2018

Patterns in Road Maintenance: An Analysis of San Diego Roads

Kendall Caitlin Alexandra Mahavier
Oberlin College

Patterns in Road Maintenance: An Analysis of San Diego Roads

Oberlin College
Economics Honors Seminar (2017-18)

Kendall Mahavier
May 15, 2018

I Introduction

In 2011, the city of San Diego, CA, performed a condition survey of all the streets the city maintains. Each street segment was assigned an Overall Condition Index (OCI)[1], which is a number between 0 and 100, and was labeled as Poor, Fair, or Good, based on that OCI number[2]. Only 33.5 percent of the streets were found to be in Good condition, while 43.9 percent and 22.5 percent were in Fair and Poor conditions, respectively[3].

In 2015, the city reassessed the streets. This time, the majority of the streets—59.7 percent—were considered to be in Good condition and only 5.7 percent were in Poor condition, with the remaining 34.6 percent in Fair condition. Figure 1 below shows examples of San Diego streets and their assigned OCI value. Between 2011 and 2015, the city clearly made an effort to improve the condition of its streets. The question is how exactly did San Diego decide what streets to fix? Decisions could have been made based on the traffic levels of the road or based more on the type of traffic rather than the amount. It could have been the road's proximity to key locations, or how many businesses depend on the road. Perhaps the most interesting underlying question to answer, however, is whether the socioeconomic characteristics of an area affected road improvement rates. This paper will examine what impacts road improvement patterns in San Diego, with a particular focus on how the income or racial demographics of an area may affect the area's roads.

Figure 1: Images of San Diego Streets

Pictured, from left to right, are Akins Avenue, Jacumba Street, and Bonswall Street. These streets have OCIs of 100 (Good), 50 (Fair), and 30 (Poor), respectively.
Images drawn from the city of San Diego's 2015-2016 Pavement Condition Assessment.

[1] See Appendix A for an explanation of the OCI rating system.
[2] Good streets have an OCI between 70 and 100, Fair streets have an OCI between 40 and 69, and Poor streets have an OCI between 0 and 39.
[3] Numbers are rounded and therefore do not sum to 100 percent.

This paper examines the period from 2011 to 2015, but it is important to understand the history of road repair in San Diego prior to 2011. For several years prior to 2011, San Diego's roads were suffering. In 2004, road repairs nearly ground to a halt due to political and financial troubles. The situation deteriorated from there, with nearly 66 percent of the roads rated as unacceptable in 2007. Former San Diego City Manager Jack McGrory said that maintenance is often what gets cut when city officials are looking to cut taxes while still expanding other services. In 2009, things finally began to turn around. The city borrowed $100 million to start repairing roads; however, starting the construction was a slow process (Dillion 2011). This period of time from 2011 to 2015 therefore is an interesting time frame to study how San Diego prioritizes road repair, as this period follows shortly after a period of minimal repair.

Previous literature on public service distribution—discussed further in Section 2—has not recently examined the distribution of road maintenance services, and the existing literature also bears marked differences from the approach used in this paper. George Antunes and John Plumlee studied the quality of neighborhood streets in Houston—however their methodology involved mainly comparing the mean roughness of streets in different neighborhoods at one fixed point in time (Antunes and Plumlee 1977). My approach looks at changes over time and also includes a larger sample size. Further explanations of how my paper differs from prior literature can be found in Section 2. Section 3 will describe my data, which includes information on the demographics of various areas in San Diego, but also includes information on road traffic levels, road classifications, the proximity of the road to certain locations (such as schools), the length of the road, and more. Section 3 also notes the sources of the data used, and explains how traffic data was calculated for the road segments. Section 4 explains the empirical methodology of the paper.

Section 5 presents the results. Many of the variables included in the model ended up being significant in predicting road condition and improvement. Median household income ended up having a positive coefficient in all regressions run—roads in rich areas started better in 2011 and were also more likely to be improved between 2011 and 2015 than roads in poorer area. Other findings include: a census block with a high business density had better roads than census blocks with a smaller business density; certain City Council districts have roads that are significantly better or worse than other districts (for example, City Council District 8, which is on the international border, has roads than are better maintained than some other districts); and the racial makeup of an area had little impact on the road quality of that area. These findings, and others, are discussed further in Section 5. Section 5 also includes information on robustness checks.

Section 6 discusses some limitations of this paper, and Section 7 concludes the paper by restating key findings and offering thoughts for future research in this area. Appendix A and B follow the conclusion; Appendix A explains the OCI system and Appendix B contains has the correlation matrix for the independent variables used in the models in this paper.

2 Literature Review

The literature on road quality is relatively sparse, and the literature that does exist does not always provide a comprehensive review of what impacts road maintenance. Additionally, the methodology and data in the existing literature differ in several ways from my own model. Therefore, the following literature is used mainly to help inform about the nature of public service distribution and the conclusions that others have drawn, and to generally provide ideas for my model as to what variables may impact road maintenance, such as income and race.

As mentioned in the Introduction, Antunes and Plumlee studied the quality of neighborhood streets in Houston by comparing the mean roughness of a random selection of streets across a collection of census tracts that were either predominantly black or predominantly white. The authors compared the mean street roughness by race of neighborhood and found that although black neighborhoods did have a higher mean roughness, the difference was statistically insignificant at the .05 level. Similarly, poorer neighborhoods also had a higher mean roughness than richer neighborhoods, but again the difference was statistically insignificant (Antunes and Plumlee 1977). The authors only used one point in time, and used a random selection of streets, whereas I am covering a four-year time span and have a much larger sample size of over 28,000 roads. Additionally, the measure of OCI is a little more complex than just mean roughness.

Another study around the same time period also looked at Houston, but the authors chose to look at the distribution of police services instead of road quality. They used a correlational analysis to measure differences in police response times, and the authors actually found that poorer and minority areas had faster response times (Mladenka and Hill, 1978). Since my paper is not looking at police services, this study is less relevant to my paper but still provides interesting insight into public service distribution and offers a conclusion that differs from what one might expect. Additionally, this paper and the previous one both suggest that looking at how an area's income and racial demographics impact public services is a worthwhile topic to pursue, especially since there does not seem to be a universal answer to how income and race affect public services.

A more recent study looked at changes in bicycle lanes, bus transit service, off-road trails, and parks over a 25-year period. The authors' goal was to look at how different neighborhoods change over time, and how improvements may not be equally distributed across all neighborhoods. Four different U.S. cities were examined—Birmingham, AL; Chicago, IL; Minneapolis, MN; and Oakland, CA—to get variation in terms of size, diversity, and geography (Hirsch et al. 2017). While this study, like mine, looks across time, the authors were able to use a much larger time frame than I can. Additionally, they chose to examine four cities, while I am limiting myself to just San Diego. There are potential pros and cons to both approaches, however I believe that by focusing on one city I will be able to fully understand the within-city dynamics of San Diego and not worry about between-city differences impacting the results.

The distribution of other types of public services has also been examined in various other papers. Another paper that looked at police services focused on Tuscaloosa, AL, and found little evidence of discrimination toward neighborhoods with lower incomes and racial minorities. The author's analysis was based on the idea that it's not enough to compare the level of services different areas have received, but instead look at the amount of services an area should have received. He also argues that it's important to not only look at equity of inputs, but also equity of service outputs (Coulter 1980). By looking at whether or not roads were improved, I am also focusing on service outputs (quality of roads) rather than service inputs (the amount of money spent on each road).

Another study examined park and recreation services over a period of 22 years in Chicago, marking a distinction from some other papers by using a longitudinal design instead of a cross-sectional analysis. The paper concludes that class, as opposed to race, has become a new determinant of the level of public service an area receives (Mladenka 1989). Conversely, another study—that came almost a decade later—did find patterns relating park access to the percentage of non-White residents in a neighborhood. What is interesting about those results is that these patterns in park access were different in the two cities examined: Macon, GA, and Pueblo, CO. Macon's distribution of parks actually seemed to favor non-White, lower income areas, while Pueblo's distribution favored higher income areas with more White residents (Talen 1997). These results may suggest that different cities have different patterns of unequal public service distribution, which is a caution against generalizing my results from San Diego. But the results also suggest that it is worthwhile to study cities individually because of city differences. In this regard, studying San Diego adds new information to the literature since it is not a city that has been studied previously in this regard.

Some other papers have sought less to examine why different characteristics of an area affect the level of public service distribution, but rather how different levels of public services can affect an area. One recent paper looks at how regional disparities in levels of public service distribution in China is related to other regional disparities. In their paper, they argue that equalizing public service levels helps equalize income and consumption across regions (Li et al. 2017). Li et al. examine public service distribution in terms of the entirety of the system, which is very different from examining a small part of the system such as road quality, but nevertheless their results offer a potential argument for why understanding public service distribution is important. Differences in levels of public service distribution can have much larger effects than may be realized, even if that public service distribution is something as seemingly basic as road repair.

These papers show that literature on public services covers a wide range of areas and, at the same time, does not seem to cover what I am trying to do in this paper. Prior papers do point to race and income as interesting variables to look at to predict public services outcomes, and prior papers also indicate that different cities will give different results. Additionally, the literature shows that different types of public services are not distributed in the same way, making the study of roads—a type of public service that the literature is weak on—a valuable addition to the discussion.

3 Data

This section begins by describing the different variables, and also offers rationale for why some of the more unusual variables were included in the model. The second subsection notes some summary statistics for each of my variables. The third subsection describes the sources of my data. The fourth and final subsection describes the process I used to estimate traffic data for all road segments. Additionally, Figure 2 on the next page shows a map of the city of San Diego to provide those who may be unfamiliar with the area an idea of what San Diego looks like.

