Open Data and Human-Based Outsourcing Neighbourhood Rating: A Case Study of the San Francisco Bay Area Gentrification Rate

Eleanna Panagoulia[1]

[1]University of California, Berkeley, California/USA · eleannapan@berkeley.edu

Abstract: The past decade has experienced a staggering rise of data-aided analysis that facilitate understanding the impact of socio-economical flux and socially oriented activities towards the quality and liveability of space. Evaluating urban environments is not only important from the planners' perspective, but has larger implications for the residents themselves. In this paper we argue that the liveability of a city or a neighbourhood is not necessarily described by conventional, authoritative data, such as income, crime, education level etc., rather ephemeral data layers, related to human perception, can be more effective in capturing the dynamics of space. Implementing methods that are considered disassociated with urban analytics, we attempt to go beyond the conventions in understanding the dynamics that drive socio-economical phenomena and construct lived space. Our objective is to create methods of anticipating and evaluating urban environment by re-patterning different datasets and taking advantage of their combinatory potential.

Keywords: Data-aided analysis, neighbourhood rating, open-data, human-based outsourcing

1 Introduction

Urban design practice has been significantly influenced by the development of tools and platforms that have set the scene for new ways of understanding place and space. Current data streams have allowed planners to view the city as a constantly transforming and unpredictable environment. This paper will focus on the development of alternative urban computing methods that are non-deterministic and attempt to provide new insight in mapping the complexity of the urban environment. The research builds on the term of mapping as James Corner defines it. Corner suggests ways in which mapping acts may emancipate potential and reveal hidden, invisible layers (CORNER 1999). Under this scope, we create a case study that focuses on mapping the gentrification rate in the San Francisco Bay Area; a phenomenon that establishes rapidly in the area; hence, it is considered a challenge to visualize due to its dynamical and complex character. This phenomenon is a subject of debate, as there is extended conflict and consensus whether it is beneficial or not to the broader community. Gentrification recalibrates terms such as safety, affordability, aesthetics etc. towards upgraded standards that may be regarded as improvement on the surface, but have a negative effect on lower income society. Gentrification alters not only the social fabric, but also the physical makeup of a neighbourhood, such as its identity, local culture, indigenous characteristics and traditions preserved by the local residents. In this paper, we are attempting to break this complexity down into its constituent parts and find points of leverage to map and visualize urban development. We implemented a data-driven process utilizing multiple databases that offer opportunities to author urban data through subjective observation and crowd sourced survey techniques. We formulated an accumulative analysis that consists of three methods that operate at different scales, regional and local, in an attempt to show, through the lens of new and old ideas, how the city can be better understood nowadays. In order to provide an informative

Journal of Digital Landscape Architecture, 2-2017, pp. 113-124. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-629-1, ISSN 2367-4253, e-ISSN 2511-624X, doi:10.14627/537629012.

framework for our research on data-driven mapping methods, we will analyse the groundwork of some fundamental topics and problematics that influenced our approach on data and design, by providing key terminology and by analysing precedent studies on the field. As mentioned before, this research operates at multiple scales, including that of neighbourhood. In the following section, we will analyse terms such as neighbourhood, gentrification and displacement, and attempt to provide an understanding of their interrelation. Finally, we will emphasize the role of data in mapping processes and its contribution to the current state of the art.

2 Key Terminology

2.1 Neighbourhood and Neighbourhood Change

The notion of neighbourhood is one that planners and scholars usually presuppose as consistent; however, its role in the urban environment has been debated upon in the past. The neighbourhood has come to be understood as the physical building block of the city for both "social and political" organization (SAMPSON 2011), and thus, combines physical and non-physical attributes. Early scholars defined neighbourhoods as closed ecosystems, characterized by their physical elements, such as size, density, demographics etc. that would get disrupted by external factors, such as new residents. Based on these theories, neighbourhood change was a natural process of population relocation and competition for space, until a state of new equilibrium is established. These ideas about neighbourhood presented a deterministic model, where neighbourhoods could be categorized based on simplified criteria such as their residents' financial status etc. However, neighbourhoods are not introverted, autonomous clusters and the mechanisms of neighbourhood change do not rely on exclusively external factors. According to Jane Jacobs, nowadays people identify a neighbourhood by a landmark in the city because it has become intimate from daily use or encounter (JACOBS 1961). For her, the key that creates the notion of a neighbourhood is diversity and identity. The stability of a neighbourhood relies on its capacity to absorb opportunities and sustain its diverse character. In this paper, the term neighbourhood can be described as an instance of organized complexity (JACOBS 1961), a network of numerous connections, where transformations can occur unexpectedly. A diverse mix of people and processes, with its own self-organizing dynamic.

