation grouping into nine main categories. Their translations are as follows:

- a: self-employed
  - a1: Owner and co-owner, owner, proprietor, master craftsman, self-employed individuals, entrepreneurs
  - a2: leaseholders
  - a3: administrators, directors, executive officers, senior officers and other operations managers
- b: employees and officials
  - b1. technical employees and civil servants, specialist staff
  - b2. foremen and supervisory staff
  - b3: commercial employees and administrative officials, office staff
- c: workers
  - c1. workers in professions characteristic of the respective industry
  - c2: industrial craftsmen and important auxiliary professions
  - c3: other workers.

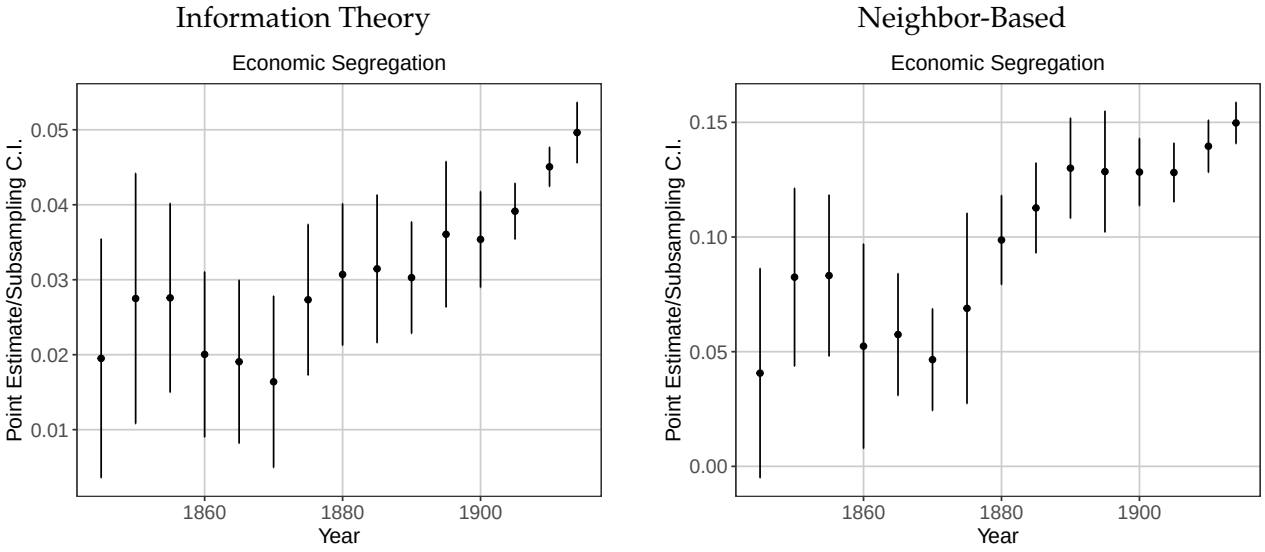


Figure A.2: Alternative Measures of Economic Segregation

Note Figure shows the evolution of the information theory (Panel A) and neighbor-based (Panel B) index of occupation-based spatial segregation over time. We classify a household as "blue collar" if it has a below-median predicted occupation score. The information theory index is then computed at the level of grid cells of 200 meters width, on the basis of the relative shares of blue collar and white collar workers in each cell. For the neighbor-based segregation index, we determine neighboring households via neighboring houses and, where applicable, floors within a house.

A.2 Geolocating Addresses: Special Cases

Figures A.3 and A.4 show the stability of house numbering: Even though Luitpoldpark was expanded after WWII, displacing Borschtallee and most houses on the Western side of Belgradstraße, which are still visible in the left column of Figure A.3, house numbering was not changed, as becomes evident in the right column: House numbers leave off at 113 in the beginning of the park (lower left corner) and resume at 169 (upper left corner) today. On the right side of the street (even numbers), house numbers remain unchanged. Figure A.4 speaks to the same point: even though houses 63–67 were torn down, house numbers were not changed, instead omitting the affected house numbers in the current street course.

Hence, we interpolate missing addresses coming from minor changes to the cityscape, distinguishing two degrees of certainty:

- If there is a gap between two existing addresses on a street, we uniformly distribute the missing house numbers between these two bordering addresses.
- If a street was cut off at the end or start, we fit a line through all existing houses on the street using OLS, and then add houses to the start or end using the average between-house-distance on that given street.[1]

Figures A.3 and A.4 demonstrate how we fill gaps in house numbers: For Belgradstraße, we place all numbers in $\{167, 165, ..., 111\}$ evenly spaced in between the location of houses 169 and 113. For Bayerstraße, we place all numbers in $\{63, 65, 67\}$ evenly spaced between 61 and 69.