3.1 Data Descriptions

Some of the variables in my model are perhaps unfamiliar to some people or, at least, the rationale for why I chose to include them in a model of road repair may not be obvious at first. This section will explain these variables and offer some thoughts on what the expected effect of those variables would be on the likelihood of road repair.

Two of the more unusual pieces of data I collected were information on San Diego's maintenance assessment districts and historical districts. Historical districts are areas established by San Diego's Historical Resources board that contain multiple objects and/or properties that have historical significance. Roads in these districts could be more likely to be improved, due to the city's desire to preserve these areas for their historical significance and to keep the areas nice for tourists. Maintenance assessment districts are areas where property owners have voted to pay additional taxes to receive services above and beyond the general services provided to other areas of the city—since these areas get more services than usual, it seems reasonable that they could experience more frequent road repair.

Three other interesting variables are the ones that I will denote as the "nearby variables". These variables were built using data on the location of schools[4], colleges, and tourist attractions, by assigning road blocks a 1 if I deemed them to be "nearby" the specified location. The distances used were 500 meters for a school, 800 meters for a college, and 800 meters for a tourist attraction. These variables were included in the model because San Diego stated in a Pavement Condition Report that they take into account a road's proximity to schools and tourist attractions when making road repair decisions (City of San Diego 2016, *2015-2016 Pavement Condition Assessment*). Colleges were included due to the possibility that they are included in the city's understanding of "schools".

I also included San Diego's nine City Council districts for a few reasons. One reason was to see if politics mattered—perhaps the political party of the councilmembers is related in some way to the likelihood of repair. But beyond simple political motivations, the City

[4] The category of schools does not include colleges. It does include private, public, and charter schools.

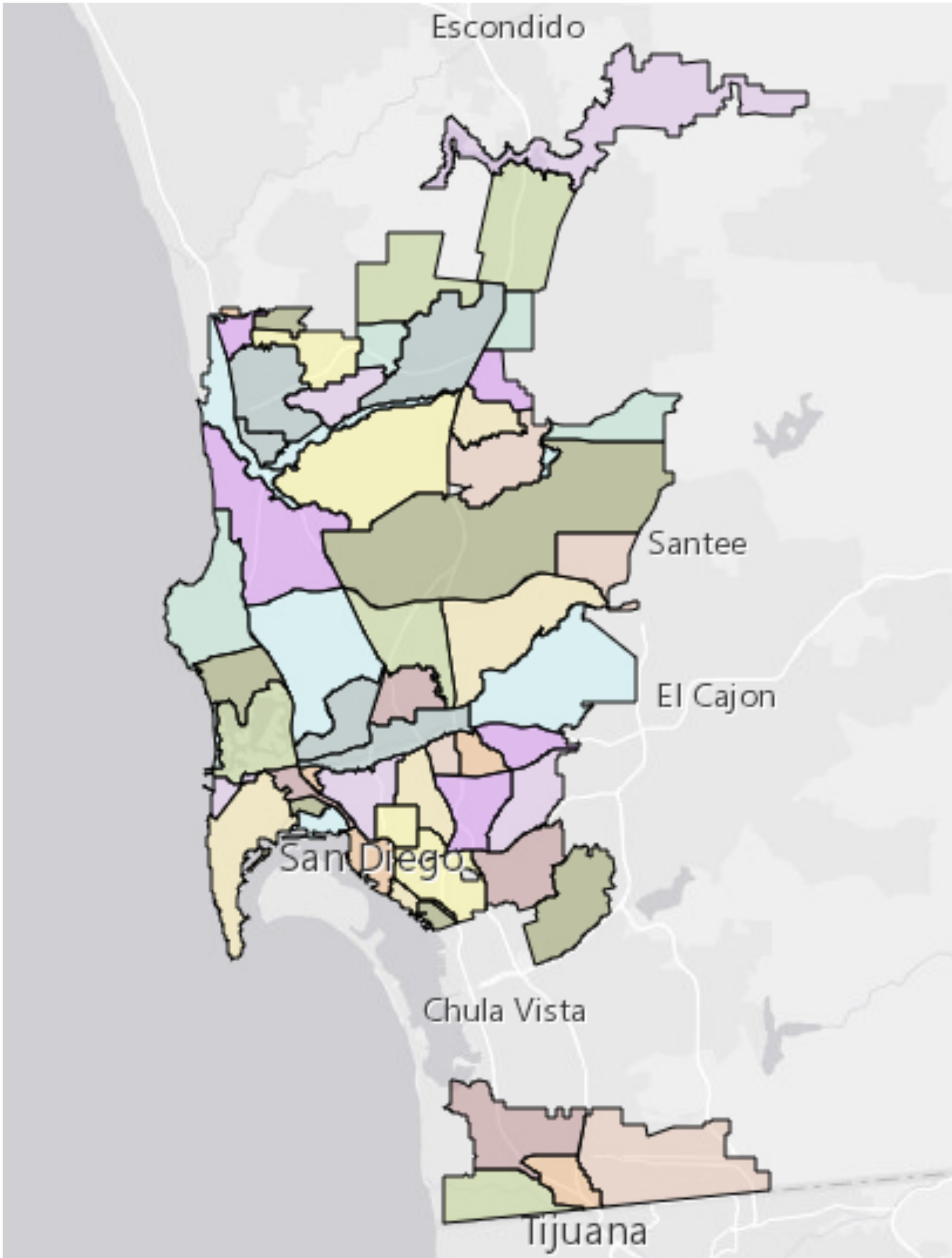

Figure 2: Map of San Diego

Each outlined area represents a community within the city of San Diego—for the purposes of this paper, these individual communities are unimportant. The point of this map is to highlight the outline of the city itself and provide readers with an idea of where San Diego is.
Map taken from the city of San Diego's map gallery.

Council districts also provide a way to divide the city into large sections to help see if the characteristics of a larger area have an impact on the roads. Some areas of the city, for example, are more tourist-heavy than others, or carry a lot of trade from the international border. Those are things that could potentially have an impact on road improvement, and the districts offer a way to sort roads by the characteristics of a larger area.

The rest of the variables in my model are, to an extent, more self-explanatory than the previous ones. I collected data on traffic levels, road type, recent construction history, and the condition in 2011, which are all things that would likely have an effect on the likelihood of repair between 2011 and 2015. Higher traffic means improving the road could benefit more people; recent construction means the road was better to start off with in 2011; and a poor condition in 2011 will likely increase the likelihood of road repair between 2011 and 2015.

The five road types included are local, residential, collector, major, and prime, which are likely unfamiliar classifications to many, especially since road classification can vary somewhat by area. Local roads primarily provide direct access to abutting property and carry low levels of traffic. Collector roads primarily provide movement between local or collector streets and streets of higher classifications and carry low to medium levels of traffic. Prime roads provide a network connecting vehicles to other primary roads and to freeways, and these roads carry heavy traffic levels. Major roads connect vehicles to major roads, primary roads, and to freeways, while also providing secondary access to abutting property. Major roads have moderate to heavy traffic. Finally, residential roads provide access to residential lots and carry low to moderate traffic. These distinctions of road types could feasibly mean that some roads are more important to maintain than others, having an effect on road improvement patterns. Local roads are the road type left out of the model (due to perfect collinearity), and so therefore the results of the model show the difference in other road types compared to local roads. Based on the description of road type, you might expect to see that all other four road types are better off than local roads, which are the roads with the smallest amount of traffic.

Data was collected on racial demographics, the population levels, and the median household income because prior literature suggests that these factors could be important. As discussed in Section 2, the results vary from paper to paper, but the fact that past papers have found significant results for the effect of income and race suggests that such variables are important to include in my model. And regardless of the results that others have found, I believe it is worthwhile to continue to study these variables to understand the biases in our government system.

Whether a road had a designated bike route was also included in the model; it is reasonable to think that roads with heavy bike traffic might need more frequent repair, since a small crack that a car would ignore could be more dangerous to a biker. Another piece of information included in the model was the business density of the census block, which served to provide information on the economy of the area, and the share of units in the census block in 2011 that were vacant, which helps give a picture of how abandoned or dilapidated an area might be.

3.2 Summary Statistics

Summary statistics for all variables included in the final specifications, as well as the related variables excluded because of perfect collinearity, are in Table 1 and Table 2. Table 1 has all the dummy variables, and notes what it means for the variable to be a 1 as well as the average value. Table 2 is all the continuous variables, and denotes the shorthand name for the variable, as well as the mean, the standard deviation, the minimum value, and the maximum value for the variable. Table 1 is below, while Table 2 follows on the next page.

Variable Name	Variable Description	Average
Improved	1 if the OCI improved by 5 points between 2011 and 2015	.612
PriorConstr	1 if the road had construction from 2008 to 2011	.254
Residential	1 if the road's functional class is Residential	.682
Prime	1 if the road's functional class is Prime	.064
Collector	1 if the road's functional class is Collector	.119
Local	1 if the road's functional class is Local	.03
Major	1 if the road's function class is Major	.105
District1	1 if in City Council District 1	.207
District2	1 if in City Council District 2	.117
District3	1 if in City Council District 3	.132
District4	1 if in City Council District 4	.113
District5	1 if in City Council District 5	.105
District6	1 if in City Council District 6	.072
District7	1 if in City Council District 7	.105
District8	1 if in City Council District 8	.089
District9	1 if in City Council District 9	.059
NearCollege	1 if near a college	.126
NearSchool	1 if near a school (not a college)	.552
NearAttraction	1 if near a tourist attraction	.09
BikeRoute	1 if road has a designated bike route	.74
MaintenanceD	1 if the road is in a maintenance assessment district	.361
HistoricalD	1 if the road is in a historical district	.024
LowTraffic	1 if the road had estimated daily traffic below 5,000	.237
MedTraffic	1 if the road had estimated daily traffic between 5,000 and 10,000	.625
HighTraffic	1 if the road had estimated daily traffic above 10,000	.138

Table 1: Dummy Variable Descriptions

Variable Name	Average	Standard Deviation	Minimum	Maximum
OCI2011	56.773	22.504	.5	100
OCI2015	70.198	18.310	0	100
OCIchange	13.425	21.7618	–93.63	96.82
Length	500.318	495.382	28.187	18791.15
Income	75.637	32.729	21.792	180.833
TotalPop	247.738	370.242	0	7910
PercWhite	50.161	32.135	0	100
PercHispanic	21.556	23.950	0	100
PercBlack	4.848	9.576	0	100
PercAsian	12.993	16.054	0	100
PercOther	10.443	4.437	0	100
OccShare	0.95	0.076	0	1
VacShare	0.046	0.074	0	1
Business Density	0.009	0.065	0	4.491

Table 2: Continuous Variable Descriptions

These tables include all the independent variables used in my models, as well as the dependent variables. The dependent variables used in my different models are *Improved*, *OCIchange*, and *OCI2011*. The meaning and purpose of the different dependent variables will be explained more thoroughly in Section 4. *OCI2015* was a variable that was used to help create the *OCIchange* and *Improved* dependent variables; otherwise, all other variables listed in the tables are independent variables.