2.2 Gentrification and Displacement

Gentrification is one category of neighbourhood change and is broadly defined as the process of improving previously deteriorated neighbourhoods by the middle or upper class, often by displacing low-income families and small businesses. The first documented use of the term "gentrification" (GLASS 1964) describes the influx of a "gentry" in lower income neighbourhoods. Owens identifies nine different types of neighbourhoods that are experiencing upgrading: minority urban neighbourhoods, affluent neighbourhoods, diverse urban neighbourhoods, no population neighbourhoods, new white suburbs, upper middle-class white suburbs, booming suburbs and Hispanic enclave neighbourhoods (OWENS 2012). Gentrification does not only rely on a singular cause, as it may emerge when more than one condition is present. It is a complicated process that does not rely on binary and linear explanations. Gentrification

does not result in negative effects, as it can also be regarded as a tool for revitalization. When revitalization occurs by existing residents, who seek to improve their neighbourhood conditions, the result can be constructive in enforcing the neighbourhood stability. This condition is called incumbent upgrading or "unslumming" as Jane Jacobs defines it (JACOBS 1961). When revitalization causes displacement of current residents and a decline in neighbourhood diversity, then neighbourhoods gradually become segregated by income, due in part to macro-level increases in income inequality as well as decline of job opportunities (FREEMAN 2004). Hence, neighbourhood stability is compromised because the opportunities have been narrowed down to a very limited range of financial status and lifestyle. Displacement is identified as the biggest negative impact of concern resulting from neighborhood revitalization and gentrification. Displacement occurs when any household is forced to move from its residence, usually because of eviction and unaffordable rent increase. However, tracking unwilling displacement can be challenging to categorize, as researchers have faced limitations regarding data availability and comprehension.

3 Precedent Studies on Gentrification Detection

There have been several precedent studies that aim at identifying gentrification rate and its consequences in several American cities since the late 1970s. In this section, we will briefly depict some main methods previous researchers used, as well as their strengths and weaknesses. Early researchers analysed gentrification under a binary, rather simplified scope, under solely macro-level capital accumulation or micro-level sociological processes of individual preferences. Yet their methods did not take into consideration politics, as they viewed the process as a natural neighbourhood succession, where property changes hands and residences become displaced. In addition, their surveys suffered from data limitation, short span timeframes and a canvas that did not convey the details and the complexity of the real situation. Research on gentrification and displacement waned in the late 1980s and early 1990s, as researchers came to study gentrification as both a revitalization process, as well as a cause of displacement. They shared methods of the previous literature, combined now with more access to detailed datasets, allowing for the introduction of more advanced statistical techniques in an attempt to tease out the independent effects of gentrification on residential displacement. Many of these studies also pay much closer attention to the impacts of displacement on neighbourhood scale rather than studying displacement of the general population. More recent analyses span larger timeframes to get a better understanding of resident movement in and out of gentrifying neighborhoods.

All surveys suffered from the fact that the results masked a great deal of heterogeneity between urban areas and even within the Census tracts. This resulted from deficiencies in the data sets and short time-scale of the analysis, factors that designated the low predictive capacity of the models and the insufficiency to fully understand neighbourhood dynamics. Although varied in their approaches, questions and results, one consistent finding across these studies is that movers in gentrifying tracts were more likely to be higher income, college educated and younger. This came down to depicting certain categories as indicative that the process of gentrification was already underway: a) shift in tenure, b) influx of households interested in urban living, and c) increase in high income serving amenities such as music clubs, coffee shops, galleries, etc. d) rise of educational level. It is important to note that the above categories summarize quantitative data sources only. Even when data sets allow track-

ing of individual households, they do not provide a sufficient response to measure displacement. For instance, the reason for a household to move to a different neighbourhood may rely on subjectivity, which is hard to quantify. Moreover, data on many of the drivers and impacts of gentrification and displacement are not regularly gathered, hence they may not capture all the changes even in the categories they represent. It is therefore important to explore the implications of the data limitations and to consider qualitative sources of information to better understand the drivers and impacts of neighbourhood change.

