THE UNIVERSITY OF CHICAGO

Gentrification In Berlin: A Novel Conceptually Driven Quantitative Approach to Evaluating Anti-Displacement Legislation

Clyde Schwab

A thesis
submitted in partial fulfillment of the
requirements for the degree of
Bachelor of Arts
(Geographical Sciences)

May 2021

Table of Contents

List of Figures

List of Tables

Acknowledgements

This project would not have been possible without a network of wonderfully supportive faculty, colleagues, peers, friends, and family.

First, I want to extend my deepest gratitude to Dr. Marynia Kolak for their continuous support, advice, and patience throughout this process. I owe my passion for geography to Dr. Kolak's incredible enthusiasm, brilliance, and kindness, and their instruction and encouragement has continually helped foster of excitement, joy, and curiosity in education while never forgetting to center justice and equity. Thank you for your help.

I would also like to thank Dr. Juliana Koschinsky for her indispensable assistance in facilitating my research, and more generally her excitement, encouragement, and support. I am also grateful to my advisor Evan Clark for his counsel and advice in developing my ideas and guiding my excitement about Berlin into a project I never thought was possible. Finally, the completion of this project was greatly assisted by Dennis Helweg, Guido Schulz, and Holger Lischke, who generously took the time to discuss their work and connect me with invaluable data sources.

Finally, I want to extend my gratitude to those who supported me personally throughout this project and in an unprecedented time — to my mother and father, who fostered my intellectual passion and curiosity, and continue to provide inspiration and guidance, to Francesca, my partner, for standing with me through the highs and lows accompanying my work this year, and to our cat, Ferguson for keeping me company during the many late nights.

Chapter 1: Introduction

As cities across the world have experienced urbanization over the past several decades, gentrification, and its accompanying social and economic tensions, has become one of the most prominent issues in geography literature. As affected citizens struggle to fight displacement and cities coordinate policy responses, the academic community has struggled to rigorously define and quantify gentrification despite the substantial attention paid in popular media to gentrification as an urban phenomenon (Preis et al, 2020). Policy responses have also been varied, with cities using a mix of rent controls, supply-side initiatives, and other regulatory measures to mitigate — and sometimes encourage —gentrification (Lees and Ley, 2008).

In western Europe, Berlin has most recently stood at the forefront of debates surrounding gentrification. In the last five years alone, rent has increased in the city by 44% (Schönball, 2020). Rent prices, which rose 6% in 2017, have even been eclipsed by property sale prices, which grew by 13.6% in the same year. The dramatic increase in gentrification has spurred confrontations, protests, and even violence across the city, where residents say their wellbeing is being sacrificed to wealthy, out-of-town investors (Schönball, 2018). The city made headlines for its recent rent freeze in 2020, known as the *Mietendeckel* (Eddy, 2020), and again in April 2021 for its reversal by the German Federal Court of Justice (the highest court). This measure, which took effect in January 2020 after lengthy challenges in German courts, froze rents to the current level for the next five years. Evidently, Berlin faces a substantial housing crisis.

Nonetheless, researchers have begun exploring systematic quantification of gentrification in Berlin and the open data environment available to aid in the identification of gentrification hotspots in the last 10 years (Holm & Schulz 2018, Schulz 2017, Helweg 2016, Ulbricht & Döring, 2018). However, as of yet, no attention has been paid to the role of *Erhaltungsgebiete*, or conservation regions, and their relationship to gentrification in Berlin. Generally, this legislative tool — colloquially known as *Milieuschutz* (community defense) — allows district governments[1] to grant certain regions a protective status on a sub-neighborhood level. However, several aspects of the

[1] There are 12 *Bezirke* (districts) in Berlin: Charlottenburg-Wilmersdorf, Friedrichshain-Kreuzberg, Lichtenberg, , Marzahn Hellersdorf, Neukölln, Pankow, Reinickendorf, Spandau, Steglitz-Zehlendorf, Tempelhof-Schöneberg, and Treptow-Köpenick.

Erhaltungsverordungen (conservation ordinances) may serve to exacerbate gentrification by disincentivizing and thus reducing the net supply of housing in popular areas of Berlin and increased gentrification in neighboring areas.

In this paper, I seek to answer both (1) how we can operationalize a *conceptually driven*, *quantitative*, and *useful* model of process-based gentrification and, with that model, (2) what effect community protection ordinances have on gentrification in Berlin. Implicit to the framing of this research design is also the question of whether the operationalization can be sufficiently successful to actually evaluate policy on a sub-neighborhood level. First, I situate community protection areas in the history of housing and gentrification in Berlin and existent measures designed to control rent in the background section. Second, I contextualize my research in the field of urban studies and geography in a three-part literature review: (I) defining gentrification, (II) previous quantitative analysis of gentrification, and (III) previous analysis of the relationship between preservationist public policy and gentrification. Third, I explain the datasets used for my analysis, including demographic and real estate data from 2000-2019, to develop a more nuanced diagnosis of gentrification in Berlin on the *Planungsräume* level. Fourth, in my methods section I describe how 1. I develop a categorical system of designating differing stages of gentrification in the city and 2. I conduct exploratory analysis to understand the spatial and temporal relationship between the process of gentrification and designation as a *Erhaltungsgebiet* using survival analysis and difference-in-difference (abbr. DID) design. Finally, in my results and discussion, I examine the observed effects of protection areas on gentrification, the assumptions and underlying each analysis, and potential steps for further research with spatial design.

Chapter 2: Background: Housing in Berlin

Much of Berlin's contemporary urban residential fabric was constructed in the late 1800s after German unification, when Berlin became the capital of the newly formed German government (O'Sullivan, 2020). It grew to be the world's third largest city in 1921 and was home to some of the largest German industrial corporations, such as Siemens and Borsig (Hake, 2008). However, Berlin's modern housing history largely begins Post-WWII, when housing stock in the city was substantially destroyed. East and West Berlin took different approaches to rebuilding housing: slab-housing apartment blocks were built in the East while the West's housing stock was mostly rebuilt to its prewar state, most commonly five-flats built as tenement housing in the late 1800s, known as *Mietkasernen* (rental barracks) O'Sullivan, 2020). Following WWII, these rental barracks were popular locations for squatters and grew to support a vibrant alternative scene in Berlin, but these same apartments have becoming increasingly popular and expensive because of their high ceilings, polished exteriors and shared backyard (Ibid.).

Despite the increasing social capital of Berlin, its isolation from the rest of Germany left it severely economically disadvantaged compared to the rest of the nation, with sluggish job grown and a lower average GDP than the rest of the nation (Instituts der deutschen Wirtschaft, 2016). The city is famously known as *arm, aber sexy* — poor, but sexy — reflecting its contrasting cultural and economic capital. After German reunification in 1991 Berlin returned as the capital of Germany, but the city's economy floundered during the aughts as industry in the nation had largely moved to western and southern Germany. Berlin has been home to dramatically fewer large corporations than other European metropolises such as London or Paris despite its size and maintains an unemployment rate 6-7% above the national average (OECD 2010).

Schulz (2016) notes that due to a range of factors including the cultural assets of the city and cheap and available housing, Berlin began to rebound in the early 2000s with falling poverty and unemployment rates. Now largely a service-sector economy, Berlin has caught up with the rest of Germany in the past five years and seen strong growth in creative, education, and research industries (Wirtschafts- und Innavationsbericht 2020). The global financial collapse in 2007-2008 resulted in low interest rates which spurred the conversion of rental units to condominiums, and the city began dramatically gaining population in the early 2010s, both of which were accompanied by dramatic rises in rent (Schulz 2016). The average Berlin household is still relatively poor compared to the rest of the country, and the demand for housing has quickly eclipsed the supply

— it is estimated that nearly 200,000 new homes will need to be built by 2030 to accommodate Berlin's population (O'Sullivan, 2018). Also of note is that Berlin is a city of renters, with about 80% of the population choosing to rent.

There have been several substantial policy initiatives in recent years which have sought to slow rising rents. In 2015 Germany passed what is known as the *Mietpreisbremse*, or rent price break, which instituted a national rent index in which rent can only be increased by 2% yearly to account for inflation and rents cannot go beyond 20% of the rent for equivalent properties according to a nationwide rent index (*Mietspiegel*). This measure has been largely ineffective in slowing rent increases, particularly in Berlin (Thomschke, 2016). The most recent development in rent control is the *Mietendeckel* (rent cap) which froze rents at levels in the national rent index in June 2019. While there are some notable exceptions, including social housing (which is already rent controlled) and buildings built after 2014, the law affects 90% of the housing in Berlin (Senatsverwaltung für Stadtentwicklung und Wohnen, 2020). While the *Mietendeckel* was the result of hard-fought activist campaigns, concerning side effects were observed: Immoscout24, one of the largest real estate portals in the country, reported that the supply of apartments in Berlin fell by 41% while demand has increased by 172% (Engelbrecht and Krone, 2020). With the recent defeat of the rent cap in German high courts, landlords are holding tenants accountable for back rent owed had the law not been in place — on average €6 per square meter higher than the legal rent, which itself is €7 per square. (Knight 2021, Engelbrecht and Krone, 2020).

Erhaltungsgebiete represent a somewhat different legislative model for mitigating gentrification than rent control. In protected areas, landlords and tenants are prohibited from modernizing, demolishing, or converting units to condominiums or other uses without a permit. Moreover, when attempting to sell a property within one of these regions, the city's government may exercise the right of first refusal (known as *Vorkaufsrecht*) and buy the property themselves to convert it to affordable housing. These regulations may appear to help prevent gentrification: according to Berlin's city website, "Social conservation ordinances are intended to prevent the composition of the resident population from changing due to displacement through expensive modernization measures" (Bezirksamt Mitte, Berlin.de, translated from German via Google Translate). *Erhaltungsverordnungen* have existed in Germany since the early-1970s, but their deployment has increased dramatically in the past ten years (O'Sullivan, 2017). The first protected area was created in Berlin in 1991, and several more were introduced in the late 1990s. As of 2015 there were only 15 such zones, but currently there are 72. These areas are designated on the sub-neighborhood level (e.g. *Stephankiez* in the district of Mitte) in a process initiated at the request of any concerned citizen(s) or public, and instituted after review by the district assembly based on the housing characteristics of a neighborhood.

Chapter 3: Literature Review

3.1: Defining Gentrification

The term gentrification generally refers to the process whereby people with higher income move into a lower income neighborhood and, because of rising property values, lower income residents are displaced. This phenomenon was coined in England by sociologist Ruth Glass in 1964 and is also associated with a change in the neighborhood's perceived character (Barton, 2014). A central theory in understanding the causal framework of gentrification is that a rent gap is created in a certain region, whereby the ground rent currently captured diverges from the maximum potential rent in an area, which leads to increased investment and eventual pricing-out of the original residents (Preis et al, 2020). However, while gentrification can be understood as a purely economic and social process, other authors have frequently emphasized the cultural changes that surround the process of gentrification through qualitative data (Barton, 2014). Though a universal definition of gentrification has yet to be determined, in this project it is generally taken to mean the confluence of increased real estate evaluation, displacement, and cultural change in a specific neighborhood.

One popular theoretical framework for understanding gentrification is the invasion-succession cycle. First coined by Park (1952), this theory was adapted from theories describing ecosystem change in nature by the University of Chicago human ecologists Robert Park and Ernest Burgess. In their theory of concentric zones, Park and Burgess argued that the growth of a city involves the successive displacement of wealthier residents from the outer zones of a city, in what they dub an "invasion" poorer minority residents. In this model, as poorer residents continually replace wealthier residents, wealthier residents move to the outside of the city, and the poorer residents gradually gain economic standing and become integrated into American society, before eventually themselves being "invaded" by a new class of poor residents in the form of immigrants. However, other sociologists have harshly criticized this model as being overly naturalistic and lacking a description of the factors causing displacement, such as political economy (Guterbock, 1980). Moreover, the casual framework of this model can exclude factors of political economy, such as discriminatory lending practices, local governmental ordinances, and state-encouraged investment. (Lee, 2007).

However, beginning in the late 20th century, the invasion-succession model was adapted by sociologists to describe gentrification, in what represents a reversal of Park and Burgess's model. According to the invasion-succession framework, gentrification is a process where poorer, inner-city residents are displaced by wealthier "invaders." Moreover, this vision of gentrification sees actors defined by demographic and behavioral characteristics intercede at different points in the process. This paper adopts the invasion-succession model of gentrification in its descriptive capacity[2] particularly based on Dangschat (1988), Kecskes (1994), Atkinson et al (2011), and Blasius et al (2016). In his text on gentrification in Hamburg, Dangschat adapted Park's invasion-succession model to describe the reversal of dynamics in inner-city areas that were previously experiencing blight or deterioration. Dangschat's model of gentrification includes two immigration phases and two displacement phases: first, younger, socially mobile, often well-educated but lower-earning people such as artists and students move into a neighborhood because of affordable rents — these are known as the "pioneers." This prompts the first stage of initial displacement of the original inhabitants of a neighborhood. Next, "A 'scene' with shops, bars and restaurants related to the needs of the new residents is forming" (Kecskes, 1994, p. 28, translated with Google Translate). As the original inhabitants move out and the neighborhood gains recognition as an interesting or attractive place, older and wealthier people are attracted to the neighborhood and move in, representing the second immigration. This population is referred to as the "gentrifiers". Increased demand for property and increased price flexibility of the newcomers results in a wave of real-estate investment from local, national, or multinational property managers. In the final stage, life neighborhood is dominated by high-earning families and property owners and both the original residents and the pioneers are priced out.

[2] The criticisms of the Chicago-school invasion-succession model also hold for this model of gentrification — namely, that factors such as political economy are ignored.

