Andrew Yen
AS 430.604.81
Spatial Analytics Final Report

Outlining an Apparent Affordable Housing Development Paradigm for Baltimore City:

Background

Baltimore City has a long history of segregation that extends as far back as the Reconstruction Era (MacGillis, 2016; Power, 1983), and is still felt by residents today (Lee & Rios, 2015; MacGillis, 2016; Willeboordse, 2017). Segregation in Baltimore has played a dramatic role in imbalances in home equity, access to resources, and is further exacerbated by the segregation of affordable housing away from wealthier, more convenient neighborhoods (Faber & Ellen, 2016; Rothstein, 2018).

I wanted to use publicly available demographic, economic, and crime data to determine the degree to which contemporary affordable housing projects are built in less desirable parts of the city compared to market rate housing. By firming up suspicions of inequality with data that demonstrates inequality, we can be better equipped to challenge the political and economic mechanisms that work to perpetuate the imbalance.

Affordable housing in particular is surrounded by a polarized discussion, between local government, intractable NIMBYs (Not-In-My-Backyard), affordable housing developers, and price market advocates, making the placement of projects challenging (Cholo, 2016). However, many urban planning and urban sociology experts agree on what locations constitute effective placement for affordable housing. In my research, I will test sites based on their access to businesses, public transit, and areas with low rental affordability index scores — in short: the least affordable and most desirable places to live. I plan on constructing an effectiveness rating index based on the above criteria to assign to affordable housing projects in the final stage of the efficiency analysis.

Corporate interests and elected officials (rather than citizens) typically make development decisions — that require a significant amount money and land — often without adequate consultation with those who will be directly impacted by them (Jacobs, 1961; Scott, 1998). I hope to better understand the factors that govern decisions on urban development projects, and assess the effectiveness of these decisions. Knowing how our cities grow and change will improve our understanding of how to change current practices to better match the desires of the people who use the city.

The conclusions drawn in this analysis are meant to be understood by anyone with a stock in remediating housing inequality in Baltimore. The accompanying web app will feature data that would be familiar to any resident of Baltimore, from public transit routes to neighborhood blocks.

Introduction

I lived in Baltimore City for four years and have seen segregation and neighborhoods ravaged by vacancy concurrently with gentrification and rapid urban redevelopment in the Inner Harbor and Harbor East. If my experience of living in Baltimore is any indication, much needs to be done to correct the imbalance in access to resources and equity for the city's residents — particularly as construction has ramped up to accommodate the swelling population of the more expensive downtown area ("World Population Review", 2018).

In my analysis, I compare characteristics of affordable housing and market rate housing to assess the efficiency of affordable housing projects in Baltimore City. I define efficiency as the degree to which the demographic, economic, and criminological characteristics of affordable housing projects vicinity represent the typical circumstances experienced by those living in market housing.

I use the same criteria used to assess placement efficiency to select an ideal site for a future, hypothetical affordable housing construction project. These criteria include: proximity to emergency services (two minute driving time from fire stations and hospitals), work and retail amenities (quarter mile walk from commercial or office zoned areas), and Baltimore's Circulator bus route, a free bus service that has routes throughout the main downtown area (quarter mile walk from any Circulator stop). The ideal site could not be located in a food desert (areas of the cities longer than a quarter mile walk from a supermarket — this definition from the Baltimore City Health Department). The final site was chosen based on which of the remaining areas had the highest median sale price of homes, a proxy for neighborhood desirability (Valdez, 2018).

The last stage of the analysis addresses the effectiveness of the certain demographic, economic, and criminological variables at predicting housing type across all housing points in the dataset (market rate and affordable housing) using Forest-based Classification and Regression. The explanatory variables used were: median household sale price, average days on the market for homes for sale, the number of homes sold in 2016, percentage of homes owner occupied, vacancy rate, mortgage affordability index score, rental affordability score, unemployment rate, number of businesses, percent population black, percent population white, median household income, percent of households with an annual income higher than $75,000, percent of households under the poverty line, violent crime rate, and broken street light rate. All attributes were joined to house point attribute tables based on which community statistical area they were located.

This research operates under a number of assumptions that may introduce error in the results. First of all, not all of the data are contemporary with one another. All of the statistical data stored in Baltimore community statistical areas (CSAs) is from 2016, while the construction project points vary from 2008 to 2018, as some projects in the dataset have been completed while others are still in progress or planned. The second important assumption is that the demographic, economic and criminological data can be used to characterize individual housing projects even though the data has been summarized by much larger CSAs.

