Article

Coupling data science with community crowdsourcing for urban renewal policy analysis: An evaluation of Atlanta's Anti-Displacement Tax Fund

Jeremy Auerbach (ORCID)
Colorado State University, USA

Christopher Blackburn
Georgia Institute of Technology, USA

Hayley Barton
Duke University, USA

Amanda Meng and Ellen Zegura
Georgia Institute of Technology, USA

Abstract

We estimate the cost and impact of a proposed anti-displacement program in the Westside of Atlanta (GA) with data science and machine learning techniques. This program intends to fully subsidize property tax increases for eligible residents of neighborhoods where there are two major urban renewal projects underway, a stadium and a multi-use trail. We first estimate household-level income eligibility for the program with data science and machine learning approaches applied to publicly available household-level data. We then forecast future property appreciation due to urban renewal projects using random forests with historic tax assessment data. Combining these projections with household-level eligibility, we estimate the costs of the program for different eligibility scenarios. We find that our household-level data and machine learning techniques result in fewer eligible homeowners but significantly larger program costs, due to higher property appreciation rates than the original analysis, which was based on census and city-level data. Our methods have limitations, namely incomplete data sets, the accuracy of representative income samples, the availability of characteristic training set data for the property

Corresponding author:
Jeremy Auerbach, Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, CO, USA.
Email: jeremy.auerbach@colostate.edu

tax appreciation model, and challenges in validating the model results. The eligibility estimates and property appreciation forecasts we generated were also incorporated into an interactive tool for residents to determine program eligibility and view their expected increases in home values. Community residents have been involved with this work and provided greater transparency, accountability, and impact of the proposed program. Data collected from residents can also correct and update the information, which would increase the accuracy of the program estimates and validate the modeling, leading to a novel application of community-driven data science.

Introduction

As the back-to-the-city movement continues across the United States, cities have undertaken major urban improvement and renewal projects. While investment in key infrastructure can create new resources and more desirable living areas, it can also result in increased rents and property appreciation. For example, home value appreciation due to new rail networks has been observed in Chicago (McMillen and McDonald, 2004), Boston (Armstrong, 1994), and Portland (Chen et al., 1998). These increases in property values have been perceived to cause displacement of long-term and low-income residents, along with changing neighborhood demographics and characteristics. Scholars observe positive correlations between displacement and negative effects such as homelessness, health issues, violence, and reduced academic performance (Bartlett, 1997; Hartman, 2003; Hope and Young, 1986). Clearly, much is at stake when city planners and officials design and implement urban renewal projects that increase property values, potentially causing displacement. Planners need to be aware and take into consideration the costs to community residents associated with increased property values due to urban renewal projects.

We have partnered with members of the Westside Atlanta Land Trust (WALT), a program of HELP ORG INC in Atlanta, Georgia, that focuses on permanently affordable housing for current, low-income residents amidst significant urban renewal and development. The goal of one project emerging from this partnership is to estimate and share the eligibility and cost of the Anti-Displacement Tax Fund (ADTF) with community members. The ADTF is a privately funded program designed to offset future property tax increases for qualifying nearby neighborhood homeowners from the construction of a stadium in Westside Atlanta (the qualifications are described in more detail in the next section). Figure 1 provides a map of Metropolitan Atlanta including the stadium, the original four Westside neighborhoods in the program area, the Old Fourth Ward neighborhood used for the methods, and the Atlanta BeltLine, an unused railway corridor encompassing the metropolitan center being converted into a multi-use trail.

An initial attempt to predict program cost estimation relied on data aggregated at the census and city-level, ignored eligibility requirements, and made several inexact modeling assumptions (Bedsole, 2017). To assess the total cost of the ADTF, we estimated the number of eligible homeowners and the future property tax assessments for their properties. We first estimated household-level family size and income with machine learning approaches applied to publicly available household-level data. With help from community members, we also estimated the presence of liens for properties in the eligible neighborhoods. We then
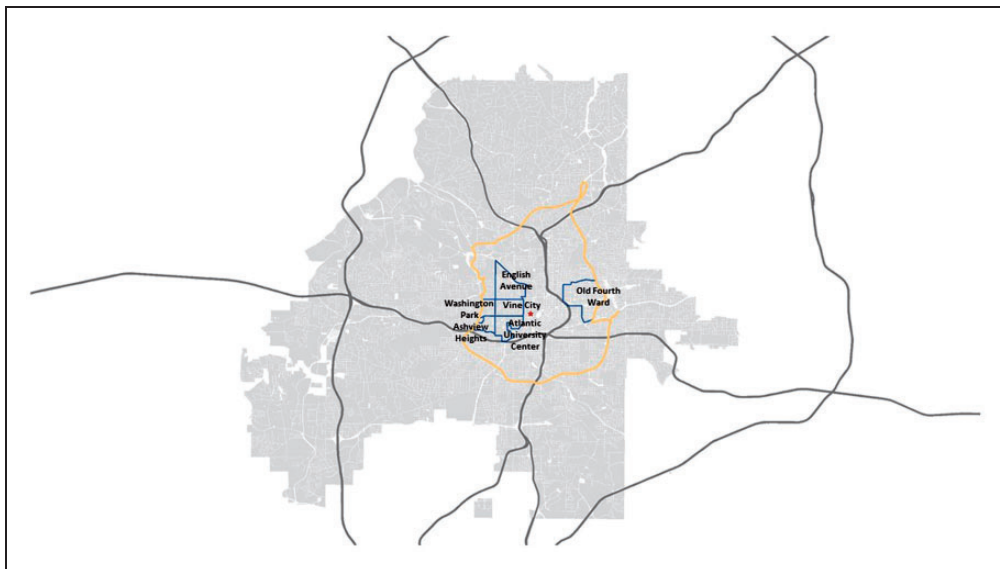
Figure 1. Map of the Atlanta study area. The four original Westside neighborhoods in the program area, the Washington Park neighborhood (which was included in the study at the request of the community residents), and the Old Fourth Ward neighborhood, which is used for the program analysis methodology, are labeled and displayed with blue boundaries. The stadium is shown with the red star, parcels are shown in light gray, interstates are shown in dark gray and provided for reference, and the Beltline is the orange line encircling downtown Atlanta.

forecasted future property value appreciation in the Westside neighborhoods with machine learning techniques applied to historical tax assessment data from a socioeconomically similar neighborhood in Atlanta that was in comparable proximity to an urban renewal construction project. We used these projections and household-level eligibility to estimate the costs of the program for different eligibility scenarios. Hedonic models have traditionally been used to estimate future property values but these methods have significant issues when used for individual home value projections (Diewert, 2003; Shöni, 2013). Machine learning techniques can be a powerful tool for planners and program analysts (see, for example, Tribby et al. (2017) and Zhang et al. (2017)), yet they have not been used extensively in these domains. We find that our household-level data and machine learning techniques result in a fewer number of eligible homeowners yet a larger program cost, due to higher property appreciation rates, compared to the initial project analysis.