This theory of gentrification was further expanded by Blasius et al (2016) in a systematic, quantitative study of gentrification in Cologne, Germany. To further understand the demand-side of gentrification on a micro-level, Blasius et al categorize actors in the gentrification process based on previous literature and implement this classification system in a large panel study (N=1009) in two neighborhoods. To these ends, several methods of classifying different actors are reviewed, including profession and social group, attitudes (particularly risk-aversion), and sociodemographic characteristics. The authors review the utility and drawbacks of each identification method based on both data availability and diagnostic utility: for most professions it is impossible to infer income, age group, or familial status (e.g. "artist; pioneers are characterized as childless), risk-aversion and social attitudes are not collected in any national census in Germany, and cultural and economic capital is difficult to data in sufficient capacity on for rigorous study. Blasius et al conclude that the most accessible framework is socio-demographic indicators and design a classification scheme based on existent literature and case studies; the descriptors used in this framework are shown in table 1.

Table 1: Gentrification Typology in Blasius et al (2016), p. 57.

Table 1. Classification of groups (typology).

Characteristic	Pioneers	Early gentrifiers	Established gentrifiers	Others	Elderly
Age	≤35 years	≤45 years	≤45 years	≤64 years	>65 years
Years of schooling	12 years	No definition	No definition		
Household size	Any	2 pers., max. one child	2 pers., max. one child	No pioneers or gentrifiers	By age criteria, no pioneers or gentrifiers
Children	No	Max. 1	Max. 1		
Income*	<1.500 €	≥1500 to <2500 €	≥2.500 €		

*Household equivalence income, calculated by OECD scale: first adult = 1.0, other persons ≥ 15 years = 0.5, below 15 years = 0.3.

The authors implement this framework based on a survey conducted in 2010 in two neighborhoods in Berlin with 1009 respondents, and conclude, based on a series of two-way variance tests, that further research is necessary to account for the complex multi-stage nature of gentrification with an expanded typology. Notably, the authors also recommend further implementation of protection areas to prevent condominium conversions seen in the more-gentrified neighborhood.

3.2: Quantifying Gentrification

Quantitative approaches to gentrification have also varied significantly. Perhaps best describing the heterogeneity of quantitative approaches to evaluating gentrification on the city-level, Preis et al. (2020) examines the variability between governmental approaches to identifying gentrification. The authors select four different methods of identifying gentrifying census tracts drawn from the city governments of Portland (OR), Seattle, Philadelphia, and Los Angeles, and applied these methods to Boston, MA. Particular attention is paid to methods of mapping gentrification produced by city governments because of their potential influence on public policy and resource allocation. Outlining the variables used by the four different methodologies, the authors delineated variables measured on the individual-level (e.g., % nonwhite population), household-level (e.g. household income), and neighborhood-level (e.g., proximity to transit). In applying these different methods to Boston, the authors dramatic differences between the four models. The most lenient method identified at-risk 119 census tracts (Seattle) where the most conservative identified only 25 at-risk tracts (Philadelphia). The average percentage of tracts retained by another model is 63%, and only seven tracts were identified by all four models as at-risk. The difference in the models resulted from different variables being used by each method — for instance, the Philadelphia model does not include race as a variable, and thus excluded large portions of the city with the highest percentage of African American households. This paper demonstrates not only the heterogeneity of methods of modeling gentrification, but also the extent to which these methods represent conflicting theories of gentrification — how does one understand issues such as redlining and other forms of discrimination in the context of gentrification? This question is further complicated in light gentrification appearing in cities across the world, where different historical contexts result in heterogeneous processes and perceptions around gentrification.

In the context of Berlin, the most substantial quantitative analysis of gentrification is Holm and Schulz's paper 2018 text on their *GentriMap* project. *GentriMap* is described as an analytical tool which seeks to create a method of empirically identifying gentrification and displacement in cities across the globe, through the lens of political economy. The authors construct a real-estate index and a social index, and then use the combination of the two to form a single gentrification index. In Berlin, the two indexes were based on data from years 2007-2014 from the online housing portal Immobilienscout24 and welfare-recipient data respectively and implemented on the prognosis-area level (N=60). Some results of the indexes generally correspond to popular views on gentrification in Berlin, with Mitte, Kreuzberg, Charlottenburg, and Neukölln demonstrating the most dramatic real-estate value increases, while other results, particularly in the social index, show trends of gentrification in areas on the outskirts of the city. The combined gentrification index shows medium to high gentrification occurring in nearly the entire central city. Despite this text's value as a basis for quantifying gentrification in Berlin, analysis was limited to the *Prognoseräume* (prognosis areas) (n=60), rather than the *Planungsräumenn* (n=448).

A similar study conducted in 2017 by one of the lead researchers for *Gentrimap*, Guido Schulz, examined the relationship between displacement, measured by the proportion of welfare-recipients in an area, and increased real-estate valuation, measured by the yearly rate of change of inflation-adjusted median rent price offerings. The period of examination is 2007-2012 on the *Planungsräume* level (n=429 at time of study). Areas with high real-estate valuation increases are compared to control areas with similar poverty levels and rent offering prices at the start of the examination period. According to Schulz (2017), about 10.5% of Berlin is classified as gentrified, and of other areas 32% selected as controls. Interestingly, all areas classified as gentrifying were in the top-half of the distribution of poverty rates across Berlin and had rental-price offerings were in the middle price-segment. Gentrified areas had 70% higher migration rates than control areas and, after controlling for demographic differences, Schulz found a regression coefficient of .135 and R^2 of .85, indicators which "confirm the presumed connection that the emigration rate increases as real-estate values increase" (Schulz, 2017, p. 66, translated by Google Translate). This text is valuable in its novel analysis of gentrification on a granular level and its comparison of gentrified areas to control areas.

Helweg (2018) represents an exhaustive exploration into the possibilities of analyzing gentrification in Berlin using "big-data" and machine learning. While the methods used in Helweg (2018) are outside the purview of this project, the author goes into detail regarding the availability, structure, and integration possibilities of numerous publicly available datasets. Helweg, in contrast to the previous two texts mentioned, focuses on the relationship between social status of a neighborhood and the *Angebotsstruktur*, or amenity/offerings structure (e.g., number of cafes, restaurants, fast-food restaurants etc.). In his analysis of available datasets, Helweg chooses the *Einwohnerreggister* (occupant register) and *Monitoring Soziale Stadtentwicklung* reports (MSS, Monitoring Social Urban Development), and OpenStreetMap (OSM) points-of-interest (POIs) because of the availability of historical data, area of analysis (LOR *Planungsräumen,* N=448*)*, and open availability. The availability of data frames the central questions of the thesis: whether there is a relationship between the amenity structure and the social status of a neighborhood and whether this relationship is temporal.

Through the use of machine-learning methods, he finds a strong relationship between change in the local amenity structure and social status on the district level (n=12), and correlation between restaurants with a higher social status and fast-food with a lower social status. The most novel finding of Helweg's analysis are indications that change in amenity structure seems to follow change in social status, rather than precede it, but there was no single POI type that was correlated with the dynamic of a neighborhood.[3] In the discussion section, Helweg notes some limitations to his study: OSM points are user-contributed and not centrally audited, and thus may include substantial heterogeneity in how and when they are reported.

3.3: The Effect of Preservation Policy on Gentrification

To understand the role of *Erhaltungsgebiete* in Berlin, it is necessary to look at them in the context of preservation measures at large. In McCabe's (2018) review of literature on the relationship between preservation policies and gentrification in the US, the author presents several goals articulated by these policies, ranging from honoring, rehabilitating, or preserving designated buildings or neighborhoods. To these ends, some policies target specific buildings and the type of changes that can legally be made to them and implement land-use restrictions and prevent densification on the neighborhood-level. McCabe describes a range of outcomes: bestowing honorary status on a neighborhood or set of building can attract wealthier and more educated residents and densification rules in preservation ordinances can reduce housing supply, spurring eventual neighborhood change.

There have been a number of more involved case studies of the relationship between neighborhood change and preservationist policy. Kinahan (2019) looks at the effect of the Federal Historic Rehabilitation Tax Credit (RTC), the largest national preservation program in the U.S., from 1998 to 2010 in six legacy cities across the U.S.: Baltimore, Cleveland, Philadelphia, Providence, Richmond, and St. Louis. Comparing census tracts with RTC activity, the author finds that RTCs attract higher income residents with no losses to lower or middle income or minority residents. However, the author also emphasizes that these results occur in cities with weak housing

[3] This means that while fast-food restaurants are associated with areas with a more negative social status, they are not associated with areas with a negative social dynamic; the same is true for restaurants and positive social status.

markets in tracts with many vacant or underutilized buildings; this is not the case with Berlin, one of the fastest growing housing markets in western Europe.

Glaeser (2010), writing for the City Journal magazine, provides a comprehensive albeit opinionated history of the creation of New York City's preservation districts. The New York City Landmarks Preservation Commission was created in 1965 to preserve architectural and culture heritage, with proponents arguing that unmitigated economic systems may not sufficiently value historic, aesthetic, or cultural considerations. Between 1989 and 1993, 509 acres were added to hundreds of existing acres of preserved land, where any external changes to a building must pass through the city's Historic Preservation Commission. According to Glaeser's analysis, census tracts with substantial historic designation overlap saw dramatically fewer units built, and in some cases, an overall reduction of housing stock since the 1980s. Glaeser finds that average incomes are 74% more in historic districts, and incomes were 29% higher in tracts that would become at least partially historic districts. While this article certainly establishes the possibility of a causal relationship between historic district designation and gentrification, it makes no attempt to implement a rigorous causal framework to identify historic districts as a causal mechanism of price increase or displacement.

Been et al. (2016) undertakes a more rigorous economic analysis of New York City's landmarked neighborhoods and the policy's effect on localized property values using a DID analysis. The authors explain that historic designation can have disparate impacts on real estate valuations: while landlords are discouraged from undertaking modernizations which may increase the value of a property, new investment which can increase housing supply in attractive neighborhoods is also hindered. By designing a model which includes the numerous potential effects of landmark designations, the authors find that historic designation results in reduced construction and increased valuation in areas where the value of foregone development potential is lower. While property values rose in historic districts with lower initial valuations, historic designation produced a negative effect on property values in already high value areas. Been et al. (2016) is valuable for its analysis of the economic relationship between historic designation and real estate values, but there are also substantial differences between New York City's historic designation laws that *Milieuschutz* in Berlin, including the higher rate of renters in Berlin, the possibility of utilizing *Vorkaufsrecht*, and the prohibition of conversion to condominiums.

No similar quantitative undertakings have been performed in Berlin, but several texts have examined *Erhaltungsgebiete* and their role in the urban fabric through a qualitative and theoretical lens. In a dissertation written at the Berlin Technical University, Geßner (2008) explores *Milieuschutz* as a policy tool in the context Berlin's larger city development policy and *Milieuschutz*'s successes and shortcomings using case studies of two areas, Stephankiez and Boxhagener Platz. Geßner finds that *Milieuschutz* works to protect residents in several ways: by using Berlin's rent index as grounds to evaluate whether certain renovations will increase the value of a property beyond what is manageable, unnecessary renovations are successfully avoided, which decreases the risk of price-induced displacement. Adding bureaucratic and procedural costs can hinder dubious investors. At the time of publishing, the author found no indication that investment and new building growth in the two areas had been hindered. Geßner finds generally that *Milieuschutz* is a valuable policy instrument when working alongside other policy measures, such as increased local management and urban redevelopment, to promote continued necessary renovation and repair. Geßner (2008) articulates several aspects of the *Milieuschutz*, particularly in contrast to other policies, but Berlin's housing market has developed substantially since 2008, and Lischke (2020) provides further context on *Milieuschutz* in the city using a case study of the aforementioned Stephanerkiez. The author concludes that other urban-development programs implemented in the area serve to increase demand, which protected status is only marginally able offset with anti-displacement measures. Finally, in a report by the Institute of German Economics (Institute der Deutschen Wirtschaft) sponsored by the Berlin homeowner association (Verein zur Förderung von Wohneigentum), the authors argue that protected status results in short-term protection against displacement but a longer-term reduction in housing stock investment, leading to decreased affordability. The authors make no attempt to substantiate these claims through any analysis beyond cursory graphs, but their claims represent hypothetical adverse effects seen in other preservation legislation and provide the basis for this paper's evaluation of protected areas.

In this literature review, numerous reasonably-successful attempts to evaluate gentrification as-process and quantify it have been discussed, but currently there has been no attempt to operationalize a process-based approach on a spatially aggregated level and evaluate policy measures based on this operationalization; this paper will attempt to fill this hole in the literature.

Chapter 4: Data Generation

In this section I describe data sources I use to implement my evaluation of gentrification in Berlin. To provide a rough overview, I synthesize bi-yearly demographic and real-estate valuation data on the LOR planning region level (n=448) and to generate a classification of each area bi-yearly from 2001-2019 based on both the change in select variables and the previously assigned classification. Lastly, I perform tests of global spatial autocorrelation on the results of the data generation process to detect the presence of spatial dependence or spatial heterogeneity, which could affect the program evaluations performed in chapter 5.

4.1: Data

To implement my gentrification classification system, I draw on demographic data from two sources: Berlin's resident register (transl. *Einwohnerregister*) and the Berlin Senate's Administration for Urban Development and Housing (abbr. SenSW, transl. *Senatsverwaltung für Stadtentwicklung und Wohnen*) in their report Monitoring Social Urban Development (abbr. MSS, transl. *Monitoring Soziale Stadtentwicklung*). The resident register is published yearly on the LOR planning region level. Data includes the number of residents in each age group, number of foreigners by home country, and, for each year after 2008, the number and proportion of residents who have been in the neighborhood for five years and ten years. The MSS report series was first published in 2000 and covers the period from 1997-2019. In this report, analysts at the SenSW use several indicator variables to evaluate the status and dynamic of areas across Berlin to assist governmental resource allocation. Variables include the proportion of people on welfare, proportion of people from the EU, the unemployment rate, and the average proportion of migration (positive values indicate increased population, negative values indicate population loss).[4] An example of the data produced by this report is available in appendix B. MSS reports published from 2007 onwards are aggregated on the LOR planning region level, but years 2001 through 2005

[4] While the methodology of data collection and aggregation has changed over the past 20 years, changes are accounted for in data cleaning and mitigated by the repeated processing of data over nine time periods. Of note is the selection of welfare payment indicators which results in a dramatic difference in the distribution of welfare recipients by planning area from 2011-2013.

are published in the traffic-cell format (transl. *Verkehrzellen*) and interpolated to the LOR planning region level using areal weighted interpolation.

