Data & Civil Rights: Housing Primer

by Alex Rosenblat, Kate Wikelius, danah boyd, Seeta Peña Gangadharan, and Corrine Yu
Produced for Data & Civil Rights Conference / October 30, 2014

Access to housing is central to economic and personal security and to social inclusion, yet remains shaped by illegal housing discrimination throughout the United States – both intentional discriminatory acts and facially “neutral” policies that may limit housing opportunities based on race, color, national origin, religion, or sex or for families with children and people with disabilities. Affordable housing is in critically short supply.¹ Communities throughout the United States remain marked by a high degree of racial segregation and concentrated poverty, such that housing-related issues furthers inequality in access to education, employment, and healthy public spaces, and perpetuating gaps in opportunity for successive generations.² These inequities were exacerbated by the economic downturn, and in particular, by the impact of predatory lending practices and residential foreclosures on minority communities.

Data has always played an important role in housing policies, practices, and financing. Housing advocates worry that new sources of data are being used to extend longstanding discriminatory practices, particularly as it affects those who have access to credit for home ownership as well as the ways in which the rental market is unfolding. Open data practices, while potentially shedding light on housing inequities, are currently more theoretical than actionable. Far too little is known about the ways in which data analytics and other data-related practices may expand or relieve inequities in housing.

Background: Discrimination in Housing

Major civil rights concerns and themes in U.S. housing:

- **Housing Segregation and Discrimination:** Most housing discrimination goes unreported, but it is estimated that 4 million instances of housing discrimination occur every year just in the rental and real estate sales markets.³
- **Foreclosures and Predatory Lending:** Continued residential segregation and the exclusion of racial minorities from access to quality mortgage credit created model conditions for predatory lending to poor households in communities of color.⁴ The abusive practices have led to the massive loss of wealth built over generations in communities of color.⁵
- **Bank-Owned Properties:** In an investigation into the maintenance and marketing practices of Real Estate Owned (REO) properties by banks, the National Fair Housing Alliance found that major banks around the nation maintain and market REO homes in White

Data & Civil Rights :: http://www.datacivilrights.org/
communities significantly better than in communities with higher concentrations of racial minorities.\(^6\) Failures by banks to maintain and market properties bring down neighboring home values and devastate the recovery in entire communities, and encourage investor purchasers over owner-occupant purchasers of those homes.

**Discrimination in Home Ownership**

Home ownership is privileged in American society over renting: since the Great Depression, the U.S. government has designed subsidy programs and other policies to help people attain and sustain home ownership.\(^7\) Historically, subsidies benefit higher-income households.\(^8\) Renters, which feature much lower incomes in comparison to homeowners, comprise 35% of all U.S. households as of 2012. A quarter of the renter population earns an annual income of $15,000 or lower, compared to 13% of homeowners.\(^9\)

Inequity in home ownership in the United States stems from a controversial and discriminatory history of mortgage lending practices. In the first half of the twentieth century, banks systematically excluded African Americans from access to legitimate mortgages at fair rates of interest. The Federal Housing Administration enshrined and ensured a process of residential segregation by refusing to insure mortgages in Black neighborhoods that contained Black residents, identifying and labeling these neighborhoods in red on real estate investment maps.\(^10\) This practices, which researchers and activists later called redlining, resulted in disinvestment in Black neighborhoods, and the value of African-American-owned homes dropped precipitously in redlined areas.\(^11\)

Today, differences in the wealth in White and non-White neighborhoods continue, though causes of racial segregation and housing discrimination have become more complex. Black residents tend to live in neighborhoods comprised of communities of color, while White residents live primarily in White neighborhoods.\(^12\) Homes in neighborhoods that are at least 10% Black appreciate in value at lower rates than homes in White neighborhoods.\(^13\) Meanwhile, the benefits of gentrification accrue disproportionately to lower-income White neighborhoods. For example, neighborhoods experiencing early signs of gentrification generally continue on that path if they are, at minimum, 35% White. Gentrification slows or stalls in neighborhoods that are 40% or more Black. In general, gentrification affects mixed-race neighborhoods, as opposed to White neighborhoods.\(^14\)

Since the homes are where households build equity over time, the networked effects of disinvestment in neighborhoods subject to redlining and segregationist practices reverberate over generations. Lower rates of home ownership and home equity means that members of minority groups may have fewer assets with which they can collateralize loans and consequently lower
Data & Civil Rights

Data Analytics

A history of low-tech racial profiling in credit distribution systems forms the basis for concerns for digital redlining and the use of data mining tools to mask or amplify discriminatory practices in the mortgage and rental markets. Civil rights concerns center on three major aspects of the big data phenomenon: prejudicial risk-assessment algorithms, tech-driven circumvention—unintentional or otherwise—of existing regulatory frameworks for targeted marketing, and fair housing.