More granular data would have been useful for better defining the experience of residents of affordable and market rate housing. This would have also been the case for the work/retail proximity section of the efficiency analysis. Large blocks of the city would be zoned into a generalized “commercial” district without much indication of how well the businesses in these areas are performing, or more to the point, how well these businesses satisfy the needs of those living nearby. Baltimore has a noteworthy problem with vacant houses which, based on my personal experience, extends to retail spaces in many areas as well.

Image 1 *Shuttered blocks like this are a common sight in Baltimore’s many distressed communities. The problem extends to residential blocks as well.*

Data

The analysis will revolve around demographic and economic data from the Baltimore Neighborhood Indicators Alliance (BNIA; https://bniajfi.org/), a research institute focused on strengthening Baltimore City neighborhoods. BNIA publishes “Vital Signs” statistics that range from income and race, to newborn birth weights and street light outages. These neighborhood indicators (summarized by community statistical areas) will be cross-referenced with upcoming and ongoing construction projects in Baltimore, available through EconView data prepared by the Baltimore Development Corporation.

BNIA Data:

The BNIA maintains an ArcGIS Online Data portal with Vital Signs shapefiles summarized by CSA, grouped thematically into several categories (crime, demographics, housing, etc.). The range of demographic information is very extensive and provides a satisfying foundation for profiling areas. Relevant indicators were extracted from different shapefiles and merged to generate a custom vital signs dataset. For the above preprocessing task, a file geodatabase in ArcGIS Pro was used to store and manage data. BNIA provides a full glossary of indicator definitions and has been responsive to data inquiries via email.

The dataset is comprised of several polygon feature layers packaged into categories (e.g. Housing and Development, Demographics, etc.). Each feature layer has 55 polygon features, each corresponding to a community statistical area (CSA) of Baltimore City[1]. I relied on four of these curated shapefile packages: Housing, Economics, Demographics and Crime. All were available through BNIA’s ArcGIS Online data portal.

EconView Data:

EconView point shapefiles are available via Baltimore City’s website (http://cityview.baltimorecity.gov/econview/). This dataset is not as large or diverse as the BNIA Vital Signs data and did not require intensive preprocessing. I extracted affordable housing construction from other residential projects to conduct more in depth analysis about how affordable housing fits into Baltimore’s landscape. I treated the EconView data as a representation of current urban development trends in Baltimore. The data present a number of accuracy issues: the construction projects detailed in the dataset are in varying stages of completion. Some of the projects have been already been underway from as early as 2008, and others are either completed or only planned. Whether or not these projects (and prospective projects) can paint a full picture of urban development in Baltimore is impossible to tell without more comprehensive and multitemporal data. Even so, the dataset has 560 fairly well distributed data points and a fair diversity of project types to work with. There is no metadata

[1] See BNIA documentation for details at:
https://bniajfi.org/wp-content/uploads/2014/04/CSA_Tracts_2000.pdf

available for the EconView, but Baltimore City's Department of Planning have been available for data inquiries.