To measure change in real-estate valuation, I draw on data provided by the Berlin SenSW's Advisory Committee (transl. *Gutachterausschuss*) on the average square meter property sale price from 2000-2021 on the block level.[5] This data is processed as points, averaged to the nine observation periods over three-year intervals (year 2003 is calculated by averaging property sale prices in 2002, 2003, and 2004). Because sale prices are not available for every block (not every block had a house that was sold), the prices are interpolated using to a raster covering Berlin by cells that are 16 square hectares, selected because the average block size in Berlin is 4.5 hectares (Atlas of Urban Expansion, 2016), using kriging. Cell values are extracted and aggregated to the LOR planning area level. The two sources of demographic data and the real-estate value data are combined to a single data frame for the nine periods of examination and subtracted from each other to obtain the change between each biyearly period. The variables available each year are described in appendix C.

4.2: Methods

While several authors have used analogous variables to quantify gentrification in Berlin using real estate valuation increases, social change or displacement, change in amenity structure, or some combination of the three, I hope to expand on these attempts using the model of gentrification as a process developed by Dangschat (1998). As described in the literature review, Dangschat's gentrification process begins with (1) an increase in pioneers who are younger and frequently more educated, higher income, and of a different ethnic or racial demographic background than the inhabitants. (2) This results in the first stage of price increase and ensuing displacement and social change. (3) Next, a new wave of gentrifiers move in who are older and wealthier than the pioneers. (4) The final stage of gentrification is signified by drastic increased

---

[5] Property sale data, while not a direct substitute for rent price changes, are the best approximation for real-estate value change; other sources, including publicly available data from an apartment rental portal *Immobilienscout24* were considered but ultimately rejected due to a lack of sufficient spatial granularity, lack of sufficient time scale (data is only available from 2007-2013), and the potential for bias introduced by the use of other rental portals by various demographics in Berlin.

investment and real estate valuation and the displacement of pioneers and the remaining working-class inhabitants of the neighborhood.

To adapt the scheme proposed by Dangschat (1988) and Blasius et al's (2016) to Berlin, I design a function to classify each of the planning areas successively in each bi-yearly period according to the previously mentioned demographic and real-estate valuation variables and the areas previous designation.

Table 2: Gentrification Classification System

Stage/ Filter	**Description**	**Variable indications**	**Variable indications description**	**Assignment after filter**
1	*Pre-gentrification*	`gaa < 60th quantile`	Real estate value (`gaa`) is below the 75th quantile. This is designed to screen out areas that are already among the most highly valued in the city, and thus not susceptible to gentrification.	1 (pass) 0 (fail)

2	*Pioneer I*	`E_18U25 > 0`
		`eu > 0`
		`PDAU10 < 0`
			- Increase in 18–25-year-olds (`E_18U25`), people from the EU (`eu`): this reflects a younger group, likely with higher cultural capital than existent residents.
			- Decrease in the proportion of people who have lived in the area for more than 10 years (`PDAU 10`): as there are more new-comers, the percent of people who have lived in the area for more than 10 years should decrease.
				2 (pass)
				1 (fail)

3	*Pioneer II*	`gaa > 0`
		`welf < 0`
		`aus_noneu < 0`
		`DAU10 < 0`
			- Real estate value increase: the pioneers have caused the neighborhood to become more trendy, and thus the first wave of displacement occurs
			- Decrease in the proportion of people on welfare (`welf`), foreigners not from the EU (`aus_noneu`), and the number of people who have lived in the area for 10 years (`DAU10`): these groups, including those in poverty and ethnic minorities, experience some amount of displacement
				3 (pass)
				2 (fail)

4	*Gentrification I*	`E_25U55 > 0`
		`E_0U6 > 0`
		`PDAU5 < 0`
		`PDAU10 < 0`
			- Increase in 25-55-year-olds (`E_25U55`) and children aged 0-6 (`E_0U6`): the first families come to the area after displacement of poorer residents occurs.
			- Decrease in the proportion of people who have been in the area for 10 years, as even the pioneers are now likely well established in the area.
				4 (pass)
				3 (fail)

5	*Gentrification II*	`gaa > 60th quantile`
		`E_18_25 < 0`
		`DAU5 < 0`
		`DAU10 < 0`
			- Increase in valuation above 60th percentile: the area has become one of the more expensive in the city.
			- Decrease in number of 18-25-year-olds and people who have lived in the area for 5 or 10 years: the final stage of displacement occurs, where remaining longer-term residents move out, along with the younger pioneers who can no longer afford to live in the area.
				5 (pass)
				4 (fail)

The method this classification system uses is best illustrated by an example of some area$_i$. Before any areas are classified, a new variable called `gcode` (short for gentrification code) is created in for 2001, and all areas are assigned `0` (here, zero refers to a somewhat ambiguous pre-pre-gentrification stage). When 2001 areas and their accompanying variables are passed through the filter, the average sale price of residential properties in $area_i$, stored in variable `gaa`, is checked, and, if it is below the 60th percentile, the value of variable `gcode` in row *i* is set to `1`, and otherwise remains `0`. The `gcode` assigned to each area is joined to the next year, 2003, by the area ID, and the data from 2003 is passed through the same filtering mechanism. First, row *i* (containing $area_i$) is checked to see its gentrification status is, and, because $area_i$ was last classified as 1, it is passed through filter 2. This filter then checks whether $area_i$ is experiencing change symptomatic of being in the early pioneer stage of gentrification, and, if it passes through this filter of variables (shown in row two, column three of table 2), variable `gcode` is set to `2`. This process is repeated each year bi-yearly from 2001 to 2019. Additionally, each yearly dataset is checked on whether variables DAU5 … PDAU10 are available to test, and filters with more than two conditions allow areas that meet *1 — n(conditions)* to pass.

A core concept underlying this classification strategy is that an area can only be tested for each successive phase of gentrification if it has reached the one previous. To illustrate this with another example, if $area_j$ has been tested in 2001 and was assigned 1, tested again in 2003 and assigned `2`, and tested again 2005 and again assigned `2` (meaning that it did not pass filter 3), in 2007 it will again be passed through filter 3. Extending this, the highest filter any area can pass through in 2007 is filter 3, meaning that the highest value post-classification in 2007 is `4`. This also underscores another essential side-effect of this classification system, that no areas can go backwards in their classification. The functionality of these assumptions is discussed in-depth in section 6.2, but broadly speaking the construction of a more flexible classification system was beyond the scope of this paper.

4.3: Results

The results of this process are illustrated below in figure 1, a map of Berlin showing the categorization of each planning area bi-yearly from 2001 to 2019. Adjacent to this image is figure 2, a map of Berlin by districts, which can be understood roughly as neighborhoods. Additionally, table 3 below contains the number of areas placed into each stage. While this cartographic display of the data generation process does not necessarily facilitate rigorous insight, the above map provides several interesting insights into gentrification in Berlin via the data generation process. In first map in 2001, it is evident that substantial portions of the city are not classified as gentrifiable because they are above the 60th percentile of land value, notably the central Berlin district of Mitte, almost entirely classified as stage 0 in 2001.

More generally, the transition from areas in different districts from stage 0 to stage 5 is remarkably heterogeneous — by 2019, each district contains at least one stage 4 or 5 area. Perhaps

Figure 1: Cartographic Graph of Generated Gentrification Status by Year

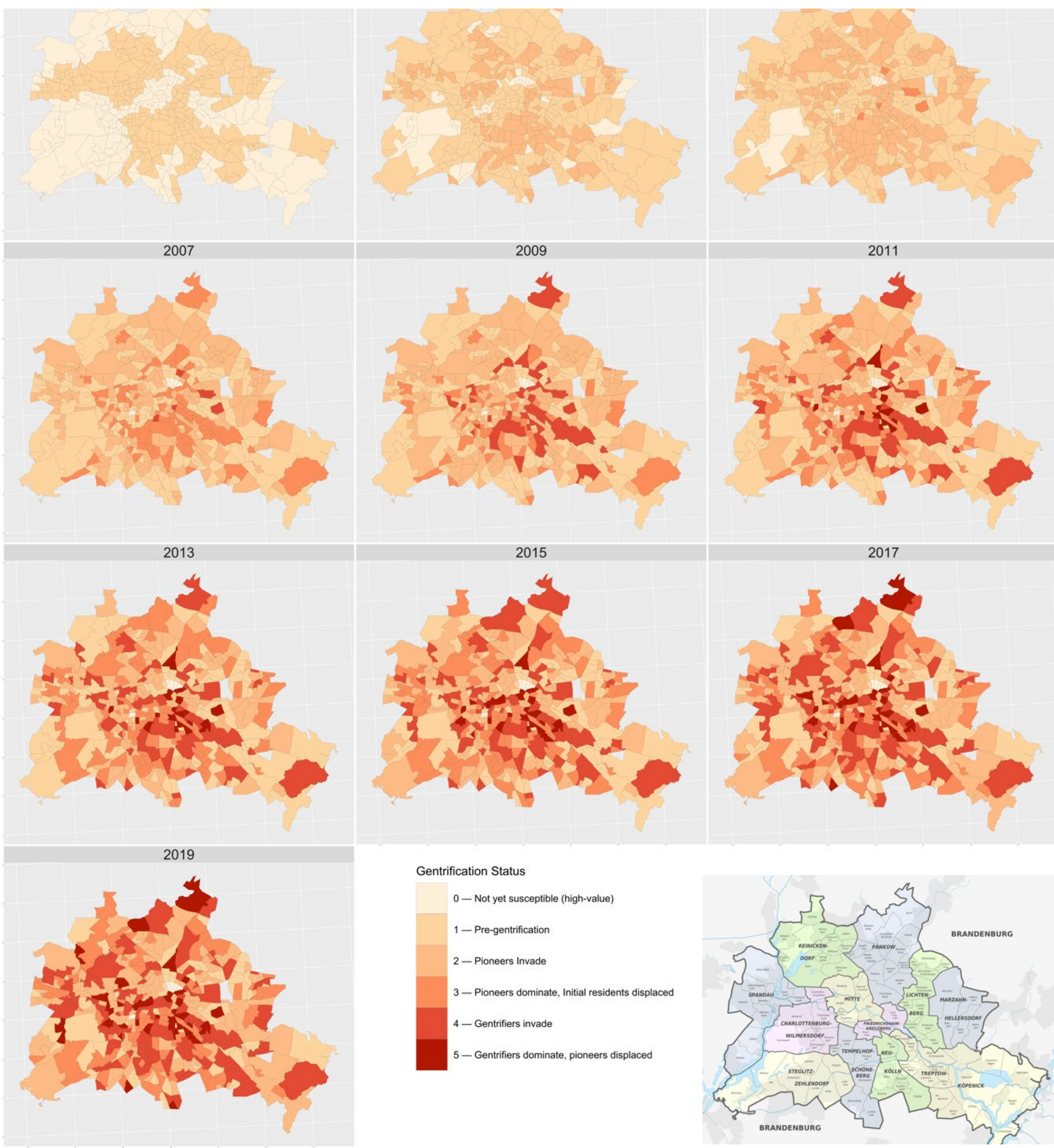


Image Source: Wikipedia Commons

the most concentrated band of gentrification begins in Mitte and moves southeast to the district of Kreuzberg-Friedrichshain, a neighborhood known popularly for its vibrant history cultural history and contemporary struggles with gentrification (Wilder, 2017).

A frequency table of the number of areas classified in each stage in each year is shown below in table 3; at the start of the process, 60% of the areas are classified as pre-gentrification, with the remaining 179 being assigned as `1`, not-yet susceptible to gentrification. The following year, a substantial portion of those initially assigned to `1` are reassigned to `2`, the pre gentrification stage, and almost 30% pass through the filter 2. The more stringent filter is filter 3, with only five units assigned to `3`. By the end of the filtering process in 2019, 8% of planning areas are classified as fully gentrified. Particularly interesting are perhaps the five areas which, even after 2007, remain in gentrification stage 0, meaning that at no point do they fall below the 60th percentile of sale prices for residential properties. While these areas certainly merit further examination beyond the scope of this paper, it is particularly surprising that four of these areas were classified as protected in 2014 when only 20 total protected areas existed in Berlin.

Table 3: Frequency of Classifications by Year

	Stage *[start] →* *0: Not yet gentrifiable*	*→* *1: Pre-gentrification*	*→* *2: Pioneer I*	*→* *3: Pioneer II*	*→* *4: Gentrification I*	*[end]* *5: Gentrification II*
2001	179	268				
2003	41	273	132		*Not yet eligible*	
2005	12	229	200	5		
2007	5	176	171	89	5	
2009	5	133	204	50	54	0
2011	5	123	144	103	56	15
2013	5	99	135	113	75	19
2015	5	80	132	117	88	24
2017	5	68	133	100	111	29
2019	5	57	132	89	127	36

The data generation process results in the classification for each of the 447 planning areas bi-yearly from 2001-2019, data which forms the basis of the program evaluation in chapter 5. The data generated by this classification system results also in several other helpful variables, including the year each area receives each classification, the years each area spends in each stage, and the variables used to classify each area. Finally, it is pausing to consider what level of measurement the generated gentrification-status variable falls into. Gentrification-status is almost certainly properly understood as ordinal — there is a strong implication of ordering, and it is unclear whether the distance between ungentrified and pioneer II (stage `1` to `3`) is the same as the distance between pioneer I and gentrification I (stage 2 to 4). However, as will be explored in later chapters, it is possible (and at times necessary) to understand gentrification-status as a continuous variable that is only observed during its manifestation in discrete values.