Coarse-grain analysis that relies on geographical data makes it easy to reinforce inequalities in the housing market. For example, people who navigated Wells Fargo’s online site to browse for homes for sale were directed to neighborhoods with a similar racial make-up based on their own zip codes. The example of targeted marking in subprime lending serves as case in point. Due to housing policies and practices that over time concentrated Black residents in neighborhoods, neighborhood data correlate strongly with race. As demonstrated in research by Linda Fisher, lending and other institutions compiled and merged data profiles on localities along shared attributes in order to target sub-prime borrowers in a given area. For example, mortgage brokers purchased detailed lists that combine public records data, like recent house sales, and detailed consumer mortgage data, including the foreclosure status of a given home, in order to target potential customers who might be interested in second mortgages or in refinancing packages. Though lenders did not use race as a factor in targeting per se, it became a factor. Behavioral advertising firms identified leads on customers who might be ripe for a sub-prime mortgage, and they were targeted for ads; the targeted customers were predominantly low income, Black and Hispanic. The subprime crisis revealed that African Americans and Hispanics acquired subprime mortgages a much higher rate than Whites. For example, Blacks represent 54.7% of the people of all races who received high-cost loans, while Whites make up 17.2% of the total.

While not all subprime products qualify as predatory, the sub-prime lending boom of the 2000s demonstrated their profound impact on communities of color. Mortgage brokers systematically steered Blacks and Hispanics into sub-prime mortgages, even if their credit information indicated they were eligible for prime rate mortgages. Lower-income borrowers in particular struggled to repay credit at higher interest rates than higher-income borrowers receive. According to estimates by the Urban Institute, between 2007-2010, Black families lost 31% of their wealth; Hispanic families, many of whom bought homes at the height of the housing bubble, lost 44%; and White families lost 11%.

Risk-assessment algorithms used in mortgage lending reveal a similar problem as in targeting: though racial discrimination does not factor into the identification and categorization of the

Data & Civil Rights :: http://www.datacivilrights.org/
creditworthiness of borrowers, algorithmic analysis may differentially evaluate creditworthiness according to race because of the confounding nature of neighborhood data. For instance, middle class Black homeowners do not rank as well as middle class White homeowners due to residential location. Commenting on work by Patricia Sharkey, Ta-Nehisi Coates writes, "Black families making $100,000 typically live in the kinds of neighborhoods inhabited by White families making $30,000." In a simplified scenario, if a calculation determining eligibility for a prime-rate mortgage excludes race but includes data on income and neighborhood, the strong correlation between race and neighborhood will have a differential effect on applicants according to race, even if Black applicants belong to the same class as White applicants.

Though the processes for assessing credit scores – and of underwriting and conducting pricing analysis for loans – are computerized and automated, and can guard against human error or bias in the mortgage industry, they raise other questions and concerns. As explained by researchers Solon Barocas and Andrew Selbst, datasets that contain prejudicial bias from which algorithms infer rules about what qualities correlate to higher or lower risk loan candidates can have a disparate impact on protected classes. Algorithmic systems that profile and assess applicants can lead to the rationalization of discriminatory outcomes when they lack intentional motives of bias or harm against particular populations. Some might, for example, rationalize that the use of neighborhood and income in risk-assessment algorithms presents a legitimate business consideration that overrides concern for adverse disparate impacts on communities of color. And while others might advocate for the use of more granular data, in order to categorize applicant eligibility according to different criteria, its implementation may replicate the problem: many factors can act singularly or in combination with each other, effectively creating proxies for protected-class variables like race. A mortgage lender may design (or contract a third party to design) algorithms in ways that comply with laws protecting civil rights in the housing markets, but which nevertheless have an unequal effect.

