Residential Rooftop Solar Equity in Washington State

Trisha Ray

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science in Electrical Engineering

University of Washington

2021

Reading Committee:

Daniel Kirschen

Baosen Zhang

Program Authorized to Offer Degree:
Electrical and Computer Engineering

University of Washington

Abstract

Residential Rooftop Solar Equity in Washington State

Trisha Ray

Chair of the Supervisory Committee: Dr. Daniel Kirschen

Residential rooftop solar has not been deployed equitably in Washington State. This is consistent with various national studies that found socioeconomic barriers to solar access related to income, race, etc. This work focuses specifically on analyzing solar disparities in Washington state. A study by Tony G. Reames found significant differences in the socioeconomic factors influencing solar adoption in four US cities [1]. For this reason, analyzing Washington state specifically can provide the Washington Department of Commerce with better insight into existing solar disparities. I model the relationship between solar per capita and socioeconomic factors in each census tract using a zero-inflated Poisson regression model. The model shows that socioeconomic disparities in Washington differ across the state as well as from national findings. Having a college education has the strongest positive correlation with the amount of solar per capita out of all socioeconomic factors considered. My model also confirms inequities in household income and race, similar to national findings.

TABLE OF CONTENTS

LIST OF FIGURES

GLOSSARY

AREA MEAN INCOME (AMI):

LMI: Low-to-moderate Income

DAC: Disadvantaged Community

NREL: The National Renewable Energy Labratory

PV: Photovoltaic

CETA: Clean Energy Transformation Act

REPLICA: Rooftop Energy Potential of Low-Income Communities in America

ACS: American Community Survey

SASH: Single-Family Affordable Solar Housing program

DGEN: NREL's Distributed Generation Market Demand Model

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to University of Washington, where she has had the opportunity to work with members of the Renewable Energy Analysis Lab as well as Dr. Daniel Kirschen, who advised me throughout my time here.

Chapter 1

INTRODUCTION

The path to a renewable energy future will play a critical role in mitigating anthropogenic carbon emissions to slow climate change. According to the Intergovernmental Panel on Climate Change (IPCC), the power sector needs to completely decarbonize from fossil fuels by 2050 to stay below the 1.5° C limit. Recently, the power industry has seen a rise in distributed energy resources (DERs), which includes solar photovoltaic (PV), fuel cells, microturbines, and energy storage. These clean, small-scale energy sources are shifting the traditional top-down approach of distributing electricity from power plants to consumers. A major contributor to the rapid growth in DERs is distributed solar. As of 2020, more than 2.3 million solar generators have been installed into the U.S. distribution system. This increasing percentage of DERs contributes heavily to our ability to decarbonize the electricity grid.

In the early stages of distributed solar deployment, the high cost of photovoltaics (PVs) made rooftop solar only accessible to affluent households. However, with the falling cost of PVs and the potential financial savings from installing rooftop solar, the question of solar equity arises. Equity is defined as the fair treatment, access, opportunity, and advancement for all people, while at the same time striving to identify and eliminate barriers that have prevented the full participation of some groups. In the case of solar equity, this means identifying barriers to solar access that have prevented a certain group from adopting solar. These barriers can range from insufficient financial incentives to a lack of language translations in solar marketing. One way to identify barriers to solar access is by analyzing solar disparities of a particular region. In this study, I look at Washington state specifically to identify any existing disparities that would assist the Washington Department of Commerce in making plans to make an equitable transition to clean energy.

1.1 *Washington State's Equitable Clean Energy Transformation*

On May 7, 2019 Governor Jay Inslee signed the Clean Energy Transformation Act (CETA) into law, committing Washington to clean electricity by 2045. As an accountability measure, the law places milestones along the way that are as follows:

- By 2022, all utilities must have a clean energy implementation plan.
- By 2025, all utilities must eliminate coal-fired electricity from their state portfolios.
- By 2030, electricity must be greenhouse gas neutral.
- By 2045, electricity must be 100% renewable or non-emitting without the help of carbon offsets.

With these strict deadlines in place, residential rooftop solar will have to play a critical role in supplying energy as an alternative to fossil fuel plants. Currently, the electricity resource mix is shown in Figure 1.1. The majority of Washington's electricity supply comes from hydropower which is renewable. However, 13.15% comes from coal and 10.73% comes from natural gas, both of which are significant contributers to Washington's electricity supply. Meanwhile, only 0.29% of the electricity mix comes from solar.

As a result of CETA, in December 2020, the Washington State Department of Commerce released the "Washington 2021 State Energy Strategy: Transitioning to an Equitable Clean Energy Future." The goal of this strategy is to provide a roadmap for Washington State to stay on track of the deadlines set by CETA. Five main areas of focus are outlined in the strategy: communities, transportation, buildings, industry, and electricity. The two main areas that tie closely to the issue of equitable rooftop solar deployment are communities and electricity.

The communities section of the report is titled "Build an Equitable, Inclusive, Resilient Clean Energy Economy." What does it mean to transition to an equitable, inclusive, and

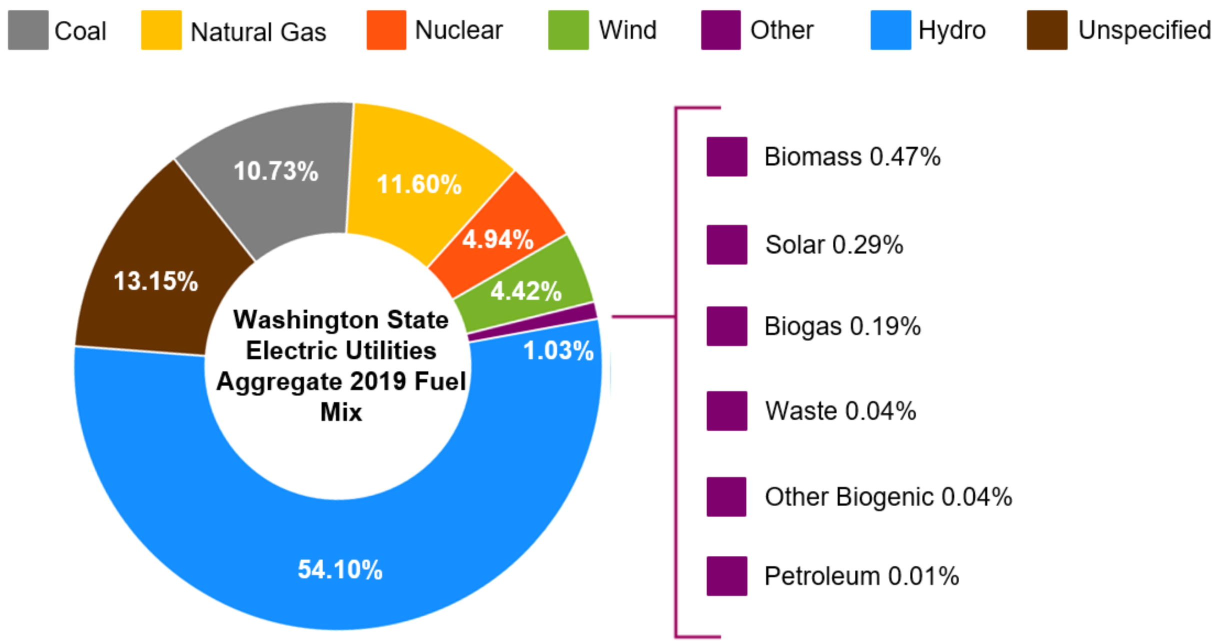

Figure 1.1: Washington State Electric Utilities Aggregate 2019 Fuel Mix. Image edited from [2]

resilient energy economy? An equitable transition to clean energy acknowledges that there are high impacted populations that are disproportionately affected by climate change. Meanwhile, an inclusive transition to clean energy means that decisions are made collectively by holding public meetings and seeking comments from historically excluded groups. Lastly, a resilient transition to clean energy means planning for increased natural disasters as a result of climate change. Furthermore, the communities section outlines the three dimensions of environmental justice work: 1) structural equity, 2) procedural equity, and 3) distributional equity. My analysis in Chapter 3 will focus on distributional equity, which focuses on the fair distribution of resources and benefits. However, all dimensions of environmental justice will be required when drafting equitable solar policies.

The electricity section of the report is titled “100% Clean Electricity to Meet the Needs of a Decarbonized Economy.” As expected from the title, this section outlines the need to invest in renewables, to built a smart and flexible grid, and to facilitate community deployment of renewable and grid services. One of the key actions within this section is to “develop distributed energy resources along with smart grid capabilities.” As mentioned earlier, distributed energy resources predominantly consists of rooftop solar. Combining this key action with the environmental justice framework discussed in the communities section of the report, rooftop solar must also be deployed in a equitable, inclusive, and resilient way. Equitable means that communities can install solar panels despite their income, race, location, age, etc. Inclusive means that government officials are seeking comments from groups with disproportionately less solar installed to understand barrier to adoption. Resilient mean that rooftop solar is being strategically installed to provide protection and stability from natural disasters by reducing dependence on large power plants.

1.1.1 King County Climate Action Plan

So far we have discussed statewide goals that support an equitable adoption of rooftop solar. If we now narrow our focus to King County, where the greatest number of residential rooftop solar panels in the state resides, we see a similar story. The King County 2020 Strategic Climate Action Plan presents a roadmap with three goals: (1) reducing greenhouse gas emissions, (2) using a Sustainable & Resilient Frontline Communities (SRFC) framework, and (3) preparing for climate change. Similar to the Washington 2021 State Energy Strategy, King County is centering their plan around equity and inclusion. The purpose of discussing the King County Climate Action Plan is (1) to introduce a more concrete way to prioritize community interests and (2) to discuss historic and current social inequities that create solar barriers.

One way King County is able to receive input from frontline communities is through their Climate Equity Community Task Force (CECTF). This task force consists of 22 community leaders that represent many different identities within frontline communities. The SRFC framework provides a community-driven framework for addressing climate change using the following key strategies:

1. Build King County and community capacity to prioritize climate equity.

2. Prioritize collaborative language access in partnership with trusted community partners.

3. Advance frontline community leadership by investing in long-term community and tribal partnerships, community capacity development, and improved infrastructure for community driven policy and decision-making.

4. Address root causes of climate vulnerability by prioritizing comprehensive solutions co-developed with frontline communities that reduce systemic inequities and have co-benefits.

5. Advance an equitable climate future and outcomes by investing in climate solutions and opportunities with and for frontline communities.

6. Align with and elevate actions in related County plans and programs that support frontline communities and climate resilience.

Another way King County is prioritizing community interests is by surveying community members on rooftop solar. Surveys found that there was strong interest for small-scale residential, business, and community solar power. In addition, community members were interested in financial incentives that expanded access to solar programs for BIPOC communities and low-income communities.

Furthermore, the King County 2020 Strategic Climate Action Plan acknowledges that climate change can worsen social and economic inequities. Historically, state residential solar energy incentive programs have seen greater participation from higher-income communities. Meanwhile, historic redlining and marginalization have left parts of southwest King County, where a higher percentage of people of color and low-income families live, with less green space and access to public transit. The plan states, "As we work toward clean energy solutions, some of the solutions and benefits such as solar panels or electric vehicles have been out of reach for people living with low incomes, renters, and more." The only way these solutions can become more accessible is by addressing the unique barriers each disadvantaged community faces.

1.2 Overview of Thesis

Both the Washington 2021 State Energy Strategy and the King County 2020 Strategic Climate Action Plan provide motivation for analyzing the equity of residential rooftop solar installations in Washington State. As I will also discuss in the next chapter, regional disparities in solar adoption exist that can differ from national findings. For these two reasons, my thesis focuses on rooftop solar equity in Washington state. The rest of the thesis will be broken down as follows. Chapter 2 will provide a literature review of national solar disparities that have been identified by several papers. Chapter 3 will then provide an analysis of solar disparities in Washington state. From this analysis, Chapter 4 will then discuss potential solutions to increase solar equity based on various studies as well. In this thesis, the terms "rooftop solar" and "solar" are interchangeable with "residential rooftop solar."

Chapter 2
LITERATURE REVIEW

2.1 Socioeconomic Disparities in Rooftop Solar Adoption

The purpose of this section is to explore existing literature on equity in rooftop solar. There are many categories in which equity issues in rooftop solar can exist. A well explored equity issue is income disparities in rooftop solar. Since rooftop solar has a high upfront cost, the question of who can afford rooftop solar naturally arises. Less apparent are disparities regarding race/ethnicity, internet access, and English proficiency. While these disparities are less understood, some studies [1] and [4] indicate the need for closer examination of the population benefiting from rooftop solar. Identifying equity issues in rooftop solar ensures marginalized communities, who are often most in need of rooftop solar's benefits, are not further harmed.