4.4: Spatial Autocorrelation Tests

Before I begin using the generated data for program evaluation, I perform one final piece of exploratory data analysis on gentrification in Berlin. The utility of this test is both in an exploratory capacity, to test whether the process of gentrification in Berlin is spatially dependent, and as a precursor to the analysis performed in chapter 5, where spatial dependence or heterogeneity may affect assumptions of independence integral to most statistical analysis. The central question framing this section is whether each area's gentrification status is related to the status of the areas surrounding it.

4.4.1: Methods

To examine the possibility of spatial dependence, I perform a Global Moran's I test. Developed by statistician Patrick Moran in 1948, the Moran's I statistic is used to describe the relationship between the distance between spatial units and the value of an attribute of those units. Broadly, the Global Moran's I statistic tests whether the pattern of attribute values across space compared to a null hypothesis of spatial randomness, meaning that there is no association between distance and similarity in attribute value in a dataset. By comparing each unit's deviation from the mean to the deviation values of neighboring units and obtaining cross-products of this comparison, the Moran's I statistic can show whether values autocorrelate either via clustering, meaning nearby units have more similar values, or dispersion, meaning the values of nearby units are more dissimilar (Anselin, 1996). Notably, there is some variation in the specifics one can use to perform a Global Moran's I test: in defining the spatial relationship between units, one can use k nearest neighbors, contiguity, or inverse distance weighting. To test for spatial autocorrelation against the null hypothesis, one can compare the distribution of attribute values to a normal distribution or perform a series of randomized simulations of attribute values to obtain a

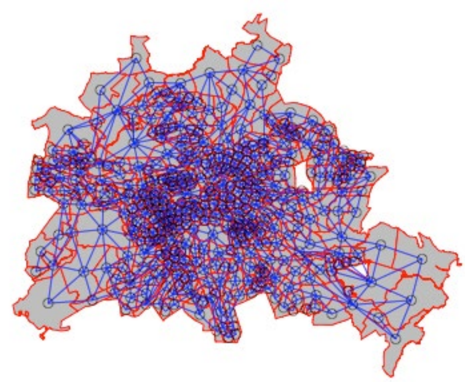


Figure 2: Contiguous neighbors

spatially random distribution which one can compare to the true distribution to obtain a pseudo p-value (Anselin, 2020).

While a visual analysis of figure 1 appears to show some clustering of gentrification status, it is also unclear the extent to which this clustering changes from 2001-2019. As such, I conduct a Moran's I for each year bi-yearly from 2001-2019 on the gentrification status of each area in each year. For each year, I conduct a test defining neighboring polygons by queen contiguity, an illustration of which is shown in figure 3, and a secondary test defining neighbors using inverse-distance weighting according to the distances represented in figure 4. Finally, for all the resulting tests I perform a Monte Carlo permutation test, which represents the second method to obtain a reference distribution of the null hypothesis and thus a pseudo p-value. Using this simulation, I obtain a Global Moran's I for the spatial autocorrelation for each year. Notably, in this exploratory data analysis of autocorrelation, gentrification status is treated as a continuous variable, as no appropriate statistic exists for analyzing ordinal spatial autocorrelation.

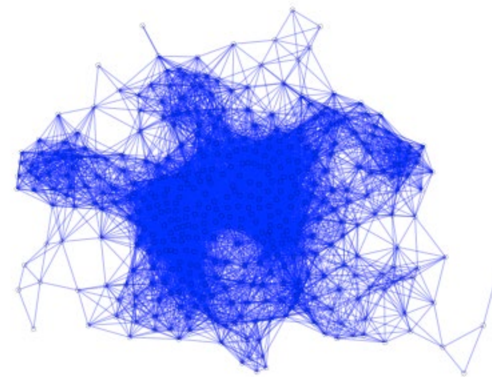


Figure 4: Distances for Inverse-Distance Weighting

4.4.2: Results

The results of the first set of tests by queen-contiguous neighbors are available below in table 4, and the results of the inverse distance weighted tests are shown in table Y. Each year tested achieves a pseudo p-value below 0.01, which passes the conventional threshold to reject the null hypothesis of spatial independence of gentrification in Berlin. Interpretation of the Global Moran's I is based on its nature as a cross-product of the deviation from the mean and the deviation values of neighboring units means that, if overall the statistic is positive, neighboring areas are similarly distant from the mean, compared to a negative statistic which would be generated by more dissimilar neighboring values (with 0 representing the null hypothesis) (Anselin 2020). Broadly, these statistics can be interpreted as confirming that the process of gentrification in Berlin is not spatially independent, and instead that there is the presence of spatial clustering, which in turn implies the presence of spatial dependence in gentrification categorization. Interestingly, the Global Moran's I statistics generated using contiguous neighbors are substantially larger than those produced by the inverse-distance weighted neighbors. The distance between the two autocorrelation statistics is most pronounced during the earliest years, 2001, 2003, and 2005, when the only existent values of gentrification status were 0-3. This could suggest substantially localized clustering effects that quickly decay beyond immediate contiguous neighbors, which is confirmed by the apparent increase in spatial heterogeneity in a visual analysis of later years.

Year	Global Moran's I Statistic	Pseudo p-value	alternative		Year	Global Moran's I Statistic	Pseudo p-value	Alt. Hypothesis:
2001	0.5198	0.005	greater		2001	0.2126	0.005	greater
2003	0.2375	0.005	greater		2003	0.1043	0.005	greater
2005	0.1445	0.005	greater		2005	0.0682	0.005	greater
2007	0.1338	0.005	greater		2007	0.1011	0.005	greater
2009	0.1577	0.005	greater		2009	0.1344	0.005	greater
2011	0.1598	0.005	greater		2011	0.1467	0.005	greater
2013	0.1382	0.005	greater		2013	0.1233	0.005	greater
2015	0.1183	0.005	greater		2015	0.1181	0.005	greater
2017	0.1107	0.005	greater		2017	0.0995	0.005	greater
2019	0.0976	0.005	greater		2019	0.0888	0.005	greater
Average	0.1818	0.005	greater		Average	0.1197	0.005	greater

Global Moran's I for contiguous neighbors *Global Moran's I for IDW neighbors*

Table 4: Global Moran's I Tests

Based on the positive Global Moran's I statistics, spatial dependence must be at least acknowledged as a factor effecting the results of the program evaluation in the following chapter.

Chapter 5: Program Evaluation

To analyze the effects of protection area designation, I run several exploratory analyses and discuss their insight and shortcomings: a DID model, a longitudinal study, and survival-analysis of time spent on each stage of gentrification. This chapter is presented not as a rigorous attempt to ascertain the treatment effects of community protection areas, a task that is complicated by several substantial methodological challenges, notably the presence of spatial dependence and a small sample size of treated areas. Instead, the methods deployed can better be understood as a quasi-exploratory analysis of the utility of the generated gentrification data possible. The challenges to rigorous evaluation and potential future extensions of this data generation process are further discussed in chapter 6.2.

5.1: Data

The data used in this chapter is the data generated in chapter 4, and a full description of the generation process can be found there; to provide a brief overview, each of the 447 areas in Berlin were passed through a series of filters to detect change in demographics or real estate valuations according to which state of gentrification an area was previously classified as. Additional data generated includes the year each area is classified in each stage that it reaches, and, based on this, the amount of time each spent in each stage.

Table 5: Community Protection areas by year

Protected:	No	Yes
2003	440	7
2005	437	10
2007	436	11
2009	436	11
2011	436	11
2013	436	11
2015	427	20
2017	415	32
2019	394	53

To evaluate the effects of community protection areas according to their stated goal of preventing change in the composition of residents due to increases in rent, I download a dataset of each community protection area in Berlin. Luckily, these areas are declared on the planning-area level, allowing for a relatively seamless designation of areas as treated. Table 5 shows the total number of community protection areas present in each year from 2001 to 2019; from this table, it is evident that, in the majority of the period of observation from 2001-2019, only approximately ten areas were designated as protected. A further illustration of the gentrification status of protected areas each year 2001-2019 is shown in table 6 below. Notably, four new areas designated in 2015 are in already very highly valued areas. Nonetheless, there is a relatively even distribution of areas across each year, particularly in 2015-2019. The treatment data illustrated in these tables is treated as generally considered as the independent variable which may have some effect on the dependent gentrification-classification by year.

Year	Stage 0		Stage 1		Stage 2		Stage 3		Stage 4		Stage 5	
	Number	*Percent*	*Number*	*Percent*	*Number*	*Percent*	*Number*	*Percent*	*Number*	*Percent*	*Number*	*Percent*
2001	3	42.9 %	4	57.1 %								
2003	2	28.6 %	4	57.1 %	1	14.3 %			*Not eligible for designation*			
2005	0	0 %	6	60 %	2	20 %	2	20 %				
2007	0	0 %	4	36.4 %	2	18.2 %	3	27.3 %	2	18.2 %		
2009	0	0 %	4	36.4 %	2	18.2 %	1	9.1 %	4	36.4 %	0	0 %
2011	0	0 %	4	36.4 %	1	9.1 %	2	18.2 %	2	18.2 %	2	18.2 %
2013	0	0 %	4	36.4 %	1	9.1 %	2	18.2 %	2	18.2 %	2	18.2 %
2015	4	20 %	5	25 %	2	10 %	5	25 %	2	10 %	2	10 %
2017	4	12.5 %	5	15.6 %	5	15.6 %	7	21.9 %	5	15.6 %	6	18.8 %
2019	4	7.5 %	9	17 %	9	17 %	11	20.8 %	12	22.6 %	8	15.1 %

Table 6: Gentrification Stages of Protected Areas by Year

The final substantial component underlying both methods of program evaluation is the presence of spillover effects. In contemporary literature on evaluation of programs target to spatial units, there are several distinct ways of conceptualizing spillover effects which may violate key assumptions underlying any quantitative analysis (Butts, 2021). Examples of treatment spillovers are discussed further in section 3 of this chapter in the context of difference-in-difference analysis, but broadly speaking, spillover effects accounted for by the presence of an indicator variable for whether an area is within 500 meters of a treated area. Before analysis, the dataset is transformed from a wide format consisting of one row for each area to a long format, also known as panel data, where each row represents the variables associated with a particular area at one of the 10 periods of observation.

5.2: Survival Analysis

Survival analysis attempts to measure the expected duration of some time until some event and how it may be affected by covariates, where substantial portions of the data are in some way censored. Censorship occurs when units it is unknown when units experience this event for a variety of possible reasons. In this section, I apply a survival analysis framework to measure whether treatment affects the time any area takes to progress to the next stage of gentrification, or whether treatment results in increased rates of non-progression (in this context, a positive sign for the effectiveness of the program).

5.2.1: Methods

Survival analysis originated in biostatistics and public health to deal with, classically, measurements of survival rates in terminal diseases such as smallpox, dating back to the 17th century (Anderson and Keiding, 2014). Adapted to medical research and public health in the 20th century by the famous Kaplan-Meier study in 1958, survival analysis is perhaps best illustrated in a classic example of time to death or remission in the case of cancer, where the actual time many study participants will experience these events is unknown, either because the event took place sometime before change was observed, or the final event state is not reached by the end of the study period. The censoring of substantial portions of the observations, censoring which may itself be analytically valuable, mean that uncensored data likely underestimate the true time-to-event and assumptions of normality underlying regression analysis no longer hold. As such, new methods to infer the probability of survival based on the length of time to an event were discovered, perhaps most significantly the Kaplan-Meier survival estimate (Kaplan and Meier, 1958), a statistic which shows the probability that an individual will survive from the starting time to a future time, a survival curve, which plots the probability of survival against increasing time.

To evaluate the effects of protected areas in Berlin, I utilize a slightly more complicated extension of survival analysis called a multi-state model following the framework developed by Therneau and Grambash (2000), as described in a vignette accompanying the `survival` R package (Therneau et al, 2021). While multi-state models can model a range the time to a variety of different related events, such as progression from healthy to diseased to deceased versus straight from healthy to deceased, for this analysis I consider the transition between gentrification stages 1-5[6] and the length of time spent in each state. Using an indicator variable for whether an area was designated as protected while that area was been stage i and $i + 1$ — which I refer to as treatment — I compute two separate Kaplan-Meier survival curves of the effect of time on the probability of being in any one stage and making it to the subsequent stage. Using data visualizations and exploratory data analysis I assess general trends in the relationship between treatment and survival time.

Treatment	*Stage*	*# total*	*# event*	*Restricted Mean*
Untreated	1	1180	0	5.64986805920748
Untreated	2	1180	380	11.0931025185106
Untreated	3	1180	245	1.63694298649564
Untreated	4	1180	158	0.541795961490054
Untreated	5	1180	34	0.0789749369929679
Treated	1	60	0	10.1281417267618
Treated	2	60	4	4.38807586309297
Treated	3	60	7	2.86147382778849
Treated	4	60	5	1.35365723254769
Treated	5	60	2	0.269335812505853

Table 7: Kaplan Meier Survival fits split on treatment

5.2.2: Results

From both the numerical summary of the survival curves and the graph of each curve split into treatment and non-treatment groups, it is evident that, for each successive stage, the probability of moving to the subsequent stage as a function of time decreases. Also apparent from figure 4 is that, for each stage except stage 1-2, the probability of treated areas increasing to the subsequent stage as a function of time is greater than untreated areas, which suggests that treatment is, except in stage 1-2, associated with slower gentrification. Finally, the pronounced stepwise shape of the treated curve for the lower righthand plot merits skepticism but can be explained by the "# *event*" column which shows only two treated areas transitioning from stage 4 to 5 during the period of study. A final plot from the survival analysis is shown on the left in figure 5. This plot can be understood

[6] For this analysis, stage 0 to stage 1 is not considered, as it is not actually a stage of gentrification.

as a more elegant summary of the four plots shown in figure 4, with the addition of the red band titled “event,” corresponds to censoring at any stage.