Online advertising is one higher-tech forum where data analytics strives to achieve personalization based on data, but not necessarily on context, and the result can be unequal targeting. The advent of the Internet and digital marketing harbored a potential boon for financial inclusion: ads for financial services could reach underserved communities with ease, and applicants could navigate them without face-to-face interactions that engender prejudicial bias. Financial services companies comprise one quarter of the top 50 spenders on online advertising. As legal scholar Lori Andrews notes, the mortgage lenders Countrywide Financial and Low Rate Source made the top 10 list of companies which spent the most money in online advertising in July 2007. Keywords like “mortgage” and “refinance” earned $20 to $30 per click it received for Google, which earns most of its revenue from the income generated by ads, leading financial blogger Faisal Laljee to conclude that Google was the real profiteer in sub-prime boom.
Research by former FTC Chief Technologist Latanya Sweeney raises the possibility that online ad delivery services that use data analytics to deliver different financial services ads to different segments of customers could replicate a channeling effect that creates an adverse disparate impact on protected groups. Data collection and analysis supports consumer differentiation, which enables targeted marketing practices that adversely impacts members of protected classes, even if they are seemingly neutral. Online ad-delivery systems target certain demographics, and some websites have audiences that are almost exclusively of a particular race, or another protected-class characteristic. Advertisers do not need to know a lot of details about consumers to identify that they are probable consumers of the services or products they offer. As data scientist Claudia Perlich observes, ad-delivery systems that predict which customers are likely to click on a given ad can rely solely on their browsing histories rather than on their data profiles or protected-class information. However, ads that nudge certain segments of consumers towards particular products based on which website they visited can reflect a prejudicial weblining effect and an adverse disparate impact on protected classes, even if there is no broker with a particular prejudicial bias steering consumers towards a lower-quality product. That outcome is not likely to be the result of intentional discrimination based on protected class categories. When it comes to compiling and sorting data into aggregates (e.g. demographics of consumers who share certain purchasing habits) or personalized categories (e.g. based on the browsing data of an individual), social sorting mechanisms like consumer differentiation are inherently discriminatory, but they do not necessarily invoke the categorical notions of discrimination against protected classes.

Networked Harms: Steering in the Rental Market

The development of new technologies has made it clear that there are lots of new data, and such data is more publicly accessible than it was in a pre-digital, pre-datafied stage of our networked society. The availability and accessibility of the data, and the ways that it is organized, produce new types of discriminatory harms, even if they are not incorporated specifically into ‘Big Data’ analytics.

Data on people and their neighborhoods comes from both from governmental sources, like the Census Bureau, and from private enterprises, like apps that geo-tag locations, like restaurants and shops, that people visit. Data analytical tools make it easy and efficient to sort, label, and interpret to myriad streams of information, including for the end user interested in housing markets. Whether these new kinds of products racially steer consumers—to rent or buy in Black neighborhoods or to stay away from them—remains an open question.

According to Linda Provost’s article on data-driven home searching, some companies may be sorting and disclosing neighborhood data in ways that could undermine the fair housing laws that prohibit racial and other types of steering. Suburban Jungle Realty Group operates ambiguously as both a brokerage and a neighborhood guide for city-dwellers moving to the suburbs, but their
staff is only partially comprised of licensed brokers, presumably so their other staff can disclose information freely. Some brokerage sites, like Movoto.com display ZIP code data on race, languages and foreign-born households directly on their sites. Relocality.com pairs people with neighborhoods based on their Facebook data: the site’s algorithm determines matches after ranking their priorities, as evident in their personal interests and photos. Clearly, there is significant tension between what data analytics offers (personalized matches) and concerns that personalization apps will re-enforce or amplify existing biases that spur patterns of residential segregation.