3.3 Data Sources

I used a few data sources to collect all the data in my analysis. Much of the data comes from the San Diego city government and other local government sources. San Diego has data online of the overall road conditions by block for 2011 and 2015 (City of San Diego 2016, "Streets Overall Condition Index"). I have also drawn other data about the classification of the road, the length of the road segment, the date the road segment was added, the census block and tract, and the driving speed based on the segment class from the city of San Diego (City of San Diego 2016, "Road Lines"). The overall condition data also included some construction history information from 2008 onwards (City of San Diego 2016, "Streets Overall Condition Index"), and there was also data on the presence of bike routes on the streets (City of San Diego 2016, "Bike Route Lines").

Data on the location of schools, colleges, and tourist attractions in San Diego came from the San Diego Association of Governments' (SANDAG) Regional GIS Data Warehouse (San Diego Association of Governments 2012). Data on the location of businesses in San Diego was also drawn from that same SANDAG collection of GIS maps (SANDAG 2012). Racial demographics data was drawn from a different SANDAG webpage and was available for every census block (SANDAG 2010). Data on the median household income of an area came from the American Community Survey's 2007-2011 5-Year Estimates (United States Census Bureau 2011).

Other data on the area in which a road exists came from the San Diego city government, such as data on historical districts (City of San Diego 2017, "Historical Districts") and data on maintenance assessment districts (City of San Diego 2017, "Maintenance Assessment Districts"); both districts will be described in the following section. Additionally, data was also collected on San Diego's nine City Council districts, which were most recently redrawn in 2011 (City of San Diego 2016, "City Council Districts"). Finally, data for San Diego's zoning designations was also taken from the city of San Diego's website (City of San Diego 2016, "Zoning"). The zoning data, however, was not used in the final model.

The last piece of data collected was data on the daily traffic volume on roads. This data, however, was not available for all roads, and the data that did exist spanned several years. San Diego reportedly collects data for specific blocks because the blocks are believed to be representative of a longer segment, and the traffic counts are taken depending on the needs of local government staffers. San Diego notes that the reported average daily traffic counts are likely to be good representations of the number of vehicles on the road on an average weekday. This data was drawn from the city of San Diego's data website, like much of the other road data (City of San Diego 2016, "Traffic Volumes"). This data covers several thousand blocks, and so I used this data to extrapolate values for the rest of the road blocks. More detailed information on that process can be found in the next section, Section 3.4.

There are, in total, 28,263 road blocks[5] for which I developed complete data. Throughout the data merging process[6], a few hundred road blocks were lost through lack of complete data.

3.4 Traffic Data

The data on daily traffic volumes was only available for a subset of the data—less than 10,000 road blocks. In order to keep the full set of 28,263 road blocks in my model, I ran a regression using the traffic data I did have as the dependent variable, so I could then try to predict the traffic values for all the other road blocks. This regression uses different variables that are not specified elsewhere.

The variables used in this regression include an average driving speed for the road that is determined by segment classification, as well as a dummy variable for a road that was one-way. There were also several dummy variables for road classification that are more

[5] A "road block" simply means a block of a road: a stretch of road between two intersections. OCI values are assigned per block, and therefore the unit of analysis is a road block.

[6] Data was combined using both ArcGIS and Microsoft Excel, depending on the type of file the data was originally found in.

specific than the road types used in the main regression. These road classifications included subtypes based on the number of lanes on the road, as well as other more specific and varied road types. The final variable included was the width of the right-of-way for the road block.

After running this regression, I had Stata predict traffic values for all 28,263 road blocks using that regression. I then created a new variable that used the actual traffic values for the road blocks the data was available for, and then used the estimated values for the rest of the road blocks. After creating this variable, I then used it to create three dummy variables to indicate whether the road had low, medium, or high traffic levels. I defined low traffic volume as below 5,000 vehicles per day; medium was between 5,000 and 10,000 vehicles per day; and high was above 10,000. This created the variables *LowTraffic*, *MedTraffic*, and *HighTraffic*, which took on a 1 if the daily traffic was judged to be low, medium, or high, respectively. *MedTraffic* and *HighTraffic* were used in all regressions, with *LowTraffic* omitted due to perfect collinearity.

4 Methodology

My methodology included three main regressions, as well as an additional regression to develop estimates of the traffic levels on the road blocks that I did not have actual traffic data for, and this section will cover all of these regressions. Section 4.1 explains a regression that uses a dummy variable—*Improved*—as the dependent variable. Section 4.2 explains the next regression, which uses a continuous variable—*OCIchange*—as the dependent variable. Section 4.3 discusses the last of my main three regressions, which uses *OCI2011* as the dependent variable to explain what the roads looked like at the start of the time period I'm examining.

All regressions were run using an OLS regression with robust standard errors. Probit and logit models were also considered, however the results did not drastically change, and so OLS was chosen for simplicity and ease of interpretation. Robust standard errors were used to help minimize heteroskedasticity problems.

4.1 *Improved* Regression

This regression has 31 independent variables, and the dependent variable is the dummy variable *Improved*. *Improved* is a 1 if the road block's OCI improved by at least 5 points between 2011 and 2015, and 0 if not[7]. This regression serves to show what independent variables could have had a significant impact on whether or not a road block was repaired during this four-year period, and, since *Improved* is a simple dummy variable, this regression treats all improvement the same. A change from an OCI of 50 to an OCI of 55 is treated the same as a change from an OCI of 50 to an OCI of 90.

The following equation shows specifically what variables were used in this regression.

[7] A positive change of 5 points was used to account for any measurement error in the OCI and ensure that the road was, in fact, improved over this time period.

$$
\begin{aligned}
Improved = \; & \beta_0 + \beta_1(OCI2011) + \beta_2(PriorConstr) + \beta_3(Residential) + \beta_4(Prime) + \beta_5(Collector) \\
& + \beta_6(Major) + \beta_7(MHI) + \beta_8(TotalPop) + \beta_9(PercHispanic) + \beta_{10}(PercBlack) \\
& + \beta_{11}(PercAsian) + \beta_{12}(PercOther) + \beta_{13}(VacShare) + \beta_{14}(Length) \\
& + \beta_{15}(District1) + \beta_{16}(District2) + \beta_{17}(District3) + \beta_{18}(District4) + \beta_{19}(District5) \\
& + \beta_{20}(District6) + \beta_{21}(District7) + \beta_{22}(District8) + \beta_{23}(NearCollege) \\
& + \beta_{24}(NearSchool) + \beta_{25}(NearAttraction) + \beta_{26}(BikeRoute) + \beta_{27}(MaintenanceD) \\
& + \beta_{28}(HistoricalD) + \beta_{29}(BusinessDensity) + \beta_{30}(MedTraffic) + \beta_{31}(HighTraffic)
\end{aligned}
$$

The results from this regression can be found in Section 5.1.

4.2 *OCIchange* Regression

This regression, like the previous one, has 31 independent variables. In fact, this regression is identical to the *Improved* regression except for the fact that the dependent variable is different. Here, the dependent variable is the continuous variable *OCIchange*, which represents the change in the OCI value of a road block from 2011 to 2015. A positive value for *OCIchange* means that the road block's OCI got better, while a negative value means the condition got worse.

This regression, unlike the previous one, treats different levels of improvement differently. A change from an OCI of 50 to an OCI of 55 will be treated differently than a change from an OCI of 20 to an OCI of 90. So, the previous regression could show that a certain type of road was more likely to be improved, but this regression could then reveal that the amount the roads were improved by was a very small amount. By having a continuous variable as the dependent variable instead of a dummy variable, this regression also offers more variation in the dependent variable.

The equation below depicts this regression, although again, it is identical to the previous regression in terms of its independent variables.

$$
\begin{aligned}
OCIchange = \; & \beta_0 + \beta_1(OCI2011) + \beta_2(PriorConstr) + \beta_3(Residential) + \beta_4(Prime) + \beta_5(Collector) \\
& + \beta_6(Major) + \beta_7(MHI) + \beta_8(TotalPop) + \beta_9(PercHispanic) + \beta_{10}(PercBlack) \\
& + \beta_{11}(PercAsian) + \beta_{12}(PercOther) + \beta_{13}(VacShare) + \beta_{14}(Length) \\
& + \beta_{15}(District1) + \beta_{16}(District2) + \beta_{17}(District3) + \beta_{18}(District4) + \beta_{19}(District5) \\
& + \beta_{20}(District6) + \beta_{21}(District7) + \beta_{22}(District8) + \beta_{23}(NearCollege) \\
& + \beta_{24}(NearSchool) + \beta_{25}(NearAttraction) + \beta_{26}(BikeRoute) + \beta_{27}(MaintenanceD) \\
& + \beta_{28}(HistoricalD) + \beta_{29}(BusinessDensity) + \beta_{30}(MedTraffic) + \beta_{31}(HighTraffic)
\end{aligned}
$$

The results from this regression can be found in Section 5.1.