4 Recent Computational Tools – The Role of Data

In the early 2000s, several urban analytics models incorporated computational tools that introduced automation, in order to simulate relationships among the urban space. Tools are divided into the ones focusing on representing the movement of individuals and households into spatial patterns of settlement tend to be specified through "agent-based models," also referred to as "multi-agent systems", and into the ones focused on capturing inter-related patterns of change among spatially fixed entities, such as housing units or entire neighborhoods, tend to be specified through cellular automata. Urban simulation models are guided by a set of specified rules that simulate decision making, that perform in a simplified environment disconnected from real facts, thus they may not capture complex gentrification dynamics. One explanation for this is the difficulty of adequately incorporating the breadth of social theory needed to account for the range of gentrifying mechanisms. For instance, even the simulation of the relationships that occur in a park of a business district neighborhood during day and night time, quickly becomes a complicated problem to simulate. These models are constrained by their inability to theoretically ground mechanisms of neighborhood change and translate them into a set of rules. They are limited by lack of empirical detail, both in their specifications of agent attributes, as well as in their specification of neighborhood choice and parcel change mechanisms. As cities are becoming more networked, more data is being generated about the urban environment and its residents, allowing urban designers to access the local scale fabric of the city, opening up new research directions for understanding the city. Going beyond traditional data sources, such as Census, which is fairly static and updated only every 10 years, we encourage designers to engage with other types of real life data that capture the ephemeral side, such as, people's desires, problematics, trends etc. It is important to recognize the opportunities for making better sense of public space through technology. One of the key benefits of adopting a data-driven approach to urban analytics surveys is the ability to see a combination of datasets in context with each other, and to detect temporal and spatial patterns. The following section describes the case study that attempts to visualize neighbourhood change and gentrification in the San Francisco Bay Area.

5 Case study

5.1 Introduction

As mentioned in the main introduction, this case study will focus on visualizing the gentrification rate in the San Francisco Bay Area. The case study involves a combination of three different methods in an attempt to provide a holistic understanding of the flux in the urban environment. This case study will use a series of maps at different, suitable scales to visualize

all the data collected from all three methods. We employed James Corner's principals in combining data sets in a creative way that could uncover realities previously unseen even in exhausted, over-studied areas, such as the Bay Area (CORNER 1999). In order to get a better understanding of the overlapped data sets, we first created a geo-located 3D space in the software processing, where multiple data sets can be displayed at the same time. The visual communication of the information was based on Tufte's principal regarding the clarity of information displayed on the map. He argues against using excessive decoration in visual displays of quantitative information and encourages the use of data-rich illustrations that when examined closely, gives a value to every data point, but when looked at more generally, only trends and patterns can be observed (TUFTE 1990). In particular, changes of degree in a factor are displayed with a gradient of the same colour, changes of type are displayed with different colours, and the general vocabulary of visual styles is communicated with dots, lines and areas.

5.2 Authoritative Data Approach

The first method is based on a Census GIS data analysis to identify the areas that have altered their character in the last 10 years, based on authoritative parameters associated with gentrification, such as tenure status, land value, income and employment rate. The Geographic Information System, or GIS, allows for very fast accumulation of Census data that represent multiple categories relevant to our study. However, for domain specific categories, such as population, income, educational level, transport etc., surveys are conducted every few years, using a limited spatial and temporal sampling framework. As a first step, we identified all the green areas, parks etc. in the entire Bay Area and excluded them from the calculations, as they would have compromised the results of the survey. The initial survey was done for the county of San Francisco based on an assumption that most changes would occur there. We created a series of maps that show a range of household income, a range of home value, owner occupied housing, vacant lots and the ratio of unemployed population to the total population. Soon it became apparent that the most suitable scale for this kind of data set display is an urban scale, that of the entire Bay Area, because this data has low spatial resolution and hence, refers to large-scale surveys, where comparison would make more sense. As a second stage of the process, we re-collected data for the entire San Francisco Bay Area. The Census data that we collected consists of a combined data set from 2000 to 2012 that compares tenure, median household income, median home value and employment rate. Through calculations, we generated the delta of these values and remapped the values in a greyscale range of 0-255. In order to enrich the process, we added an additional layer of information, that of artists' employment rate. Artists' communities are considered highly associated with gentrification rates. Previous surveys in the field have established artists as agents of urban gentrification, for the reason that low-income artists tend to revalorize unproductive spaces, since they are affordable, and, as a result, increase the attractiveness of urban space. Artists make the first move into post-industrial, post-welfare neighbourhoods. Soon they attract the hipster movement before eventually being displaced by them and their new middle-class neighbours. Both participate in the cycle of exploring and developing new potential sites for capital investment. Hence, the combined data set of the other categories is overlapped with artist employment rate Census data. All the relevant data was collected from government websites in .csv format, then imported to Microsoft Excel for the calculations to be performed (delta calculation) and then re-exported in .csv spreadsheets that were imported in Grasshopper and visualized in the Processing model space created for this purpose (Figure 1).