Community residents also expressed a desire for a tool that would help them determine eligibility and quantify the expected program results. Our estimates were incorporated into such an interactive tool for residents to determine their program eligibility and view their expected home value appreciation. Community organization members can update the household information while canvassing neighborhoods, which would in turn provide better estimates for the program cost. The tool that we created demonstrates how community engagement and data science can be combined to examine proposed urban programs and policies, and most importantly, how community members can be informed and involved in this process to increase transparency and accountability.

We first provide a review of the literature on property appreciation and displacement, which describes the lack of evidence for homeowner displacement due to increased property

values. We then introduce a proposed anti-displacement policy in the Westside of Atlanta and describe our partnership with community members to evaluate the costs and strengths of the policy. We discuss our methodology in comparison to the original program study and describe the tool built for community members to search eligibility and update records. We conclude with a summary of future work.

Urban renewal and displacement

Potential displacement of low-income residents is an issue at the center of many debates about gentrification and its effects. As property values appreciate due to the influx of wealth and resources in a gentrifying neighborhood, the concern arises that longtime, low-income residents will be forced to move, especially elderly residents or people on fixed income who may not be able to afford these increases (Levy et al., 2006; Marcuse, 1985). However, several empirical studies have shown that displacement of vulnerable homeowners due to increased property taxes is no more common in gentrifying areas than in non-gentrifying neighborhoods (Ellen and Ondende, 2011; Freeman and Braconi, 2004; McKinnish et al., 2010; Martin et al., 2016). Still, there is evidence that gentrification and rising property values directly displace renters (Martin et al., 2016), affect different socioeconomic groups disproportionately (Ding et al., 2016; LeGates and Hartman, 1982), and change the composition of a neighborhood over long-time periods (Ellen and Ondende, 2011). As property values rise, there are fewer housing options available to low-income residents in these neighborhoods, and even if current residents are not displaced directly, there is a future reduction of low-income movers into the affected areas (Freeman and Braconi, 2004).

Many cities in the US have proposed a variety of policies to combat potential involuntary displacement of low-income residents and to maintain affordable housing options in increasingly desirable areas. These programs include property tax circuit breakers, community land trusts, preserving or creating affordable housing units, inclusionary zoning, and linkage fees. For a review of anti-displacement programs, see reports by the Metropolitan Area Planning Council (2015) and Levy et al. (2006). For our work, we estimate the cost and impact of the ADTF, a policy similar to a property tax circuit breaker, for the Westside of Atlanta which is undergoing large-scale urban renewal projects.

Atlanta's urban renewal and the ADTF

Atlanta is currently undergoing a city-wide transportation revitalization project through the conversion of an underutilized rail-line into a 22-mile multi-use bike and pedestrian trail, called the BeltLine (Atlanta BeltLine, Inc., 2017). Since the construction on part of Atlanta's Eastside BeltLine has been completed, increased rents and property appreciation have been observed in the nearby neighborhoods (Immergluck, 2009). Properties adjacent to the BeltLine have experienced a 40–68% increase in value during the 2012–1017 time period (Immergluck and Balan, 2017). As construction continues on the Westside BeltLine segments, nearby residents are concerned about a similar increase in rents and property taxes and possible displacement. In addition to development spurred by the BeltLine, Atlanta's Westside neighborhoods are adjacent to the newly constructed Mercedes-Benz Stadium, a multi-billion-dollar project. Residents question the use of public funds for a recreational facility and have already experienced displacement, as the development required the demolition of two historically black churches (Belson, 2017). In response to these concerns, the city, along with the private sector, made promises to prevent future displacement of low-income residents. The ADTF, a result of this promise, attempts to offset the increase in

property taxes for eligible homeowners in six neighborhoods on Atlanta's Westside as home values rise due to the stadium. Table 1 summarizes the eligibility requirements for the ADTD: homeowners must occupy the residence in the English Avenue, Vine City, Atlanta University Center, Ashview Heights, Booker T. Washington, and Just Us Westside Atlanta neighborhoods; they must enroll prior to 15 May 2018; their income must be below the Area Median Income for their household size; and they cannot have any outstanding debts or liens on the property (not including mortgages). The ADTF is a collaboration between the city and the non-profit Westside Future Fund (2017), which is funded by a collection of local Fortune 500 companies administering the program and has contracted a local consulting group to aid in its execution. The ADTF ultimately aims to ensure that long-term, low-income homeowners remain in place and benefit from the urban renewal projects (Dastrup et al., 2015; Lester and Hartley, 2013).

While other regional planners have tried to predict displacement and gentrification, including Los Angeles and San Francisco (Los Angeles Innovation Team, 2016; Institute of Governmental Studies, 2017), there has been very little work examining the cost and effectiveness of suggested anti-displacement programs. In Atlanta, the only public estimate of the cost and scope of the ADTF predicted that 400 homeowners in the Westside neighborhoods would participate, 165 of those would remain in their homes at the end of 20 years, and the program would cost $5 million (USD) over 20 years (Bedsole, 2017).

The original estimates for the program eligibility and cost of the ADTF used aggregated Census data, city-level statistics, ignored the lien eligibility requirement, and made several assumptions when data were not available (Bedsole, 2017). To estimate the number of owner-occupants in the area, the model used the rate of residents claiming a homestead tax exemption, which requires the homeowner to also be the occupant of the residence. The homestead exemption rate for the Westside neighborhoods was 79% in 2016. This rate was also used as the program participation rate and the model assumed a 5% annual dropout rate without justification. As we show in the following section, with the use of granular data, we find that liens are common, the exemption rate significantly overestimates the actual owner-occupancy rate, and the dropout rate is unsupported.