A visual analysis of figures 4 and 5 also raise interesting and significant questions. It is evident from both figures that for each successive stage, probability over time decreases substantially for both treated and untreated areas. Additionally, figure 5 provides valuable visual context into the relationship between treatment and faster progression to the next stage — the right-hand green, blue, and magenta bands are certainly larger than those to the left, but the difference is far less pronounced than that of the red and mustard bands representing censoring and transition from stage 1-2 respectively. As such, figure 5 suggests that community protection areas slow the transition from pre-gentrification to the first pioneer stage, and general prevent areas from transition to the next stage, as reflected by the substantially increased incidence of censoring by time.

Figure 3: Plots of Survival Curves by Stage

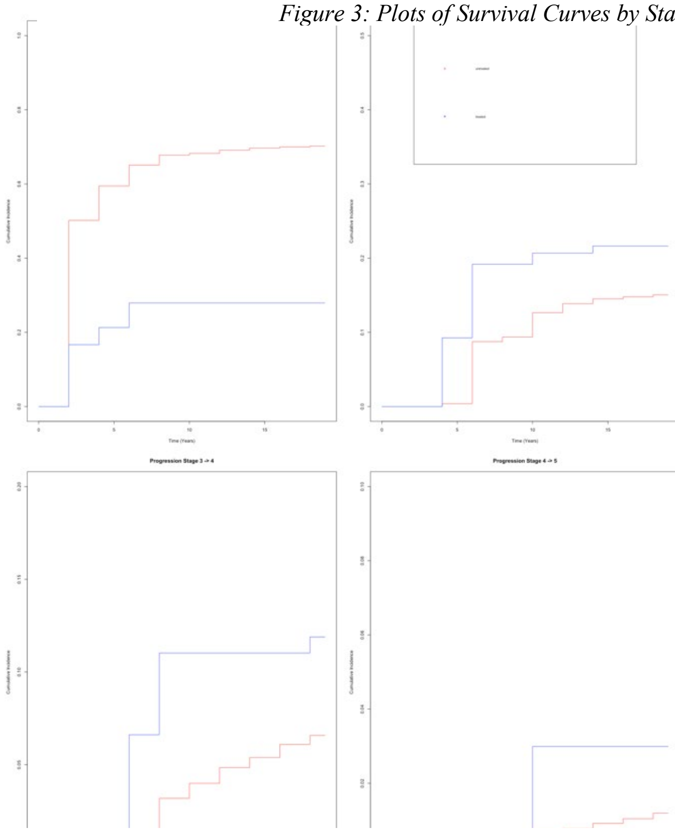


Figure 4: Competing Risk Plot for Multi-State Analysis

5.3 Difference-in-difference

In the second section of my program evaluation, I use a quasi-experimental difference-in-difference (DID) setup to estimate the effect of treatment on gentrification status and account for spillover effects on non-treated and treated units according to recent literature.

5.3.1: Methods

DID estimation, the most popular research design in economics and quantitative social sciences, utilizes real world interventions as a substitution for a randomized experiment (a method which emerged 85 after the invention of DID) to ascertain the effect of some treatment on the outcome of those treated (Cunningham, 2021). Theoretically, the effect of treatment on units could be given by the difference between the observed outcome of treated units and the counterfactual outcomes of those same units had they not been treated. Because the true value of this counterfactual is necessarily impossible to observe, one instead designates a set of control units who never receive treatment to substitute for this counterfactual, and then accounts for any existent differences between the control group and the treatment group exogenous to treatment by taking the difference between the average outcome variable of interest in the control units and the treated units before treatment and after treatment. The difference between control and treatment groups before and after treatment constitutes the average treatment effect on the treated (ATT).

The use of control groups as to model the counterfactual outcomes of treated units means that, if the distance between the control group and the treatment group in the variable of interest changes between the pre-treatment period and the post-treatment period due to unknown factors, the estimated ATT will instead reflect some unknowable portion of this change. As such, the validity of ATT relies on the "parallel trends" assumption, which states that some unobserved variable which changes over time results in control units and treated converging or diverging in the outcome variable of interest. Strongly related is the strict assumption of exogeneity in treatment assignment, meaning that units are not assigned treatment because of some unobserved characteristic.

In the past several decades, researchers have become increasingly interested in causal inference methods such as DID in spatial contexts, where treatment is assigned not to individuals but to areal units. This emergent wealth of methodological literature explores how to estimate and account for spatial relationships among observations that may impact the stable unit treatment value assumption (SUTVA) (Rubin, 1974), a crucial assumption which posits that the outcome of any unit is unrelated to the treatment of any other unit (Angrist et al, 1996). Examples of violations of this assumption in spatial contexts are hardly difficult to imagine, chiefly in the form of spillover effects. Illustrations of spillover effects beyond those specific to the evaluation of community protection areas in Berlin are abundant but, following Butts (2021), can generally be reduced to the presence of two additional variables in the mathematical estimation of treatment effect. First, treatment in one location can affect those surrounding it, such rural residents who may travel several counties over to receive discounted primary care due to a new government program, as in Bailey and Goodman-Bacon (2015). In this example, if the outcome of interest is mortality or disease rates, counties adjacent to treated areas may also have benefitted from the program and thus bias the calculation of ATT towards underestimation. To illustrate the bias treatment spilling over onto other treated areas, one can take the classic example of a spatially designated manufacturing program which produces regional agglomeration effects which act as a further boon to the regional economy. In this example, increased concentration of treated areas produces even stronger treatment effects could lead to the overestimation of treatment effects.

To account for these and other potential issues created by spatial interaction between units in causal inference literature, researchers have proposed a variety of methods. Kolak and Anselin (2020) provide a thorough overview of contemporary causal inference methodologies and how spatial dependence (SD), meaning the increased relatedness of nearby units versus further units, and spatial heterogeneity (SH), meaning the increased dissimilarity of nearby units, can violate important underlying assumptions. The authors review contemporary literature and provide a range of methods to explicitly account for spatial effects in treatment assignment and relevant variable outcomes. Examples include Delgado's (2015) proposed an extension of DID design which decomposes treatment effects into direct effects on treated units and indirect effects on neighboring units and the addition of state and year based fixed effects on Du Mouchel, Williams, and Zador (1987) evaluation of state-level alcohol purchase age laws.

Butts (2021) argues that an indicator variable for being within *x* distance of a treatment area can estimate spillover effects on control and treatment units while correcting ATT, provided the indicator variable correctly captures how spillover effects vary by intensity. Butts evaluates various methods for constructing this indicator variable using Monte Carlo simulations, finding that a set of concentric ring indicators providing the best balance of measuring spillover effects while correcting the ATT estimate, and articulates criteria on which researchers can base their decision to use additive or binary proximity-to-treatment indicator variable. Lastly, Butts discusses the complications posed by multiple treatment groups and variation in treatment timing and provides thorough documentation of the best practices for creating indicator variables which correct for spillover effects. Butts' discussion of variable treatment timing builds off recent work by economists Callaway and Sant'Anna (2019), who provide a rubric to working with multiple time periods and variation in treatment timing where simple comparisons in outcomes cannot deliver an ATT. The authors also provide an R package `did` to easily deploy their methodology with little additional data wrangling.

Evaluating the effects of community protection areas on gentrification in Berlin using my typologized data, I draw heavily on Butts (2021) and Callaway and Sant'Anna (2019). As shown in chapter 5.1, substantial variation exists in the years that protection status was adopted by different areas in Berlin, making the classical two-period DID structure unfeasible. In constructing my indicator variable, I follow Butts (2021) argument that use of a continuous indirect-treatment indicator is not possible for variable treatment timing models such as Callaway and Sant'Anna (2019), and instead create a categorical indicator which, in the longitudinal dataset of each area in Berlin in each of the 10 observation periods from 2001-2019, reflects whether an area is (1) a control area not within 500 meters of a treatment area, (2) a control area within 500 meters of a treatment area, or (3) a treatment area within 500 meters of a treatment area. In order to increasing the odds of agreeing with the parallel trend's assumption, I follow Kline and Moretti (2014) and run a logistic regression to predict being a treated area at some point from 2001-2019 using data from the starting period 2001 and drop observations below the 25th percentile of predicted probability. The product of these methodological decisions allows me to evaluate the ATT of community protection areas in Berlin according to the gentrification data I produced, while also controlling for bias induced by spillover effects onto control and treatment areas and increasing the chances of the non-violation of the parallel trend's assumption.

5.3.2: Results

The results produced by this DID estimation of ATT are best illustrated visually in figure 6 below, though a summary of the first DID test performed using the original treatment status and an indicator for being within 500 meters of a protection area as a covariate is also pictures is shown in table 8 to the left. Examining table 6 to the right reflects how the `did` function calculates the treatment effect of each group over each time interval with sufficient observations. The average treatment effect is for each group is shown in the third column, values which are largely clustered around 0 with a small smattering of slightly significant groupings, shown by an asterisk on the rightmost side of the table.

The plot below provides further insight into how spillover effects might influence the value of ATT, but the substantial heterogeneity and lack of a substantial pattern in either biasing the ATT estimate up or down provides little information into how spillover effects work in this context. Broadly speaking, the summary of the test and the graph reflect a distribution that, despite remarkably high variance, can hardly be said to deviate substantially from the null hypothesis that ATT is non-zero. It is safe to say that, based on this DID analysis, there is insufficient information to determine if community protection areas play any role in gentrification status.

Group-Time Average Treatment Effects:

Group	Time	ATT(g,t)	Std. Error	[95% Simult.	Conf. Band]	
2014	2007	-0.6811	0.1466	-1.0645	-0.2978	*
2014	2009	-0.5743	0.1774	-1.0380	-0.1105	*
2014	2011	-0.2617	0.1093	-0.5475	0.0241	
2014	2013	-0.0431	0.0518	-0.1786	0.0925	
2014	2015	-0.0352	0.0417	-0.1443	0.0738	
2014	2017	-0.1498	0.1095	-0.4362	0.1367	
2014	2019	-0.2084	0.1156	-0.5106	0.0938	
2015	2007	-0.1538	0.2520	-0.8128	0.5053	
2015	2009	0.5545	0.2377	-0.0670	1.1761	
2015	2011	-0.2694	0.0349	-0.3608	-0.1781	*
2015	2013	0.8482	0.0243	0.7847	0.9118	*
2015	2015	0.0688	0.2404	-0.5598	0.6974	
2015	2017	-0.0446	0.2335	-0.6551	0.5659	
2015	2019	-0.1513	0.2362	-0.7689	0.4663	
2016	2007	0.1058	0.3038	-0.6885	0.9001	
2016	2009	-0.0812	0.2184	-0.6523	0.4899	
2016	2011	0.5035	0.1804	0.0318	0.9751	*
2016	2013	-0.1134	0.0381	-0.2129	-0.0139	*
2016	2015	-0.0924	0.0271	-0.1634	-0.0215	*
2016	2017	-0.1120	0.0267	-0.1819	-0.0422	*
2016	2019	-0.2004	0.0371	-0.2974	-0.1033	*
2017	2007	0.3540	0.1975	-0.1625	0.8705	
2017	2009	0.2883	0.2189	-0.2842	0.8608	
2017	2011	0.4416	0.2086	-0.1038	0.9869	
2017	2013	-0.1518	0.0243	-0.2153	-0.0882	*
2017	2015	-0.1234	0.0216	-0.1799	-0.0670	*
2017	2017	-0.1134	0.0236	-0.1751	-0.0518	*
2017	2019	-0.2237	0.0344	-0.3138	-0.1337	*
2018	2007	-0.1083	0.1715	-0.5567	0.3400	
2018	2009	0.0527	0.1588	-0.3625	0.4679	
2018	2011	0.0432	0.2609	-0.6389	0.7253	
2018	2013	0.1786	0.1790	-0.2895	0.6468	
2018	2015	-0.0936	0.0254	-0.1600	-0.0271	*
2018	2017	-0.1120	0.0264	-0.1810	-0.0430	*
2018	2019	-0.0854	0.0323	-0.1699	-0.0010	*
2019	2007	0.1238	0.2035	-0.4085	0.6560	
2019	2009	0.2429	0.1991	-0.2777	0.7636	
2019	2011	-0.0749	0.2355	-0.6907	0.5410	
2019	2013	-0.0753	0.1053	-0.3507	0.2001	
2019	2015	0.1499	0.1471	-0.2347	0.5345	
2019	2017	-0.0260	0.1166	-0.3308	0.2789	
2019	2019	-0.0354	0.0661	-0.2082	0.1375	

Signif. codes: `*' confidence band does not cover 0

Control Group: Never Treated, Anticipation Periods: 0
Estimation Method: Doubly Robust

Table 8: DiD Call Summary from R

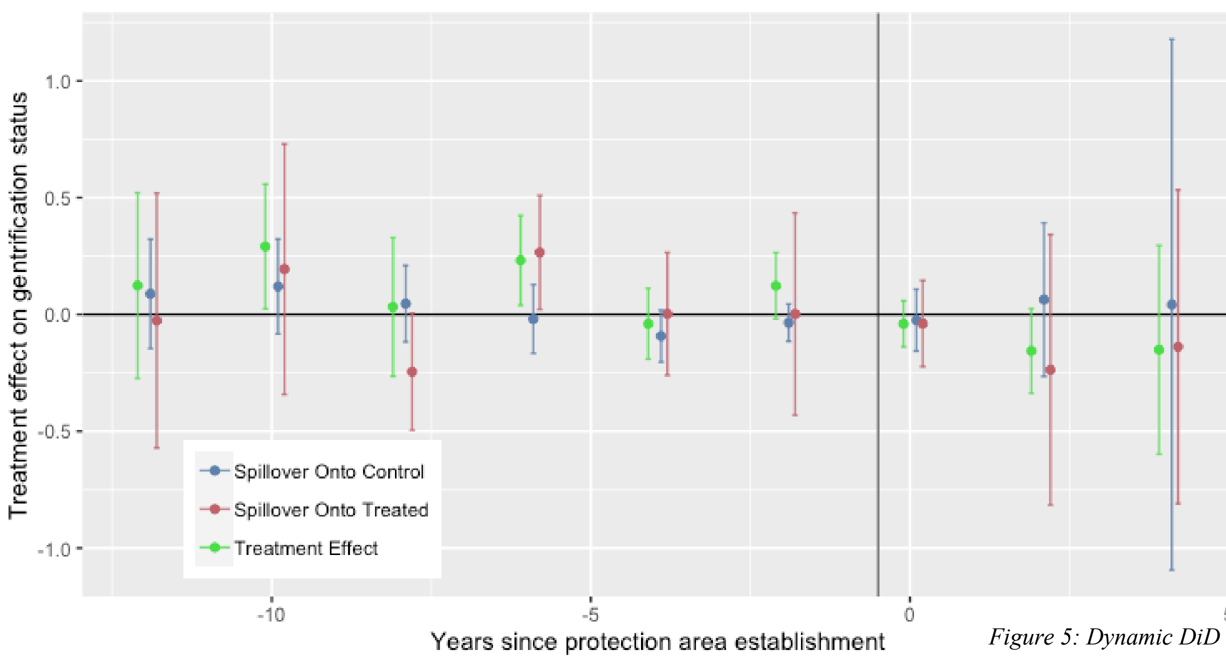


Figure 5: Dynamic DiD test results

Ch. 6: Discussion

6.1: Discussion of Data Generation

In this section I return to the data generation process described in chapter 4 to discuss how my results compare to those produced by other quantitative analysis of gentrification in Berlin and how future work can build on the conceptually driven approach I took.