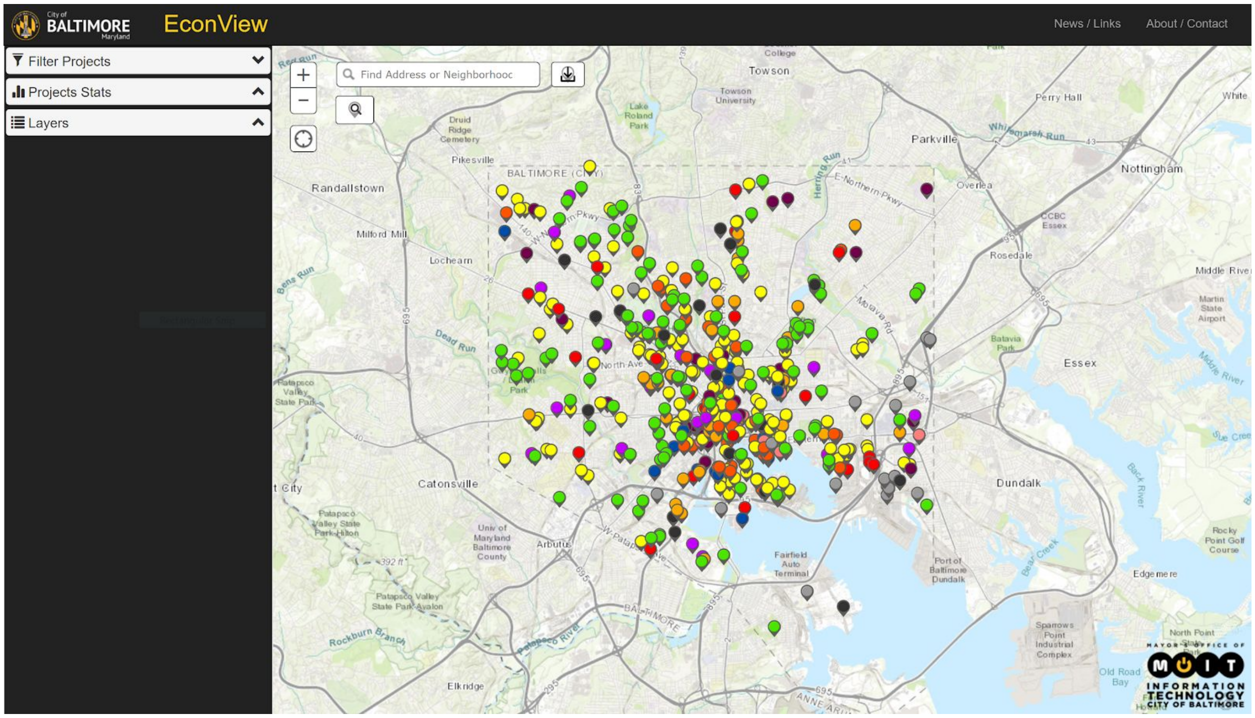


Image 2 *EconView construction data prepared by the Baltimore Development Corporation and hosted on the City of Baltimore's "CityView" data portal (http://cityview.baltimorecity.gov/econview/)*

Food Desert Data:

This dataset is a collaboration between Johns Hopkins University and the Baltimore City Health Department polygon shapefile offered through JHU's ArcGIS Online data portal (https://hub.arcgis.com/datasets/95e5df8341e1476f824a578c8fd8aa54_126)[2]. The only fields are an OID field and shape area, which was used to eliminate areas to consider for affordable public housing construction.

Hospitals & Fire Station Data:

Hospital and fire station locations throughout Baltimore City used as another factor to measure efficiency of affordable housing placement; all housing should be constructed reasonably close to emergency services in order to be considered efficient. This data is available from Baltimore City's open data hub at https://data.baltimorecity.gov/

[2] For more detailed, statewide data visit https://gis.mdfoodsystemmap.org/map/#x=-8583266.780196756&y=4688647.315011486&z=7&ll=2,3

Database Schema

Data were imported into ArcGIS Pro by creating a local folder connection. After assessing the usefulness of datasets, layers selected for analysis were preprocessed and then extracted to a feature geodatabase. Analysis products were stored in this same location.

Analysis Workflow

1. Efficiency Assessment

Efficiency was determined based on how disparate neighborhood statistics of affordable housing projects were compared to market rate constructions.

In some cases, comparisons were made based on proximity of housing to resources, namely: Emergency Services (a category that includes fire stations and hospitals), Work and Retail infrastructure, and the free Circulator bus routes.