Another major eligibility requirement, income, was estimated from census block group data with 62% of the households being eligible. However, the program allows for different

Table 1. Summary of anti-displacement tax fund eligibility requirements.

Location	Participants must reside in one of the following Westside neighborhoods: English Avenue, Vine City, Atlanta University Center, Ashview Heights, Booker T. Washington, and Just Us[a]
Enrollment period	Residents may apply for the program's first year between 12 April 2017 and 15 May 2018. Participants will need to reapply annually between 1 January and 15 March to continue receiving program funding
Income	Participants' household income must be below the Area Median Income for their household size
Owner-occupancy	Only homeowners who both currently live in their home and have lived there for at least one year are eligible. New homeowners are also eligible if the home had been previously enrolled in the program
Liens/property taxes	Participants cannot have any standing liens or debts attached to their property
Heirs	Heirs who meet all other conditions are also eligible

[a]We have also included the Washington Park neighborhood as a scenario in our analysis at the request of community members to determine the cost and feasibility of adding this area to the program boundaries.

household income requirements at different household sizes and the original analysis failed to account for this household size effect. Using household-level data, we find that more households would be eligible when household size is accounted for. We also found issues with the property value appreciation model. For property value appreciation, the original model assumed properties valued at or above $37,000 would appreciate annually at 12%, while properties valued under the $37,000 were assumed to appreciate at 50% until the value reached the $37,000 threshold, after which appreciation was set to an annual rate of 12%. This 12% was taken from the average property appreciation rates found in other Atlanta neighborhoods since 2012. This is a broad assumption as these other Atlanta neighborhoods are not socioeconomically similar to the Westside neighborhoods in the affected area and have not experienced significant development projects. Additionally, property appreciation for Atlanta neighborhoods that have undergone a recent development project have not followed these simple trends (as we show in the next section). In 2017, the Fulton County Board of Commissioners froze the property tax assessments to their 2016 levels after residential complaints of increased values with a median appreciation of 20% (Franco, 2017). The original study estimated 400 residents would be eligible and enroll with a program cost of $0.3M for the first seven years and $5M over 20 years. While these assumptions provide conservative initial program cost estimates they do not account for heterogeneity in socioeconomic conditions across the neighborhoods in the program area, the existence of liens, the empirical evidence suggesting homeowners remain in areas experiencing urban renewal or nonlinear property value appreciation. Our approach finds this modeling overestimated the number of eligible residents and considerably underestimated the program costs.

Planning scholars Berke and Conroy (2000) find that

> better information about outcomes would be useful in assessing progress that communities are making toward sustainability, and evaluating the performance of mandates, plans, and implementation efforts. Better information would also improve the ability and legitimacy of planners in promoting the more holistic sustainability concept.

We believe this to be true across disciplines, and at the behest of the WALT our work is a re-evaluation of the previous program cost and eligibility estimates by including machine learning techniques and a richer dataset for household-level eligibility predictions. In addition, the inclusion of community members in the data science process will allow us to further refine the costs and the number of eligible participants in future projections.

Community voice and community-driven data science

For this project, we have partnered with members of WALT, a community organization in Atlanta that focuses on challenges involving urban development and grassroots advocacy. Their interest in partnering with data scientists was to make use of public data to estimate and share with community members the cost and impact of the ADTF, thereby moving up the ladder of citizen participation (Arnstein, 1969). The ongoing construction of the new Mercedes-Benz Stadium, the conversion of the BeltLine trail, and Westside redevelopment more broadly has left area residents feeling excluded from planning and frustrated with the lack of meaningful consultation. While they voice their desires for development in their neighborhoods, they feel ignored by local officials and development stakeholders. Planning scholar Raymond Burby echoes WALTes call for broader participation leading to greater efficacy in land use. In his study of 60 plan-making processes, he concludes that "broad stakeholder involvement contributes to both stronger plans and the implementation of

proposals made in plans" (Burby, 2003). In today's smart city context, citizens make use of data to remain as relevant as sensors in influencing policy making (Gabrys et al., 2016; Jasanoff, 2017; Schrock, 2016). Critical data scholars, Cardullo and Kitchin (2018), apply the aforementioned ladder of participation to understand what it means to be a citizen in the smart city. They find that citizens are often data points or consumers of smart city services, but they can also be empowered to negotiate and even co-create the city through and with data. We offer this project as an example of how residents could improve decision making in policy when they are able to initiate and consult in data science projects.

In addition to estimating the total committed funds, WALT has expressed a desire for an ADTF mapping tool, a geographic information system which allows members to visualize eligible compared to ineligible households, share ADTF requirements, and collect household-level eligibility data from community homeowners. Public participation geographic information systems can promote the involvement of urban residents (Ceccato and Snickars, 2000) and have shown to be effective tools to empower community-based organizations and enhance municipal policy making (for a review of such systems, see Sieber, 2006). We developed such a tool that allows for canvassers to collect and update the household-level data used for the eligibility estimates, making the costs and impacts of the program more accurate, providing a means for validating our methodology, fostering the participation of under-represented groups in scientific research (Nature Editorial, 2018), and more importantly, advertising and informing community members about the program strength. Furthermore, we have included our property tax forecasts for canvassers to share with Westside homeowners.

Research strategy, methodology, and data

We take a two-fold approach to quantify the expected cost of the ADTF. First, we developed a method that determined the number of homeowners who would be eligible. Second, we developed a method for predicting the assessed values of the properties in each of the qualifying neighborhoods. One of the key innovations of our study is the use of publicly available data and machine learning to determine household-level income and projected property value appreciation methods that are easily accessible for analysts to replicate. This project was resident-initiated and throughout the process we consulted our community partners at WALT to ground-truth modeling assumptions and identify geographic scope, involved them as co-researchers to collect data and conducted user-experience testing of the interactive mapping application. Consequently, our study provides a core framework that can be replicated for income estimation, property appreciation, or analysis of property tax subsidization policy programs in other locales across the country.