The following studies explore socioeconomic disparities in two distinct ways. The first two studies explore solar adoption versus a singular socioeconomic factor. The first study looks at income disparities while the second looks at racial disparities. The subsequent studies explore correlations between solar adoption and a compilation of socioeconomic factors.

2.1.1 Income Disparities

Below you will see a list of terms that will be used throughout the rest of this paper. These are common terms used to analyze income equity.

Terminology for Income Equity Assessment

- Area Mean Income (AMI) – the household income for the median – or middle – household in a region.

- Low-to-moderate Income (LMI) – defined as having income levels lower than 80% of the area mean income (AMI). The LMI population makes up 43% of the total population in the United States.

- LMI Census Tract – a tract where 51% or more of the population in the tract is LMI.

There are multiple studies that analyze income trends of solar adopters. One notable study from Lawrence Berkeley National Lab [3] found that only 15% of solar adopters are <80% of the area median income (AMI). As shown in Figure 2.1 from this study, the distribution of solar adopter incomes is skewed to the right. This indicates that rooftop solar is still not being adopted by lower-income households at the same rate. However, the study also found that solar adoption is slowly moving towards lower income customers. This reflects falling photovoltaics (PVs) prices , increasing financing options, programs to incentivize LMI households, and maturing PV markets. The study also found that the degree to which the income distribution skews high changes at different spatial levels. Looking at more granular levels such as state-level, county-level, and tract-level, the study found that there was less disparity when examining smaller populations. 18% of all 2018 solar adopters were below the median income at the national level, but 31% of all 2018 solar adopters were below the median income at the census block group level. These spatial differences motivate the need to understand different socioeconomic disparities across the US. The income disparities of a specific state or county might look vastly different than the national overview.

2.1.2 Racial Disparities

Sunter *et al.* investigated racial disparities in rooftop solar adoption and found that black and Hispanic-majority census tracts show less rooftop solar installations than no-majority and white-majority census tracts [4]. A majority-tract is defined as a race having more than 50% representation in a tract. This disparity existed even after accounting for household income and home ownership. As shown in Figure 2.2, at all income levels, Black-majority census tracts and Hispanic-majority census tracts saw less solar deployment than no majority

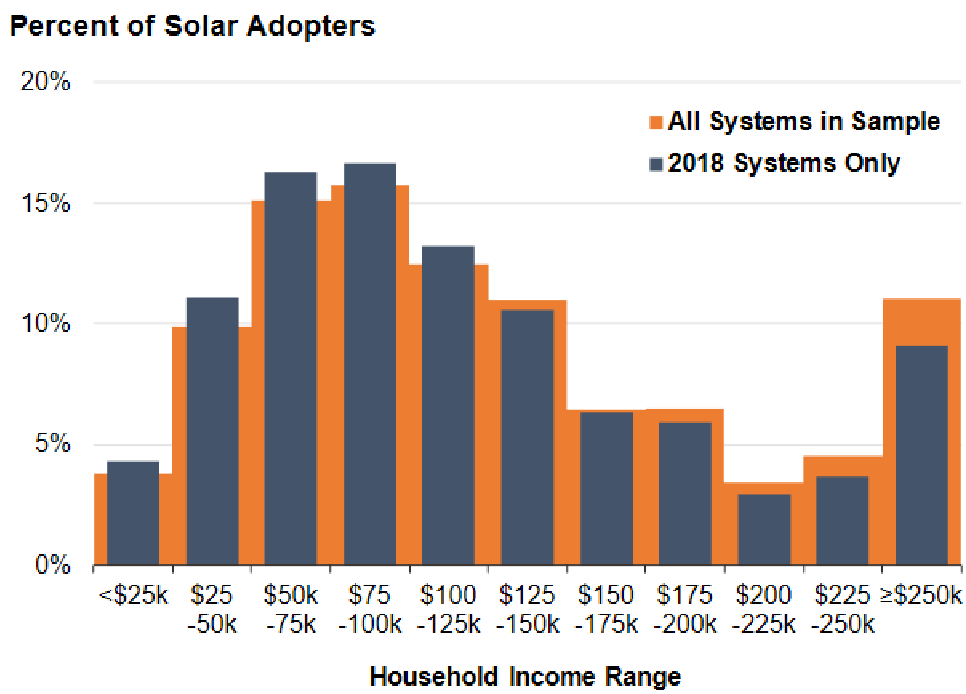

Figure 2.1: Figure taken from [3]. Household Income Range for Solar Adopters. The distribution in orange is for all solar systems within the study's sample. The distribution in gray is for solar systems installed in 2018 only.

census tracts. On the other hand, white-majority census tracts saw a greater amount of solar deployment at all income levels than no-majority census tracts. The exact same story can be seen we accounting for differences in home ownerships. Figure 2.3 shows the amount of solar deployment in different racial groups when accounting for the percentage of households occupied by renters. Overall, the study found that Black-majority census tracts deployed 69% less solar than no-majority census tracts and Hispanic-majority census tracts deployed 30% less.

The study also examined the effects of seeding. Seeding is a method in which a "seed" customer installs rooftop PV which influences their neighbors to install solar as well. This method has been found to be an effective bottom-up approach to installing more rooftop solar through the social diffusion effect. When examining the effects of seeding by race, the study found that Black-majority communities suffer from a disproportionate lack of seeding. This was determined by examining Black-majority tracts with nonzero solar installations separate from all Black-majority tracts.

2.1.3 Socioeconomic Factors Affecting Rooftop Solar Adoption

The last two studies looked at one socioeconomic factor (income or race) to understand solar disparities. Another method is to analyze a variety of socioeconomic factors using a regression model. Studies analyzing correlations between socioeconomic factors and rooftop solar adoption have varied in regions explored and methods used.

In 2012, Kwan studied the influence of local environmental, social, economic, and political variables on residential rooftop solar adoption across the United States [7]. Using a zero-inflated binomial regression model, the study found that solar insolation, cost of electricity and the amount of available financial incentives are significant factors that influence solar adoption. The methodology used in this study was my motivation for implementing a zero-inflated Poisson regression model in the next chapter.

Meanwhile in a study focused on California, Lukanov and Krieger found that socioeconomic factors such as linguistic isolation, housing burden, median household income, cardiovascular disease, asthma rate, ozone, traffic and others were found to be statistically significant at the $p < 0.001$ level [6]. This study specifically used data from CalEnviroScreen to distinguish solar adoption in census tracts with high pollution and socioeconomic burdens. CalEnviroScreen is a mapping tool that ranks census tracts in California based on cumulative impacts from pollution exposures, environmental effects, sensitive populations, and socioeconomic factors. The top 25% of these tracts are labeled as Disadvantaged Communities (DAC). Lukanov and Krieger found that solar adoption in tracts labeled as DAC had disproportionately less residential rooftop solar installed than outside of these DACs. The recent California Disadvantaged Communities Single-family Affordable Housing (DAC-SASH) Program hopes to address these existing disparities.

Finally, Reames looked at the relationship between socioeconomic factors and rooftop

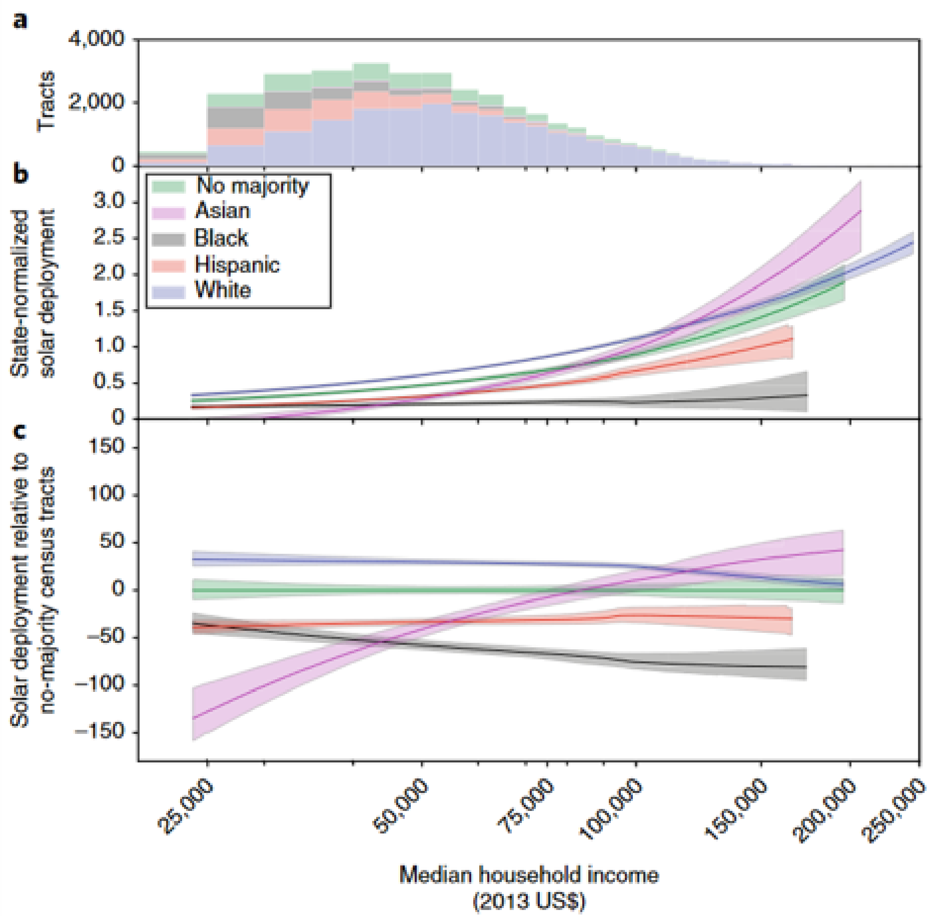

Figure 2.2: Figure and caption taken from [4]: Relationship between household income and rooftop PV installation by race and ethnicity. a. Histogram of the distribution of census tracts analysed at intervals of US$5,000. b. Rooftop PV installations relative to the available rooftop PV potential and normalized by state as a function of the median household income for majority census tracts. c. Rooftop PV installations relative to the available rooftop PV potential and normalized by state as a function of the median household income for majority census tracts and normalized relative to the rooftop PV adoption of no majority census tracts. Dark continuous curves represent the results of the LOWESS method applied to all data in each racial and ethnic majority group. Lighter shading represents the 90% CIs based on 1,000 bootstrap replications of each racial and ethnic majority group. The x axes are plotted on a base 10 logarithmic scale.

solar adoption in four different cities across the U.S: Riverside, California, San Bernardino, California, Washington, DC, and Chicago, Illinois. The study answered three questions us-

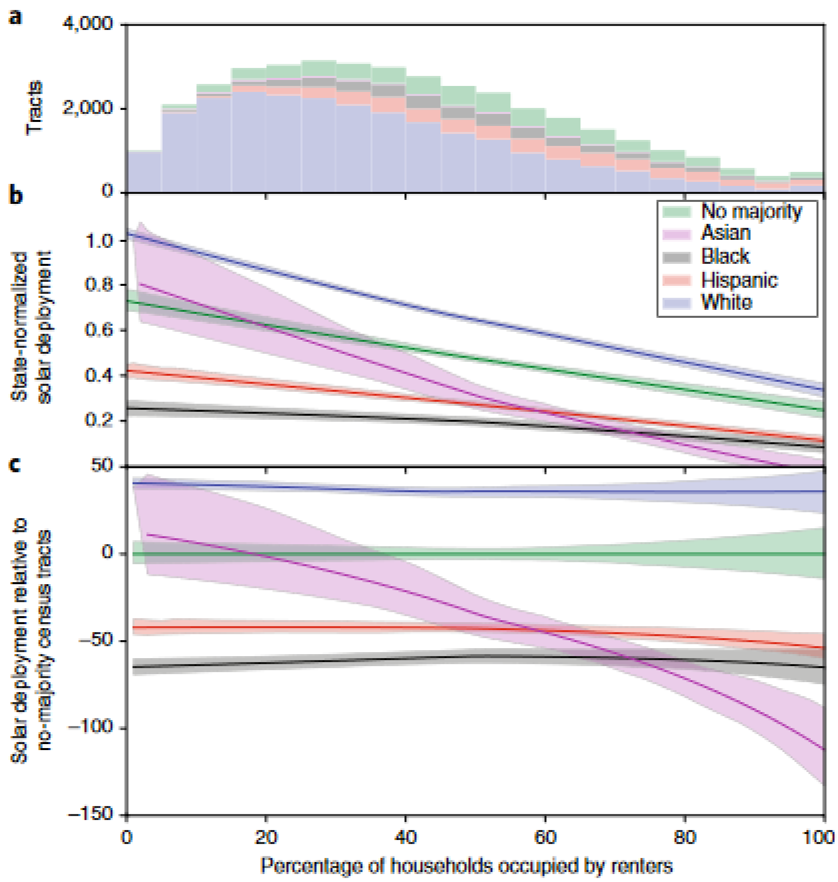

Figure 2.3: Figure and caption taken from [4]: Relationship between home ownership and rooftop PV installation by race and ethnicity. a. A histogram of the distribution of census tracts analysed at intervals of $5,000. b. Rooftop PV installations relative to the available rooftop PV potential and normalized by the state as a function of renter-occupied households. c. Rooftop PV installations relative to the available rooftop PV potential and normalized by the state as a function of renter-occupied households and normalized relative to the rooftop PV adoption of no majority census tracts. Dark continuous curves represent the results of the LOWESS method applied to all data in each racial and ethnic majority group. Lighter shading represents the 90% CIs based on 1,000 bootstrap replications of each racial and ethnic majority group.

ing data from these four cities: (1) how are spatial distributions of rooftop potential and penetration similar and different across cities?, (2) how is rooftop penetration distributed across non-LMI and LMI communities in different cities?, and (3) how do the relationships between rooftop penetration and local socioeconomic and demographic characteristics, identified as barriers to solar adoption, differ? Overall, the study highlighted the importance of understanding how socioeconomic factors affect a specific community compared to nationally. The study found that potential socioeconomic barrier to rooftop solar adoption vary between the four cities. While all four cities found correlations between rooftop solar and higher income census tracts, factors such as English proficiency, education, race, internet access, and house age and value influenced rooftop solar adoption differently in each city. The study also found that high rooftop solar penetration was not always found in the tracts with the highest income. This fact further supports the need to analyze socioeconomic trends regionally.