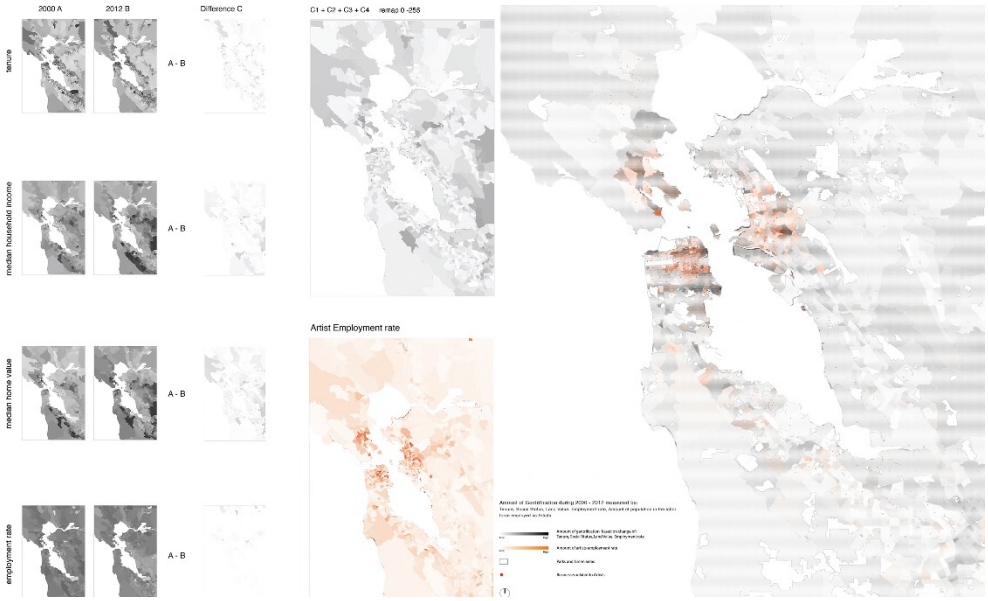

Fig. 1: San Francisco Bay Area, Census GIS data comparison from 2000 to 2012 (left) and Census data overlapped with businesses related to artists from Google Places (right)

5.3 Ephemeral Data Approach

The second method operates at a local level analysis, in San Francisco and Oakland respectively. The data resources for this research derive from open data platforms, such as Google API, Google Places and collective, open-data platforms where users post all kinds of requests (sell and buy, real estate etc.), such as "craigslist.org". Our database is articulated by tracing certain population categories that reflect potentialities about gentrification. The first category involves artists and their recent activity in the San Francisco Bay Area. The same logic as in the first method is applied in this method as well. The artist population is considered as the frontline of gentrification. The second category involves the homeless population rates in the same period. Gentrification rate is also highly associated with eviction rate, since it caters to an environment that is affordable only by higher-income clientele, leaving out those individuals and families who face eviction and live on the brink of homelessness, applying for shelters in those areas. In gentrified areas, low-income families soon face a significant rise in rent cost, combined with reduced chances of job advancements. The main difference of this method of comparison to the previous one is that the data accumulation derives from open data platforms by defining an equivalent keyword query. Although we are dealing with the same group of people (artists), the data come from an entirely different type of source. We argue that for the artist community particularly, this data source describes more effectively the activity of this group, as most of the people are freelancers or unemployed, however they actively pursue real estate for their studio or advertise artwork exhibitions etc. This activity would be completely masked by the Census data set, however it is revealed at this stage of the process, since Google places and "craigslist.org" allow for every request is geo-located.