Data

We relied on publicly available data from various local and national government agencies and Zillow.com to conduct our program analysis. First, we used the historical Fulton County Tax Assessor data for individual home characteristics and historical tax assessments from 2005 to 2016 (Fulton County Board of Assessors, 2017). Second, we used data collected from the Georgia Superior Court Clerkso Cooperative Authority (2017) database to determine homeowner liens in the program area. Third, our income prediction model utilized data from the U.S. Bureau of Labor Statistics' Consumer Expenditure Survey (CEX) for the years 2013, 2014, and 2015 (the most recent years available), scraped 2017 Zillow rent estimate data, and the 2015 American Community Survey neighborhood population

estimates (Bureau of Labor Statistics 2016; United States Census Bureau, 2017; Zillow, 2017). The CEX was a random survey of residents across the U.S. and the dataset includes the following household-level attributes that were used in the modeling: before-tax-income, monthly rent payments, the number of bedrooms in the house, the number of bathrooms, the number of rooms, and the age of the house. The Zillow rent estimate data were collected in July 2017 and represents the company's best estimates for monthly rental payments at the household level. We used the low-rent estimates from Zillow since the homes in the program area are predominantly older and serve lower-income demographics.

Despite the richness of the data used in our study, we encountered several issues while working with this data, finding numerous discrepancies and missing values within the Fulton County Tax Assessor data, including the lack of 2009 data. This is not an unexpected hurdle as there are over 100,000 residential parcels in Fulton County and approximately 20 tax assessors. Furthermore, the lien data were stored in formats that were not machine readable and required a considerable amount of time to parse. Even with these issues, we were able to collect a substantive amount of usable data in a short time period for the analysis.

Overview and rationale of machine learning methodology

Traditionally, hedonic models have been used to estimate future property values but due to multicollinearity, nonlinearity, heteroskedasticity, and non-normality with the data, these methods have significant issues when used for individual home value projections (Diewert, 2003; Shöni, 2013). For our analysis random forests outperformed these hedonic methods.

Random forests are a collection of randomly generated decision trees. Decision trees have several nice properties: they are scale independent, robust to the inclusion of irrelevant features, and interpretable, but these come at the expense of overfitting and inaccuracy. Random forests average over the set of trees and are not subject to the biases that occur with ordinary least-squares linear regression when multicollinearity, nonlinearity, heteroskedasticity, and non-normality are present. At the cost of interpretability, random forests can be more accurate when these data issues are present. They can be utilized for prediction, which we use for the income estimates, and classification, which we use for the property appreciation modeling.

ADTF eligibility estimation

Determining the individual households eligible for the ADTF was crucial both to estimate the total cost of the program and for our interactive eligibility tool. As mentioned in Table 1, we identified the eligible households based on location, owner occupancy, the presence of liens, and income (which is dependent on the household size).

- *Location*: Using the Fulton County Tax Assessor 2017 data, we filtered out all non-empty, residential parcels which lie in Ashview Heights, Atlanta University Center, Vine City, English Avenue or Washington Park neighborhoods, which resulted in 2600 geographically eligible residential parcels. While Washington Park is not considered eligible for the ADTF, our community partners requested we include this neighborhood. The Booker T. Washington and Join Us neighborhoods were not included in the study as they were added to the program area after the analysis was conducted.
- *Owner occupancy*: From the Fulton County Tax Assessor 2017 data, we compared parcel addresses with owner addresses to determine if the household is owner-occupied. We also

considered all households who claim Homestead exemption, which requires the owner to be the occupant. About 36% of the homes were found to be owner occupied.

- *Lien status*: Lien data were gathered from the Georgia Superior Court Clerkso Cooperative Authority. Since the data were not machine readable, a random sample of 30% homeowners in the original four Westside neighborhoods and 45% Washington Park were gathered. Of the households in the four Westside neighborhoods, 59% did not have liens, nor did 58% in Washington Park.
- *Income*: We estimated household-level income by modeling the relationships between rent, home characteristics, and before-tax income for homeowners in Atlanta from the CEX with missing data first imputed with machine learning techniques. Based on ZIP code-level tax return data from 2014, we found that 90% of households in the Westside might qualify for the program based on household income which is significantly less than $47,250 (Internal Revenue Service, 2014). A simple assumption might be to consider all households eligible based on income criterion, but we attempted to predict household-level income eligibility based on observable, physical characteristics of a house and its expected rent. Since individual household income data was unavailable, we used the relationship for rent estimates and household income. For each house in the program area, we merged the house characteristics from the Fulton County Tax Assessor data and the Zillow rent estimates for each residential property in the Westside. Income estimates were derived from the relationship found between home characteristics, rent and income from the CEX and home characteristics and rent from the Zillow data (see the online Supplementary Material for a more detailed description of the methodology).

Property tax assessment forecasting

Although the new Mercedes-Benz Stadium may impact Westside Atlanta home values, previous studies of sports venues have found mixed effects on property values. A study of Texas metropolitan areas found that property values decreased after new sports venue announcements (Dehring et al., 2007). Another study found a positive price improvement near the construction of the FedEx stadium in Landover (MD) (Tu, 2005). Because of these conflicting findings, we excluded potential stadium-related effects in our estimates of home values and instead focused on the effect of the BeltLine trail construction, which has been shown to increase the property values in Atlanta neighborhoods around completed segments (Immergluck and Balan, 2017).

To estimate future tax assessments of homes in the Westside, we began with the assumption that the Westside would experience changes similar to the impact of completed BeltLine construction near the Old Fourth Ward neighborhood. This assumption is validated primarily by shared characteristics, including proximity to urban revitalization projects, proximity to Atlanta's downtown, proximity to industrial land use zones, and similar historic socioeconomic makeup. Due to the spatial and temporal variation and nonlinearity in the tax assessment time series trends, using a single static trend statistic, as the 2017 study did, was considered inappropriate. With clustering techniques several property appreciation time series trends were indicated and the most influential housing characteristics were identified for each cluster (see Figure S1, Figure S2, and Table S4 in the online Supplementary Material). The most important features that determine property appreciation trends were the distance to the BeltLine, property value in 2005 (the year before construction of the BeltLine) and age of the home. To forecast the future tax assessment values for homes in the Westside, we used the corresponding time series trends from the Old Fourth Ward neighborhood clusters after removing the recession time period data (2008–2012), which resulted

in projections to 2024. Several models were tested for the forecasting, including Multiple Linear Regression, Ridge Regression, Lasso Regression, Decision Trees, and Random Forests (see Table S5 in the online Supplementary Material for model validation results).