2.2 Summary

Each of these studies influenced my study of Washington in the following section. I included income and racial socioeconomic factors to compare my results with those of [3] and [4]. I built my own Zero-Inflated Poisson Regression model based on [7] to analyze relationships between rooftop solar adoption and socioeconomic factors. I incorporated the Washington Environmental Health Disparities Map, similar to CalEnviroScreen, in my analysis of rooftop solar adoption in disadvantaged communities. Lastly, I compared socioeconomic disparities in urban and rural census tracts in Washington to analyze spatial differences in socioeconomic influences. My full methodology will be discussed in the following section.

Chapter 3

ANALYSIS OF ROOFTOP SOLAR EQUITY IN WASHINGTON

3.1 Introduction

The previous section provided an overview of the correlations between socioeconomic factors and rooftop solar adoption. In this section, I will analyze how disparities in Washington State agree or disagree with finding in the literature. I will use multiple methods to analyze socioeconomic factors that might influence rooftop solar adoption. These methods include: geospatial analysis using GIS, Spearman's correlation coefficient, a multiple regression model, and two poisson regression models. First, I will briefly discuss how I gathered data for this analysis. Next, I will analyze results from the methods mentioned above. Lastly, I will discuss regional differences in results in Washington and overall conclusions.

3.2 Overview of Open-Source Datasets

When looking at tract-level data on residential solar installations, I found three open-source options. These are the Deep Solar dataset from Stanford University [8], the Project Sunroof dataset from Google [9], and the Tracking the Sun dataset from the Lawrence Berkeley National Lab [10]. Deep Solar and Project Sunroof both use similar methods to identify solar installations. These datasets use machine learning models to identify and classify rooftop solar from satellite imagery from Google. While Deep Solar covers every tract in Washington state, Project Sunroof only covers a portion of the state. For this reason, Deep Solar was chosen over Project Sunroof, but the Project Sunroof dataset was used to validate findings.

Meanwhile, the Tracking the Sun dataset is an ongoing collection of solar installations sourced from state agencies, utilities, and other organizations that administer PV incentive programs, solar renewable energy credit (SREC) registration systems, or interconnection processes. Due to imperfections in data collection, the Tracking the Sun dataset roughly covers 81% of all distributed solar installations. The open-source version of this dataset does not include tract-level data, but zipcode level data. Lawrence Berkeley National Lab provided me with the census-tract level data.

One important distinction between these three datasets is the time in which the data was collected. While all three datasets plan to be continually updated, none of them are completely up-to-date to 2021. Project Sunroof is currently using satellite imagery taken between 2012-2014 while Deep Solar is using imagery from around 2015. The most recent solar installation dates in the Tracking the Sun database is April 2019. However, the Tracking the Sun database appears to be incomplete in earlier rooftop solar installations with only one solar installation reported in 2012 and none before. For this reason, I use the Deep Solar dataset for my analysis and compare results with the other two datasets to ensure there are no significant discrepancies.

In order to analyze socioeconomic disparities in Washington State, I merged data from three open-source datasets: NREL's Rooftop Energy Potential of Low-Income Communities in America (REPLICA), Deep Solar, and the United States Census Bureau's ACS Community Survey (2011–2015, 5-year).

The REPLICA dataset provides estimates of the residential rooftop solar technical potential in each census tract. This data set is the result of a study by NREL that looked specifically at the rooftop solar potential in LMI households [11]. The dataset is the compilation of 11 smaller datasets:

- LMI PV Rooftop Technical Potential from NREL
- Residential Electricity Expenditures from Low Income Energy Affordability Data (LEAD)
- Electric Utilities from EIA 861, 2016
- State Residential Solar Incentives from DSIRE

- Demographics from the American Community Survey, 2011-2015
- Air Quality Index from EPA
- Heating and Cooling Degree Days from the National Solar Resource Database (NSRDB)
- Climate Zone from IECC/ASHRAE
- Locales from National Center for Education Statistic
- Public Housing from the National Housing Preservation Database (NHPD)
- Low Income Tax Credit Qualified Tracts from the Department of Housing and Urban Development (HUD)

The main contribution from [11] was the first dataset ”LMI PV Rooftop Technical Potential from NREL.” The first dataset provides the number of households in each tract that are low-to-moderate income as well as their rooftop solar potential. I use this information to identify correlations between rooftop solar adoption and income group.

The ACS Community Survey (2011–2015, 5-year) provides tract-level socioeconomic, demographic, and housing data. (ACS, 2011–2015, 5-year) Both the REPLICA dataset and Deep Solar dataset are merged with relevant data from the ACS Community Survey to simplify the data gathering process. This final merged dataset allowed me to perform census-tract analyses to determine any relationship between the census tracts with a greater solar per capita and census tracts with a greater percent of a certain demographic.

3.3 Geospatial Analysis in King County

First, I will visually analyze where solar installations are more prevalent in Washington state. At the tract-level, as shown in Figure 3.1, it is hard to distinguish which tracts possess the most solar. From Figure 3.2, you can see that the majority of solar installation is in King

County and surrounding counties. This indicates that statewide, and even county-wide, there are very concentrated levels of solar in specific regions.

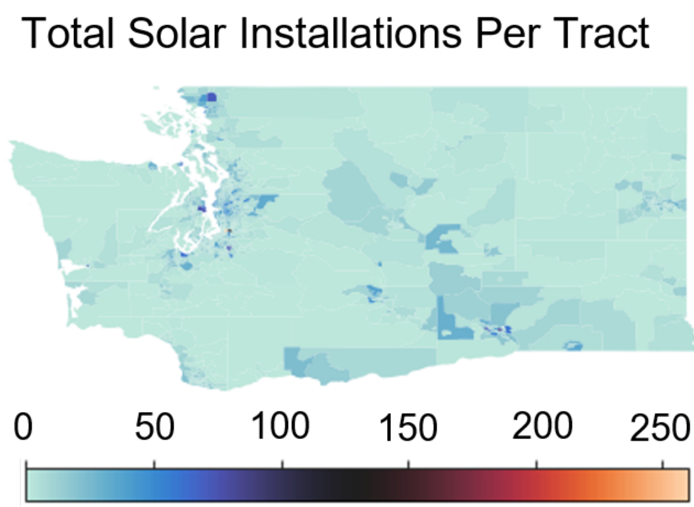

Figure 3.1: Total number of residential rooftop solar systems in each Census Tract. Data from Deep Solar database.

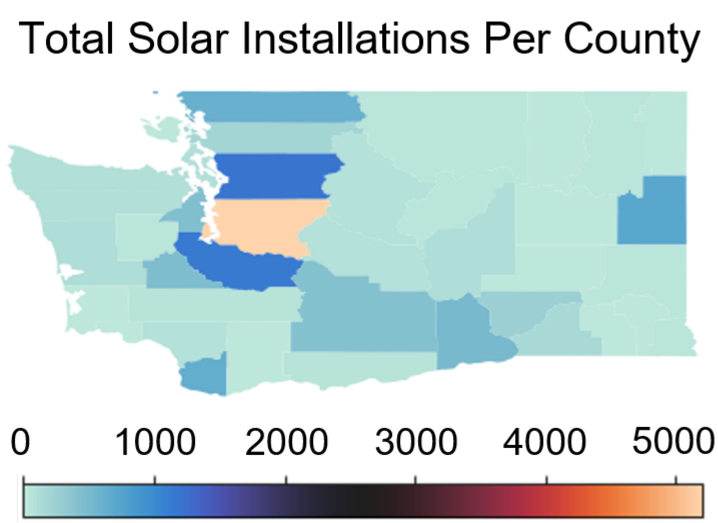

Figure 3.2: Total Solar Installations Per County in Washington State. Data from Deep Solar database.

A likely reason there is more solar installations in King county and surrounding counties is due to higher population density. For this reason, Figure 3.4 shows the rooftop solar installations per capita.

As we can see when comparing Figure 3.2 and Figure 3.3, solar is not being installed where there is higher solar irradiation in Washington state. This goes against what would make economic sense since solar is more profitable in regions like California that receives significant sunlight throughout the year. Yet, this fact does not hold true in Washington State. Solar is concentrated in areas with the least amount of sun in Washington state. This is an indication that economics might not be the strongest motivating factor for installing solar in Washington state.

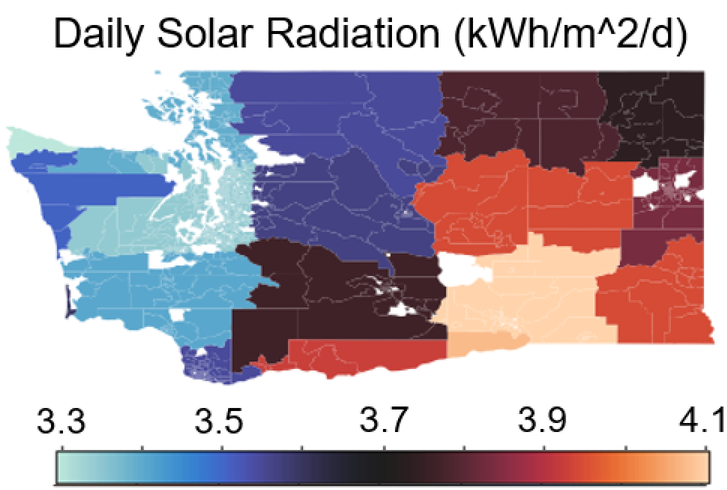

Figure 3.3: Solar irradiation in Washington State

As shown in Figure 3.1, it is hard to detect which tracts have a higher percentage of solar. For this reason, in the next part of this analysis, I look specifically at King County since it is the county with the most solar installations. Figure 3.5 shows six different maps of King county at the tract level. The top left map depicts the spatial distribution of solar per capita. The other five of the maps depict the spatial distribution of a different socioeconomic

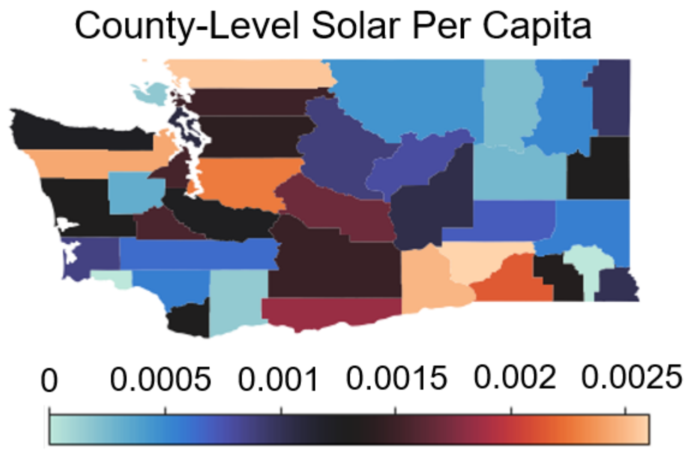

Figure 3.4: Solar Per Capita in Each County in Washington State. The units are the number of solar installations per person in each county.

factor. Since the numerical ranges of all six graphs differ, a qualitative scale on the right of the maps is used where light orange indicates the highest percentage and light blue indicate the lowest percentage.