4.3 *OCI2011* Regression

This regression is slightly different from the previous two regressions. There are 30 independent variables, and the dependent variable is the continuous variable *OCI2011*, which means the OCI of the road block in 2011, the beginning of the time period of interest.

This regression serves to develop a baseline of what the condition of roads was like in 2011. If roads in a certain City Council district were significantly worse off in 2011, for example, that could explain why they were significantly more likely to be improved prior to 2015. The regression is very similar to the other two, with the only difference being that *OCI2011* has been moved from the right-side of the equation to the left-side.

The equation below shows this regression.

$$
\begin{aligned}
OCI2011 = \; & \beta_0 + \beta_1(PriorConstr) + \beta_2(Residential) + \beta_3(Prime) + \beta_4(Collector) + \beta_5(Major) \\
& + \beta_6(MHI) + \beta_7(TotalPop) + \beta_8(PercHispanic) + \beta_9(PercBlack) + \beta_{10}(PercAsian) \\
& + \beta_{11}(PercOther) + \beta_{12}(VacShare) + \beta_{13}(Length) + \beta_{14}(District1) + \beta_{15}(District2) \\
& + \beta_{16}(District3) + \beta_{17}(District4) + \beta_{18}(District5) + \beta_{19}(District6) + \beta_{20}(District7) \\
& + \beta_{21}(District8) + \beta_{22}(NearCollege) + \beta_{23}(NearSchool) + \beta_{24}(NearAttraction) \\
& + \beta_{25}(BikeRoute) + \beta_{26}(MaintenanceD) + \beta_{27}(HistoricalD) \\
& + \beta_{28}(BusinessDensity) + \beta_{29}(MedTraffic) + \beta_{30}(HighTraffic)
\end{aligned}
$$

The results from this regression can be found in Section 5.1.

5 Results

In this section, I begin in Section 5.1 by discussing the various regression results and offering explanations for the significance of key variables. Section 5.2 details some robustness checks and explains why certain variables were not included in the regressions.

5.1 Regression Results

Tables 3–6 show the results from all three regressions. The tables have three separate columns for the three specifications, with the specific dependent variable of the regression at the top of the column. Coefficients are reported with robust standard errors in parentheses below. The R^2 values and total number of observations are reported at the bottom of each table. The reason that there are four tables for the results is simply because of the large number of variables; the results are spread across four tables to aid in organization and readability. All four tables, Tables 3–6, include results from the same three regressions, and the tables can be found below and on the next three pages.

It is also important to note that the three dependent variables have different ranges, and therefore the sizes of the coefficients in each regression will differ greatly. *Improved* is a dummy variable, and therefore all the coefficients in that regression will look rather small. Meanwhile, *OCIchange* runs from -93.63 to 96.82, so coefficients in this regression will be a lot a larger than the *Improved* regression. Similarly, *OCI2011* has values from 0 to 100, so again, those coefficients will be larger than the *Improved* ones.

Of the 31 independent variables in the *Improved* and *OCIchange* regressions, 19 variables were significant in the *Improved* regression and 21 variables were significant in the *OCIchange* regression. The *OCI2011* regression had 21 significant variables out of its 30 independent variables. Several of the significant variables were significant across two or all

Independent Variable	Dependent Variable		
	Improved	**Change in OCI**	**OCI 2011**
Income	0.0008801***	0.0569873***	0.0938464***
	(0.0001115)	(0.0041552)	(0.0053849)
TotalPop	0.0000231***	0.0008004***	0.002126***
	(0.000007)	(0.0002555)	(0.0003198)
PercHispanic	–0.0000332	0.0076232	–0.0102614
	(0.0001518)	(0.0061246)	(0.0079677)
PercBlack	–0.000216	–0.0159343	0.0589619***
	(0.0003006)	(0.0119287)	(0.0166814)
PercAsian	0.0007325***	0.0231509***	0.0510303***
	(0.0002001)	(0.0075012)	(0.0093799)
PercOther	–0.000518	–0.0134989	–0.0355837
	(0.0005696)	(0.0245923)	(0.0286564)
VacShare	0.0358538	3.450907**	–12.46056***
	(0.0330319)	(1.476857)	(1.723899)
Observations	28,263	28,263	28,263
R^2	0.2994	0.4726	0.1931

* p<0.1, ** p<0.05, *** p<0.01

Table 3: Regression Results Part 1

The rest of the results can be seen in Tables 4–6

three regressions, although there were some variables that were also only significant in one regression. Perhaps the most interesting finding was that an area's median household income seemed to reasonably predict the condition of the area's roads.

Median household income was positive and significant for all three regressions, which would suggest that roads in census tracts with a higher median income were better off in 2011, and also more likely to be improved and to have their OCI improve across the time-period of interest. The coefficients were small across all three regressions, but even though the income variable was already in thousands of dollars, the small coefficients quickly become large when you consider the differences between neighborhoods. The minimum median household income in one of San Diego's census tracts was approximately $22,000, while the highest was approximately $181,000. So, going from the poorest area to the richest has a difference of over $150,000. When you consider the coefficient in the *Improved* regression, which was 0.0008801, and multiply it by 150, the resulting value is 0.132015. The roads in the richest area were around 13 percent more likely to be improved than roads in the poorest district. Additionally, since the coefficient on median household income was

Independent Variable	Dependent Variable		
	Improved	Change in OCI	OCI 2011
OCI2011	–0.0124088***	–0.7089901***	---
	(0.0001035)	(0.0053758)	---
PriorConstr	0.0919731***	8.283216***	18.31113***
	(0.0058652)	(0.192451)	(0.2766369)
Residential	0.0598833***	4.060557***	–3.949534***
	(0.0147517)	(0.5191619)	(0.7118664)
Prime	0.0579502***	4.673323***	–3.33463***
	(0.0174449)	(0.6343452)	(0.8467395)
Collector	0.085309***	6.304961***	0.4799
	(0.0164677)	(0.5947782)	(0.8130964)
Major	0.0936969***	6.57914***	–1.561107*
	(0.016772)	(0.5976718)	(0.8342085)
Length	0.0000419***	0.0011374***	–0.0000524
	(0.000005)	(0.0001769)	(0.0002513)
MedTraffic	–0.0034595	–0.3176418	0.6495091**
	(0.0061095)	(0.2462379)	(0.3155427)
HighTraffic	0.0337868***	0.6105532*	–0.9115542**
	(0.0088846)	(0.351559)	(0.4615945)
Observations	28,263	28,263	28,263
R^2	0.2994	0.4726	0.1931

* p<0.1, ** p<0.05, *** p<0.01

Table 4: Regression Results Part 2

The rest of the results can be seen in Tables 3, 5, and 6

0.0938464 in 2011, that means roads in the richest area already had conditions that were around 14 points better than the poorest area. When you put the coefficients into this perspective, it suddenly seems like a much bigger difference. But it is important to note that areas that are much closer in income will have far less of a difference in their roads, and that it really only is when you're comparing the poorest to the richest areas that the coefficients are large enough to suggest any actual economic significance.

The fact that roads in richer areas were better off to start with in 2011 should suggest that they would have been less likely to be improved between 2011 and 2015, but the opposite result was found. These findings could suggest that San Diego is in some way biased to fix roads in richer areas before roads in poorer areas, perhaps because people in

Independent Variable	**Dependent Variable**		
	Improved	**Change in OCI**	**OCI 2011**
District1	0.014334	–1.473641***	0.1467719
	(0.0133971)	(0.5666833)	(0.7016796)
District2	0.0401907***	–1.473641	0.7417257
	(0.0135909)	(0.5739399)	(0.7232283)
District3	0.0122369	–0.3047714	–2.065468***
	(0.0125649)	(0.5462658)	(0.6926355)
District4	0.0120581	–0.7137776	1.305014*
	(0.0134044)	(0.5736745)	(0.7037322)
District5	–0.040851***	–3.861455***	3.478923***
	(0.0151991)	(0.6044713)	(0.7598309)
District6	–0.0505944***	–6.182216***	0.239679
	(0.0155922)	(0.6261857)	(0.7557537)
District7	0.0372323***	–1.228113**	–2.347684***
	(0.0131551)	(0.5655311)	(0.7056344)
District8	0.042637***	0.7455259	7.621802***
	(0.0141148)	(0.5752164)	(0.7133974)
MaintenanceD	0.0288407***	1.890721***	3.004481***
	(0.0064822)	(0.2572105)	(0.3416796)
HistoricalD	–0.0222773	0.741441	–1.039414
	(0.0146775)	(0.658264)	(0.9514075)
Observations	28,263	28,263	28,263
R^2	0.2994	0.4726	0.1931

* p<0.1, ** p<0.05, *** p<0.01

Table 5: Regression Results Part 3

The rest of the results can be found in Tables 3, 4, and 6

richer areas are more likely to call and complain (the "squeaky wheel gets the grease" theory). If it is due to that—meaning the idea that the roads the city hears the most about will be the ones that are fixed—that still suggests some bias on San Diego's part, as roads need repair regardless of how many people call to complain about the roads. This paper will be unable to fully explain why this relationship exists, but the fact that my analysis has shown that the relationship does exist is an important finding, and could have important policy implications if cities want to prevent such a relationship from occurring.