In detail about the methodology itself, using Google API and "craigslist.org", we performed multiple requests at a daily basis (1000 requests per 24 hours), in order to collect all the necessary data. The keyword queries were related to temporal requests and offers regarding real estate for artists' studios, gallery spaces, events, artists' resources, artwork sale, exhibitions, FAQ etc. Regarding the homeless population queries, those involved shelters, community amenities, technology stations, hygiene stations, food supply stations etc. The data accumulated was formatted in .csv format, the same as with the Census data process and visualized as nodes of the same context (Figure 2).

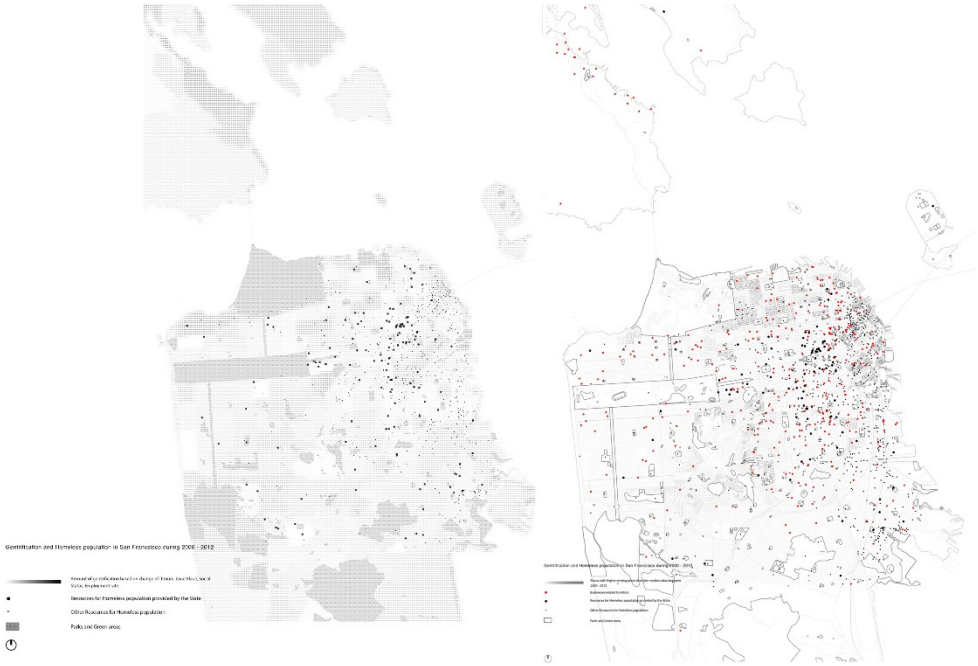

Fig. 2: San Francisco (left), Census GIS data overlapped with resources related to homeless population and San Francisco (middle) data overlap of: businesses related to artists, resources for homeless population, high priced real estate and public transport

5.4 Empirical Data Approach

The third method operates at neighbourhood level. This method embraces an empirical data approach, where human perception and subjectivity are considered a qualitative source of data that can unveil qualities that the other processes overlook. In order to allocate a group of people for crowd-sourcing, we utilized a human-based outsourcing platform called Amazon Mechanical Turk. Amazon Mechanical Turk is a crowdsourcing Internet marketplace, operated by Amazon, enabling individuals to coordinate the use of human intelligence to perform tasks that computers are currently unable to do. It is an on-demand large sample of users that executes large assignments over a given period of time. In our case, a large group was given two different sets of questions. The first set targets human subjectivity, where the users were asked subjective questions in order to rate certain neighbourhoods based on

Google Street View viewpoints (Figure 3). This research takes advantage of human subjectivity when it comes to rating an area based on personal interpretation of safety, affordability and infrastructure condition, qualities that vary significantly even among neighbouring blocks, however the amount or the frequency of variation may have a significant role (Figure 4, 5). The second set targets the collection of detail features (e. g. the presence of: expensive loft housing, abandoned buildings, industrial buildings, bikes, public stairs etc.) that are encountered in the areas of interest using the same Google Street View viewpoints. These features are time consuming to collect manually therefore; this tool is proven convenient as it succeeds in collecting this information in short time. The areas of interest were Tenderloin in San Francisco and Emeryville in Oakland. Tenderloin was chosen because despite the fact that it is adjacent to already gentrified areas, it has a different character, whereas Emeryville was chosen because it is transforming from a crime area into an urban, entertainment and commercial attractor point. The questions were submitted to Amazon Mechanical Turk through a template in .json format. The questions were structured in a way that the answers would be easy to process and to visualize, such as numerical (scale 1-10), binary (yes/no) or choice (tick the box), while we completely avoided answers text form. The answers we received were in .json format so we transformed them into .csv format and then imported to Grasshopper and Processing as in the previous two methods.