In the "Solar Per Capita" map, census tracts with a higher solar per capita are clustered in the northwest region of King County near the Greater Seattle Area. There are a few outlier tracts in orange to the south of this region as well. Hoewever, when cross-referencing these outliers with the Project Sunroof and Tracking the Sun datasets, it is likely these outliers are actually errors in data collection since the other two datasets did not show these outliers. Meanwhile, tracts on the east side of King County show little to no solar deployment.

In the "Percent of LMI Households" map, census tracts with a higher percentage of LMI households are clustered to the south of the Greater Seattle area, at the southwest part of King County. When comparing this map to the "Solar Per Capita" map, this southwest region is also where there is less solar. King County has identified this region of the county to face greater marginalization due to historic redlining as noted in the previous chapter and [12].

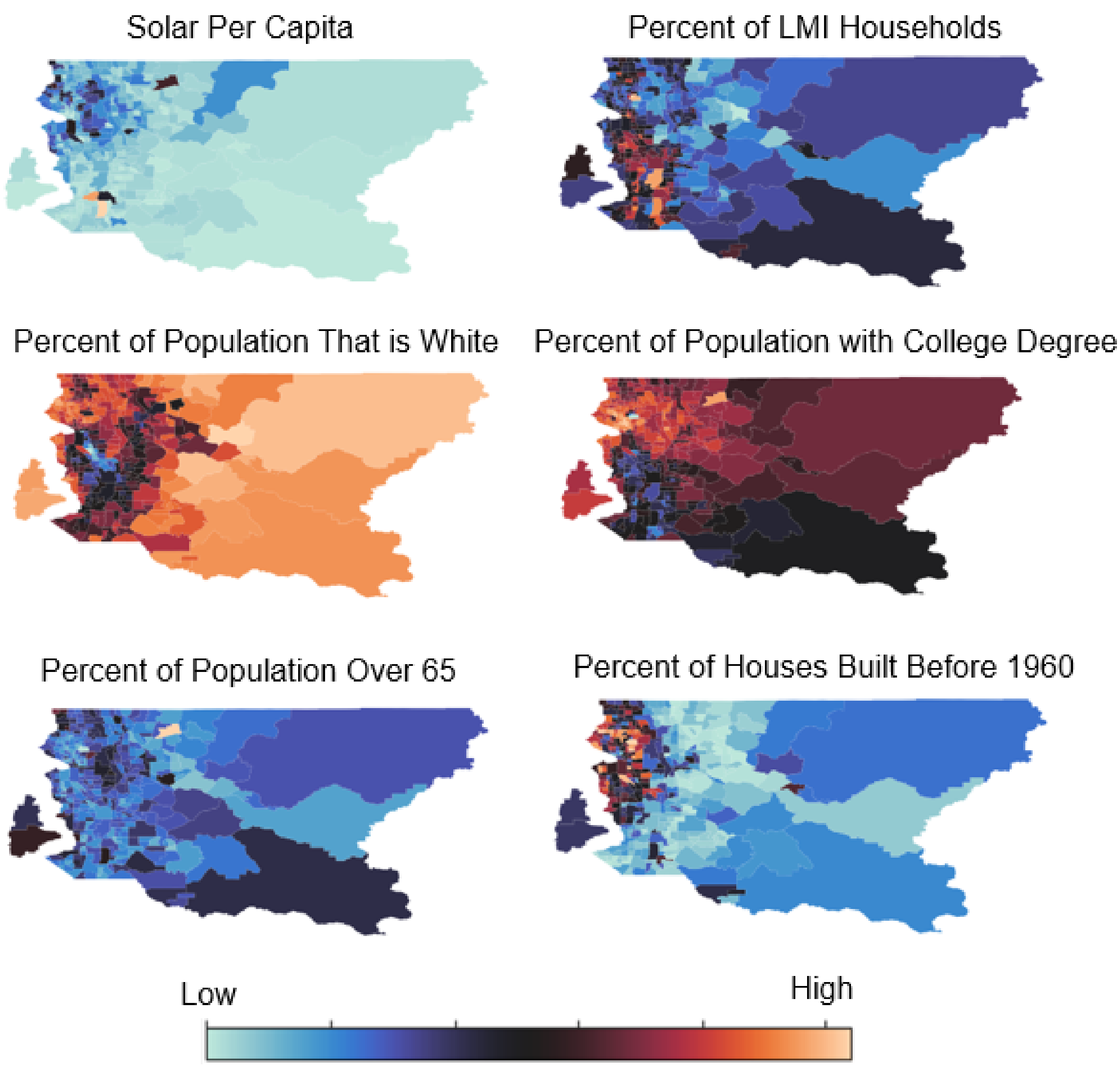

Figure 3.5: Six different maps of King County providing tract level information of the solar per capita or the following socioeconomic factors in the following categories: income, race, education, age, and house age.

In the “Percent of Population That is White” map, census tracts with a lower percentage of Caucasians are again clustered in the southwest portion of King County. This means that people of color are often the ones faced with greater socioeconomic barriers, including access to solar. Most of the eastern part of King County is almost entirely made up of white people. In the Greater Seattle area, there is a much higher white population than tracts to the south of it. When comparing this map to the “Solar Per Capita” map, there isn’t a clear overall relationship. Less solar has been deployed in the east where there is a strong white majority. Likewise, less solar has been deployed in the southwest region where there is the lowest percentage of white people in the county. However, there is a higher white majority in the Greater Seattle area where there is the most solar.

In the “Percent of Population with College Degree” map, about half of the population in census tracts to the east of the county have a college education. This differs from many tracts in the northwest part of the county where closer to 80% of the population has a college education. Meanwhile, a higher number of tracts in the southwest region have populations where only 10-20% of people have a college education. When I compare this map to the “Solar Per Capita” map, the majority of tracts with a greater amount of solar installations are in the same areas with a high percentage of college educated people.

In the “Percent of Population Over 65” map, no overall trend is apparent. 5-20% of the population across King County is over 65. With no spatial trend in the percent of population over 65 at the tract-level, I can’t determine if senior citizens are more or less likely to adopt solar.

In the “Percent of Houses Built Before 1960” map, the majority of houses in the Greater Seattle area were built before 1960. For the rest of King County, the majority of houses are newer. When we compare this map to the “Solar Per Capita” map, we can see that census tracts with a greater number of older homes were the same census tracts with a greater amount of solar installations. This correlation is likely because these are more older homes and more solar in the Greater Seattle area. Older homes are unlikely to attract more solar installations than newer homes.

3.4 *Bi-variate Analysis using the Spearman Correlation Coefficient*

Examining correlations between variables provides insight into the strength and the direction of their relationship. A correlation coefficient tells us whether the variables have a negative or positive relationship and how strong that relationship is. The value of the coefficient can range from -1 to 1, where -1 indicates a strong negative relationship and 1 indicates a strong positive relationship. In order to calculate this correlation coefficient, I will first define covariance.

Covariance is the tendency for two variables to vary together. [13] It is represented by Equation 3.1 where μ_x is the mean of X and μ_y is the mean of Y.

$$\text{Cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)(y_i - \mu_Y) \tag{3.1}$$

One of the major drawbacks to using the covariance to understand the relationship between two variables is that the units of the covariance is the product of the units of X and Y. This makes the covariance hard to interpret. One solution to this problem is to divide the covariance by the standard deviations of both variables to obtain a standard score. The Pearson Correlation Coefficient and the Spearman Correlation Coefficient implement this solution. The Pearson Correlation Coefficient is defined in Equation 3.2.

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \tag{3.2}$$

There are two types of correlation coefficients: the Pearson Correlation Coefficient and the Spearman Correlation Coefficient. The Pearson Correlation Coefficient requires the distribution of the variables be approximately normal. If the distributions of the variables are non-normal, the distributions of the the standard scores will also be non-normal. Meanwhile, the more robust Spearman Correlation Coefficient can be used for non-normal distributions. The Spearman correlation coefficient is determined by transforming all values to their percentile rank. As shown in 3.3, the main difference between the equation for the Pearson correlation coefficient and Spearman correlation coefficient is that the X and Y values are converted to rank values.

$$\rho_{r_X,r_Y} = \frac{\mathrm{Cov}(r_X, r_Y)}{\sigma_{r_X}\sigma_{r_Y}} \tag{3.3}$$

I will now use the Spearman Correlation Coefficient to determine the relationship between residential rooftop solar adoption and socioeconomic factors. In Figure 3.6, the correlation coefficient is given for increasing levels of education, house age, age, and household income. The color gradient indicates the relative strength and direction of the linear relationship. Each colored square shows the correlation coefficient between the percent of a specific group of people in each census tract and the solar per capita in every census tract. For example, the correlation coefficient between the percent of high school students in each census tract and solar per capita in each census tract is -0.27.

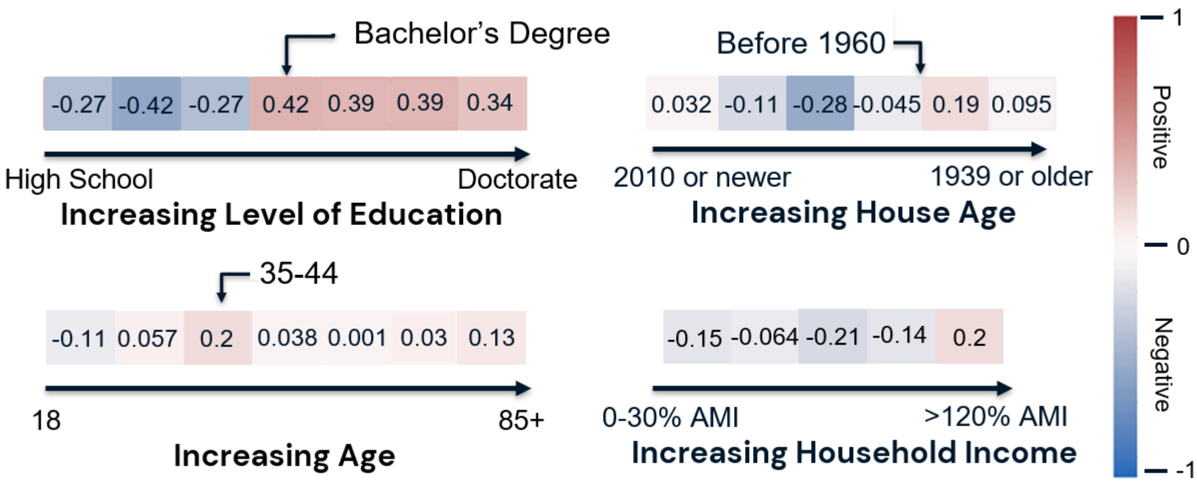

Figure 3.6: The Spearman correlation coefficients for four socioeconomic categories. Each category is broken into subgroups. For example, the increasing levels of education include the percent of the population with only a high school education to the percent of the population with a doctorate degree. The four categories explored are education, house age, age, and household income.

When looking at the correlation coefficients for increasing levels of education in Figure

3.6, there is a clear contrast between census tracts with a higher percentage of people with less than a bachelor's degree and census tracts with a higher population of people with at least a bachelor's degree. These correlations show that having a college education increases the likelihood of a person adopting solar. This suggests solar disparities in education level might exist within Washington.

When looking at the correlation coefficients for increasing levels of house age as well as age in Figure 3.6, the story is not as clear. Many of these correlations are close to zero, which indicates that there are no clear disparities visible at the census tract level. As the geospatial analysis showed, the slightly positive correlations with census tracts with more older homes and solar per capita is likely due to the higher concentration of older homes in the Greater Seattle Area. In terms of correlations between age, the strongest correlation out of all the age groups is the census tracts with more people aged 35-44. This is what I'd expect given the common age people buy houses and rooftop solar.

When looking at the correlation coefficients for increasing percentile of the area mean income (AMI) in Figure 3.6, all income brackets except for the top income group, which earns more than 120% of the AMI, show a negative correlation with the solar per capita in a census tract. This matches what the Berkeley Lab paper on income disparities found [3].

Next, Figure 3.7 shows the relationship between race and solar per capita. In this analysis, I examine the relationship between the percentage of white people, black people, Hispanic people, Asian people, and Native American people in a census tract and the solar per capita. First, I examine the relationship when including all tracts as shown in the top of Figure 3.7. I then compare this to the relationship when only including tracts with existing residential rooftop solar.

One notable difference between these two analyses is that the sign of the correlation coefficients for the percentage of people in a census tract that are white and black switch from the first analysis to the second. There is a slightly negative relationship between the percentage of people that are white and the solar per capita in a census tract when including all tracts, but a more positive relationship when excluding tracts without solar.