Independent Variable	Dependent Variable		
	Improved	Change in OCI	OCI 2011
NearCollege	0.0036176	1.068234***	–0.6717828*
	(0.0077466)	(0.3179186)	(0.4018866)
NearSchool	–0.0124226**	–0.7084443***	0.1759109
	(0.0051903)	(0.199301)	(0.257837)
NearAttraction	0.006485	0.1851617	–1.400705***
	(0.0094263)	(0.3868022)	(0.4916989)
BikeRoute	–0.0139573**	–0.5924341**	0.8229106**
	(0.0061466)	(0.2459315)	(0.3204079)
BusinessDensity	0.0148517	2.491381**	4.221451**
	(0.038298)	(1.025278)	(1.773396)
Observations	28,263	28,263	28,263
R^2	0.2994	0.4726	0.1931

* p<0.1, ** p<0.05, *** p<0.01

Table 6: Regression Results Part 4

The rest of the results can be found in Tables 3–5

Of the population and racial demographics variables, only the total population and the percentage of Asians in the census block are significant across all three regressions. Both the population and percentage of Asians have positive coefficients across the three regressions. More people in an area means better roads to start with and a greater likelihood of improvement, which makes sense. If a city is trying to allocate its resources effectively, it makes sense to fix the roads that will be used by more people (meaning the roads in areas with a higher population). The significance of the percentage of Asians is a little odd, since one might not expect that more Asians in an area would lead to an increase in road quality. The percentage of Asians is positively correlated with the median household income at .218, so perhaps that relationship is what is causing *PercAsian* to be positive and significant. This could be further evidence that richer areas have better roads. One final point of interest in the racial variables is that the percentage of Black people in a census block is significant in the *OCI2011* regression (coefficient of 0.05896). In 2011, roads in areas with a larger black population were more likely to better off, however being in a black-heavy area had no significant impact on the rate of improvement between 2011 and 2015. Therefore, I think that the only really important racial variable in these regressions is the *PercAsian* variable, which is possibly only important due to its relationship to median household income.

While race and income are very interesting variables to study, they were hardly the only variables of interest in this regression. Maintenance assessment districts, in fact, also had interesting results that could have implications for understanding the positive coefficients on income as well. These districts, as a reminder, are districts where property owners have voted to pay extra taxes to receive extra services. It seems reasonable to suspect that this type of district will likely receive extra attention from the city in all matters and therefore have better roads. The results confirm this hypothesis, with positive coefficients for this variable across all three regressions. These results, in combination with the income results, could also suggest that there is a larger relationship between taxes and public services that goes beyond just maintenance assessment districts. The reason that richer areas get more frequent road maintenance could be because richer residents pay more in taxes than poorer residents—property taxes can make up a lot of a city's revenue, and richer people are generally going to have more expensive houses than poor people. These results therefore could support the argument that, in San Diego, higher taxes lead to better roads, if you take the significant coefficients on maintenance assessment districts as being indicative of a larger relationship between taxes and roads, which is also arguably reflected in the significance of median household income.

Other variables in the regressions, however, are perhaps much less politically interesting but still important to understand the distribution of road maintenance. The *OCI2011* variable is only found as an independent variable in the *Improved* and *OCIchange* regressions, of course, because it is the dependent variable in the final regression. In the two regressions where it is an independent variable, however, it is highly significant and negative. It has coefficients of -0.01241 and -0.70899 in the *Improved* and *OCIchange* regressions, respectively. What those negative coefficients signify is that the better the OCI was for a road block in 2011, the less likely the road would be improved and the more likely the road's OCI would fall between 2011 and 2015. This makes logical sense, since roads that were worse off should be more likely to be repaired than roads that were better off.

The *PriorConstr* variable, which is a dummy variable that notes whether there was construction on the road since 2008, is significant across all three regressions. Understandably, having recent construction improved the OCI in 2011, as the variable is positively significant at 18.311 in that regression. Strangely enough, however, the variable also had significant positive coefficients in the *Improved* and *OCIchange* regressions (0.09197 and 8.2832, respectively). This means that although roads with recent construction were better off in 2011 than other roads, these roads were still more likely to be improved and to see their OCI value rise from 2011 to 2015. Perhaps these roads are ones that just generally fairly important for the city to maintain, or it could also be the case that the recent improvements they had that improved their OCI prior to 2011 were short-term fixes for larger issues that got fixed between 2011 and 2015, once the city had more money to spend on road repair.

The four road type variables included in the model—with local roads being the type left out, due to perfect collinearity—were all significant and positive across the *Improved* and *OCIchange* regressions, while only residential, prime, and major roads were significant in the baseline *OCI2011* regression. Residential, prime, and major roads all had negative coefficients in the *OCI2011* regression, so it could be the case that since these roads were bad to start off with, that's why the roads saw positive improvement between 2011 and 2015. What is interesting though is the differences in coefficients for the road types. Major roads had the largest coefficients in the *Improved* and *OCIchange* regressions, followed by collector roads, and then prime and residential roads[8]. This suggests that there is an important distinction between road types in terms of what roads are more important to the city.

The share of vacant units in a census block was significant for the *OCIchange* and the *OCI2011* regressions with coefficients of 3.45 and -12.46, respectively. So, the areas with a large share of vacant units were worse to start off with, but they were significantly more likely to experience a positive change in their OCI from 2011 to 2015. This makes the variable rather uninteresting, as it suggests whatever caused areas with lots of vacant units to have worse roads initially was a fluke, given that the streets in these areas experienced positive improvement over the studied time period.

The length of a road had a positive impact on the likelihood of improvement and the overall change in the OCI value across these four years of interest. The coefficients on the length were very small, at 0.000042 for the *Improved* regression and 0.00113 for the *OCIchange* regression, but the units on the road length variable were in meters, meaning the small coefficients are to be expected. Length is somewhat positively correlated with major roads and with high traffic levels, meaning the positive coefficients on length could be due to a relationship between the length of a road and the amount and type of traffic on the road.

Of the eight City Council districts included in the three regressions, only District 5 and District 7 had significant coefficients across all three regressions, while Districts 6 and 8 were significant in two of the three regressions. Districts 1, 2, 3, and 4, however, were each significant in only one regression, and those districts, since they seem to have minimal impact on road repair, will not be discussed further. The districts can be seen on the next page in Figure 3.

District 5, which was significant across all three regressions, has a simple explanation for its significance—the coefficient for the district in *OCI2011* was positive, and the coefficients for the district in *Improved* and *OCIchange* were both negative. This suggests that, for some reason, the roads in district 5 were better to start off with in 2011, and so therefore the roads were less likely to be improved and were more likely to have their OCIs fall across this four-year period. Given that the roads started off well, that makes logical sense. It is curious that the roads started well off in 2011, but that was likely due to some random chance. It does not seem that San Diego is favoring these roads, since they were more likely to not be improved.

[8] The coefficients on prime and residential roads could never be judged as significantly different from each other, with F-values of 0.99, 2.24, and 1.51 across the *Improved, OCIchange*, and *OCI2011* regressions, respectively. The probabilities of seeing higher F-values were 0.3190, 0.1342, and 0.2196, respectively.

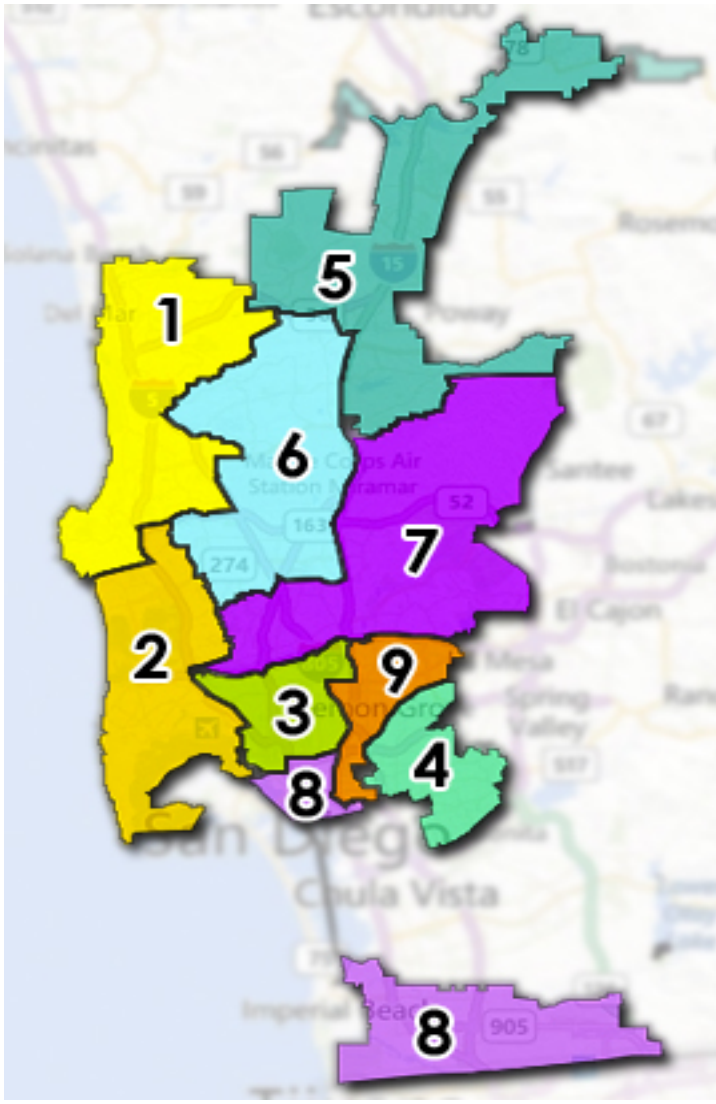

Figure 3: San Diego City Council Districts
Map taken from the City of San Diego's "City Council Offices" webpage

District 6 was significant in the *Improved* and *OCIchange* regressions with coefficients of -0.044 and -6.186, respectively. The negative coefficients mean that road blocks in District 6 were less likely to be improved and more likely to have a fall in their OCI value. The coefficient on District 6 was positive in the *OCI2011* regression, but not significant, so there is no evidence to support the claim that the roads were initially better off. It is the case that District 6 has a high percentage of Asians, as the two variables are correlated at 0.4264. It could be that District 6's negative coefficients are due to the positive coefficients on *PercAsian*. If what makes *PercAsian* positive is the relationship between being Asian and having a higher income, and if there are a lot of Asians in District 6 but the District is not particularly wealthier than other districts, then perhaps the negative coefficients on the District 6 variable are meant to balance out the positive coefficients on the Asian variable. Or, of course, there could be some other entirely unknown characteristic of District 6 that is causing the negative coefficients.