ADTF cost estimation

We conducted program costs estimates for several different scenarios: (i) accounting for lien rates and with and without residential parcels in Washington Park, (ii) disregarding liens rates with and without Washington Park, and (iii) including a 5% dropout rate and a 79% enrollment rate for all of the previous scenarios (which the original study incorporated and we considered for comparison). The Fulton County and City of Atlanta 2016 Millage rates were applied to tax assessment estimates after individual household exemptions were identified. Final program costs were calculated by summing all of the future property taxes minus the 2017 taxes for eligible households over the seven-year period. To simulate liens for households which we did not have data, we used simple random sampling with the sample lien rates for the neighborhoods. The distribution of program cost estimates was normal so the average program cost estimate was used in the final results. Simple random sampling was likewise used for estimating the program costs when household dropout and partial enrollment were considered.

Community eligibility and method validation tool

One of the primary goals of this research is to ensure that Westside community members are aware of the eligibility requirements of the fund. To assist residents in making an informed decision on participating in the fund, we developed an interactive web tool for Westside homeowners to view eligibility information and projected property taxes over the next seven years. Figure 2 shows the online mapping tool and the eligibility information for an example residence in the Westside. Information available with the tool includes owner name, income estimate, Homestead exemption, owner-occupancy, the presences of liens, forecasted home value, and overall eligibility. Canvassers will assist us in introducing the tool and the information it provides to residents in order to spread awareness about the program's impact. Through this canvassing, they can also update eligibility information within the tool to provide better program estimates and validate the modeling. Issues with residents sharing personal information and validating any responses are still expected, but are not different than those encountered with professional data (Specht and Lewandowski, 2018).

Using and sharing personal data of community members, even if they supply it willingly, raises several ethical issues. Resolving security issues with personal data are paramount as individuals or organizations may use this data for non-intended or personal use. For example, how do we ensure canvassers who have access to the tool use it responsibly? What are the rights of residents who share their information with respect to how it is stored and used? Should they be able to request the removal of information? Should the data and website be password protected to ensure third parties cannot use the information, such as predatory lenders? To provide security, we have password locked the tool and only supplied this password to canvassers.

Results

Estimated program enrollment and costs from utilizing our methods under different scenarios: (i) 79% enrollment with 5% annual dropout (which the original estimate from Bedsole (2017) used), (ii) full enrollment with 5% annual dropout, and (iii) full enrollment
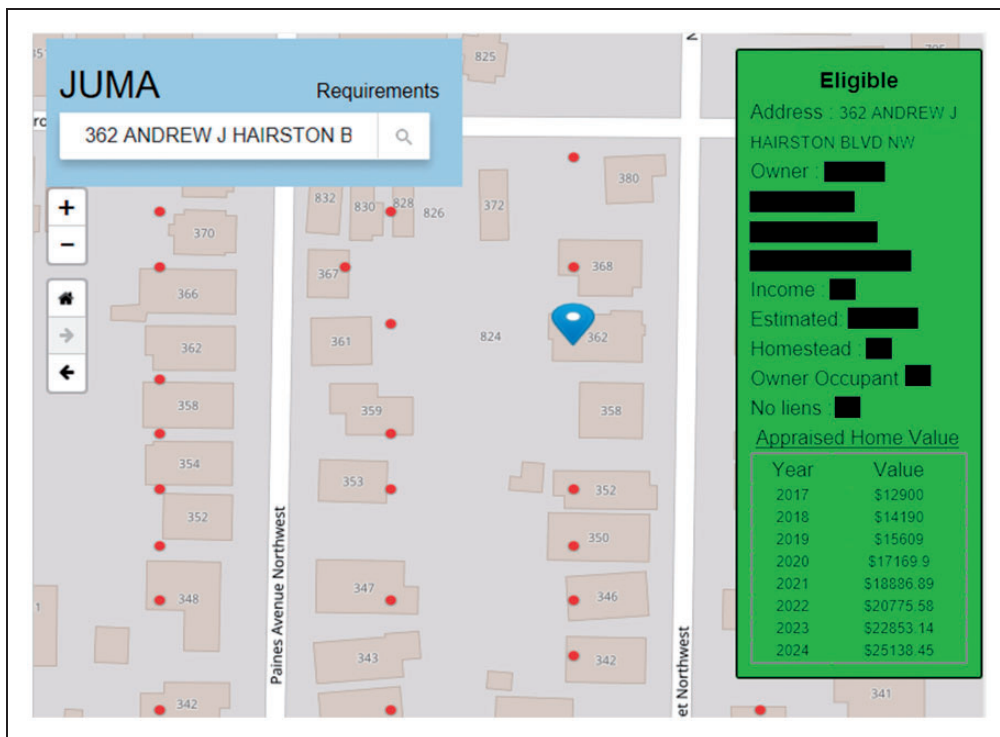
Figure 2. Web Mapping tool used to inform residents on their eligibility status and projected home value appraisals to the year 2024.

Table 2. Estimated seven-year costs of the program for different scenarios and the original program cost estimate (Bedsole, 2017).

	Eligible households		Total program cost	
	With liens	Without liens	With liens	Without liens
Without dropout and full enrollment				
With Washington Park	489	702	2.7M	4.0M
Without Washington Park	372	560	2.1M	3.2M
With 5% dropout and full enrollment				
With Washington Park	489 → 339	702→487	2.0M	3.1M
Without Washington Park	372 → 257	560 →389	1.6M	2.4M
With 5% dropout and 79% enrollment				
With Washington Park	384 → 268	555→385	1.6M	2.5M
Without Washington Park	294 → 203	442→307	1.3M	1.9M
Original estimate	–	400→275	–	0.3M

Note: The original estimate did not consider liens.