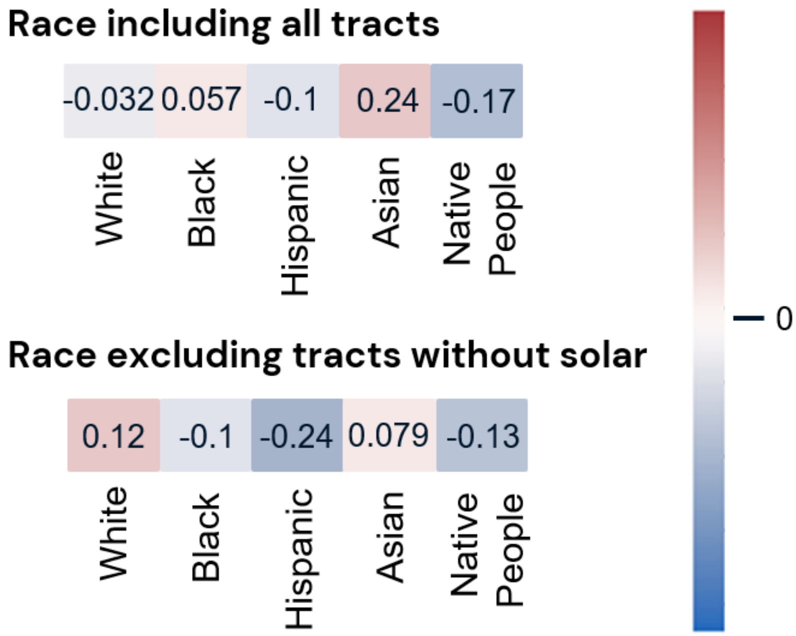

Figure 3.7: The Spearman correlation coefficients for solar per capita and five racial groups. These coefficients are calculated for when all census tracts are included (top) and when only tracts with solar are included (bottom).

One possible explanation for this discrepancy is that many rural tracts with a predominantly white population do not have solar yet. Therefore, when these rural tracts are excluded, the correlation coefficient switches from slightly negative to positive. For the same reason, the correlation coefficient for the percentage of people that are black switchers from slightly positive to negative when excluding tracts without solar.

When examining the differences between the correlation coefficients between the two analyses for the Asian racial group, there is a much stronger positive correlation when including all tracts than when excluding tracts without solar. This is likely due to a concentration of the Asian population in the Greater Seattle area where there is more solar. Therefore, the correlation coefficient is stronger when including all tracts because this analysis includes white majority tracts in rural Washington with no solar.

The opposite is true when examining the differences between the correlation coefficients between the two analyses for the Hispanic racial group. The analysis excluding tracts without solar shows a stronger negative correlation than when all tracts are included. This is likely because of the lack of solar in communities of color in more urban tracts. Including predominantly white census tracts with no solar hides this racial disparity. Meanwhile, the correlation coefficient between solar per capita and Native people does not change significantly in the two analyses. This is likely because the Native population is mostly concentrated to a few census tracts, so excluding a small portion of tracts with no solar doesn't change the coefficient value by much.

3.5 Multiple Regression Model

In the last section, I measured the strength of the linear relationship between solar per capita and individual socioeconomic factors. Spearman's correlation coefficient is limited to a bi-variate analysis where socioeconomic factors are analyzed in isolation with the solar per capita. However, as the heat-maps above showed, these socioeconomic factors are correlated with each other. For this reason, I next looked at a multiple regression model. A multiple regression model looks at all the socioeconomic factors together in comparison with the dependent variable, solar per capita. The model reveals the impact of each socioeconomic factor in comparison to each other.

A multiple regression model can be expressed as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \epsilon \tag{3.4}$$

where X_j represents the jth predictor and β_j is the coefficient quantifying the relationship between the variable and the response. As the equation shows, each of the independent variables are combined in the model to determine the value of Y.

The only unknowns in Equation 3.4 are the coefficients that are determined using the ordinary least squares method. This method chooses the coefficients that minimize the residual sum of squares (RSS) as shown below:

$$\min_{\beta} RSS = \sum_{i=1}^{n} (y_1 - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - ... - \beta_p x_{ip})^2 \tag{3.5}$$

Once β has been minimized, the optimal value β^* can be plugged back into Equation 3.4 to get a predicted value of solar per capita denoted as $\hat{y}$. The difference between the predicted value $\hat{y}$ and the actual value y is called the residual.

Furthermore, a multiple regression model must adhere to the same assumptions as a linear regression model. The assumptions are as follows:

1. Linear Relationship: There is a linear relationship between the dependent variable and independent variables.

2. Independence: The values of the residuals are independent.

3. Homoscedasticity: The variance of the residuals is constant.

4. Normality: The values of the residuals are normally distributed.

Appendix A checks the model assumptions of the multiple regression model in detail. In this section, I will only discuss the independence assumption. Collinearity ensures that the variables included in the model are linearly independent. Collinearity checks how correlated two or more of the dependent variables are with respect to each other. If a variable can be calculated from a weighted sum of the other variables, it is not independent. As a result, collinearity can hinder our ability to draw conclusions from the model's coefficients.

In order to check for collinearity, I create a correlation matrix with the socioeconomic factors of interest. The correlations between solar per capita are shown in Figure 3.8, where red corresponds to a positive correlation and blue corresponds to a negative correlation. The boxes with a white X signify a high correlation, which means both variables should not be included in the regression model to avoid collinearity. To account for variables with high collinearity, I exclude the renter population, white population, and daily solar radiation. In addition, due to the lack of significant trend in age, I also excluded age from the regression model.

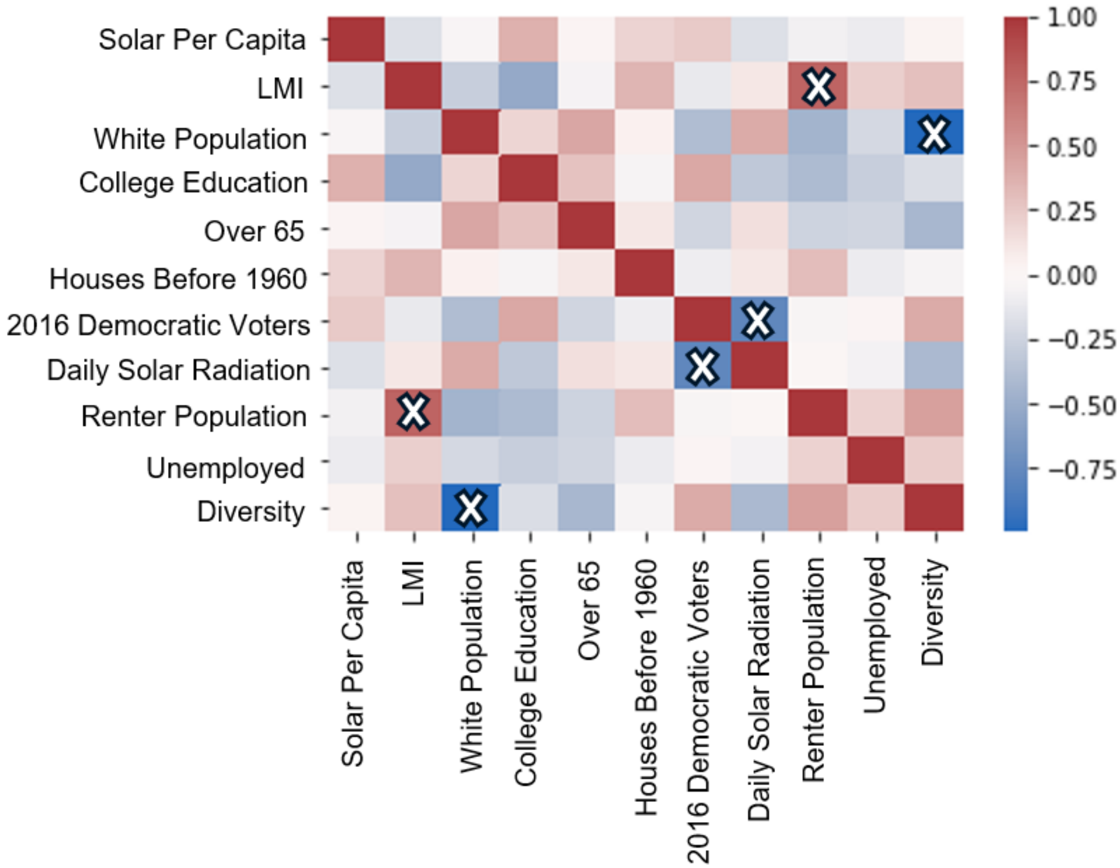

Figure 3.8: The Spearman correlation coefficient matrix for the solar per capita and the socioeconomic factors of interest for the regression model.

In order to make the dependent variable, solar per capita, better fit the model, I log transformed the dependent variable. This helps create a more normal distribution. However, the census tracts with zero solar installations have to be removed before log transforming the data since the log of zero is undefined. For this reason, the multiple regression model only considers census tracts with solar installations. Table 3.1 shows the results from the multiple regression model. For each of the independent variables included in the model, the table provides a coefficient value, a standard error values, and a p-value. The coefficient value provides the relative strength and direction of the dependent variable's influence on the solar per capita. The standard error value measures the accuracy of the coefficient value by providing the average distance the observed values stray from the regression line. Lastly, the p-value helps determine if the coefficient values are significant. The p-value tells you the probability that the null hypothesis is true; in this case, the mull hypothesis is that there is no correlation between the dependent variable and the independent variable.

There is also an overall p-value for the entire regression model, which is zero in this model. This means I can reject the null hypothesis, which says that there is not correlation between the dependent variable and the independent variables. The model also has an overall R-squared score of 0.224. This indicates that this model would not be very accurate in predicting the solar per capita in a census tract. I expected this low R-squared value since accurately predicting the amount of solar in each census tract would require more factors besides just socioeconomic factors. The purpose of this regression model is not for prediction purposes but to understand the relative influence of different socioeconomic factors on the solar per capita of a census tract. Therefore, the low R-quared score is acceptable.

Table 3.1: Multiple Regression Results

	Coefficient	Standard Error	P-value
Intercept	-8.18	0.18	0.000
LMI (%)	-0.42	0.22	0.053
Population with College Education (%)	2.55	0.28	0.000
Diversity (%)	-0.24	0.18	0.176
Houses Built Before 1960 (%)	1.08	0.11	0.000
Democratic Voting Percentage in 2016 (%)	0.30	0.25	0.224
Unemployment Percentage (%)	0.9549	0.821	0.245

To interpret the coefficients from the regression model, the coefficients need to be exponentiated to account for the logged dependent variable. The percent of the population with a college education has the strongest positive effect on the solar per capita in a census tract. A 1% increase in the percent of the population with a college education corresponds to a 1180% increase in the solar per capita. Since the average solar per capita in all the census tracts in the model is 0.0017, a 1180% increase would increase the average to 0.026. While a 1180% increase seems unrealistically large, this is mainly because the solar per capita in each census tract is very low. In addition, the main purpose of conducting a regression model is to understand the relative influence of all socioeconomic factors of interest. Therefore, it is more important to note how education plays the largest role in determining solar per capita than to note exactly how large that role is. Figure 3.9 depicts the coefficient values of the socioeconomic factors based on how they change the average statewide solar per capita (0.0017) given a 1% increase in that socioeconomic factor when all other factors are kept constant.

From the results, it is clear the percentage of the population with a college education has the strongest positive correlation with solar per capita. This indicates that there might be barriers to solar access for those with less than a college education. An unexpected result I find is that there is a positive relationship between unemployment rate and solar per capita. This most likely means that the correlation between unemployment and solar per capita is close to zero when only including tracts with non-zero solar installations. Overall, the multiple regression model reveals that disparities do exist in solar deployment in Washington state. However, the multiple regression model is limited to only analyzing census tracts with a non-zero solar per capita. In order to incorporate all census tracts, I use a Poisson regression model in the next section.

3.6 Poisson Regression Model

The most common way to model count data is with a Poisson regression model. A Poisson regression model assumes that the response variable has a Poisson distribution with a probability mass function as shown in Equation 3.6.