- *Distance to Resources:*

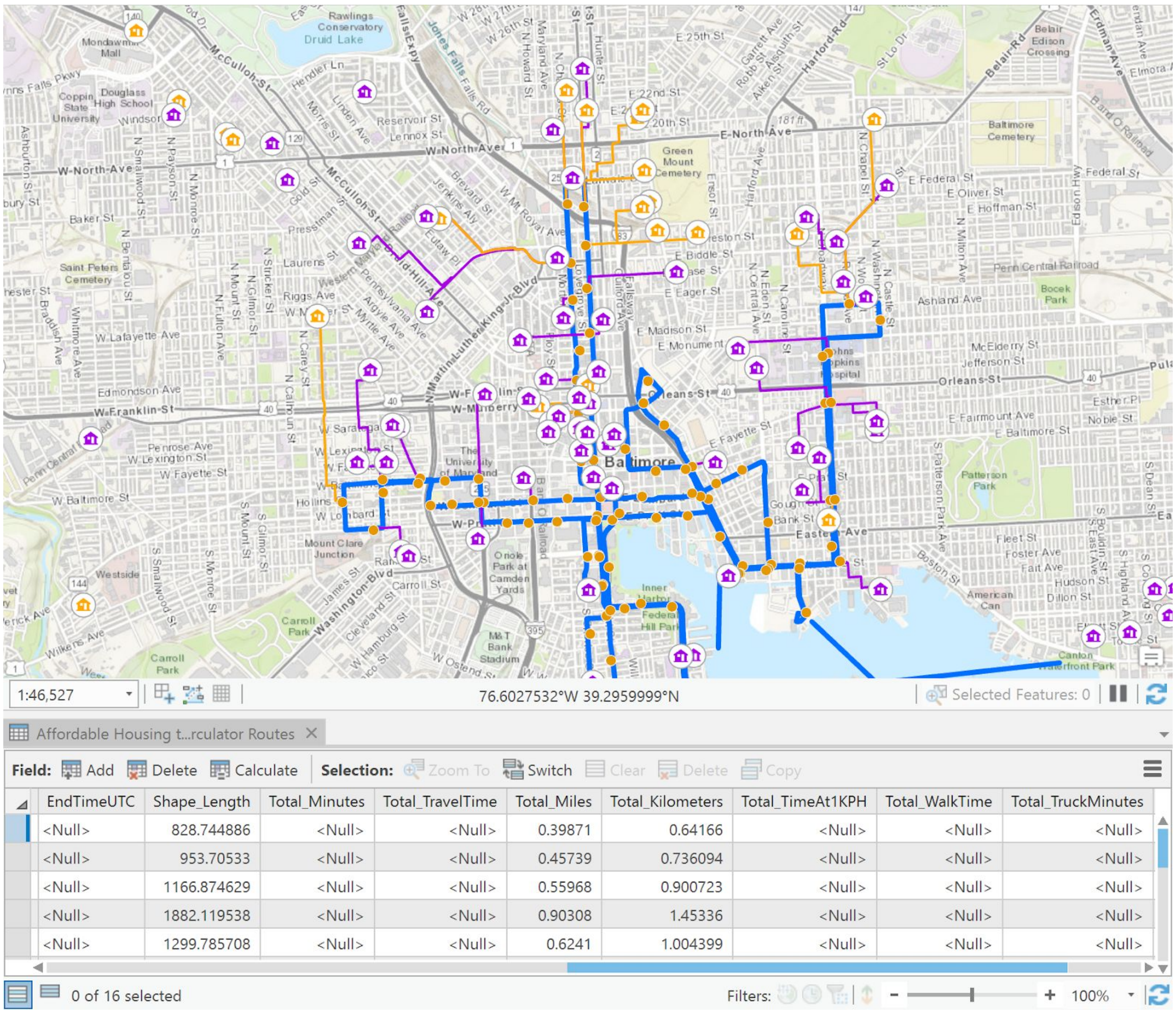
For these comparisons, the Network Analysis tool, *Closest Facility* in ArcGIS Pro, was used to calculate the appropriate cost for affordable housing and market rate housing separately. Housing points were treated as the “Incident” layers for the the tools, while hospital, fire station, and bus stops were treated as “Facilities”. In order to find the distance from work and retail infrastructure from housing, it was necessary to convert the original zoning polygons into points on the network using the *Feature to Point* tool.

The Closest Facility solver was parameterized appropriately for each Facility layer:

- Distance between housing and emergency services facilities was measured in terms of drive-time from facilities, with no cutoff, and search for only one facility per incident.
- Because I assumed that most people would prefer to be within walking distance of retail outlets and their work (rather than strictly driving distance), the solver for this relationship was parameterized to measure distance in terms of walking distance towards facilities, with no cutoff, searching for the first facility.[3]
- The closest Circulator bus stops to houses were searched for by walking distance as well. It was necessary, however, to assign a cutoff of one and a half miles from facilities because the Circulator only provides service to the downtown

[3] With more granular data (block-level detail) I would have liked to clarify and disaggregate the work/retail polygon features in order to measure distance from work or retail separately. Adding weights to offices or retail outlets with more employees would have been a nice feature too.

area of Baltimore — prohibitively far for many living in housing points in the rest of the city. The reason I chose to focus the analysis on the Circulator was two-fold: one, because it is free and provides access to the most desirable neighborhoods of the city to those who don't use cars, and two, because although Baltimore City's paid bus system is fairly comprehensive in terms of coverage, it is much less frequent and reliable than the Circulator routes.



EndTimeUTC	Shape_Length	Total_Minutes	Total_TravelTime	Total_Miles	Total_Kilometers	Total_TimeAt1KPH	Total_WalkTime	Total_TruckMinutes
<Null>	828.744886	<Null>	<Null>	0.39871	0.64166	<Null>	<Null>	<Null>
<Null>	953.70533	<Null>	<Null>	0.45739	0.736094	<Null>	<Null>	<Null>
<Null>	1166.874629	<Null>	<Null>	0.55968	0.900723	<Null>	<Null>	<Null>
<Null>	1882.119538	<Null>	<Null>	0.90308	1.45336	<Null>	<Null>	<Null>
<Null>	1299.785708	<Null>	<Null>	0.6241	1.004399	<Null>	<Null>	<Null>

Image 3 *Example of the results of the Closest Facility solver for distance from housing to Circulator bus stops.*

- *Access to Food*:

  Comparing housing types relative access to groceries was done by performing a *Spatial Join* between housing points by intersection with a food desert polygon layer. I then added a new categorical field that would indicate whether a housing project was within or without a food desert by referencing the "Join_Count" field (points with a join count of 1 would have this new field calculated as "Inside Food Desert", points where join count =

0 would be calculated as “Outside Food Desert”. I took the additional step of selecting housing points by join count and then extracting the selection into a new layer for each housing type in order to symbolize housing points inside/outside food deserts differently.