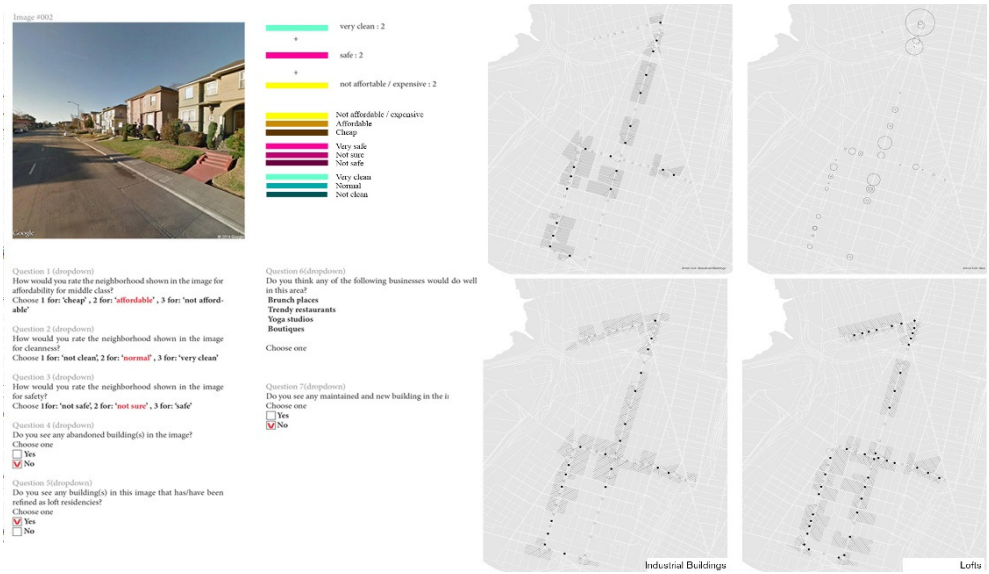

Fig. 3: Amazon Mechanical Turk submitted questionnaire (left) and features maps (right)

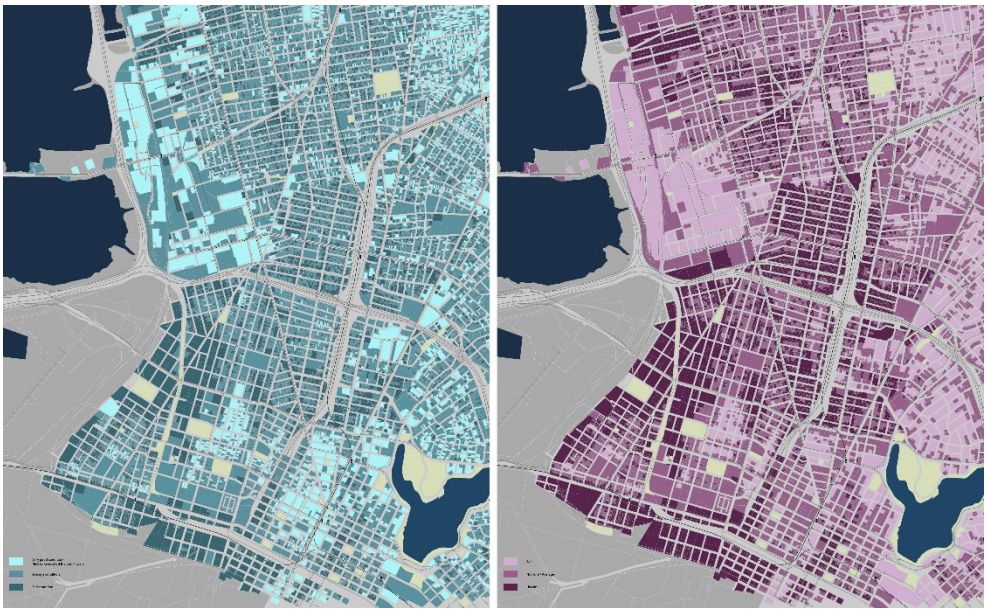

Fig. 4: Amazon Mechanical Turk neighbourhood rating: Neighbourhood infrastructure evaluation (left), Neighbourhood safety evaluation (right)