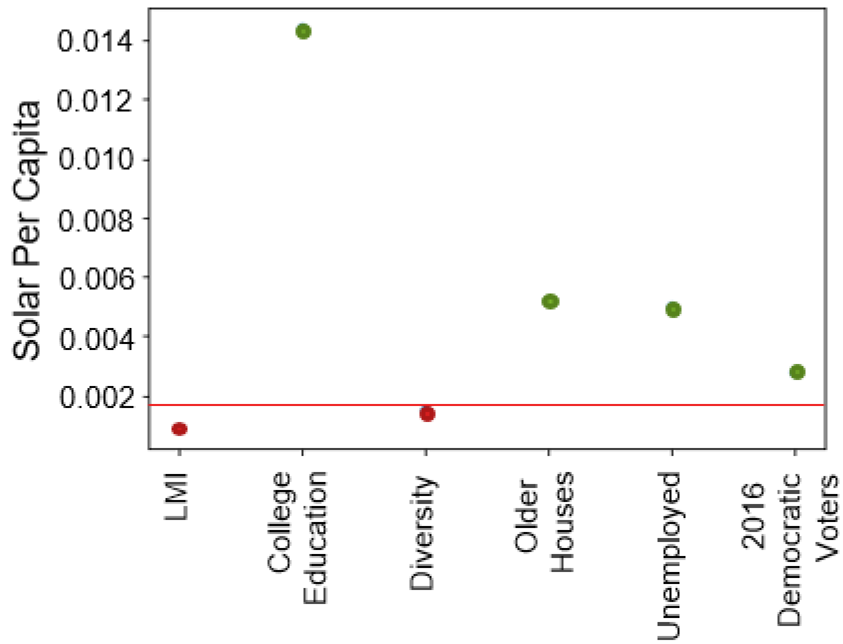

Figure 3.9: The coefficient values for the independent variables in the multiple regression model. The red line indicates the average solar per capita. If the coefficient value was zero, the corresponding dot would be on this red line.

$$\Pr(Y = y \mid \lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!} \quad (y = 0, 1, 2, \ldots) \tag{3.6}$$

where Y is a discrete random variable, y is the number of occurrences, and λ is the mean and variance of the distribution. Figure 3.10 shows the Poisson distribution for different λ values. As the value of λ decreases, the distribution looks closer to the distribution of the solar per capita in each tract in Washington, as shown in the right graph in Figure 3.10.

The assumptions for the Poisson Regression model are as follows:

1. Poisson Response: The dependent variable is a count per unit of time or space and has a Poisson distribution.

2. Independence: The observations must be independent from each other.

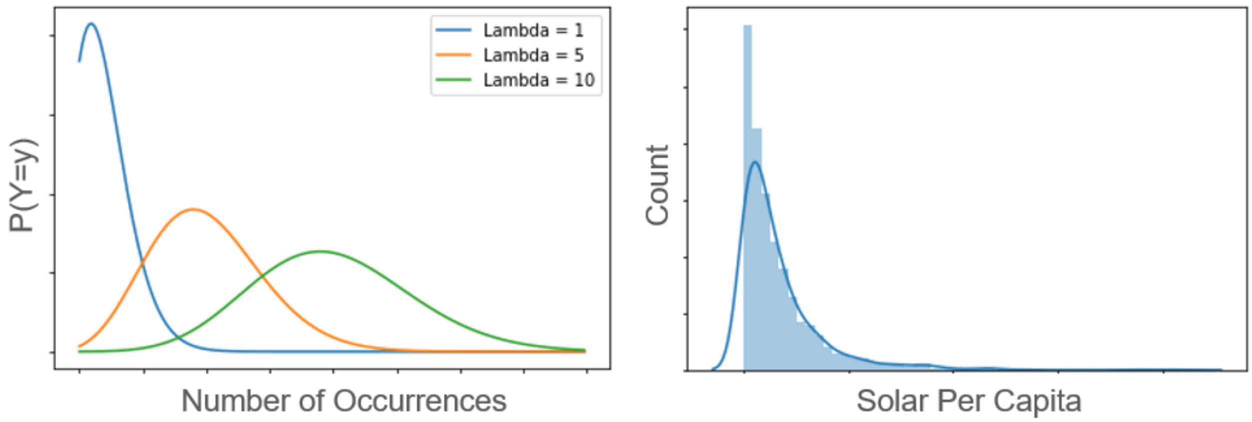

Figure 3.10: (Left) The Poisson distribution for different values of Lambda, which is the mean of the distribution. (Right) The distribution of solar per capita for all census tracts in Washington.

3. The mean of the dependent variable is equal to its variance.

4. Linearity: The log of the mean rate must be a linear function of x.

It is easy to check assumptions 1 and 3. From Figure 3.10, a Poisson random variable with a low λ value resembles the solar per capita variable. In addition, the solar per capita variable is a solar installation count per unit of space, so it is appropriate to assume it has a Poisson distribution. When checking assumption 3, I found the variance (0.002) of this variable is almost equal to the mean (0.0017). Assumption 2 and 4 are also required for the multiple regression model, so these assumptions are checked in Appendix A.

The Poisson regression model can be expressed as Equation 3.7

$$\lambda = t\, exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p) \tag{3.7}$$

where λ is the average number of solar installations, t is the exposure, the β's are the coefficients, and the x's are the socioeconomic factors. Similar to the multiple regression model, the only unknowns in this equation are the β's. Unlike the multiple regression model that solves for these coefficients using the ordinary least squares method, the Poisson regression model uses the Newton-Raphson method.

Another difference between the multiple regression model and the Poisson regression model is the exposure term t. The exposure term allows us to model rate data in a Poisson regression model. One of the requirements for the Poisson regression model is that the dependent variable is count data. However, in this case, I am modeling the amount of solar per capita, which is a rate. As shown in Equations 3.8 - 3.11, if I transform the rate into a count variable, I get an extra term t that is multiplied with the exponential.

$$log(\frac{X}{t}) = \beta_0 + \sum_i \beta_i X_i \tag{3.8}$$

$$log(X) - log(t) = \beta_0 + \sum_i \beta_i X_i \tag{3.9}$$

$$X = \exp(log(t) + \beta_0 + \sum_i \beta_i X_i) \tag{3.10}$$

$$X = t\exp(\beta_0 + \sum_i \beta_i X_i) \tag{3.11}$$

The results of the Poisson regression model are shown in Table 3.2. In order to interpret the results, the coefficient values have to be exponentiated, just like the multiple regression coefficients. Figure 3.11 shows the effects of a percent increase in each socioeconomic factor on the solar per capita when keeping all other socioeconomic factor values constant. The standard error and P-value columns in Table 3.2 can be interpreted the same way as the multiple regression model results. For a Poisson regression model where the dependent variable is discrete, it is not possible to obtain a standard R-sqaured score. McFadden's Psuedo R-squared score provides an alternate way to interpret the fit of the model. The McFadden's Psuedo R-sqaured score for this Poisson regression model is 0.18. The Pseudo-Rsquared score is calculated by comparing the log likelihood of the fitted model versus the log likelihood of the null model, which includes no independent variables in the model. While this value might seem low for a normal R-squared score, as stated in [14], R-squared values of 0.2 to 0.4 represent an excellent fit.

Table 3.2: Poisson Regression Results

	Coefficient	Standard Error	P-value
Intercept	-7.72	0.07	0.000
LMI (%)	-0.66	0.09	0.000
Population with College Education (%)	2.38	0.11	0.000
Diversity (%)	0.1	0.07	0.166
Houses Built Before 1960 (%)	1.20	0.04	0.000
Democratic Voting Percentage in 2016 (%)	0.45	0.10	0.000
Unemployment Percentage (%)	-1.42	0.38	0.000

Looking at the results, the percentage of the population with a college education has the strongest influence on the solar per capita on a tract. I saw this same result in the multiple regression model. The two main differences I find between the Poisson regression model and the multiple regression model is with diversity and unemployment rate.. One main reason these differences exist is because I am including all tracts in the Poisson regression model. Instead of the relationship between between diversity and solar per capita being slightly negative as shown in the multiple regression results, the relationship is slightly positive. This discrepancy is due to a higher white population in the rural tracts with no solar. In terms of the unemployment rate, the results from the Poisson regression model matches our expectations, unlike what the multiple regression results showed. Since the coefficient for unemployment is only slightly negative, it appears that unemployment is not strongly correlated with the solar per capita in a census tract. Overall, including census tracts with no solar slightly changes our interpretation of existing solar disparities. For this reason, the last regression model I will explore will look at these differences in socioeconomic influences between tracts with solar and tracts without solar.

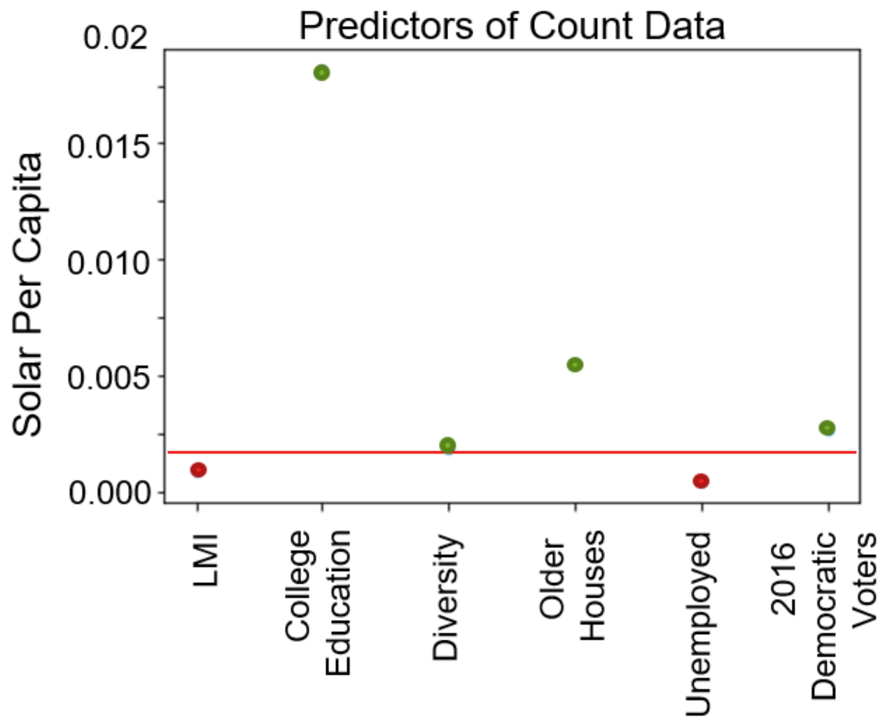

Figure 3.11: The coefficient values for the predictors of count data. The red line indicates the average solar per capita. If the coefficient value was zero, the corresponding dot would be on this red line.

3.7 Zero-Inflated Poisson Regression Model

There are cases when an excess of zeros make the Poisson distribution less ideal. In Washington, many census-tracts have no solar installations, which leads to an excess of zeros in the data. One way to account for these excess zeros is using a zero-inflated Poisson (ZIP) regression model, which assumes there is a separate process leading to excess zeros than the count data. This means that some independent variables might help explain the inflated number of zeros while other independent variables might help explain the count data. It can also mean that the same independent variable has a different coefficient value for the two processes. More specifically, analyzing predictors of excess zeros allows me to identify socioeconomic differences between tracts with solar and tract without solar. Analyzing predictors of count data allows me to identify socioeconomic differences between tracts with more solar and tracts with less solar.

The probability distribution for the zero-inflated Poisson regression model is defined in Equation 3.12, where μ_i is the mean and variance of the Poisson distribution. This probability distribution assumes there are two cases. In case 1, the solar per capita is always zero with probability π. In case 2, the solar per capita follows a Poisson distribution with probability $(1-\pi)$. Therefore, as shown in Equation 3.12, the probability that the solar per capita is zero equals the probability π from case 1 plus the probability of a zero for a Poisson random variable times the probability $(1-\pi)$ for case 2. The probability a Poisson random variable is zero equals $\exp(\mu_i)$ since 0! equals 1 and μ_i^0 equals 1. Next I look at the probability that the solar per capita is greater than zero in Equation 3.12. When $j > 0$, the probability equals the probability distribution function of a Poisson random variables times the probability of case 2, which is $(1-\pi)$.

$$\Pr\left(y_i = j\right) = \begin{cases} \pi_i + (1-\pi_i)\exp\left(-\mu_i\right) & \text{if } j = 0 \\ (1-\pi_i)\,\frac{\mu_i^{y_i}\exp(-\mu_i)}{y_i!} & \text{if } j > 0 \end{cases} \tag{3.12}$$

The probability π is the logistic link function as defined in Equation 3.13.

$$\pi_i = \frac{\lambda_i}{1+\lambda_i} \tag{3.13}$$

where

$$\lambda_i = \exp\left(\ln\left(t_i\right) + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \cdots + \gamma_m z_{mi}\right) \tag{3.14}$$

In Equation 3.14, t_i is the exposure time, which is the total population in each census tract. The γ's are the coefficients the model solves for and z's are the socioeconomic predictors of the excess zeros.