without any dropout are provided in Table 2. The original program estimates are included for comparison. We show that the initial study of the ADTF overstated the number of eligible participants and significantly underestimated the total program cost, even though we include a large lien rate, which disqualified 40% of the otherwise eligible homeowners.
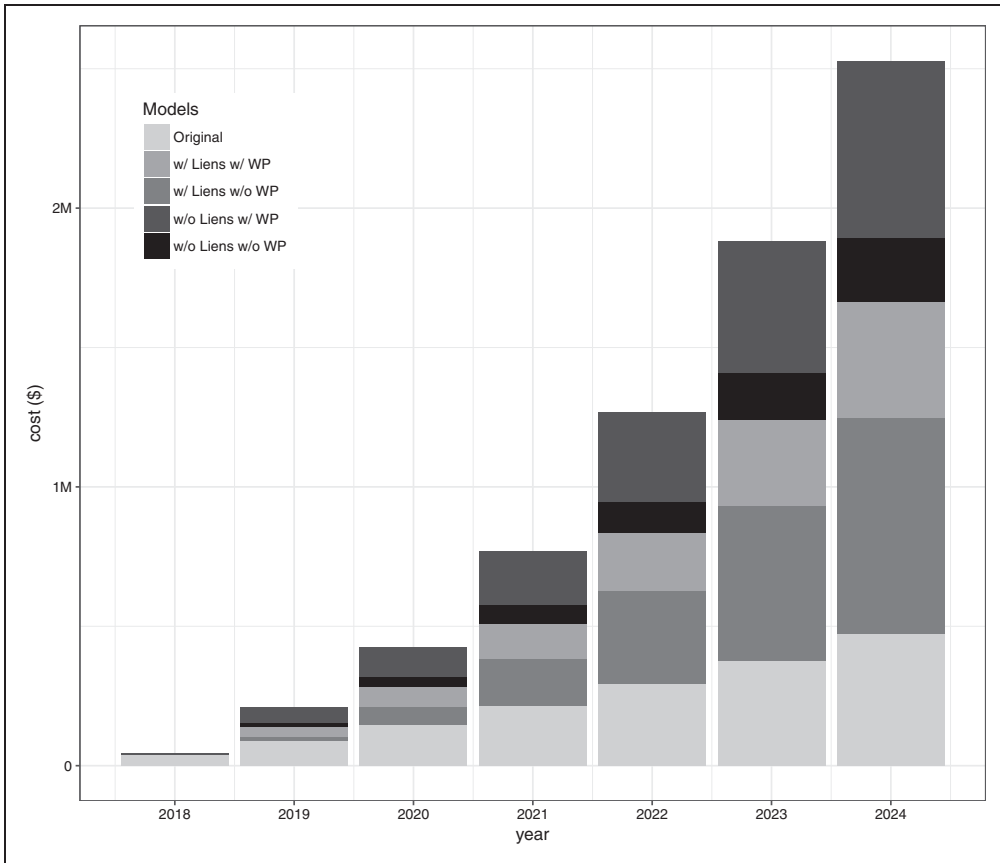
Figure 3. Total program costs for several models, with the 5% dropout rate and 79% enrollment which the original model included, during the first seven-year period (Bedsole, 2017).

Figure 3 presents the annual projected costs of the program for the first seven years under the different scenarios and includes the original cost estimates for comparison. While our estimates are limited to seven years due to our home value appreciation modeling, we find the larger property value appreciation rates in our evaluation will drastically increase the program cost over 20 years compared to the initial projection even with our lower number of eligible homeowners. Furthermore, the recent inclusion of the Booker T. Washington and Just Us neighborhoods will increase the number of eligible participants and the total program cost compared to our original estimates. If community members can convince the ADTF to include the Washington Park neighborhood, which is an expressed hope of Washington Park residents, the number of eligible participants and the total program cost will also be significantly larger.

These differences in program cost estimates are primarily due to our use of granular data; using individual household-level data for a program aimed at individual homeowners is more appropriate for this type of program evaluation. Identifying which homes were occupied by the homeowner had a significant impact on the number of eligible homes. By hand-checking a sample of households for lien data in the affected neighborhoods, we were able to identify a very large lien rate for homeowners in the Westside neighborhoods which

the original study ignored without justification. Using machine learning techniques on the filtered CEX data and Zillow rent estimates for homes in our study, we could estimate the household-level income for homeowners in the Westside compared to the aggregated ACS data used in the original study.

Our program estimates also benefited from the use of more advanced statistical techniques that are generally not applied to program evaluation. Our clustering of the tax assessment time series data revealed significant variation within property value appreciation for the Old Fourth Ward neighborhood after the announcement and during construction of the nearby BeltLine. We could identify that urban revitalization in the area considerably weakened the impact of the recession on neighborhood property values, albeit with different effects for homes with different characteristics, such as home age or distance to the urban revitalization project. We identified these pertinent home characteristics from the clusters, even though there were several classic statistical issues with the data, such as strong correlations between home features. Our results found proximity to the urban revitalization project, the value of the home prior to construction of the project, and the age of the home were the most important home characteristics for increased home value appreciation. These findings resulted in property appreciation rates that were larger than those in the original program study and we observed nonlinear trends over the time period, where the original analysis used flat rates.

Value of community-driven data science for program evaluation in planning

This project demonstrates the value of community-driven data science for academia, city planners, and community-based advocates alike. In comparison with the original study, our findings show that the estimated cost of the tax fund could be significantly higher than expected. Our model predicts that fewer homeowners will be able to qualify for the program than the previous study found, yet we predict similar costs for the program during the first seven years. This is due to larger tax assessment appreciation rates and we expect the program to ultimately cost more than the previous projection. Accordingly, we find machine learning techniques to be valuable tools for quantifying similar anti-gentrification initiatives. This work highlights that machine learning techniques and data-driven program evaluation can be valuable for measuring the impacts of urban projects. This work highlights that machine learning techniques and data-driven program evaluation could equip policy makers with more information on program outcomes as they consider and compare policies and programs that will have the greatest desired impact. City planners are not often trained in these fields and where these skills are lacking, cities can benefit from academic partnership.

While this exercise in data science proves useful to planners and local philanthropists, our primary partners are community residents. The results of this project support the community in three ways. First, Atlanta's top officials, planning commissioner, and philanthropic community celebrated the announcement of the ADTF as a solution to curb displacement, but our exercise in data science objectively legitimizes the community's instinct that the ADTF is insufficient. Second, the ADTF becomes much more transparent. Residents now know the estimated cost of the fund for the next seven years and can use the visual eligibility tool to determine if they are eligible and make an informed decision to participate. Not only is the process of qualification more transparent, but the research we conducted to inform our modeling will also inform the community of what the ADTF entails.