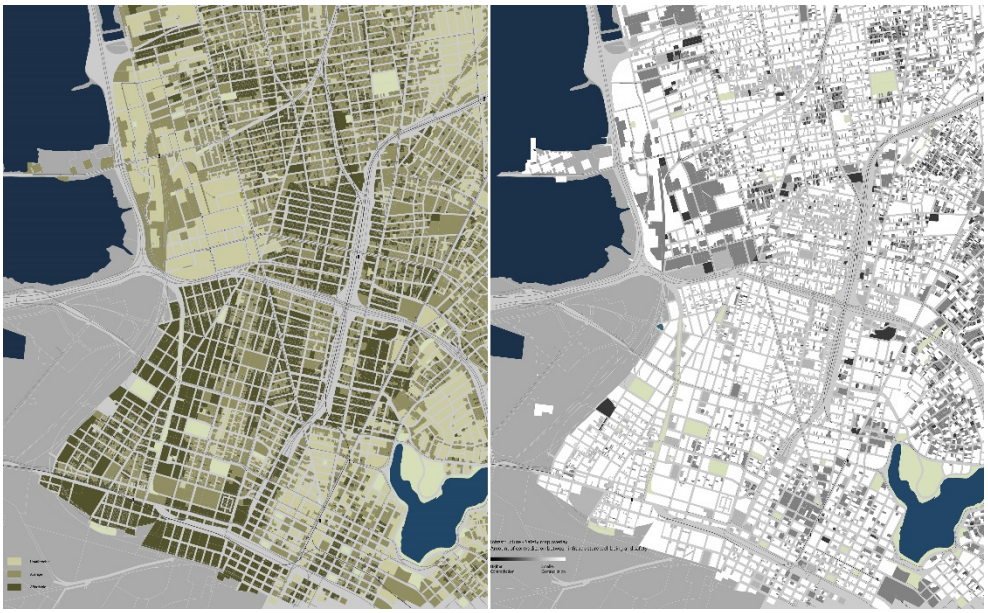

Fig. 5: Amazon Mechanical Turk neighbourhood rating: Neighbourhood affordability evaluation (left), Neighbourhood infrastructure and safety comparison (right)

5.5 Evaluation

The results of the three surveys were overlapped and weighted in order to produce a series of maps at different scales that visualize gentrification in the Bay Area. Each method presents certain advantages. The Census data analysis provides an overview of the context over a significant time span (2000-2012) and helps us understand major socio-economic shifts. The open data analysis depicts the ephemeral layer of relationships that take place in the urban environment, which is impossible to be described by authoritative data, however it is more relevant to the actual conditions. The third method enriches the process with user personal feedback about ranking the environment of a neighborhood. The project aims to provide a calibrated understanding of the multiple grains of constructed space through top down and bottom up methods, as well as to offer a tool of visualizing dynamical characteristics of the urban environment. Our research balances the traditional Census data analysis with more dynamic layers of collective platforms and crowdsourcing. Whichever method is considered, being more or less descriptive of the reality, it is worth examining all the conduits and corridors available to us, by which this change is being delivered. Looking at urban issues through maps can give us several hints about spatial and social transformations, in which we can think upon, as visualized information provokes feedback, either logical or emotional (CORNER 1999). Throughout this entire process, we can assess certain findings:

1. Based on the Census data search, nearly half of Bay Area census tracts are undergoing some form of neighborhood transformation and displacement.
2. The data accumulated from the ephemeral data research depict a significant artists' movement regarding art studio rent requests, artwork sale and creative services in general in the entire Bay Area and especially in San Francisco and Oakland. As it is also evident from the maps, Oakland has historically been overshadowed by the San Francisco arts scene; however, in combination with the staggering rise of rent in San Francisco, we can anticipate that the artist movement will intensify in East Bay in a short timeframe.
3. Studying Oakland at a local street view scale, we can assess that the area is undergoing disperse development that presents high contradictions related to infrastructure condition, affordability and safety (Figure 6). The results from the crowdsourcing survey vary significantly in building block scale, therefore any sense of continuity of the same character because of proximity is not necessarily a criterion to rely upon.
4. Moreover, certain re-developed areas have uniform functional identity, such as Emeryville, as they present excessive duplication of the most profitable uses (malls, restaurants), while San Francisco and Oakland downtowns present excessive duplication of financial functions (bank district).
5. We notice significant contradictions on the results of the crowed sourced research regarding infrastructure condition, safety and affordability perception of the participants. Some of the findings depict areas of new development (last 3-4 years) that are yet islanded off because the surrounding area is significantly undermined. However, this contradiction reveals certain dynamics regarding the future, further re-development of the area, as well as the areas that accumulate similar features.

This new establishment of relationships is replacing almost entirely the previous condition of gradual displacement and gentrification. It evolves rapidly, and although it looks more

orderly, visually, this aesthetic ordering might not have a social correlation. Social structure and social stability are inversely proportional to visual order. This condition is known to be establishing in Oakland, which was significantly undermined in the past few years, however the challenge is not only to identify the problem, but also to find the ways to analyze its characteristics by mapping and communicate it visually to its extents. Although understanding the shifts of urban space and finding the patterns that drive them is a big challenge, we support that close engagement with technology leads us to explore numerous research methods, which have a way of contributing to meaningful connections inside data networks. As some of the above methods open the possibility of operating at a fine spatial scale, examining the city building by building, they provide the context for a more fine-tuned understanding of neighbourhood characteristics, conflicts and relationships that reveal the heterogeneous characteristics of the city. We find inspiration in the combination of the traditional ways of space categorization by investigating the relationship of home value, income, transportation, etc. with a bottom-up, participative approach in which individuals provide more ephemeral social elements of neighbourhoods. We believe that the composite association between them leads to more informed decision-making and a more qualitative image of the city that reflects subjective aspects of urban analysis (BATTY 2013).

Fig. 6: Oakland, Artists activity and high priced real estate (left), Oakland, left map overlaid with Neighbourhood infrastructure and safety comparison from Mechanical Turk study result (right).

6 Future Work

Future development of the project would be to find ways to enrich the process with user feedback data that will improve decision-making. One way to achieve this is the incorporation of social media feeds, such as Facebook and Twitter that would refine the tool by adding the feedback of targeted users and potential residents of the area. Social media is makes feelings, thoughts and intentions about the city explicit and thus, creates new opportunities for improving the existing mapping tools, as most social media is geo-located. Almost all the main social network providers allow access to data feeds via an Application Interface (API), so the data cities and individuals emit can be collected and filtered, and opens up the possibility of a real-time view of the city. In the years to come, it is vital to understand that as technology improves, the amount of data increases, and designers should problematize on the cases where data provides unique understandings they could not have had otherwise, or cases where data increase creates confusion that hinders designers' perceptions. The main challenges would be to identify whether we have enough data to create assumptions, whether we have the right type of data to support our claim and whether we can visualize urban space in ways that are perceived by everyone.

Acknowledgement

This research was conducted as part of the Studio One program 2014/2015 at the University of California, Berkeley, Department of Architecture, College of Environmental Design. The study was a teamwork of Eleanna Panagoulia, Shima Sahebnassagh and Namju Lee under the supervision of Prof. Kyle Steinfeld.

References

CORNER, J. (1999), The Agency of Mapping: Speculation Critique and Invention. Reaktion Books, London.

FREEMAN, L. (2004), Gentrification and Displacement: New York City in the 1990s. Journal of the American Planning Association, 70 (1).

GLASS, R. (1964), London: Aspects of Change. MacGibbon & Kee, London.

JACOBS, J. (1961), The Death and Life of Great American Cities. Vintage Books, New York.

OWENS, A. (2012), Neighborhoods on the Rise: A Typology of Neighborhoods Experiencing Socioeconomic Ascent. City & Community.

SAMPSON, J. (2006), Great American City: Chicago and the Enduring Neighborhood Effect. The University of Chicago Press, Chicago.

SASSEN, S. (1998), Globalization and its discontents. New Press, New York.

TUFTE, E. (1990), Envisioning information. Graphics Press LLC, Connecticut.