- *Relative Residential Real Estate Value:*

I used the median house sale price per CSA as a proxy for the relative desirability of neighborhoods; as living spaces become more convenient, safer, or better equipped, housing prices tend to climb in response. In order to illustrate differences between discrete housing points, I used performed a *Spatial Join* of the CSA housing data (which included the median house sale price statistic) and housing points. Then created histograms to see the distribution and mean of housing prices for both housing types.

2. Ideal Site Selection

The ideal site area for a hypothetical affordable housing project was accomplished by narrowing the Baltimore City area step by step based on distance to/away from resources or whether or not features fell within disqualifying areas like food deserts.

- *Distance from Resources:*

Unlike the previous network analysis workflow performed for the efficiency assessment, the *Service Area* solver was used in order to generate acceptable areas where the prospective affordable housing project could be built.

The service areas were calculated for emergency services and Circulator bus stops, and were parameterized appropriately for the mode of transportation and cost unit of measurement:

- Emergency services service areas were defined by the driving time from facilities, with a cutoff of two minutes.
- Circulator service areas were defined by a walking distance of a quarter mile.

- *Work and Retail Access:*

I elected not to run the *Service Area* network analysis for office and commercial zoning points on the network for two reasons: one, that the network points for work and retail locations represented much larger polygons, meaning that the “service area” would have been extremely underestimated, and two, that these commercial and office zoning areas did not preclude the construction of residential projects, being generalized zoning areas.

Having near instant access to sundries, or ones place of work would be very convenient for residents.

Instead, the commercial and office zoning areas were *Intersect*ed with the acceptable areas generated by the service area analyses.

- *Food Access:*

  Food desert polygons were treated as a mask that disqualified any portion of the acceptable service areas that it covered. The *Erase* tool was used for this procedure.

- *Relative Residential Real Estate Value:*

  The final stage of the site selection involved referencing the median household sale price contained in the CSA polygon layer. After masking unacceptable areas and intersecting all of the acceptable areas, I used the *Dissolve* tool to combine the areas. I made sure to specify the median household sale price as the one of the *Dissolve Fields*. My reasoning was the after the areas were selected by their proximity to resources, the portion of area within the CSA with the highest median household sale price would be the final candidate area.

  The final acceptable area was adequately accessible to resources and in the 90^{th} percentile of house sale prices (rank 6 of 55).

3. Regression Analysis

The workflow at this point has focused on confirming a presumption of difference affordable and market rate housing projects based on comparisons of single variables. I used *Forest-based Classification and Regression* to see the degree to which the housing attribute data could be used to tell the story of the disparity between housing types. I parameterized the tool using a feature class with all housing points containing CSA attribute data from all the relevant packages (demographics, economics, etc.), and assigned the categorical field, 'HousingType' as the *Variable to Predict.* I selected explanatory variables that seemed to best match my preconceptions of what causes home equity disparity — and the segregation of affordable housing placement. I was interested any statistics that seemed to tell a story about housing demand in neighborhoods (vacancy rate, number of houses sold, etc.), stark income divides (household annual incomes above $75,000 or households below the poverty line), cost of living (rental affordability index), racial composition, or quality of life (violent crime, street light outages).

I performed a number of iterations of the supervised classification, one for each of the indicator groups listed above to see if anyone group of explanatory variables was exceptionally good at predicting the correct housing type. Predictions were compared to actual housing type by joining the ‘HousingType’ field to the output tables. Using this information I selected correct and incorrect predictions for confusion matrices to assess the classification accuracy for each iteration.