For example, property lien status is key component of the eligibility requirements. Although our results are preliminary, our research and modeling reveal that many of the homes within the five neighborhoods have liens associated with them. This dramatically reduces the number of residents that are eligible to receive help from the fund, which in turn decreases the overall value and scope of the tax fund. With this knowledge, the community residents could advocate that part of the tax fund be used to pay off liens within a specified threshold. Additionally, we have given examples of how to meaningfully include residents in data science. Through this experience we offer the following set of recommendations for those interested in practicing community-engaged data science. First, all data science projects begin with understanding the problem space. This stage offers rich opportunity to learn from residents' situated knowledge. Second, we advise consulting or ground-truthing modeling assumptions. Involving community partners in data collection as we did with lien data improves the completeness of the dataset available to data scientists and we also suspect it could improve data capacities on the part of community partners.

Lastly, the tool has been publicized to residents within the Westside neighborhoods during community meetings. It has been shared with city officials and affordable housing advocates with the goal of furthering the discussion of what else could be done to prevent displacement. Quantifying the program also allows residents to offer alternative solutions. The community has already voiced two primary concerns for why the ADTF will not fully address displacement. Quantifying the fund with reliable data on who qualifies allows for greater legitimacy as they continue to advocate for ways to improve the impact of the fund or alternatives like permanently affordable housing through the community land trust model. The tradeoff between how many residents are actually eligible to participate and the projected length of the program are also factors to consider. It needs to be decided whether longevity for fewer residents is more important, or if having as many residents as possible participate for a shorter time is the best solution. Either way, the data and visual eligibility tool become resources for community groups to continue the dialogue on how to achieve development without displacement for as many residents as possible.

Future work

Our work offers a descriptive example of how to engage community members in civic data science projects and based on our results we urge urban planners to make use of community-engaged data science in the decision making and planning process. Our predictive modeling is limited by time and the availability of data. Following the inauguration of the ADTF in 2018, much more data should be available about the number of enrolled households,[1] and program cost and scope will be easier to predict. Recalculating the projected costs of the program each year is essential to avoid surprises and keep the program functioning for as long as possible. Additionally, it is vital that future predictions be made public and accessible to community members so that they can advocate for their rights and be more aware of how the tax fund and other anti-displacement measures are shaping their rapidly transforming community.

A member of the Westside Atlanta Land Trust was trained to collect additional lien data from the County Clerk's records. If the mapping tool is used to advertise the ATDF and collect more accurate eligibility information from homeowners in the affected neighborhood, these data could be used to identify if there are significant differences in lien rates between neighborhoods, update the eligibility estimates, and recalculate the cost of the program. Data collection can go beyond the scope of program evaluation. Canvassing could also collect information used for longitudinal studies of urban revitalization projects

and their relationship to displacement, such as knowing where previous residents have moved to and why. Household size estimates can be improved with the techniques provided in Talent (2016) if sample household surveys are conducted in the program area.

Acknowledgements

Declaration of conflicting interests

Funding

ORCID iD

Jeremy Auerbach http://orcid.org/0000-0003-0061-7943

Note

1. Through communication with WALT, as of April 2018, it is estimated that 40 homeowners applied and were found to be eligible.

References

Armstrong RJ Jr (1994) Impacts of commuter rail service as reflected in single-family residential property values. *Transportation Research Record* 1466: 88–98.

Arnstein SR (1969) A ladder of citizen participation. *Journal of the American Planning Association* 35(4): 216–224.

Atlanta BeltLine, Inc. (2017) The Atlanta Beltline in 5. Available at: https://beltline.org/about/the-atlanta-beltline-project/atlanta-beltline-overview (accessed 4 December 2018).

Bartlett S (1997) The significance of relocation for chronically poor families in the USA. *Environment and Urbanization* 9(1): 121–132.

Bedsole M (2017) *Developing a charitable and targeted property tax relief fund: An anti-displacement initiative for Atlanta's Westside neighborhoods.* Master's Thesis, Georgia Institute of Technology, Atlanta, GA.

Belson K (2017) Building a stadium, rebuilding a neighborhood. Available at: https://www.nytimes.com/2017/01/12/sports/football/atlanta-falcons-stadium-arthur-blank-neighborhood.html (accessed 4 December 2018).

Berke PR and Conroy MM (2000) Are we planning for sustainable development? An evaluation of 30 comprehensive plans. *Journal of the American Planning Association* 66(1): 21–33.

Burby RJ (2003) Making plans that matter: Citizen involvement and government action. *Journal of the American Planning Association* 69(1): 33–49.

Bureau of Labor Statistics (2016) Consumer expenditure survey. Available at: www.bls.gov/cex/ (accessed 4 December 2018).

Cardullo P and Kitchin R (2018) Being a 'citizen' in the smart city: Up and down the scaffold of smart citizen participation in Dublin, Ireland. *GeoJournal* 83(339): 1–13.

Ceccato VA and Snickars F (2000) Adapting GIS technology to the needs of local planning. *Environment and Planning B: Urban Analytics and City Science* 27(6): 923–937.

Chen H, Rufolo A and Dueker K (1998) Measuring the impact of light rail systems on single-family home values: A hedonic approach with geographic information system application. *Transportation Research Record: Journal of the Transportation Research Board* 1617: 38–43.

Dastrup S, Ellen I, Jefferson A, et al. (2015) The effects of neighborhood change on New York City Housing Authority residents. Technical Report, NYC Center for Economic Opportunity, May.

Dehring CA, Depken CA and Ward MR (2007) The impact of stadium announcements on residential property values: Evidence from a natural experiment in Dallas-Fort Worth. *Contemporary Economic Policy* 25(4): 627–638.

Diewert E (2003) Hedonic regressions: A review of some unresolved issues. National Bureau of Economic Research Working Paper. Available at: www.nber.org/CRIW/papers/diewert.pdf (accessed 4 December 2018).

Ding L, Hwang J and Divringi E (2016) Gentrification and residential mobility in Philadelphia. Technical Report, Federal Reserve Bank of Philadelphia.

Ellen IG and Ondende KM (2011) How low income neighborhoods change: Entry, exit, and enhancement. *Regional Science and Urban Economics* 41(2): 89–97.