District 7 had significant negative coefficients in the *OCIchange* and *OCI2011* regressions, which would suggest that the roads in this district started off worse and tended to experience falls in their OCI values (although these roads were not less likely to be improved). However, it is also the case that District 7 had a significant positive coefficient in the *Improved* regression. This presents a confusing puzzle.

The final City Council district of interest was District 8, which had positive coefficients of 0.0426 and 7.6218 in the *Improved* and *OCI2011* regressions, respectively. The coefficients appear very different in size, but remember that *Improved* is either a 0 or 1 while the OCI value in 2011 could be anything from 0 to 100. Therefore, .0426 is actually a fairly large coefficient for the *Improved* regression: what it suggests it that being in District 8 instead of another district increased a road's chances of being improved by around 4 percent. Roads in this district clearly started off well and faced a large likelihood of being improved between 2011 and 2015. Why? District 8 is on the border of the United States and Mexico, and as such, is home to a large amount of international trade. San Diego is a major trading city, as San Diego and Tijuana are the largest metropolitan area on the U.S.-Mexico border and are home to the busiest border crossing in the world. Trade is vital to the economy of San Diego, and there is heavy and important traffic, likely including many large trucks—which will wear down the roads faster—traveling through the 8th district. Therefore, it makes sense for road blocks in this district to be regularly maintained.[9]

Of the three nearby variables—*NearCollege*, *NearSchool*, and *NearAttraction*—only one actually really proved to be of interest, and that was *NearSchool*. What is surprising is that *NearSchool* had significant negative coefficients in both the *Improved* and *OCIchange* regressions, even though San Diego stated in its Pavement Condition Report that proximity to schools is something the city considers when determining what roads to fix (City of San Diego 2016, *2015-2016 Pavement Condition Assessment*). Yet these regressions seem to suggest that roads near schools were less likely to be improved, going against what the city stated. Perhaps the *NearSchool* variable captures too wide a range of roads, or just generally doesn't properly portray the roads that San Diego considers to be near schools.

Having a bike route on the road block was significant across all three regressions, and the coefficients suggest an easy explanation for this significance. In the baseline *OCI2011* regression, the coefficient was 0.8229, while the coefficients in *Improved* and *OCIchange* were -0.01395 and -0.5924, respectively. This suggests that roads with bike routes started off well in 2011, and therefore were just less likely to be improved before 2015 since the road blocks already had better OCIs than other road blocks. Bike routes do not actually affect road maintenance patterns, but rather roads with bike routes simply

[9] Separating District 8 into its southern and northern sections also showed that the roads in the southern sections were in fact the ones causing the positive coefficient.

happened, by chance, to be generally better in 2011 and so they were less likely to be improved before 2015. Another less important result was historical districts. Historical districts were only significant in the *Improved* regression, suggesting that such districts have a minimal impact on road repair.

The density of businesses in the census block was significant in two out of the three regressions, specifically the *OCIchange* regression and the *OCI2011* regression at 2.4914 and 4.2215-, respectively. These positive coefficients suggest that the presence of more businesses in an area leads to better roads for that area. The roads started off better originally and were also more likely to experience a positive change in their OCI. Businesses help provide tax revenue for the city and keep the economy running (and it is also possible businesses are likely to complain about bad streets outside their stores), which could explain why San Diego is seemingly motivated to keep roads that serve a lot of businesses in good condition.

5.2 Robustness Checks

Throughout the course of my analysis, I checked to see how important various variables were to the overall model and sought to determine the best set of variables to include in my main regression models. Various sets of variables were all tested using F-tests to see if the inclusion of the variables improved the model. For these testing purposes, the model used had *Improved* as the dependent variable, since that regression was the one I was originally the most interested in when I began my analysis. Testing the inclusion of the City Council Districts resulted in a significant F-value of 10.798, so those variables were still included. Similarly, testing the racial variables, the traffic variables, and the road type variables resulted in F-values large enough to justify their inclusion in the model (values of 4.404, 4.903, and 10.359, respectively). But there were a few variables that did not make it into the model.

The main set of variables that were not included the model were zoning codes. As mentioned in Section 3, I collected data on the zones of San Diego, which are broken up into a few different categories, including zones such as Commercial, Residential, and Industrial. Including them in the model, however, resulted in insignificant coefficients on the five zoning codes included—the lowest p-value for an individual zone variable was 0.226, for open space zones. Additionally, when testing the joint significance of the variables using an F-test, the variables were not found to significantly improve the model (F=1.066). Therefore, these variables were not included in the final regressions run.

Another check I did was checking that the importance of City Council Districts was unrelated to the party affiliation of the City Councilmember for that district. Over this time period, there was very little change in the party of councilmembers, so it is therefore possible to imagine that the party of the member was somehow important to the amount of road maintenance received by the district. San Diego local elections are technically nonpartisan, but it is nevertheless fairly easy to determine the party affiliation of local politicians. I ran the regression with the variable *Democrat* in place of the City Council Districts, and found that the variable was insignificant (t=0.53). This insignificance made me feel justified that the importance of City Council Districts was unrelated to party, and that I therefore should keep the dummy variables for the districts themselves instead of the *Democrat* variable[10].

I also tried including a variable for the population density in the full model, however the variable had a t-value of -1.06 with a p-value of .290, which is insignificant. Other various models were run with various subsets of the variables to see if the population density was possibly being impacted by another variable, but the population density was insignificant in every model run. Some other p-values included .301, .298, .345, .895, and .731. The variable never had any effect, but the total population variable was sometimes significant, and so therefore I made the decision to include the total population variable instead of the population density one.

These checks on my variables helped me arrive at the model used in the main regression. Despite these tests, however, the model had its imperfections. Section 6 discusses these limitations.

6 Limitations

Most of the limitations of my research have been discussed where applicable earlier in the report, however this section will still discuss these limitations again, as well as mention a couple more. The purpose of this section is to help guide future research by pointing out the flaws in this approach, as well as to caution readers against generalizing the results of this paper too broadly.

The first limitation is the narrow focus of this paper on a four-year time span in San Diego. The results are unlikely to be generalizable to other American cities, due to the distinct differences in city size, weather, traffic volumes, economy, and other key characteristics. There is also the possibility that what matters for road repair changes over time, and so how San Diego prioritized its roads from 2011–2015 could be very different from what the city will do even a few years later. That being said, the results still do provide an interesting insight into the decisions that go into public service distribution in San Diego, while also providing a foundation for future research in this area.

Another limitation is the lack of traffic data for all road segments. If there was specific traffic data for each road block, the results could possibly have shifted and, at the very least, would provide a more accurate picture of the importance of traffic levels to road maintenance. Another data-related limitation was the lack of data on the age of all road segments. The data that San Diego had on the ages of roads appeared to only go back around 37 years, making it impossible to judge the age of older roads and difficult to assess the

[10] As the Democrat variable is directly related to and built from the City Council District variables, it would be impossible to include both the individual district variables as well as the Democrat variable in the model, due to perfect collinearity.

importance of age on road maintenance. Age is one variable that could have reasonably had a significant impact on the state of San Diego's roads and therefore have affected the likelihood of road repair.

Similarly, the lack of available 311 reporting data was a data limitation. Having 311 reporting data would give an idea about what people are calling about road problems in what areas of San Diego, which would have helped possibly prove or disprove the "squeaky wheel gets the grease" theory and could have changed the significance of the median household income variable. Unfortunately, this data was unavailable for the time period in question.

Finally, the way that the "nearby" variables were created is also a limitation of my model. As described in Section 3.2, these variables were generated in ArcGIS by matching all roadblocks within a certain radius to the various locations (schools, colleges, and tourist attractions). However, that approach provides a less than ideal depiction of what roads are actually used to get to these places. There may be local roads that are near tourist attractions but that are never actually used to get there; conversely, there may be roads that are further away from these locations but are the main roads that connect traffic from certain areas of the city to these locations.

These limitations are unfortunate; however, the scope of my research did not allow me to come up with satisfactory solutions to any of these problems. Despite these issues, I still believe the results to be reasonably strong, and that the results offer a unique look into what is important for road maintenance practices. Future research in this area, however, would do well to consider these limitations and learn from them.

7 Conclusion

This paper has looked at the maintenance practices of the city of San Diego from 2011 to 2015 in an attempt to determine how the characteristics of a road and the area the road is in determine the maintenance levels the road receives. While San Diego is only one example of a city, the findings nevertheless contribute to a broader literature on bias in public service distribution in cities across the United States.

One of the key findings in this paper is the relationship between the condition of a road and the median income of an area. This finding could be an example of the "squeaky wheel gets the grease" idea, but then again, the finding could also indicate that levels of public services are distributed in relation to the size of an area's tax base. Understanding the source of the positive relationship between income and roads is beyond the scope of this paper, but the regressions in this paper suggest that such a relationship exists. Determining whether such a relationship should exist is a normative question and, again, beyond the scope of this paper, but future research should dig more deeply into questions of how and why cities may treat richer areas more favorably than poor areas, and what the consequences of that are.