Rental markets are also being influenced by data analytics. More public records data is available on potential tenants, and landlords have easy access to them. While landlords can run search queries for the names of their prospective online and potentially discriminate against them based on protected class characteristics, some data is both harmful to protected classes and is legitimate for landlords to use. For example, in an effort to achieve crime-free zones, over 100 cities in Illinois have passed ordinances that permit, and even encourage, landlords to evict tenants whose homes become hot-spots of criminal activity. These types of ordinances have created precarious circumstances for the victims of domestic violence, who are disproportionately women and particularly, Black and Native American women, because they fear eviction if they call the police to protect them from abusers. Landlords are being instructed by some cities to summarily evict disorderly tenants, regardless of the cause; in inner-city Black neighborhoods, women are evicted at significantly higher rates than men. Moreover, they can turn away potential tenants on the basis of eviction records. Although some cities and states have recently retracted ordinances or introduced new laws in co-ordination with the U.S. Department of Housing and Urban Development to prohibit evictions that stem from domestic violence, at least one city has replaced its policy against ‘disorderly households’ with an ordinance against tenants who are ‘disruptive’.

The Potential of Open Data

A great deal of the data used in housing apps that can influence steering practices is publicly available. The obvious benefit of this data is that it creates a visible, explicit market. Anyone can view what the prices are for a home, and what the proscribed benefits or disadvantages are. Generally, prices are not openly displayed on units for rental or sale on a sign outside a home; outside of prices displayed online, on sites like Streeteasy.com, one generally has to contact a real estate broker, a landlord, and sometimes an owner directly. This interaction could result in discriminatory steering practices if the seller’s bias reduces their willingness to interact with, or develop a relationship with, a given buyer or renter, or vice versa. Similarly, if there is publicly available data on what kinds of loans, and at what rate, people of similar credit scores and consumer profiles are receiving, this data could be used to empower loan applicants who negotiate with banks for their rates. This increase in visibility could strengthen competition between banks to gain market shares with marginalized demographics.
A lot of the benefits of open data, and housing data in general, have not been explored thoroughly, and there is a lot of room for possibilities. For example, regulatory authorities can use that data to more efficiently mine through the complaints they receive on unfair housing practices, or to anticipate or observe patterns that raise red flags. How can housing data be used to illuminate trends in racial segregation, and how can that data be used to empower marginalized groups?

Open data could also be used to devise new representations of neighborhoods that could deter, mitigate, or prohibit prejudicial bias. For example, more granular data on a neighborhood might reveal that local libraries offer many public services or that a community has hidden gems, like community gardens. New metrics like ‘how easy is it to share childcare responsibilities with your neighbors?’ could spur new interest in overlooked areas that have extensive childcare options. What kinds of data do buyers need to optimize for other values, beyond racial bias, which depart from the traditional metrics used to assess what a neighborhood is comprised of? When does race act as a proxy for other indicators, like poverty or crime rates, and how can more granular data reduce the role of race as a proxy?

Questions for Data, Civil Rights, and Housing

1. How can data analytics be used to identify patterns of unfair housing? Who is best equipped to do this kind of work?
2. How can analytics techniques undermine efforts to restrict the dissemination of information about neighborhoods that implicates protected-class data?
3. When does data profiling in housing reflect discriminatory biases? When does its usage become predatory?
4. How do we account for the discriminatory effects of redlining in online advertising (“weblining”) related to housing services, products, or availabilities if ad-delivery services do not want, use, or have knowledge of protected-class information when they send an ad to a consumer?
5. How can we ensure that consumers know that they are being presented with advertisements/mortgage offers/housing options that are equivalent to those presented to other consumers?
6. How can publicly available data on housing be used to create greater levels of accountability and advance civil rights issues?
7. How can data be presented or organized to reduce the role of prejudicial bias in residential decision-making processes?
8. How can we use open data to reduce inequality, and to reduce the accumulation of disadvantage?
8 Chase, “Rethinking the Homeownership Society,” 67.
http://www.theatlantic.com/features/archive/2014/05/the-case-for-reparations/361631/
11 Coates, “The Case for Reparations.”
13 Brown, “Home Ownership Keeps Blacks Poorer.”
15 Emily Maltby, “Minority Entrepreneurs Still Face Bias, Prof Says,” WSJ Blogs - In Charge, March 10, 2011.
19 Coates, “The Case for Reparations”; commenting on research by Jacob S. Rugh.
https://www.brooklaw.edu/~/media/PDF/LawJournals/JLP_PDF/jlp_vol18i.ashx.

28 Coates, “The Case for Reparations”; commenting on research by Patricia Sharkey.


30 Coates, “The Case for Reparations.”


37 Sweeney, “Online ads roll the dice.”


42 Prevost, “Using Data to Find a New York Suburb.”