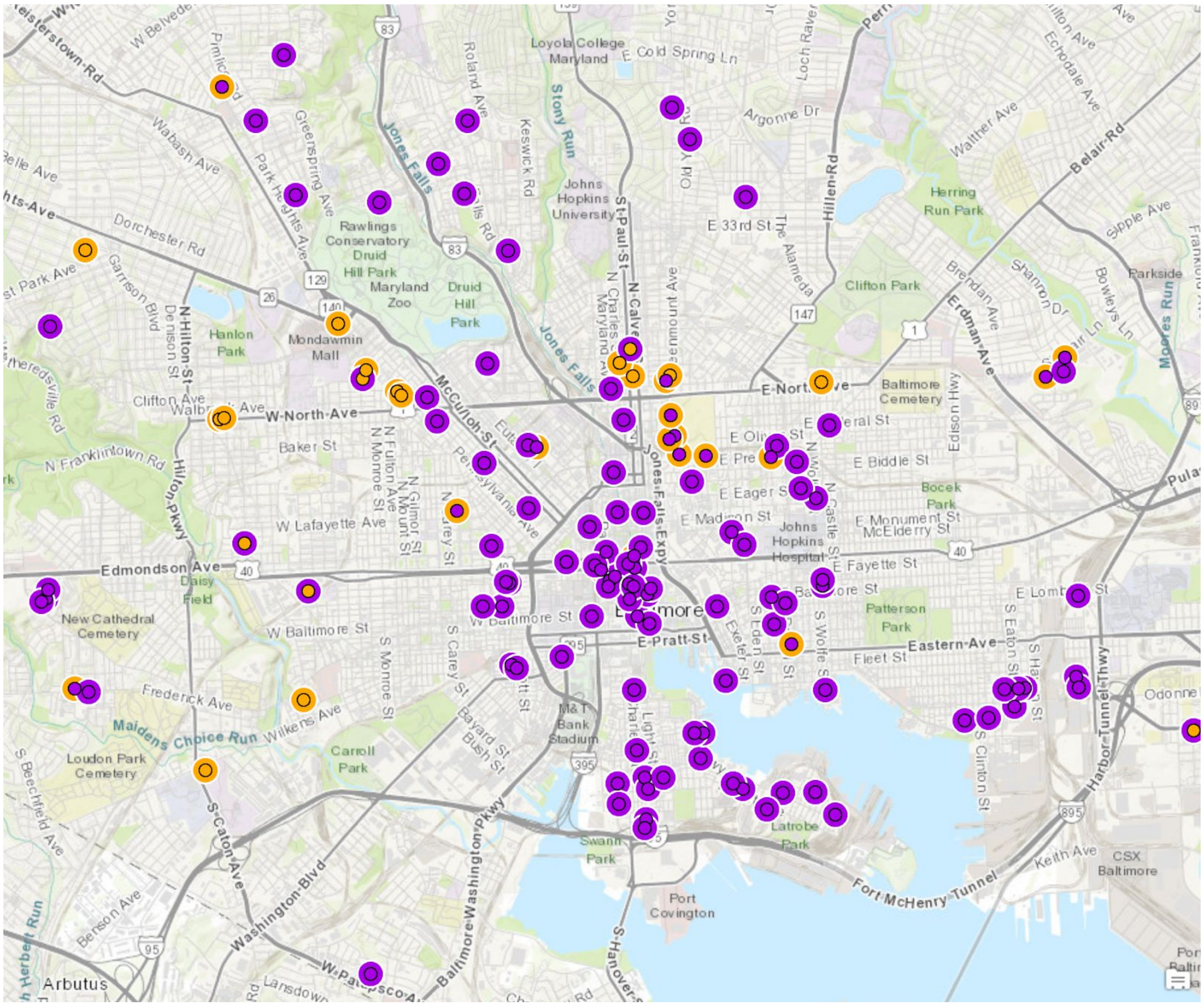


Image 4 *Visual comparison of predicted housing type versus actual housing type.*

Data Output

Overall, the results confirm a distinct disadvantage to living conditions in affordable housing compared to market rate housing. For the ideal site selection, the final location was located amidst a host of market rate housing construction projects in the major downtown area of Baltimore.

1. Efficiency Assessment

Efficiency assessments were conducted for a number of variables comparing the two housing types. For each variable (fire station proximity, hospital proximity, Circulator stop proximity, food desert coverage, and median house sale price) a map layout was generated with a map containing the relevant features along with a chart frame for each housing type. Each layout also has a description of the results in the form of a qualitative comparison of the charts.

All of the variables assessed by their proximity over a network were presented with histograms showing the distribution and mean value of distances/travel times. The layout illustrating the significance of median house sale price across housing types also contained a histogram showing the distribution and mean price). The food desert charts are bar charts that compare the number of housing projects in/out of food deserts with a guide demonstrated the percentage of projects that fall within food deserts.