The findings also indicate that areas of relative economic importance—such as City Council District 8, which lies on the international border—are more likely to have their roads in good condition. This finding does seem to speak well of the management of road repair and the priorities of those in charge of these decisions. Solid infrastructure is crucial for the economy of a city, particularly when the city is an important international trading hub like San Diego. One would hope that extra care is taken to maintain the roads at the busiest land border in the world, which seems to be what is happening.

Future research should ideally look over longer periods of time in multiple cities to help build a broader picture of road maintenance practices. But research should also go beyond just looking at the determinants of road maintenance, and instead look at the impacts of different levels of road maintenance as well. If a city fixes roads in certain neighborhoods first, does that have an impact on home values or the overall attitude towards those neighborhoods? How does road maintenance impact the economy of a city? What is a fair way to determine what roads get fixed? How do road maintenance levels affect citizens' feelings of efficacy? How does it affect their transportation choices? These questions, among others, are interesting ones to consider in the world of urban economics, politics, and planning.

Hopefully future research will help show cities how they may be unknowingly making certain discriminatory decisions in their public service distribution choices, and how those choices lead to economic impacts on the city as a whole. Regardless of what future research shows, however, what this paper shows is that variables like income are able to predict road maintenance levels, which is something that San Diego, and truly all cities, should reflect on. The future of San Diego's roads is unknown, but hopefully the city will continue to work to improve roads in low-income areas and other areas of the city that have had their roads ignored up until now. While it is important to make economically efficient decisions when deciding how to allocate resources, it is also important to ensure that all residents are receiving appropriate levels of public services. Having reliable roads is vital for a city's success, but San Diego's roads are, unfortunately, not always reliable.

Appendix A

The OCI rating system was developed by the U.S. Army Corps of Engineers and is made up of a Pavement Condition Index and a Ride Condition Index. The Pavement Condition Index is determined using the ASTM Standard D 6433-11 and rates the amount of distresses on the road. The Ride Condition Index is determined using a laser profiler and measures the roughness of the road. Both indexes are measured on a score from 0 to 100 and are combined together to form the OCI—the Pavement Condition Index accounts for 60 percent of the OCI and the Ride Condition Index accounts for the other 40 percent (City of San Diego 2016, *2015-2016 Pavement Condition Assessment*).

Some factors for determining the OCI include the type of street, traffic levels, age, and quality of ride. Other factors include the type and size of cracks, number of potholes, oxidation, and deterioration rate. An OCI score of 100 represents a pavement with a perfect surface condition. An OCI score of 0 represents a street that is beyond repair and requires complete reconstruction (City of San Diego 2016, *2015-2016 Pavement Condition Assessment*).

Appendix B

This appendix shows the correlation matrices for all relevant variables in my models. Some of the larger values have been bolded to make it easier to see the correlated variables. Also note that some variables have shorter names in these matrices than in the paper, which was necessary to help the matrices fit in the margins. The matrices follow in Tables 7–12.

	Length	Resident	Prime	Collector	Local	Major	PriorConstr
Length	1.0000						
Residential	–0.1302	1.0000					
Prime	0.0316	–0.3992	1.0000				
Collector	0.0102	–0.5458	–0.0886	1.0000			
Local	–0.0098	–0.2792	–0.0453	–0.0620	1.0000		
Major	0.1757	–0.4908	–0.0797	–0.1090	–0.0558	1.0000	
PriorConstr	–0.0051	0.0051	–0.0220	0.0458	–0.0308	–0.0205	1.0000
OCI2011	0.0126	–0.0179	–0.0079	0.0137	0.0153	0.0109	**0.3703**
Income	0.0605	0.0915	0.0368	–0.1921	0.0061	0.0298	–0.0613
TotalPop	0.0645	0.0269	–0.0250	–0.0526	–0.0162	0.0456	0.0707
PercHisp	0.0025	–0.0860	–0.0292	0.1180	0.0692	–0.0095	0.1101
PercWhite	–0.0228	0.0587	0.0726	–0.1306	–0.0389	0.0117	–0.1489
PercBlack	0.0110	–0.0690	–0.0733	0.1810	–0.0054	–0.0231	0.0477
PercAsian	0.0277	0.0588	–0.0498	–0.0396	–0.0244	0.0067	0.0892
PercOther	0.0141	0.0027	–0.0172	0.0168	–0.0071	–0.0034	–0.0102
VacShare	–0.0167	–0.0704	0.0206	0.0367	0.0159	0.0428	–0.0397
OccShare	0.0167	0.0704	–0.0206	–0.0367	–0.0159	–0.0428	0.0397
District1	0.0308	0.0796	0.0373	–0.1505	–0.0016	0.0075	–0.0755
District2	–0.0412	–0.0539	0.0915	–0.0651	0.0110	0.0726	–0.0515
District3	–0.0774	–0.0867	–0.0828	**0.2665**	–0.0325	–0.0632	0.0363
District4	0.0033	–0.0144	–0.0942	0.1593	–0.0310	–0.0537	0.0814
District5	0.0516	0.0392	0.0054	–0.0894	–0.0234	0.0449	–0.0526
District6	0.0031	0.0222	0.0450	–0.0586	–0.0354	0.0125	0.0030
District7	0.0299	0.0535	0.0290	–0.0835	0.0134	–0.0263	–0.0084
District8	–0.0007	–0.0178	0.0276	–0.0548	0.0869	0.0125	0.1220
District9	–0.0045	–0.0491	–0.0648	0.1079	0.0235	0.0009	–0.0303
NearCollege	–0.0064	–0.0719	–0.0300	0.1185	0.0233	–0.0042	–0.0070
NearSchool	0.0277	–0.0629	0.0141	0.0700	0.0015	0.0115	0.0037
NearAttract	–0.0204	–0.0964	–0.0078	0.1129	–0.0262	0.0520	–0.0440
BikeRoute	–0.1628	**0.3490**	–0.1256	–0.1716	0.0371	**–0.2855**	–0.0050
MaintenD	0.0001	0.0062	–0.0330	–0.0203	–0.0383	0.0633	0.0060
HistoricalD	–0.0191	–0.0514	0.0093	0.0513	–0.0203	0.0306	–0.0293
BusiDens	–0.0051	–0.1276	–0.0120	0.1187	0.0513	0.0902	0.0023
LowTraffic	–0.0062	–0.1459	0.1248	0.1489	0.0138	–0.0426	–0.0124
MedTraffic	–0.0869	**0.3678**	–0.1191	–0.2105	0.0144	**–0.2649**	0.0244
HighTraffic	0.1377	**–0.3546**	0.0121	0.1163	–0.0397	**0.4511**	–0.0199

Table 7: Correlation Matrix Part 1

The rest of the correlations can be seen in Tables 8–12.

	OCI2011	Income	TotalPop	PercHispanic	PercWhite	PercBlack
OCI2011	1.0000					
Income	0.1473	1.0000				
TotalPop	0.1167	0.1615	1.0000			
PercHispanic	0.0048	**−0.5474**	−0.0090	1.0000		
PercWhite	−0.0866	**0.4174**	−0.1045	−0.7499	1.0000	
PercBlack	0.0031	**−0.3182**	0.0258	0.2027	−0.4888	1.0000
PercAsian	0.1510	**0.2069**	0.1788	−0.1621	−0.4323	0.0236
PercOther	0.0054	0.0409	0.0597	−0.1404	−0.1041	0.0367
VacShare	−0.0841	−0.1070	−0.0465	0.0349	0.0462	0.0147
OccShare	0.0841	0.1070	0.0465	−0.0349	−0.0462	−0.0147
District1	0.0118	0.4206	0.0894	**−0.2964**	**0.3092**	−0.2068
District2	−0.0642	−0.0477	−0.1363	−0.1644	**0.3008**	−0.1373
District3	−0.0746	−0.1877	−0.1535	−0.0199	0.1421	−0.0144
District4	0.0148	−0.2683	−0.0125	**0.2789**	**−0.5043**	**0.5539**
District5	0.1047	0.3430	0.1001	−0.2165	0.1542	−0.1223
District6	0.0322	0.0760	0.0311	−0.1047	−0.1465	−0.0437
District7	−0.0529	−0.0142	0.0179	−0.1089	0.1371	−0.0549
District8	0.0871	−0.2527	0.0576	**0.6280**	**−0.4433**	−0.0205
District9	−0.0557	−0.2460	0.0017	0.1695	−0.1451	0.1014
NearCollege	−0.0619	−0.1756	−0.0558	0.0908	−0.0386	0.0522
NearSchool	−0.0447	−0.2492	−0.0988	0.1869	−0.1725	0.1069
NearAttraction	−0.0690	−0.0786	−0.0463	−0.1089	0.1646	−0.0413
BikeRoute	0.0299	0.0636	0.0307	−0.0366	−0.0120	0.0052
MaintenanceD	0.1351	**0.3680**	0.1481	−0.2442	0.0418	−0.1302
HistoricalD	−0.0488	−0.0859	−0.0773	−0.0034	0.0714	−0.0300
BusinessDens	−0.0407	−0.1148	−0.0462	0.0491	−0.0057	0.0200
LowTraffic	−0.0467	−0.1491	−0.0862	0.0739	−0.0260	0.0580
MedTraffic	0.0581	0.1755	0.0642	−0.0787	0.0375	−0.0765
HighTraffic	−0.0247	−0.0639	0.0185	0.0193	−0.0215	0.0371

Table 8: Correlation Matrix Part 2
The rest of the correlations can be seen in Tables 7, and 9–12.