Figure A.5 illustrate how we deal with cut-off endings of streets: It shows the start of Amalienstraße in 1938 and today. Oskar-von-Miller-Ring was paved over the first few houses, so that today Amalienstraße begins at 10 (here again the house number continuity becomes evident). We fit lines through both sides of Amalienstraße to locate 1-7 and 2-8, respectively.[2]

[1] This step is done separately for streets where we can distinguish the left and right side of the street, and for streets where we only know the overall street course. A special case is represented by houses at the same "number" (10a, 10b, ...): we place these orthogonally to the street course. Houses located on a square are distributed uniformly around the square centerpoint.

[2] The 1938 map in Figure A.5 also shows Glückstraße, which does not exist today. We manually geolocate Glückstraße.

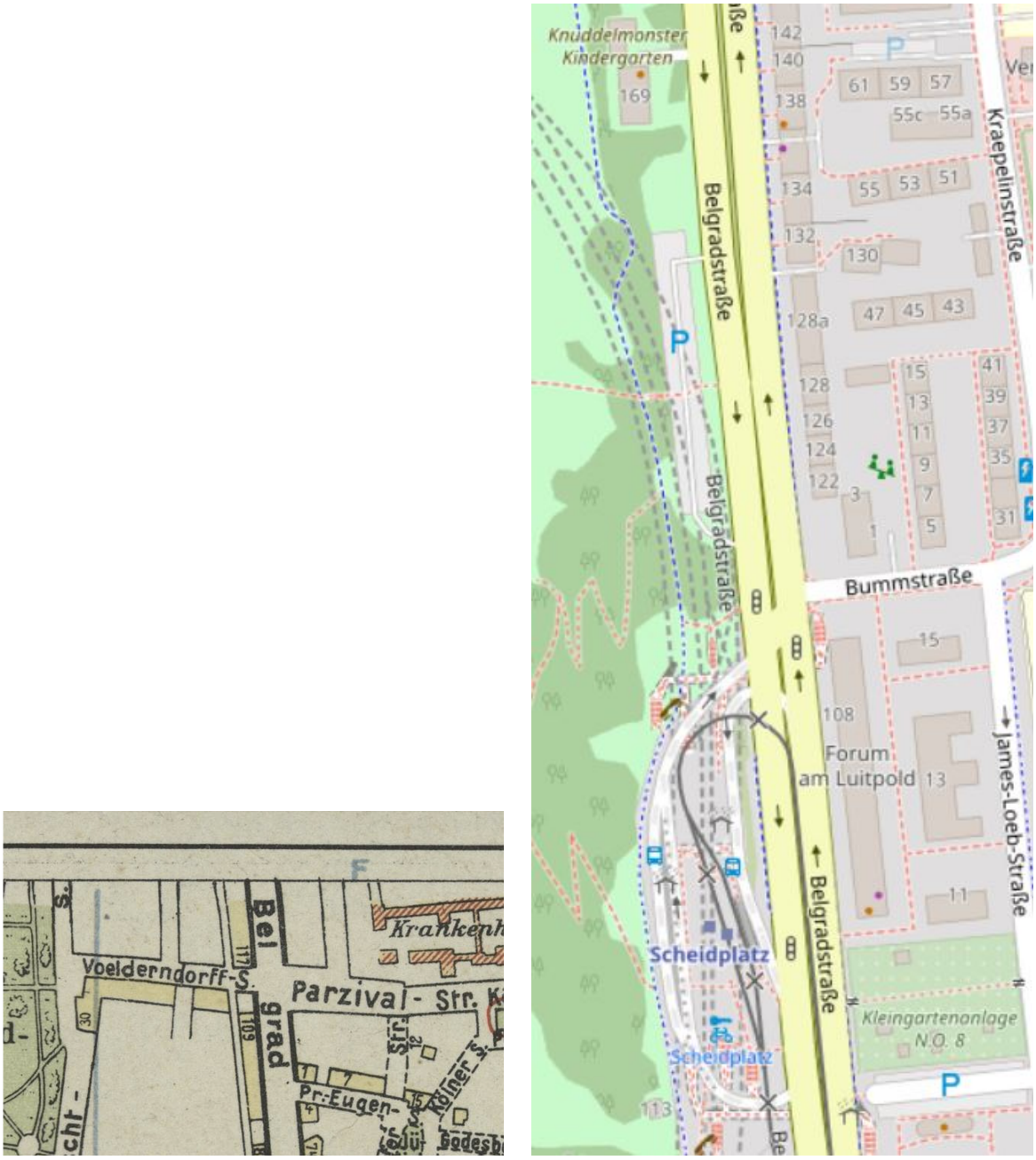


Figure A.3: Belgradstraße in 1938 (fragment which shows house 117 and Borschtallee) and today
Note Shows continuity of house numbering. Data sources: OpenStreetMaps and Stadtatlas