The results generally support the conclusion that market rate housing enjoys better access to resources compared affordable housing. Market rate housing projects on average are located half the distance to transit stops compared to affordable housing, and average travel times from hospitals to affordable housing projects were over a minute longer compared to market rate housing. The most dramatic difference between the two housing types was the median house sale price, with sale prices in neighborhoods with affordable housing were 64% of the price for neighborhoods containing market rate housing.

In some cases, the differences were fairly negligible like the relative proximity of housing to work/retail. Because work and commercial areas are fairly well dispersed throughout the city it’s not surprising that access is comparable. The issue with these results is that commercial districts are not all created equal, given that some districts suffer from rampant vacancy rates and crime. Relative to fire stations, both housing types were efficiently placed so as to experience similar response times (approximately three and a half minutes). Food desert frequency was surprisingly similar as well, 37% of affordable housing projects were located in food deserts, while market rate housing also experienced a fairly high food desert rate at 29% — an 8% difference.

2. Ideal Site Selection

A small area of Baltimore's downtown area was selected from all of the site selection criteria. This area has many opportunities for employment and entertainment in the immediate vicinity, while also being easily accessible for emergency services, and providing instant access to several Circulator routes. Condos in this same area can sell for as much as $600,000.

The results of this analysis were expected especially considering that I prioritized ready access to the city's free public transit option, which is centered in the downtown area. The Inner Harbor district is also the focal point of urban development in Baltimore City.