As described in the literature review, there have been several substantial attempts to quantitatively analyze gentrification in Berlin. Looking first to the *Gentrimap* project produced by Holm and Schulz (2018), there are several noticeable divergences in their results evident in the final maps of their gentrification index. To reiterate, the authors quantified gentrification by examining the change of real estate valuation and the proportion of residents receiving welfare from 2007 to 2014 to create a social index of change and real-estate index of change, which are combined in the gentrification-index. However, the authors utilize a different dataset to measure change in real-estate valuations, which provides data at a more spatially disaggregated neighborhood-level scale on average rent asking prices. This may affect their results, a possibility that merits future research. Visually, the difference between the two maps is striking in several ways. First, because the areas in *Gentrimap* are classified at a higher LOR level (n=60), similarity between adjacent areas is far more apparent. Additionally, where my data generation process classified areas south of Mitte as generally more gentrified and tended to regard the northwestern portion as largely in stage 0 for a substantial portion of the generation process, the gentrification status of the *Gentrimap* is reversed along the north-south axis, with large bands in the north classified as somewhat gentrified. Most noticeable is the heterogeneity shown in my classification system compared to Holm and Schulz (2018).

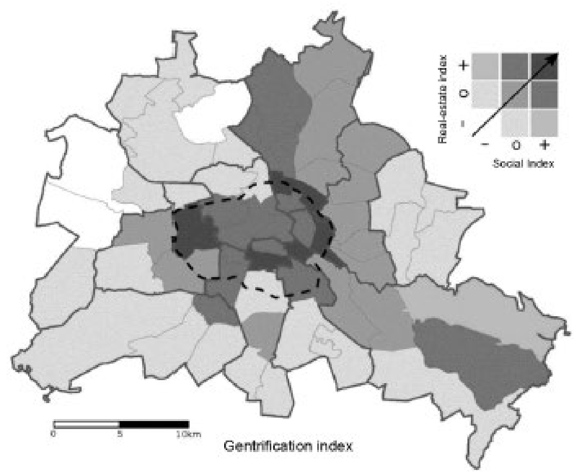


Figure 6: Holm and Schulz (2018) Gentrimap

Schulz (2017) also independently examined the relationship between price increases and displacement; while not entirely the same as gentrification, use of similar indicator variables and more involved quantitative approach, which deploys clustering and other advanced analysis techniques, merits comparison to my results. Schulz's (2017) results are similar to Holm and Schulz (2018) but display more spatial heterogeneity. Schulz (2017) also further emphasizes the gentrification status of the Mitte district, which, for substantial portions of my data generation process, is classified as ungentrifiable. Broadly speaking, this map appears more like my final version from 2019 than Holm and Schulz (2018). It is also worth noticing how Schulz (2017) and Holm and Schulz (2018) both create a continuous scale for their classification, as opposed to ordinal variables, which, while certainly easier to deploy in quantitative modeling, can also enable easier smoothing across areas and selective break-setting. Döring and Ulbrecht (2017), not pictured here, is also similar to both of the previously mentioned articles in their strong identification of Mitte as highly gentrified and increased contiguity.

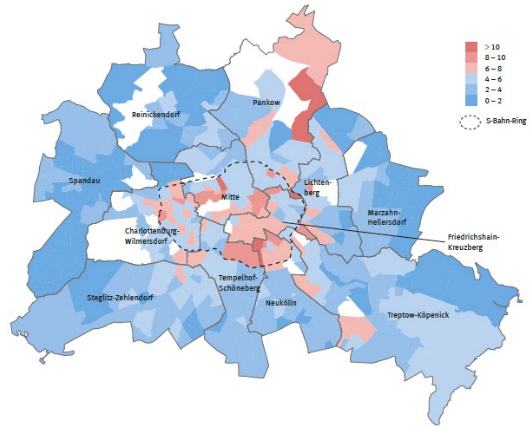


Figure 7: Schulz (2017) Displacement Map

Comparison to these different approaches of gentrification raise several interesting questions surrounding the merits and drawbacks of my data generation process. Generally, comparison to other approaches illustrates that while my approach allows for more precise identification of what stage of gentrification any area may be experiencing rather than a blanket index, the selective filtering can result in areas can be identified as not gentrified. The divergence of my model from other methods also reflects ambiguity surrounding gentrification as its popularly conceived and how it may, in truth, manifest — in other words, areas that are most popularly understood as gentrifying may not be the areas that, when all available variables are considered, are particularly characteristic of gentrification. It was this tension, in fact, that provided the motivation both to better understand gentrification in Berlin using a more articulated process-based classification system, and to explore the effectiveness of community protection areas in hindering displacement.

The substantial divergence of the results produced by my classification system also suggest underlying data quality issues; during the data acquisition, loading, cleaning, and combining process, I discovered several changes in how variables were calculated over each year, such a shift in German welfare laws in 2007 which generated far higher rates of those receiving welfare in previous years, or changes in how foreigners were defined. In chapter 4.2, I mentioned that, because each area is passed through the filtering system every two years, an inconsistency resulting in faulty data in one year will not have a particularly substantial impact on the final results. However, the opposite may also be true — because the data is only classified into five gentrification stages, lag from a single year can mean a 20% deviation from the correct classification for that year, which, if applied across the entire dataset, is hardly trivial. Moreover, the possibility exists where a hypothetical area is misclassified in 2005 but experiences sufficient demographic change over the next two years such that it no longer passes through the appropriate filter and thus correct identification is further prevented. While it is impossible to know exactly how often these misclassifications are happening, the perspective presented by other gentrification models provides a benchmark against which my method could be systematically compared.

Comparison to other methods also raises interesting possibilities surrounding alternative approaches to change — it would certainly be possible to pass datasets through the filter than measure change over larger timespans, which could provide further verification that an area has experienced some stage of gentrification, with the tradeoff of possibly missing minute but significant demographic change in between those years.

Before moving to a discussion of the results of the program evaluation, it is worth considering the divergence from other models considering the question of how useful this data generation process is in detecting gentrification, and how data can be made more robust to these changes while retaining its ability to look at how different demographic groups are affected. Broadly stated, my conclusion from this data generation process is that it represents a novel attempt at large-scale quantitative evaluations of gentrification — its value is not in the creation of an algorithm which detects gentrification with unparalleled accuracy, but rather as a first step into a more complex look at urban processes with quantitative methods that provides a foundation for future work. I look forward to discussing potential improvements of this methodology in the future, including what additional data sources can be integrated, how further emphasis can be placed on seeing which demographic groups effected during each stage, how more precise filters can be designed, and how to better compare other strategies to measure gentrification.

6.2: Discussion of Program Evaluation

To review, I used survival analysis and a difference-in-difference approach to test the effects of community protection areas on gentrification in Berlin and the utility of data generated in chapter 4. Broadly speaking, neither method provided substantial insight into if or how community protection areas affected gentrification in Berlin. While I touched on this briefly in chapter 5.1, perhaps the largest hurdle to evaluating Berlin's community protection areas, is the extremely small sample size. Though the number of areas designated has doubled since 2015 and future analysis may prove more fruitful, it is still difficult to gain insight into the effects of this policy also because of the ambiguous quality of my generated data and the lack of covariates — because I drew from all publicly available data sources on the planning-area level, the decreased variance and additional insight provided by continuous variables was unavailable. Another substantial factor affecting the validity and inferential value of both approaches is the strong presence of spatial dependence shown in chapter 4.4. This spatial clustering of gentrification can be understood as resulting from a mismatch between the spatial scale at which gentrification occurs and the units at which I measure gentrification. The inclusion of a spatially lagged dependent variable could be implemented as described by Anselin (1988).

Results from the survival analysis seemed to cautiously suggest an association between treatment status and a slower rate of transition to the next stage, but further analysis is needed to test the robustness of the models and validate of their assumptions, such as the proportional hazards assumption. The relationship between time and probability of moving into a subsequent stage shown in figure 5 raised interesting questions to me – it is unclear whether the logarithmic shape of the cumulative incidence graph is due to the change the effects of time on survival over time, whether it is due to a truncated data generation process which is too strict in its identification of later stages of gentrification, if there is a pattern of stagnation in certain stages of gentrification for certain areas, or if unobserved effects happened to hinder gentrification during the periods of observation. In future research, this dataset and program evaluation could be integrated into spatial survival analysis to identify areas with particularly extended timespans in, for example, the early stages of gentrification, or how the distribution of protection areas relates spatially to survival patterns of each stage (see Taylor, 2017).

In future research, there are numerous extensions to the survival analysis that could provide further insight into how protected status affects the rate of gentrification. Perhaps the most obvious is the use of a Cox proportional hazards model (Cox, 1972) to provide a coefficient to treatment status on the odds of survival to the next stage. The Cox proportional hazards model identifies the coefficients of covariates, such as treatment status, on a hazard function, defined as the risk of a new event being observed at any time *t*. The use of this model also requires the verification of several assumptions, most important of which is the assumption of proportional hazards, which states the effects of any covariates remain constant over time. An exciting further extension of this model would be the inclusion of covariates which could capture spillover effects of protection policies. an area of survival analysis that is currently unexplored. These new variables could include an indicator for whether $area_i$ is within 500 meters of an area in treatment and if that $area_i$ is itself in treatment or not (meaning there are three possible outcomes for any area in any year), and the number of treatment areas with 50 meters, 100 meters, and 500 meters. In the context of gentrification in Berlin, a process whose spatial dependence was demonstrated through tests for autocorrelation in chapter 4.4, an example of spillover effects from treatment onto nearby areas would be as follows: in a larger neighborhood experiencing gentrification, supply-side actors exploited unprotected areas adjacent to protected areas to capitalize on future valuation increases, thus increasing the speed of gentrification. In the context of the Cox proportional hazards model and survival analysis, this spillover effect could result in the dummy covariate for treatment being incorrectly identified as a good prognostic factor (meaning a variable which reduces the possibility of advancement to a next stage of gentrification). Additional covariates could be added to the Cox model provided interesting insight into how spatial spillover might be depicted in survival analysis, and there exists substantial potential for further research to link the nascent field of spatial survival analysis and gentrification policy interventions more explicitly.

While the DID analysis represents an exciting intersection between spatial econometrics and my original data, it is apparent from all the results of this data that the variability produced by the extremely small number of treated areas for almost every treatment group makes it extremely difficult to ascertain exactly whether protection areas effect gentrification, or if there are any agglomeration effects produced by protection areas. However, several adaptations could be made to create a model more robust to the small treatment sizes (Conley and Taber, 2011), and while I use a logit regression on treatment status to explicitly account for the parallel trends assumptions, further testing using event-study regressions could further validate this assumption (Callaway and Sant'Anna, 2020). Kolak and Anselin (2020) also provide direction on causal inference methodologies such as regression discontinuity and propensity score matching, and panel data estimators for spatial data have recently become available through the `splm` package.

Beyond regression analysis and causal inference methods, future research could use spatial clustering analysis on the gentrification data generation alongside treatment data. Potential research questions cluster analysis might include identifying clusters of particularly fast gentrification, the presence of points including amenities or crime, and spatial patterns of the gentrification process pre and post treatment.

In conclusion, the results of this paper — namely the extension of a conceptually driven, process-based approach to quantifying gentrification to the policy evaluation context vis-a-vis regression analysis and causal inference — highlight numerous opportunities to integrate exciting new conceptual and methodological approaches. Potential directions for future research include conventional regression analysis using one-way ANOVA tests, further exploration into the other covariates e.g. consumption/amenity distribution, the implementation of survival analysis to measure the effect of protected status on time spent in each stage of gentrification, or the extension of longitudinal study methods to include further time-lagged treatment variables.

Bibliography

Ahlfeldt, Gabriel M. 2010. *Blessing or Curse? Appreciation, Amenities and Resistance around the Berlin "Mediaspree."* Hamburg Contemporary Economic Discussions 32. Hamburg: University of Hamburg, Chair for Economic Policy. http://hdl.handle.net/10419/42240.