	PercAsian	PercOther	VacShare	OccShare	District1	District2
PercAsian	1.0000					
PercOther	0.1153	1.0000				
VacShare	–0.1712	–0.0595	1.0000			
OccShare	0.1712	0.0595	–1.0000	1.0000		
District1	–0.0146	–0.0031	0.0124	–0.0124	1.0000	
District2	**–0.2346**	–0.0218	0.1301	–0.1301	–0.1850	1.0000
District3	**–0.2222**	–0.0172	0.0825	–0.0825	–0.1987	–0.1135
District4	0.1980	0.0292	–0.0335	0.0335	–0.1941	–0.1304
District5	0.1002	0.0197	–0.0964	0.0964	–0.1828	–0.1228
District6	**0.4264**	0.0974	–0.0821	0.0821	–0.1388	–0.0932
District7	–0.0736	0.0478	–0.0837	0.0837	–0.1824	–0.1225
District8	–0.0646	–0.1139	0.0051	–0.0051	–0.1606	–0.1079
District9	–0.0290	–0.0370	0.0521	–0.0521	–0.1336	–0.0897
NearCollege	–0.0901	–0.0046	0.0626	–0.0626	–0.0623	–0.0050
NearSchool	–0.0218	0.0130	–0.0307	0.0307	–0.0327	–0.0496
NearAttract	–0.1221	0.0092	0.1899	–0.1899	–0.0325	0.1752
BikeRoute	0.0787	–0.0060	–0.0787	0.0787	0.0498	–0.1050
MaintenanceD	**0.3364**	0.0780	–0.0544	0.0544	–0.0387	–0.1818
HistoricalD	–0.1057	–0.0266	0.0591	–0.0591	–0.0808	0.1684
BusinessDens	–0.0717	–0.0093	0.0846	–0.0846	–0.0487	0.0497
LowTraffic	–0.0906	–0.0195	0.0522	–0.0522	–0.0775	0.0453
MedTraffic	0.0910	0.0078	–0.0842	0.0842	0.0999	–0.0744
HighTraffic	–0.0155	0.0142	0.0563	–0.0563	–0.0461	0.0509

Table 9: Correlation Matrix Part 3

The rest of the correlations can be seen in Tables 7, 8, and 10–12.

Table 10: Correlation Matrix Part 4

The rest of the correlations can be seen in Tables 7–9, 11, and 12.

	District3	District4	District5	District6	District7	District8	District9
District3	1.0000						
District4	–0.1401	1.0000					
District5	–0.1319	–0.1288	1.0000				
District6	–0.1002	–0.0978	–0.0921	1.0000			
District7	–0.1316	–0.1286	–0.1210	–0.0919	1.0000		
District8	–0.1159	–0.1132	–0.1066	–0.0810	–0.1064	1.0000	
District9	–0.0964	–0.0941	–0.0886	–0.0673	–0.0085	–0.0779	1.0000
NearCollege	0.1642	–0.0377	–0.1035	–0.0528	–0.0194	0.0388	0.1065
NearSchool	0.0924	0.0769	–0.2129	–0.0255	0.0273	0.0680	0.0767
NearAttract	**0.2325**	–0.1027	–0.0253	–0.0596	–0.0720	–0.0834	–0.0707
BikeRoute	–0.0743	0.0701	0.0562	–0.0019	0.0204	–0.0095	–0.0270
MaintenanceD	0.0285	**–0.2483**	**0.4592**	**0.2531**	–0.0430	–0.1248	–0.0530
HistoricalD	0.1142	–0.0574	–0.0540	–0.0410	–0.0539	0.0241	–0.0077
BusinessDens	0.1744	–0.0553	–0.0613	–0.0300	–0.0376	–0.0091	0.0224
LowTraffic	0.1222	0.0443	–0.0972	–0.0400	–0.0355	–0.0211	0.0830
MedTraffic	–0.1364	–0.0348	0.0920	0.0345	0.0601	0.0265	–0.1021
HighTraffic	0.0413	–0.0067	–0.0082	0.0017	–0.0425	–0.0115	0.0422

	NearCollege	NearSchool	NearAttraction	BikeRoute	MaintenanceD
NearCollege	1.0000				
NearSchool	0.0520	1.0000			
NearAttraction	0.2330	–0.0478	1.0000		
BikeRoute	–0.0436	–0.0665	–0.0638	1.0000	
MaintenanceD	–0.0759	–0.1448	0.0428	0.0317	1.0000
HistoricalD	0.0313	0.0076	0.0307	–0.1016	–0.0213
BusinessDens	0.1150	0.0591	0.1858	–0.0979	0.0025
LowTraffic	0.0568	0.0735	0.0232	–0.0218	–0.1015
MedTraffic	–0.0734	–0.0769	–0.0514	**0.2515**	0.0745
HighTraffic	0.0340	0.0171	0.0603	**–0.3458**	0.0235

Table 11: Correlation Matrix Part 5

The rest of the correlations can be seen in Tables 7–10 and 12.

Table 12: Correlation Matrix Part 6

The rest of the correlations can be seen in Tables 7–11.

	HistoricalD	BusinessDens	LowTraffic	MedTraffic	HighTraffic
HistoricalD	1.0000				
BusinessDens	0.0879	1.0000			
LowTraffic	0.0483	0.0200	1.0000		
MedTraffic	–0.0545	–0.0814	–0.7536	1.0000	
HighTraffic	0.0171	0.0948	–0.2023	–0.4913	1.0000

Acknowledgements
I would like to thank all of the Oberlin Economics Faculty for all the support and helpful advice they gave as I worked on this thesis, with special thanks to Ron Cheung and Paul Brehm. I'd also like to thank all of the other economics honors students for their useful comments and the inspiration to keep moving forward—I wouldn't have wanted to undertake this journey with any other group of students. Thank you to my friends for listening to me ramble on about roads, and thanks to my family for their support. Finally, I'd like to thank Terri Pleska for her endless kindness.

References

Antunes, George E., and John P. Plumlee. "The Distribution of an Urban Public Service: Ethnicity, Socioeconomic Status, and Bureaucracy as Determinants of the Quality of Neighborhood Streets." *Urban Affairs Quarterly* 12 (1977): 313-322.

City of San Diego. *City Council Offices*. 2013. https://www.sandiego.gov/citycouncil (last accessed December 1, 2017).

City of San Diego. *2015-2016 Pavement Condition Assessment*. 2016. https://www.sandiego.gov/sites/default/files/09_22_16_2015-2016_pavement_condition_assessment_0.pdf (last accessed November 30, 2017)

City of San Diego. "City Council Districts." *DataSD.* May 9, 2016. https://data.sandiego.gov/datasets/city-council-districts (last accessed October 28, 2017).

City of San Diego. "Bike Route Lines." *DataSD.* May 9, 2016. https://data.sandiego.gov/datasets/bike-route-lines (last accessed October 26, 2017).

City of San Diego. "Road Lines." *DataSD.* May 9, 2016. https://data.sandiego.gov/datasets/road-lines (last accessed October 26, 2017).

City of San Diego. *StreetsSD.* 2016. http://streets.sandiego.gov (last accessed October 15, 2017).

City of San Diego. "Streets Overall Condition Index." *DataSD.* September 16, 2016. https://data.sandiego.gov/datasets/streets-overall-condition-index (last accessed October 26, 2017).

City of San Diego. "Zoning." *DataSD.* June 16, 2016. https://data.sandiego.gov/datasets/zoning (accessed October 28, 2017).

City of San Diego. "Traffic Volumes." *DataSD.* May 17, 2016. https://data.sandiego.gov/datasets/traffic-volumes (last accessed November 15, 2017).

City of San Diego. "Historical Districts." *DataSD.* June 30, 2017. https://data.sandiego.gov/datasets/maintenance-assessment-districts (last accessed October 28, 2017).

City of San Diego. "Maintenance Assessment Districts." *DataSD.* June 30, 2017. https://data.sandiego.gov/datasets/maintenance-assessment-districts/ (last accessed October 26, 2017).

City of San Diego. “San Diego Communities”. *City of San Diego Map Gallery*. N.d. http://sandiego.maps.arcgis.com/apps/OnePane/basicviewer/index.html?appid=8a321161576f4c00af2c81d747efd6b7 (last accessed May 15, 2018).

Coulter, Philip B. "Measuring the Inequity of Urban Public Services: Methodological Discussion with Applications." *Policy Studies Journal* 8 (1980): 683-698.

Dillon, Liam. "For San Diego's Broken Roads, Broken Promises Too." *Voice of San Diego*, October 9, 2011.

Hirsch, Jana A., Geoffrey F. Green, Marc Peterson, Daniel A. Rodriguez, and

Penny Gordon-Larsen. "Neighborhood Sociodemographics and Change in Built Infrastructure." *Journal of Urbanism: International Research on Placemaking and Urban Sustainability* (10), 2017: 181-197.

Li, Bin, Tuo Li, Man Yu, and Bin Chen. "Can Equalization of Public Services Narrow the Regional Disparities in China? A Spacial Econometrics Approach." *China Economic Review* 44 (2017): 67-78.

Mladenka, Kenneth R. "The Distribution of an Urban Public Service." *Urban Affairs Quarterly* 24 (1989): 556-583.

Mladenka, Kenneth R., and Kim Quaile Hill. "The Distribution of Urban Police Services." *The Journal of Politics* 40 (1978): 112-133.

San Diego Association of Governments. "GIS Downloads." *SANDAG.* 2012. http://www.sandag.org/index.asp?subclassid=100&fuseaction=home.subclasshome. (last accessed October 26, 2017).

San Diego Association of Governments. *SANDAG.* 2010. "Demographics & Other Data: Census." http://www.sandag.org/resources/demographics_and_other_data/demographics/census/data.asp (last accessed October 26, 2017).

Talen, Emily. "The Social Equity of Urban Service Distribution: An Exploration of Park Access in Pueblo, Colorado, and Macon, Georgia." *Urban Geography* 18 (1997): 521-541.

United States Census Bureau. "Selected Economic Characteristics: 2007-2011 American Community Survey 5-Year Estimates." *American FactFinder.* 2011. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk. (last accessed October 28, 2017).