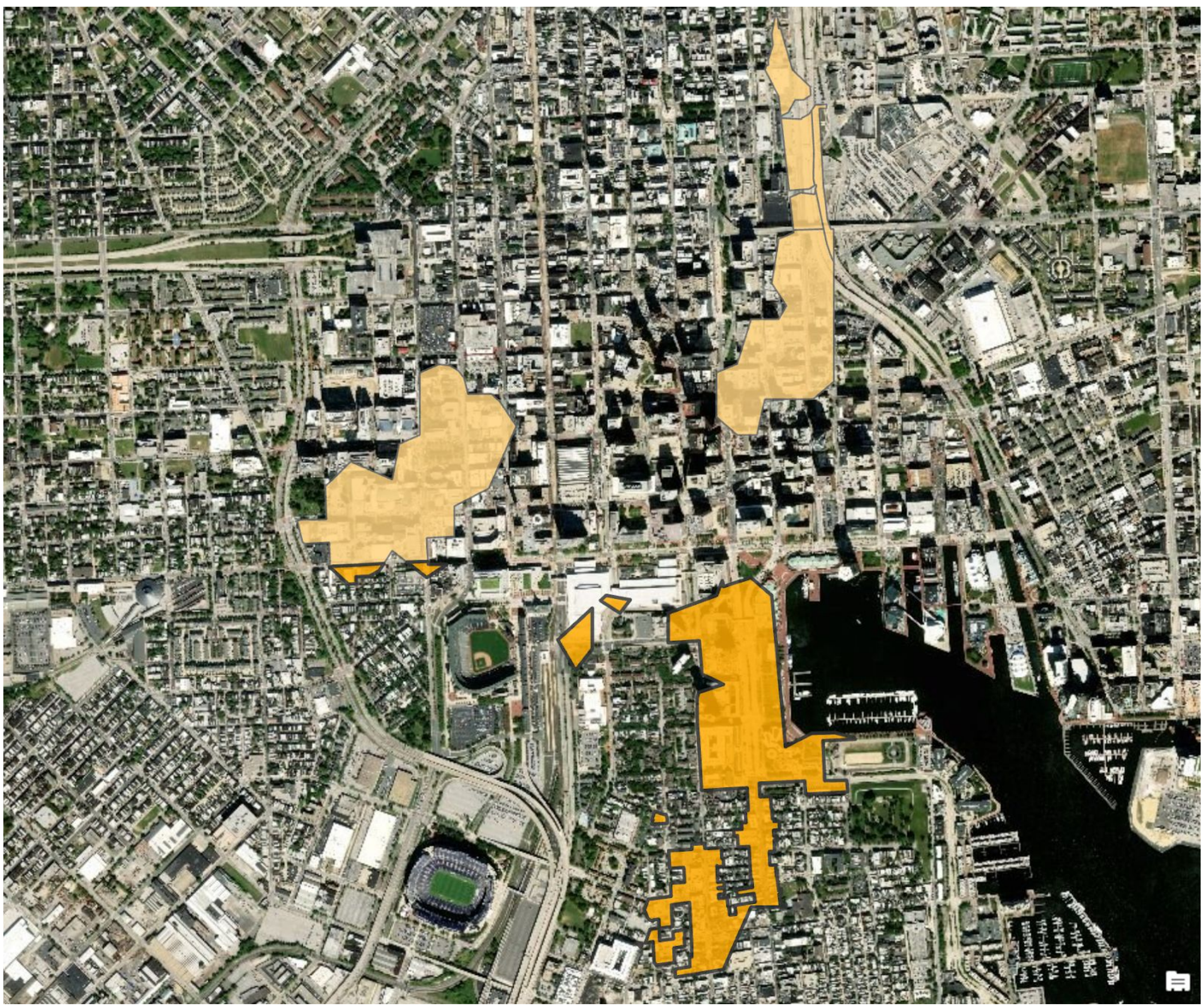


Image 5 *Acceptable site locations for a hypothetical Affordable Housing project. Areas are located in the heart of Baltimore's downtown area with ready access to employment opportunities, groceries, public transit and emergency services.*

3. Regression Analysis

The overall accuracy for all iterations of the Forest-based classification regression analysis were all fairly high (~80% accurate), however there were significant errors of commission with affordable housing being mispredicted as market rate housing (user accuracies ~50%). Market rate housing predictions were much more accurate across all iterations.

The best performing iteration used explanatory variables that (rather unsurprisingly) focused on cost of living metrics: rental and mortgage affordability index scores. The worst performing iteration was the classifier that used housing demand factors as explanatory variables. Still, the accuracy results were still fairly close across iterations ranging between 82 and 84 percent accurate.

All Indicators	Market Rate Housing (Predicted)	Affordable Housing (Predicted)	Total	User's Accuracy (Comission Errors)
Market Rate Housing (Actual)	106	5	111	95%
Affordable Housing (Actual)	18	14	32	43%
Total	124	19	143	
Producer's Accuracy (Omission Errors)	85%	73%		
Overall Accuracy:				
83.9%				
Housing Demand Indicators	Market Rate Housing (Predicted)	Affordable Housing (Predicted)	Total	User's Accuracy (Comission Errors)
Market Rate Housing (Actual)	100	11	111	90%
Affordable Housing (Actual)	15	17	32	53%
Total	115	28	143	
Producer's Accuracy (Omission Errors)	86%	60%		
Overall Accuracy				
81.8%				
Income Disparity Indicators	Market Rate Housing (Predicted)	Affordable Housing (Predicted)	Total	User's Accuracy (Comission Errors)
Market Rate Housing (Actual)	106	5	111	95%
Affordable Housing (Actual)	18	14	32	43%
Total	124	19	143	
Producer's Accuracy (Omission Errors)	85%	73%		
Overall Accuracy				
83.9%				

Image 6 *Confusion Matrices used for accuracy assessment of the supervised forest-based classification. Accuracy assessments were performed for each iteration.*

References:

Cholo, A.B. (2016). The Developer's Story: Why Affordable Housing Doesn't Get Built. In *Capital & Main*. Retrieved from: https://capitalandmain.com/the-developers-story-why-affordable-housing-doesnt-get-built-0225

Faber, J.W. & Ellen I.G. (2016). Race and the Housing Cycle: Differences in Home Equity Trends Among Long-Term Homeowners. *Housing Policy Debate, 26*(3), 456-473. http://dx.doi.org/10.1080/10511482.2015.1128959

Jacobs, J. (1961). *The Death and Life of Great American Cities.* New York: Random House.

Lee, J. & Rios E. (2015). 7 Charts Explaining Baltimore's Economic and Racial Struggles. In *MotherJones.com.* Retrieved from: https://www.motherjones.com/politics/2015/05/baltimore-race-economy-charts/

MacGillis, A. (2016). The Third Rail. In *Places Journal*. Retrieved from: https://placesjournal.org/article/the-third-rail/

Rothstein, R. (2017). *The Color of Law: A forgotten history of how our government segregated America.* New York: Liveright Publishing Corporation.

Scott, J.C. (1998). *Seeing Like A State: How Certain Schemes to Improve the Human Condition Have Failed.* New Haven, CT: Yale University Press.

Valdez, R. (2018). We Don't Need More Affordable Housing, We Need More Housing So It Will Be Affordable. In *Forbes*. Retrieved from: https://www.forbes.com/sites/rogervaldez/2018/07/16/we-dont-need-more-affordable-housing-we-need-more-housing-so-it-will-be-affordable/#6d8c403a15aa

Willeboordse, B. (2017). Economic Inequality: Baltimore, MD. In *Perspectives, Policies & Practices - Spring 2017.* Retrieved from: https://wp.nyu.edu/economicinequality/author/ilw218/

World Population Review. (2018, June 3rd). *Baltimore, Maryland Population 2018.* Retrieved from: http://worldpopulationreview.com/us-cities/baltimore/