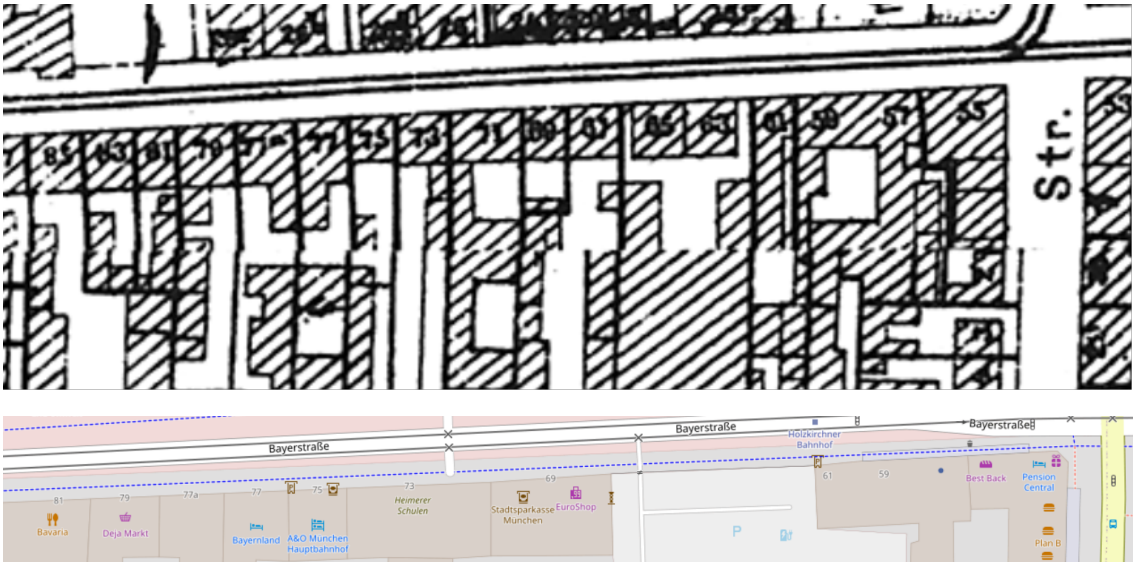


Figure A.4: Bayerstraße in 1938 and today

Note Shows continuity of house numbering. Data sources: OpenStreetMaps and Stadtatlas

Figure A.5: Amalienstraße in 1938 and today

Note Data sources: OpenStreetMaps and Stadtatlas

A.3 Predicting Occupational Titles: Detailed Approach

Step 0: Standardizing Occupation Titles. For the occupation titles listed in the directories, we first remove particles that do not indicate the occupation (like widow or retirement status). We record this information in separate indicator variables. In a second step, we expand abbreviations ("-m." becomes "-meister" etc.). Then, we standardize spelling across time, preferring the more modern spelling ("-rath" to "-rat", "controlleur" to "kontrolleur"). These steps are sufficient to reduce the amount of occupation titles by one order of magnitude.

Step 1: Decomposing occupation titles. We manually split the occupation titles listed in the 1925 occupation census into component nouns. We remove genitive-s endings where necessary (so "Stationsvorsteher" becomes "Station,Vorsteher").

For each standardized directory occupation, we iteratively apply the Python package *compound-split* (`https://pypi.org/project/compound-split/`) to produce up to 200 possible splits for each occupation.[3] We use multiple digital German dictionaries to remove genitive-s word endings. Finally, we predict the most likely split for each word. To do so, we compile three customized dictionaries of nouns in the German language: An extensive, 200,000 entry noun dictionary; one less extensive noun dictionary with frequency weights for every noun; and one occupation-specific noun dictionary.[4] We choose the best split based on the (weighted) fraction of component words that match these dictionaries.

Step 2: Component-based prediction. Based on the finely grained splits of both occupation census and directory occupation titles, we assign a score to each directory occupation title that has at least one component of overlap with the occupation census components.

To do so, we first note that occupations usually end in the most generic occupation title: A "train-help-worker" (*Eisenbahnhilfsarbeiter*) is foremost a worker, whereas the component "help" qualifies the scope of activity, and "train" specifices the industry. We hence match the longest possible connection of components, starting from the right.

In the case of the "train-help-worker", we do not find the full title in the occupation census, but we find "worker" and "help-worker". Since the latter combines more components, we consider this a match.

Next, we turn to all the unmatched prefixes of matched occupation titles (in the example case: "train"). We match prefixes to prefixes in the occupation census, finding, for example, "train-assistant", "train-secretary", "train-technician". We then calculate an occupation score multiplier for each prefix: *Removing* the prefix, how does the occupation score of the "train-assistant" compare to that of the "assistant" (and vice versa for the secretary, the technician etc.)? Averaging over these multipliers, we find the average occupation score multiplier associated with the respective prefix.

In a last step, we combine the information on matched occupation status and occupation prefix multiplier to a single occupation score. In the example, individuals working in the "train" industry receive, on average, a 10% lower occupation score than their counterparts without the prefix.

Hence, the "train-help-worker" receives an occupation status of 1.80, whereas a "help-worker" is ranked at 2.

[3] The splitter library always returns a two-way split. We take the three most likely two-way-splits and break each component up again. We repeat this step up to three "layers deep". Then, we re-assmble all possible combinations of splits that result in the full occupation title.

[4] We modify the first two dictionaries to not exclude nouns that will never occur in an occupation-specific context.