The results of the zero-inflated Poisson regression model are shown in Table 3.3. The McFadden Psuedo R-squared score for this model is 0.17. This is very close to the Psuedo

R-squared score for the Poisson regression model. Figure 3.12 visualizes these results to understand the relative influence of the socioeconomic factors under consideration.

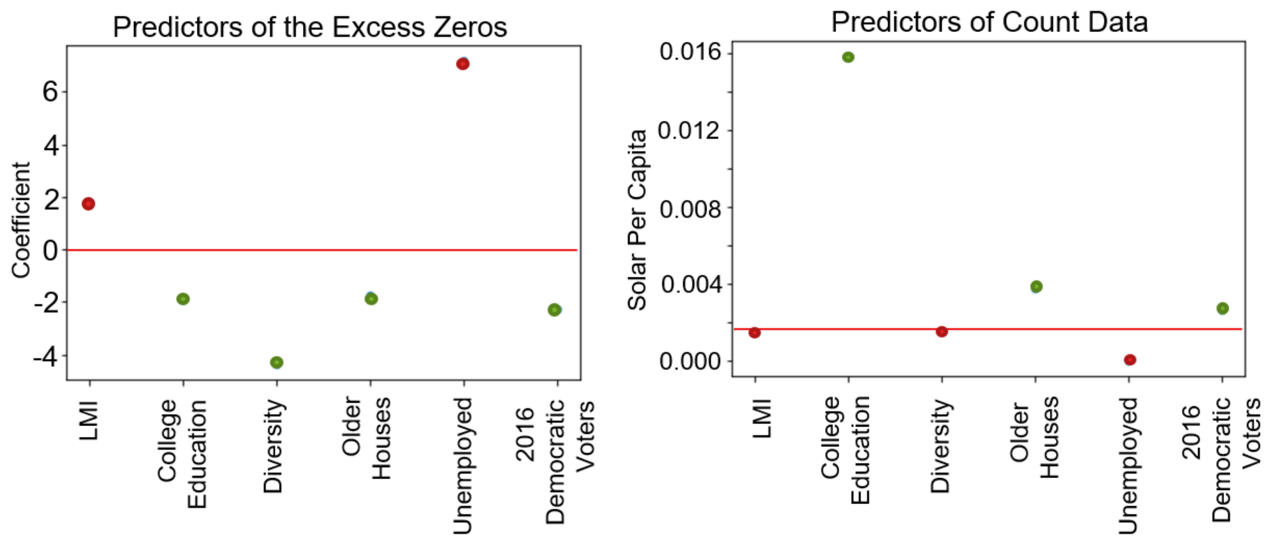

Figure 3.12: A. (left) Predictors of the Excess Zeros, B. (right) Predictors of Count Data. The red line on the right graph indicates the average solar per capita. If the coefficient value was zero, the corresponding dot would be on this red line.

All of the socioeconomic factors in Table 3.3 that start with “Inflated” correspond to the predictors of excess zeros, otherwise known as the logit coefficients. These are shown on the left of Figure 3.12 labeled “Predictors of Excess Zeros.” Interpreting the logit coefficients can be counter-intuitive, so the graphs in Figure 3.12 depict the influence of each factor visually. At a high level, the dots colored green on both graphs corresponds to a positive relationship between solar per capita and that factor; in contrast, the dots colored red on both graphs correspond to a negative relationship.

The logit coefficients measure how much the log odds of being an excessive zero would change. This means that a positive logit coefficient would increase the log odds, thereby making it more likely that a census tract has zero solar. Meanwhile, a negative logit coefficient would make it less likely that a census tract has zero solar. This is the opposite of the

Table 3.3: Zero-Inflated Poisson Regression Results

	Coefficient	Standard Error	P-value
Inflated Population Over 65 (%)	7.39	1.60	0.000
Inflated LMI (%)	2.15	0.74	0.004
Inflated Population with College Education (%)	-3.85	1.04	0.000
Inflated Diversity (%)	-3.25	0.71	0.000
Inflated Houses Built Before 1960 (%)	-1.51	0.50	0.002
Inflated Democratic Voting Percentage in 2016 (%)	-1.00	0.91	0.273
Inflated Unemployment Percentage (%)	-0.003	0.001	0.003
Intercept	0.42	0.07	0.000
Population Over 65 (%)	2.02	0.18	0.000
LMI (%)	-1.34	0.10	0.000
Population with College Education (%)	1.92	0.12	0.000
Diversity (%)	0.02	0.07	0.241
Houses Built Before 1960 (%)	0.94	0.04	0.000
Democratic Voting Percentage in 2016 (%)	0.86	0.10	0.000
Unemployment Percentage (%)	0.002	0.00	0.000

coefficients for count data where a positive value indicates an increase in the solar per capita while a negative value indicates a decrease in the solar per capita.

One potential discrepancy between the left and right graph is the percentage of diversity. As I analyzed above, the correlations between race and solar per capita differ depending on which tracts are included in the analysis. In this case, a increase in the percentage of diversity would decrease the chances that a census tract has zero solar. However, for census tracts with a nonzero solar per capita, an increase in the percentage of diversity would decrease the amount of solar in a census tract. This discrepancy is actually in line with the differences in how race affects solar per capita in urban and rural tracts, as I will examine in the next section.

For the remaining five socioeconomic factors, the sign of the predictors of the excess zeros and the predictors of count data are the same. Similar to the results from the Poisson regression model, education has the strongest positive correlation with solar per capita. On the graph for the predictors of excess zeros, I see that high unemployment rates can strongly increase the likelihood that a census tract does not have any solar in the census tract. The fact that census tracts with no solar likely have greater unemployment rates was not clear from the results of the previous two regression models. It is only by including the six socioeconomic factors in both processes that regional differences in socioeconomic influences on solar adoption become more apparent.

Since rural tracts are more likely to be the tracts with no solar and urban tracts have the most solar, I decided to explore the differences in solar disparities in these two regions more explicitly in the next section. The purpose of this analysis is to identify if any statewide disparities are different when only examining urban and rural tracts.

3.8 Regional Socioeconomic Disparities in Urban and Rural Census Tracts in Washington

In this section, I compare the socioeconomic disparities between rural census tracts and urban census tracts. I define a rural census tract as having a population density of less than 500 people per square mile. I define a urban census tract as having a population density of more than 2,500 people per square mile. Since a large portion of solar installations are in more densely populated areas such as Seattle, I wanted to rerun the same analysis while looking at only rural and urban census tracts.

From Figure 3.13, I can compare the Spearman correlation coefficients for the six socioeconomic factors of interest for all tracts, urban tracts, and rural tracts. When looking only at rural tracts, it is clear that many of the correlation coefficients are close to zero, indicating no correlation between these socioeconomic factors and solar per capita. The only non-zero correlation with solar per capita in rural tracts is with diversity. It is worth exploring why this positive correlation exists in a future study since I did not find an explanation for this. Meanwhile, the correlations in urban tracts are similar to the correlations in the statewide analysis. Due to the concentration of solar in urban areas like Seattle, it makes sense that these correlations are similar. Again, the one socioeconomic factor of interest when comparing statewide results to urban results is diversity. In this case, the cause for a stronger negative correlation when looking at urban tracts is that many of these tracts are overall more diverse than the rest of the state. Therefore, when focusing only on urban tracts, it is those tracts with a higher population of white people that see a greater solar per capita.

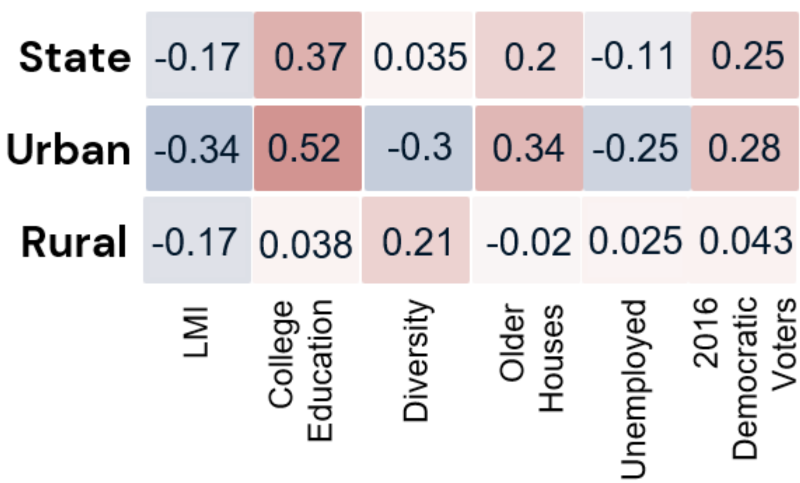

Figure 3.13: The Spearman correlation coefficients for all Washington census tracts (top), only urban tracts (middle), and only rural tracts (bottom). The correlation coefficients are given for the six socioeconomic factors of interest.

Figure 3.14 shows the results of the predictors of count data for a zero-inflated Poisson regression model on only urban tracts. As mentioned earlier, for rural tracts, the correlations between solar per capita and the six socioeconomic factors of interest are all close to zero except for diversity. Therefore, running a zero-inflated Poisson regression model for only rural tracts did not produce significant results. Comparing the results of the regression model for urban tracts versus the statewide results, I see that the signs of the socioeconomic influences do not change between the two results, but the strength of their influences do. Specifically, the percent of LMI households, percentage of diversity, and percent of unemployed people have a negative relationship with solar per capita. Meanwhile, the percent of college educated people, the percent of older homes, and the percent of democratic voters in the 2016 election have a positive relationship with solar per capita. However, when only considering urban tracts, the percent of democratic voters in the 2016 election has a much greater influence than it did statewide. This might indicate that politics influence solar adopters in urban areas more than the rest of the state.

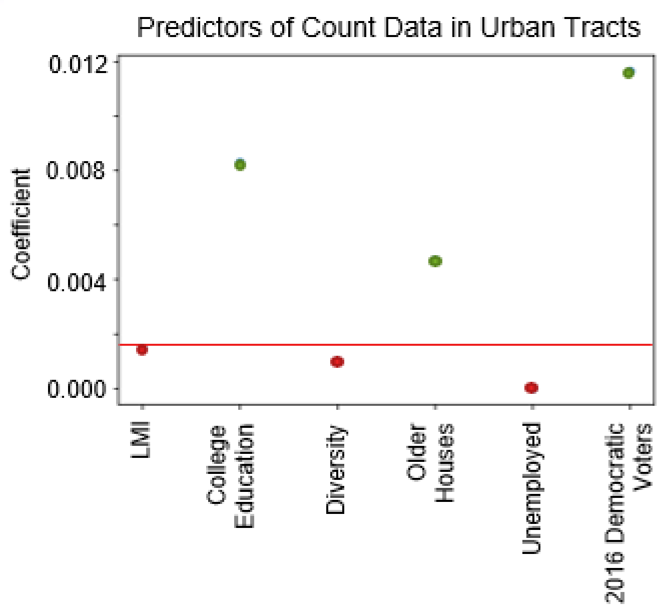

Figure 3.14: The coefficient values for the predictors of count data in urban tracts. The red line indicates the average solar per capita. If the coefficient value was zero, the corresponding dot would be on this red line.

As my results show, solar disparities exist for a range of socioeconomic factors and these disparities differ regionally. Therefore, solar equity policies, that often only address income disparities, should factor in a range of socioeconomic disparities. One potential method to achieve this would be to identify disadvantaged communities that suffer from a range of socioeconomic barriers to solar. In this next section, I will use the Washington Environmental Health Disparities Map to analyze rooftop solar adoption in disadvantaged communities.

3.9 Solar Adoption in Disadvantaged Communities

The Washington Environmental Health Disparities Map was created to identify census tracts within the state that face a disproportionately higher risk from environmental hazards and socioeconomic factors. This tool ranks the cumulative environmental health risks by census tracts to identify disadvantaged communities. The tool is modeled after CalEnviroScreen, which tracks environmental health impacts in California. In California, the top 25% of the tracts are identified as disadvantaged communities. In Washington, there is no cutoff for a disadvantaged community; however, census tracts ranked 9 or 10 are the top 20% disadvantaged census tracts in the state.