Andersen, Per Kragh, and Niels Keiding. 2014. "Survival Analysis, Overview." In *Wiley StatsRef: Statistics Reference Online*. American Cancer Society. https://doi.org/10.1002/9781118445112.stat06060.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–55. https://doi.org/10.1080/01621459.1996.10476902.

Anselin, Luc. 1996. "The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association." *Spatial Analystical Perspectives on GIS*, GISDATA; 4, , 111–25.

———. 2020. "Global Spatial Autocorrelation." Geoda: An Introduction to Spatial Analysis. October 9, 2020. https://geodacenter.github.io/workbook/5a_global_auto/lab5a.html.

"Apartment Rental Offers in Germany." n.d. Accessed March 2, 2021. https://kaggle.com/corrieaar/apartment-rental-offers-in-germany.

"Atlas of Urban Expansion - Berlin." n.d. Accessed March 10, 2021. http://www.atlasofurbanexpansion.org/cities/view/Berlin.

Baléo, Marie. 2019. "Berlin: The Challenge of Affordable Housing in a City of Low-Income Tenants." *La Fabrique de La Cité*, April 6, 2019. https://www.lafabriquedelacite.com/en/publications/berlin-the-challenge-of-affordable-housing-in-a-city-of-low-income-tenants/.

Barton, Michael. 2014. "An Exploration of the Importance of the Strategy Used to Identify Gentrification." *Urban Studies* 53 (1): 92–111. https://doi.org/10.1177/0042098014561723.

Been, Vicki, Ingrid Gould Ellen, Michael Gedal, Edward Glaeser, and Brian J. McCabe. 2016. "Preserving History or Restricting Development? The Heterogeneous Effects of Historic Districts on Local Housing Markets in New York City." *Journal of Urban Economics* 92: 16–30. https://doi.org/10.1016/j.jue.2015.12.002.

Berg, Tobias, Markus Reisinger, and Daniel Streitz. 2020. "Handling Spillover Effects in Empirical Research." SSRN Scholarly Paper ID 3377457. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.3377457.

"Bericht Monitoring Soziale Stadtentwicklung Berlin 2019 / Land Berlin." 2019. Monitoring Soziale Stadtentwicklung. Berlin: Senatsverwaltung für Stadtentwicklung und Wohnen.

Bivand, Roger, and David W. S. Wong. 2018. "Comparing Implementations of Global and Local Indicators of Spatial Association." *TEST* 27 (3): 716–48.

Bland, J Martin, and Douglas G Altman. 2004. "The Logrank Test." *BMJ (Clinical Research Ed.)* 328 (7447): 1073–1073. https://doi.org/10.1136/bmj.328.7447.1073.

Blasius, Jörg, Jürgen Friedrichs, and Heiko Rühl. 2016. "Pioneers and Gentrifiers in the Process of Gentrification." *International Journal of Housing Policy* 16 (1): 50–69. https://doi.org/10.1080/14616718.2015.1071029.

Bünger, Reinhart. 2018. "Berlin verändert Förderung für Wohnungsneubau." *Der Tagesspiegel*, January 3, 2018. https://www.tagesspiegel.de/wirtschaft/immobilien/sozialer-wohnungsbau-berlin-veraendert-foerderung-fuer-wohnungsneubau/20798178.html.

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2020. "Difference-in-Differences with Multiple Time Periods."

Campbell, John Y., Andrew W. Lo, and A.Craig MacKinlay. 1997. "Event-Study Analysis." In *The Econometrics of Financial Markets*, 149–80. Princeton University Press. https://doi.org/10.2307/j.ctt7skm5.9.

Chapple, Karen, and Miriam Zuk. 2016. "The Use of Neighborhood Early Warning Systems for Gentrification and Displacement." *Cityscape* 18 (3): 109–30.

Comber, Alexis, and Wen Zeng. 2019. "Spatial Interpolation Using Areal Features: A Review of Methods and Opportunities Using New Forms of Data with Coded Illustrations." *Geography Compass* 13 (10): e12465. https://doi.org/10.1111/gec3.12465.

Cox, D. R. 1972. "Regression Models and Life-Tables." *Journal of the Royal Statistical Society. Series B (Methodological)* 34 (2): 187–220.

Dangschat, Jens S. 1988. "Gentrification: Der Wandel Innenstadtnaher Wohnviertel." In *Soziologische Stadtforschung*, edited by Jürgen Friedrichs, 272–92. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-322-83617-5_14.

"Der Wohlstandseffekt der Hauptstadt." n.d. Accessed December 16, 2020. https://www.iwd.de/artikel/der-wohlstandseffekt-der-hauptstadt-294893/.

Dobberke, Cay. 2019. "'Vorläufiger Milieuschutz' für 40.000 Wohnungen | Namen & Neues | Tagesspiegel LEUTE Charlottenburg-Wilmersdorf." *Der Tagesspiegel*, February 1, 2019.

https://leute.tagesspiegel.de/charlottenburg-wilmersdorf/macher/2019/02/01/71048/vorlaeufiger-milieuschutz-fuer-40-000-wohnungen.

Döring, Christian, and Klaus Ulbricht. 2018. "Gentrification Hotspots and Displacement in Berlin. A Quantitative Analysis." In *Gentrification and Resistance: Researching Displacement Processes and Adaption Strategies*, edited by Ilse Helbrecht, 9–35. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-20388-7_2.

During, Rainer. 2001. "Spandau: Streit um die Wasserstadt." *Der Tagesspiegel*, September 1, 2001. https://www.tagesspiegel.de/berlin/spandau-streit-um-die-wasserstadt/252888.html.

Easton, Sue, Loretta Lees, Phil Hubbard, and Nicholas Tate. 2019. "Measuring and Mapping Displacement: The Problem of Quantification in the Battle against Gentrification." *Urban Studies* 57 (2): 286–306. https://doi.org/10.1177/0042098019851953.

Eddy, Melissa. 2020. "Berlin Freezes Rents for 5 Years in a Bid to Slow Gentrification." *The New York Times*, January 31, 2020, sec. World. https://www.nytimes.com/2020/01/31/world/europe/berlin-gentrification-rent.html.

Engelbrecht, Sebastian, and Tobias Krone. n.d. "Dauerstreit um Mietendeckel - Kampf um bezahlbares Wohnen." *Deutschlandfunk*. Accessed November 29, 2020. https://www.deutschlandfunk.de/dauerstreit-um-mietendeckel-kampf-um-bezahlbares-wohnen.724.de.html?dram:article_id=487954.

Ewert, Benjamin. 2016. "Poor but Sexy? Berlin as a Context for Social Innovation." In *Social Innovations in the Urban Context*, edited by Taco Brandsen, Sandro Cattacin, Adalbert Evers, and Annette Zimmer, 143–58. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-21551-8_8.

Fields, D., and S. Uffer. 2014. "The Financialisation of Rental Housing: A Comparative Analysis of New York City and Berlin." *Urban Studies* 53 (7): 1486–1502. https://doi.org/10.1177/0042098014543704.

Geneid, Shahinaz. 2019. "Gentrification and the Creative City: Lessons Learned from Berlin Urban Planning Policy." *SIA NYUAD* (blog). March 8, 2019. https://medium.com/sia-nyuad/gentrification-and-the-creative-city-lessons-learned-from-berlin-urban-planning-policy-79c8d98e90b4.

Geßner, Martin. 2008. "Leistungsfähigkeit des städtebaulichen Instruments Milieuschutz für die Stadtentwicklung in Berlin." *Technische Universität Berlin, Fakultät VI - Planen Bauen Umwelt*.

Doctoral Thesis, Berlin: Technische Universität Berlin. https://doi.org/10.14279/depositonce-2029.

Gimond, Manuel. 2021. *Intro to GIS and Spatial Analysis*. https://mgimond.github.io/Spatial/spatial-autocorrelation.html.

Glaeser, Edward L. 2010. "Preservation Follies: Excessive Landmarking Threatens to Make Manhattan a Refuge for the Rich." *City Journal*, 2010. https://www.city-journal.org/html/preservation-follies-13279.html.

Gornig, Martin, and Hartmut Häussermann. 2002. "Berlin: Economic and Spatial Change." *European Urban and Regional Studies* 9 (4): 331–41. https://doi.org/10.1177/096977640200900404.

Guterbock, Thomas M. 1980. "The Political Economy of Urban Revitalization: Competing Theories." *Urban Affairs Quarterly* 15 (4): 429–38. https://doi.org/10.1177/107808748001500405.

Haffner, M., M. Elsinga, and J. Hoekstra. 2012. "Access and Affordability: Rent Regulation." In *International Encyclopedia of Housing and Home*, edited by Susan J. Smith, 40–45. San Diego: Elsevier. https://doi.org/10.1016/B978-0-08-047163-1.00209-5.

Hake, Sabine. 2008. *Topographies of Class: Modern Architecture and Mas Society in Weimar Berlin*. Ann Arbor: University of Michigan Press.

Helweg, Dennis. n.d. "Untersuchung von Gentrifizierung am Beispiel Berlin mittels Big Data Analytics," 197.

Henger, Ralph, Pekka Sagner, and Michael Voigtländer. 2019. : ": Milieuschutz in Berlin." IW-Gutachten. Cologne: Institute der Deutschen Wirtschaft.

Hildburg Bruns. 2019. "At A Glance: Which Berlin Districts Are Going to Get More Expensive." *Bild Zeitung*, November 19, 2019. https://www.bz-berlin.de/bz-english-edition/at-a-glance-which-berlin-districts-are-going-to-get-more-expensive.

Hiljmans, Robert. n.d. "Interpolation — R Spatial." R Spatial — Spatial Data Science. Accessed March 26, 2021. https://rspatial.org/raster/analysis/4-interpolation.html.

Hinze, Annika Marlen, and James M. Smith. 2020. "Comparative Approaches to Development Politics and the Quest for a More Robust Theory: New Institutions of Governance in Chicago and Berlin." *PS: Political Science & Politics* 53 (1): 10–14. https://doi.org/10.1017/S1049096519001379.

Holm, Andrej. 2016. "Gentrification Und Das Ende Der Berliner Mischung." In *Wohnen*, edited by Eberhard von Einem, 191–231. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-11757-3_11.

Holm, Andrej, and Guido Schulz. 2018. "GentriMap: A Model for Measuring Gentrification and Displacement." In *Gentrification and Resistance: Researching Displacement Processes and Adaption Strategies*, edited by Ilse Helbrecht, 251–77. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-20388-7_10.

Jenkins, Blair. 2009. "Rent Control: Do Economists Agree?" *Econ Journal Watch* 6 (1): 40.

Kecskes, Robert. 1994. "Gentrification: Eine Klassifikation von Wohnungsnachfragern Auf Dem Wohnungsmarkt." *ZA-Information / Zentralarchiv Für Empirische Sozialforschung*, no. 35: 27–48.

Kettunen, Hanna, and Hannu Ruonavaara. 2020. "Rent Regulation in 21st Century Europe. Comparative Perspectives." *Housing Studies*, May, 1–23. https://doi.org/10.1080/02673037.2020.1769564.

Kinahan, Kelly L. 2019. "The Neighborhood Effects of Federal Historic Tax Credits in Six Legacy Cities." *Housing Policy Debate* 29 (1): 166–80. https://doi.org/10.1080/10511482.2018.1452043.

Kline, Patrick, and Enrico Moretti. 2013. "People, Places and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs." Working paper 19659. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w19659.

Knight, Ben. 2021. "Berlin Rent Cap Overturned by Germany's Top Court." *Deutsche Welle*, April 15, 2021. https://www.dw.com/en/berlin-rent-cap-overturned-by-germanys-top-court/a-57209268.

Koschinsky, Julia. 2009. "Spatial Heterogeneity in Spillover Effects of Assisted and Unassisted Rental Housing." *Journal of Urban Affairs* 31 (3): 319–47. https://doi.org/10.1111/j.1467-9906.2009.00462.x.

Lee, Barrett A. 2007. "Invasion-Succession." In *The Blackwell Encyclopedia of Sociology*. American Cancer Society. https://doi.org/10.1002/9781405165518.wbeosi073.

Lees, Loretta, and David Ley. 2008. "Introduction to Special Issue on Gentrification and Public Policy." *Urban Studies* 45 (12): 2379–84.

McCabe, Brian J. 2019. "Protecting Neighborhoods or Priming Them for Gentrification? Historic Preservation, Housing, and Neighborhood Change." *Housing Policy Debate* 29 (1): 181–83. https://doi.org/10.1080/10511482.2018.1506391.

Micheli, Martin, and Torsten Schmidt. 2013. "Immobilienboom Verstärkt Den Anstieg Der Mieten. (German)." *Orientierungen Zur Wirtschafts- Und Gesellschaftspolitik*, no. 136 (June): 23.

Ojiambo, P. S., and E. L. Kang. 2013. "Modeling Spatial Frailties in Survival Analysis of Cucurbit Downy Mildew Epidemics." *Phytopathology* 103 (3): 216–27. https://doi.org/10.1094/PHYTO-07-12-0152-R.

"Ordinal Logistic Regression | R Data Analysis Examples." n.d. Accessed April 23, 2021. https://stats.idre.ucla.edu/r/dae/ordinal-logistic-regression/.

"ORDINAL LOGISTIC REGRESSION | R DATA ANALYSIS EXAMPLES." n.d. UCLA Insititute for Digital Research & Education: Statistical Consulting. Accessed April 22, 2021.

O'Sullivan, Feargus. 2017. "Struggling With Rising Rents, A Berlin Borough Buys Out Private Landlords." *Bloomberg Citylab*, August 11, 2017. https://www.bloomberg.com/news/articles/2017-08-11/berlin-s-rent-control-plan-buy-out-private-landlords.