Franco G (2017) Commissioners 'freezes Fulton County property tax levels. Available at: www.fox5atlanta.com/news/commissioners-freeze-fulton-county-property-tax-levels (accessed 4 December 2018).

Freeman L and Braconi F (2004) Gentrification and displacement: New York City in the 1990s. *Journal of the American Planning Association* 70(1): 39–52.

Fulton County Board of Assessors (2017) Tax assessment data. Available at: http://fultoncountyopendata-fulcogis.opendata.arcgis.com/datasets/parcels/data.

Gabrys J, Pritchard H and Barratt B (2016) Just good enough data: Figuring data citizenships through air pollution sensing and data stories. *Big Data & Society* 3(2): 1–14.

Georgia Superior Court Clerkso Cooperative Authority (2017) Lien Index. www.gsccca.org/ (accessed 4 December 2018).

Hartman C (2003) Introduction and overview: Student mobility: How some children get left behind. *The Journal of Negro Education* 72(1): 1–5.

Hope M and Young J (1986) The politics of displacement: Sinking into homelessness. In: Erickson J and Wilhelm C (eds) *Housing the Homeless*. London: Routledge, pp.106–112.

Immergluck D (2009) Large redevelopment initiatives, housing values and gentrification: The case of the Atlanta Beltline. *Urban Studies* 46(8): 1723–1745.

Immergluck D and Balan T (2017) Sustainable for whom? Green urban development, environmental gentrification, and the Atlanta Beltline. *Urban Geography* 39(4): 1–17.

Institute of Governmental Studies (2017) What is the nature of displacement in the bay area and southern California? Available at: www.urbandisplacement.org/ (accessed 4 December 2018).

Internal Revenue Service (2014) Individual income tax statistics – 2014 ZIP Code Data. Available at: www.irs.gov/uac/soi-tax-stats-individual-income-tax-statistics-2014-zip-code-data-soi (accessed 4 December 2018).

Jasanoff S (2017) Virtual, visible, and actionable: Data assemblages and the sightlines of justice. *Big Data & Society* 4(2): 1–15.

LeGates RT and Hartman C (1982) Gentrification-caused displacement. *The Urban Lawyer* 14(1): 31–55.

Lester TW and Hartley DA (2013) The long-term employment impacts of gentrification in the 1990s. Technical Report, Federal Reserve Bank of Cleveland.

Levy DK, Comey J and Padilla S (2006) In the face of gentrification: Case studies of local efforts to mitigate displacement. Technical Report, The Urban Institute.

Los Angeles Innovation Team (2016) Los Angeles index of neighborhood change. Available at: www.arcgis.com/home/item.html?id=57e9231c3bd34d44ae49b309b0cb440e (accessed 4 December 2018).

McKinnish T, Walsh R and White TK (2010) Who gentrifies low-income neighborhoods? *Journal of Urban Economics* 67: 180–193.

McMillen DP and McDonald J (2004) Reaction of house prices to a new rapid transit line: Chicago's Midway line, 1983–1999. *Real Estate Economics* 32: 463–486.

Marcuse P (1985) Gentrification, abandonment, and displacement: Connections, causes, and policy responses in New York City. *Journal Urban & Contemporary Law* 28: 195–240.

Martin E, Cohen A, Botha J, et al. (2016) Bikesharing and bicycle safety. Technical Report, Mineta Transportation Institute.

Metropolitan Area Planning Council (2015) Managing neighborhood change: Anti-displacement strategies. Technical Report, Metropolitan Area Planning Council.

Nature Editorial (2018) Science benefits from diversity. *Nature* 558(5). Available at: https://www.nature.com/articles/d41586-018-05326-3 (accessed 4 December 2018).

Schrock AR (2016) Civic hacking as data activism and advocacy: A history from publicity to open government data. *New Media & Society* 18(4): 581–599.

Shöni O (2013) *Four essays on statistical problems of hedonic methods: An application to single-family houses*. PhD Thesis, University of Fribourg.

Sieber R (2006) Public participation geographic information systems: A literature review and framework. *Annals of the Association of American Geographers* 96: 491–507.

Specht H and Lewandowski E (2018) Biased assumptions and oversimplifications in evaluations of citizen science data quality. *The Bulletin of the Ecological Society of America* 99(2): 251–256.

Talent M (2016) Improving estimates of occupancy rate and population density in different dwelling types. *Environment and Planning B: Urban Analytics and City Science* 44(5): 802–818.

Tribby CP, Miller HJ, Brown BB, et al. (2017) Analyzing walking route choice through built environments using random forests and discrete choice techniques. *Environment and Planning B: Urban Analytics and City Science* 44(6): 1145–1167.

Tu CC (2005) How does a new sports stadium affect housing values? The case of FedEx field. *Land Economic* 81: 379–395.

United States Census Bureau (2017) Summary File. 2011–2015 American Community Survey. U.S. Census Bureau's American Community Survey Office. Available at: http://ftp2.census.gov/ (accessed June 1 2018).

Westside Future Fund (2017) Anti-displacement tax fund fact sheet. Available at: www.westsidefuturefund.org/wp-content/uploads/2017/04/Anti-Displacement-Tax-Fund-Factsheet_FINAL.pdf?preview=true (accessed 4 December 2018).

Zhang Q, Gao W, Su S, et al. (2017) Biophysical and socioeconomic determinants of tea expansion: Apportioning their relative importance for sustainable land use policy. *Land Use Policy* 68: 438–447.

Zillow (2017) Zillow rent estimates. Available at: www.zillow.com (accessed 4 December 2018).

Jeremy Auerbach is a postdoctoral research fellow in the Department of Environmental and Radiological Health Sciences at Colorado State University working to understand the relationships between a neighborhood's built environment and its resident's health and behavior.

Christopher Blackburn is a PhD candidate in Economics at the Georgia Institute of Technology with interests at the intersection of environmental economics, urban economics, and network science.

Hayley Barton is an economics student at the Duke University.

Amanda Meng is a research scientist at the Georgia Institute of Technology and engages in resident-led research and studies grassroots data activism, publishing in the fields of critical data studies and science, technology and society.

Ellen Zegura is professor and Chair of Computer Science at the Georgia Institute of Technology.