———. 2018. "Can Berlin Make Space for 200,000 New Homes?" *Bloomberg Citylab*, November 27, 2018. https://www.bloomberg.com/news/articles/2018-11-27/where-can-berlin-build-the-affordable-housing-it-wants.

———. 2020. "How Berlin's Mietskaserne Tenements Became Coveted Urban Housing." *Bloomberg Citylab*, January 13, 2020. https://www.bloomberg.com/news/articles/2020-01-13/the-story-behind-berlin-s-iconic-home-design.

Peto, R., M. C. Pike, P. Armitage, N. E. Breslow, D. R. Cox, S. V. Howard, N. Mantel, K. McPherson, J. Peto, and P. G. Smith. 1977. "Design and Analysis of Randomized Clinical Trials Requiring Prolonged Observation of Each Patient. II. Analysis and Examples." *British Journal of Cancer* 35 (1): 1–39. https://doi.org/10.1038/bjc.1977.1.

Preis, Benjamin, Aarthi Janakiraman, Alex Bob, and Justin Steil. 2020. "Mapping Gentrification and Displacement Pressure: An Exploration of Four Distinct Methodologies." *Urban Studies*, February, 0042098020903011. https://doi.org/10.1177/0042098020903011.

Reick, Philipp. 2017. "Gentrification 1.0: Urban Transformations in Late-19th-Century Berlin." *Urban Studies* 55 (11): 2542–58. https://doi.org/10.1177/0042098017721628.

Reiß, Anke, and Jeannette Raethel. 2015. "Fokus Berlin: Die „Creative Class" - Eine Unterschätzte Kundengruppe Für Die Immobilienwirtschaft? (German)." *E-Journal of Practical Business Research*, no. 17 (April): 1.

Ringelstein, Ronja. 2020. "Schärfere Regeln für Umwandlung in Eigentumswohnungen bleiben." *Der Tagesspiegel*, February 4, 2020. https://www.tagesspiegel.de/berlin/milieuschutz-in-berlin-schaerfere-regeln-fuer-umwandlung-in-eigentumswohnungen-bleiben/25507138.html.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688–701. https://doi.org/10.1037/h0037350.

Schäfer Philipp and Braun Nicole. 2016. "Misuse through Short-Term Rentals on the Berlin Housing Market." *International Journal of Housing Markets and Analysis* 9 (2): 287–311. https://doi.org/10.1108/IJHMA-05-2015-0023.

Schönball, Ralf. 2018. "Großdemo für bezahlbaren Wohnraum." *Der Tagesspiegel*, April 14, 2018. https://www.tagesspiegel.de/berlin/mietenwahnsinn-in-berlin-grossdemo-fuer-bezahlbaren-wohnraum/21174856.html.

———. 2020. "Mieten in Berlin steigen in fünf Jahren um 44 Prozent." *Der Tagesspiegel*, July 22, 2020. https://www.tagesspiegel.de/berlin/untersuchung-zum-wohnungsmarkt-mieten-in-berlin-steigen-in-fuenf-jahren-um-44-prozent/26026418.html.

Schulz, Guido. 2017. "Aufwertung und Verdrängung in Berlin - Räumliche Analysen zur Messung von Gentrifizierung," 12.

Siemer, Julia, and Keir Matthews-Hunter. 2017. "The Spatial Pattern of Gentrification in Berlin," 9.

"Soziale Erhaltungsgebiete (Milieuschutzgebiete)." 2020. November 13, 2020. https://www.berlin.de/ba-mitte/politik-und-verwaltung/aemter/stadtentwicklungsamt/stadtplanung/staedtebaufoerderung/erhaltungsgebiete/milieuschutzgebiete-492487.php.

Sturtevant, Lisa. n.d. "The Impacts of Rent Control: A Research Review and Synthesis." Washington, DC: National Multifamily Housing Council.

Terry M. Therneau and Patricia M. Grambsch. 2000. *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

Therneau, Terry M. 2021. "A Package for Survival Analysis in r." Manual. https://CRAN.R-project.org/package=survival.

Thomas Weinert. 2017. "Gentrifizierung in Berlin - Von Mieterfrust und Straßenschlachten." *Deutschlandfunk*, April 6, 2017. https://www.deutschlandfunk.de/gentrifizierung-in-berlin-von-mieterfrust-und.724.de.html?dram:article_id=387839.

Thomschke, Lorenz. 2016. "Distributional Price Effects of Rent Controls in Berlin: When Expectation Meets Reality." CAWM Discussion Paper 89. Münster: Westfälische Wilhelms-Universität Münster, Centrum für Angewandte Wirtschaftsforschung (CAWM). http://hdl.handle.net/10419/147489.

Ulrich, Paul. 2018. "Hier will Berlin fast 200.000 neue Wohnungen bauen - Ärger in Pankow." *Berliner Zeitung*, November 14, 2018. https://www.berliner-zeitung.de/mensch-metropole/hier-will-berlin-fast-200000-neue-wohnungen-bauen-aerger-in-pankow-li.7208.

"W4_Anselin1996.Pdf." n.d. Accessed May 13, 2021. https://dces.wisc.edu/wp-content/uploads/sites/128/2013/08/W4_Anselin1996.pdf.

Wensierski, Peter. 2011. "Gentrification's Victims: Berlin Fears Rise of New Slums." *Spiegel*, March 2, 2011, sec. International. https://www.spiegel.de/international/germany/gentrification-s-victims-berlin-fears-rise-of-new-slums-a-748532.html.

Werwath, Peter. 1998. "Comment on David Listokin, Barbara Listokin, and Michael Lahr's 'The Contributions of Historic Preservation to Housing and Economic Development.'" *Housing Policy Debate* 9 (3): 487–95. https://doi.org/10.1080/10511482.1998.9521305.

Wilder, Charly. 2017. "In Berlin, a Grass-Roots Fight Against Gentrification as Rents Soar." *The New York Times*, March 18, 2017, sec. World. https://www.nytimes.com/2017/03/18/world/europe/berlin-rent-fight-against-gentrification.html.

"Wirtschafts- und Innovationsbericht 2019/2020." 2020. Berlin: Senatsverwaltung für Wirtschaft, Energie und Betriebe Abteilung Wirtschaft. https://www.berlin.de/sen/wirtschaft/wirtschaft/konjunktur-und-statistik/wirtschafts-und-innovationsbericht/.

Zerofsky, Elisabeth. 2019. "The Causes and Consequences of Berlin's Rapid Gentrification." *The New Yorker*, July 19, 2019. https://www.newyorker.com/news/dispatch/the-causes-and-consequences-of-berlins-rapid-gentrification.

Zhou, Haiming, Timothy Hanson, and Jiajia Zhang. 2018. "SpBayesSurv: Fitting Bayesian Spatial Survival Models Using R." *ArXiv:1705.04584 [Stat]*, April. http://arxiv.org/abs/1705.04584.

N.d.

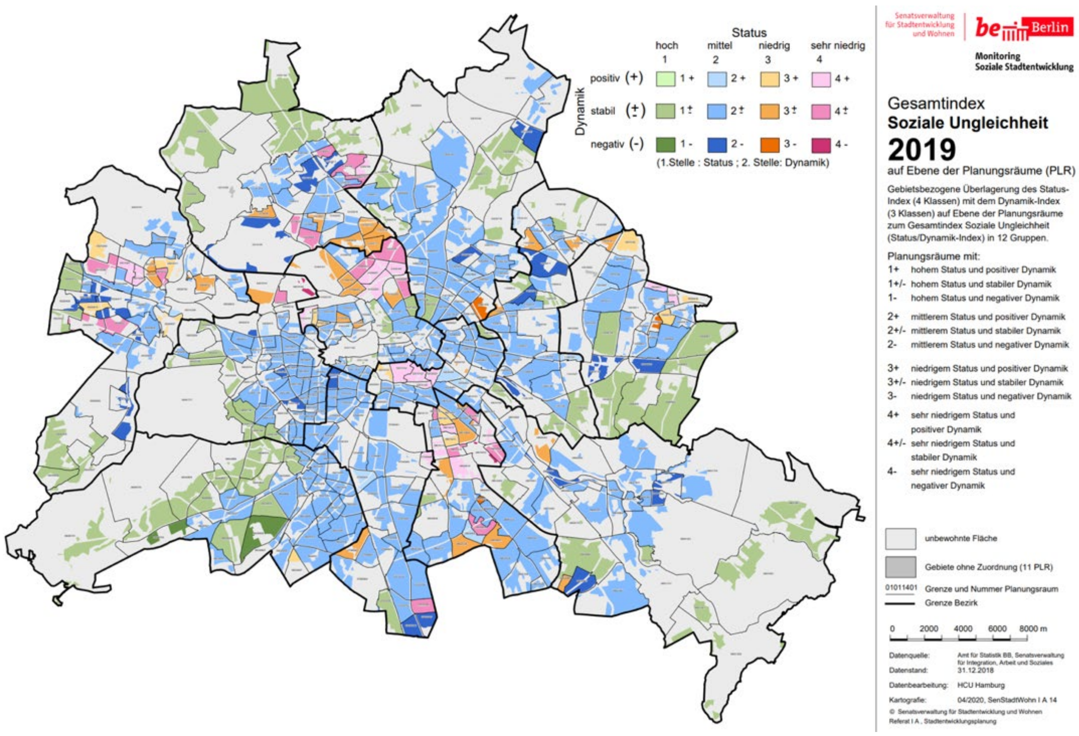
Appendices

Appendix A: Glossary of German Terms Used

- *Angebotsstruktur* — amenity structure; a term which describes the state of local commercial offerings in an area.
- *Einwohnerreggister* — occupant register; a dataset maintained by the city of Berlin detailing the address of every occupant and apartment.
- *Erhaltungsgebiet* — conservation region; a region in which a *Erhaltungsverordnung* has been applied.
- *Erhaltungsverordnungen* — conservation ordinance; policy tool which restricts building changes in designated areas.
- *Lebensweltlich orientierte Räume* — environmentally oriented areas, abbr. LOR; system of intersecting geographical areas of increasing granularity from district to planning region.
- *Mietendeckel* — rent cap; instituted in 2020.
- *Mietkaserne* — rental barracks; a popular form of housing in Berlin constructed as tenement-style apartment blocks for buildings.
- *Mietpreisbremse* — rent price break; a 2015 national law which limits rent increases per year.
- *Mietspiegel* — rent index; a tool which describes the rent per square meter of any apartment based on the age of the building, its facilities/amenities, etc.
- *Milieuschutz* — community defense/protection; colloquialism describing *Erhaltungsgebiete.*
- *Monitoring Soziale Stadtentwicklung* — monitoring social city development, abbr. MSS; a bi-yearly report detailing demographic changes on the *Planungsraum* level in Berlin.
- *Planungsraum,* plural *Planungsräume* — planning area; the smallest level of the LOR system with 488 across Berlin.
- *Senatsverwaltung für Stadtentwicklung und Wohnen* — Senate Administration for Social City Development; the governmental body in Berlin primarily responsible for urban development policy.
- *Statistische Gebiete* — statistical areas; a now defunct geographical system for governmental statistical analysis of Berlin.

- *Vorkaufsrecht* — right of first refusal/preemption; the ability of the Berlin city government to purchase a property when it is listed for sale and turn it to affordable housing before another buyer purchases it
- *Prognoseraum*, plural *Prognoseräume* — prognosis area; the least granular level of the LOR system, with 60 regions across Berlin.

Appendix B: MSS 2019 Social Inequality Map on the Planungsräume level



Translation notes:

- Status: *hoch* = high, *mittel* = middle, *niedrig* = low, *sehr niedrig* = very low
- Dynamic: *positive* = positive, *stabil* = stabile, *negativ* = negative

Appendix C: Table of Variables Available by Year

Year	n03	n05	n07	n09	n11	n13	n15	n17	n19
ID Variable	RAUMID	RAUMID	RAUMID	RAUMID	RAUMID	RAUMID	RAUMID	RAUMID	RAUMID
Population	E_E	E_E	E_E	E_E	E_E	E_E	E_E	E_E	E_E
# 18-25s	E_18U25	E_18U25	E_18U25	E_18U25	E_18U25	E_18U25	E_18U25	E_18U25	E_18U25
# 25-55s	E_25U55	E_25U55	E_25U55	E_25U55	E_25U55	E_25U55	E_25U55	E_25U55	E_25U55
# under 1s	E_U1	E_U1	E_U1	E_U1	E_U1	E_U1	E_U1	E_U1	E_U1
# 1-6s	E_1U6	E_1U6	E_1U6	E_1U6	E_1U6	E_1U6	E_1U6	E_1U6	E_1U6
Migration change	WA	WA	WA	WA	WA	WA	WA	WA	WA
% from EU	eu	eu	eu	eu	eu	eu	eu	eu	eu
% foreign not from EU*	aus_noneu	aus_noneu	aus_noneu	aus_noneu	aus_noneu	aus_noneu	aus_noneu	aus_noneu	aus_noneu
% on welfare	welf	welf	welf	welf	welf	welf	welf	welf	welf
% unemployed	unemp	unemp	unemp	unemp	unemp	unemp	unemp	unemp	unemp
Real estate value	gaa	gaa	gaa	gaa	gaa	gaa	gaa	gaa	gaa
# lived in area >10 years	NA	NA	NA	DAU10	DAU10	DAU10	DAU10	DAU10	DAU10
# lived in area >5 years	NA	NA	NA	DAU5	DAU5	DAU5	DAU5	DAU5	DAU5
% lived in area >10 years	NA	NA	NA	PDAU10	PDAU10	PDAU10	PDAU10	PDAU10	PDAU10
% lived in area >5 years	NA	NA	NA	PDAU5	PDAU5	PDAU5	PDAU5	PDAU5	PDAU5

* This variable is calculated by combining the proportion of residents that are from Turkey, Arab states, Poland, former Soviet states. It is understood as representing the proportion of vulnerable ethnic minorities in an area.