Each census tract is given a final composite score (FCS) that is used to rank each tract. The final composite score (FCS) can be calculated as the pollution burden score (PBS) times the population characteristics score (PCS) as shown in Equation 3.15. The pollution burden score (PBS) accounts for all the indicators related to environmental health risk factors. As shown in Equation 3.16, the pollution burden score is the weighted average of environmental exposures (Exp) and environmental effects (Eff). Environmental exposures are indicators such as ozone or traffic density that measure exposures to certain pollutants. Meanwhile, environmental effects are indicators such as proximity to hazardous waste generators that measure potential risk of environmental hazards in a community. The population characteristic score (PCS) is the average of sensitive populations (SP) and socioeconomic factors (SF) as shown in Equation 3.17. Sensitive populations (SP) are those that are more vulnerable to environmental hazards. Socioeconomic factors (SF) include housing burden, income, race, etc.

$$FCS = PBS \times PCS \tag{3.15}$$

$$PBS = \frac{Exp + 0.5(Eff)}{2} \tag{3.16}$$

$$PCS = \frac{SP + SF}{2} \tag{3.17}$$

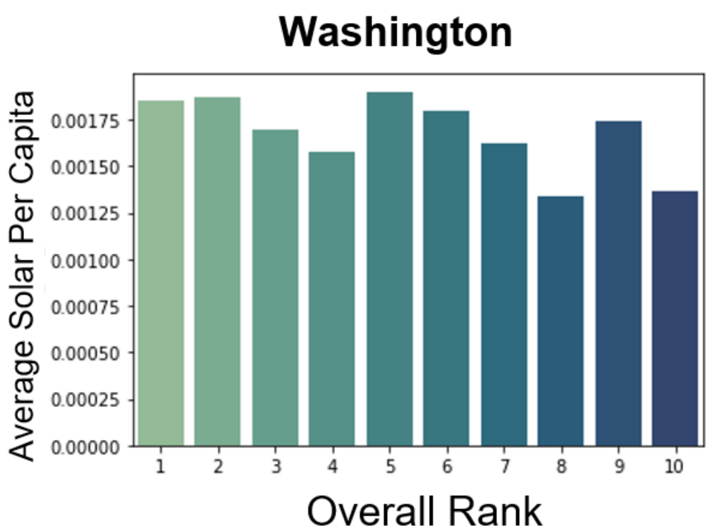

Figure 3.15: A bar graph showing the average solar per capita in each rank of the Washington Environmental Health Disparities Map

Once the final composite score has been determined for all tracts, the tracts are ranked from 1-10, where 1 is least disadvantaged and 10 is more disadvantaged. The Washington Environmental Health Disparities Map also ranks each of the census tracts based on the four categories that make up the final composite score.

From these rankings, I can analyze any trends for the average amount of residential rooftop solar installations in each rank. Figure 3.15 shows the average solar per capita in each rank in Washington. Compared to the clear negative trend in California, as shown in Figure 3.16, there is only a slight decrease in the average solar per capita in more disadvantaged communities in Washington.

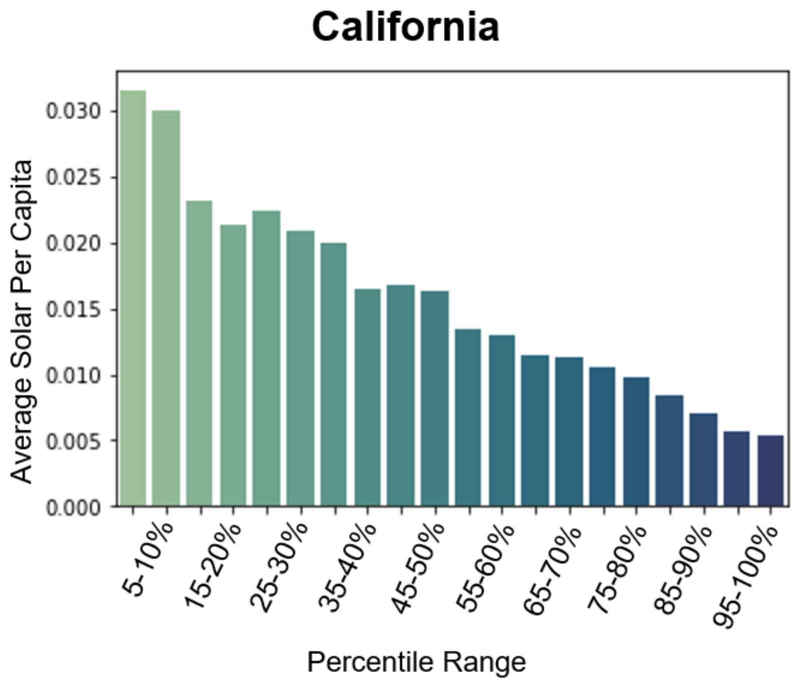

Figure 3.16: A bar graph showing the average solar per capita in each percentile rank of CalEnviroscreen, the environmental disparities mapping tool in California. A clear negative trend is visible, indicating that disadvantaged communities have less solar on average

One reason the distribution of solar inside and outside disadvantaged communities might look more equitable than in California is due to differences in trends amongst the four categories that make up the final composite score. After mapping the solar per capita with the overall rank, I graphed the average solar per capita versus the rank for the four categories that are used to determine the overall rank. As shown in Figure 3.17, there average solar per capita increases as the rank increases for the Environmental Effects and Environmental Exposures graphs. Meanwhile, the average solar per capita decreases significantly as the rank increases for Sensitive Populations and Socioeconomic Factors. Many of the census tracts with a high Environmental Effects and Environmental Exposure rank are located in highly populated areas. Looking at the maps for Environmental Effects and Environmental Exposures in King County in Figure 3.18, many of the tracts with a higher population density to the west of the county have a high rank. Meanwhile, census tracts ranking high for Sensitive Populations and Socioeconomic Factors are to the south of the Seattle area that I previously identified as a marginalized community in King County.

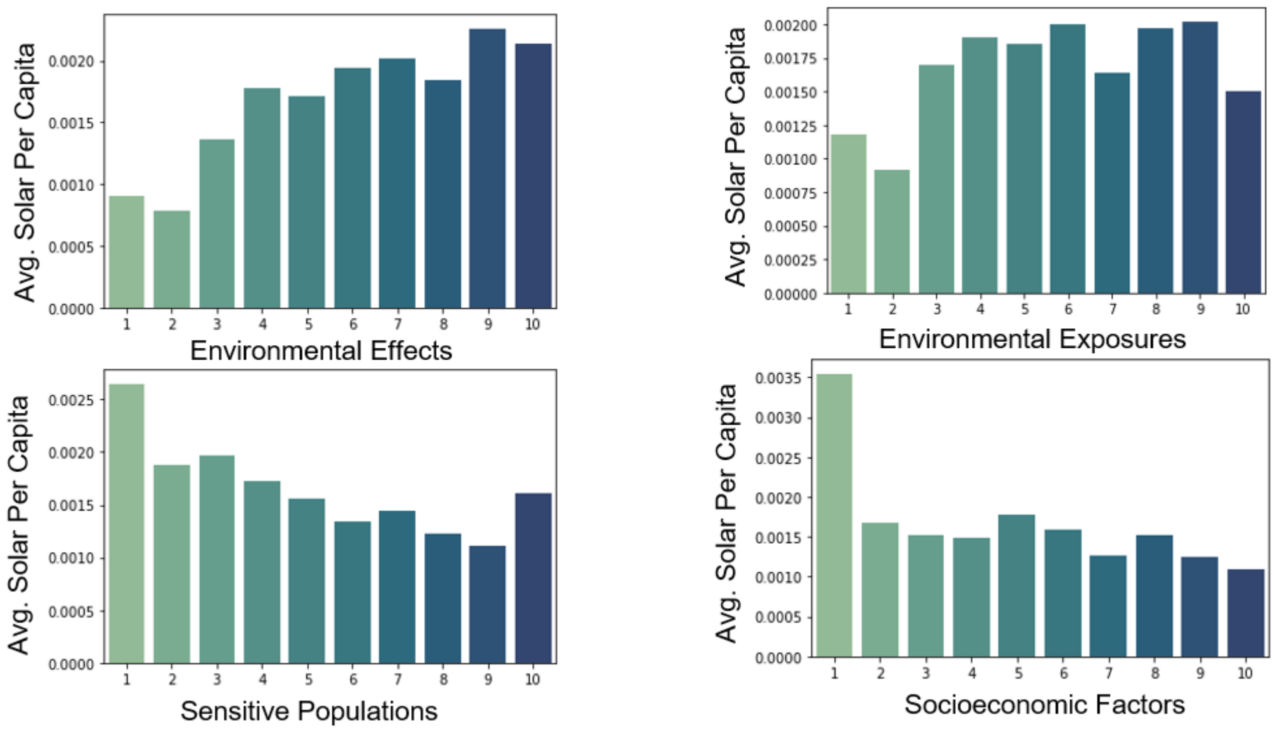

Figure 3.17: Four bar graphs showing the average solar per capita in the census tracts ranked from 1 to 10 for the following four categories: Environmental Effects (top left), Environmental Exposures (top right), Sensitive Populations (bottom left), and Socioeconomic Factors (bottom right)

This means that identifying appropriate disadvantaged communities might also vary regionally in Washington. For example, if King County wanted to ensure solar equity in the southwest region of the county, it might be more effective to identify disadvantaged communities based on the Socioeconomic Factors category of the Washington Environmental Health Disparities Map.

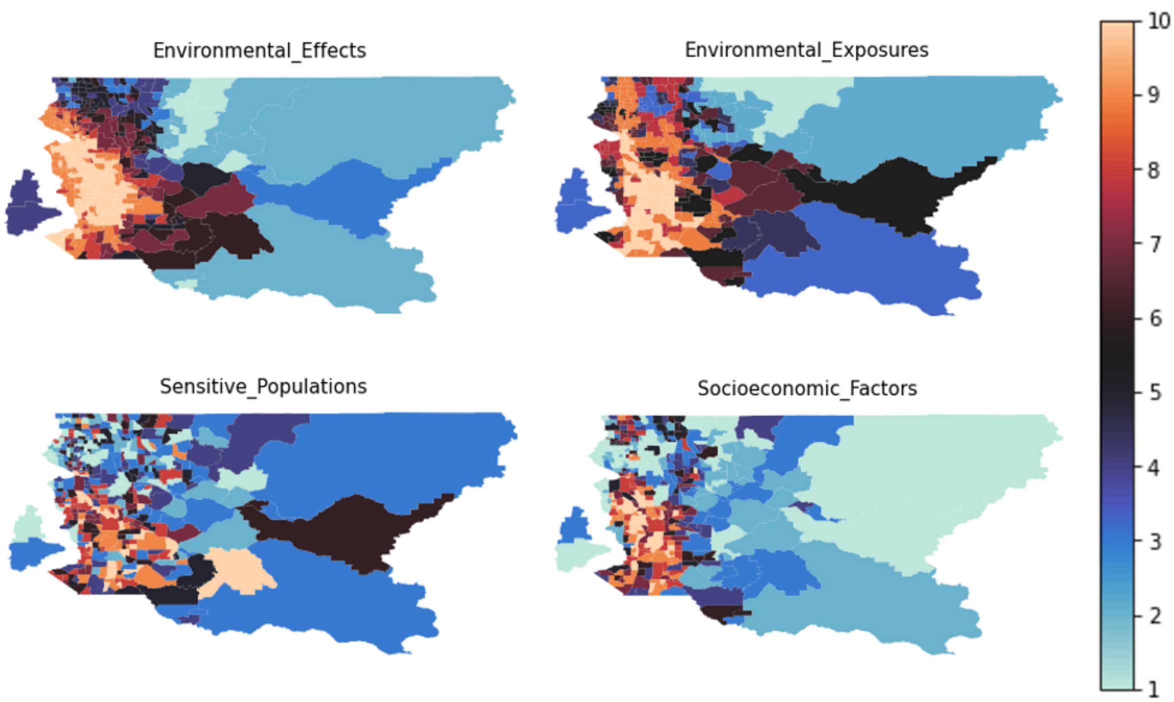

Figure 3.18: Four census tract level maps of King County. Each map shows the rank that each census tract received in four categories: Environmental Effects (top left), Environmental Exposures (top right), Sensitive Populations (bottom left), and Socioeconomic Factors (bottom right)

Chapter 4

DISCUSSION OF SOLUTIONS TO IMPROVE SOLAR EQUITY

From the previous sections, I identified existing disparities in solar adoption in Washington state as well as how those disparities vary regionally. In this section, I will discuss methods to improve solar equity based on existing literature. As mentioned earlier, much of the discussion on solar equity so far has been focused on low-to-moderate income (LMI) communities. With income disparities apparent at a national level as well as in Washington, policy intervention is likely needed to reduce current income inequities. Therefore, I will first discuss effect solar policies to increase solar adoption in LMI communities. Next, I will discuss the limitations of solar equity policies that focus solely on income, especially in disadvantaged communities. Lastly, I will provide overall recommendations to the Washington State Department of Commerce based on my analysis and the findings in this section.

4.1 Policy Incentives Targeting Low-to-Moderate Income Communities

In [5], Berkeley Lab examined five different policies to determine their effectiveness in increasing income equity in rooftop solar adoption. The five policies considered are: financial incentives, LMI incentives, solar leasing, property-assessed clean energy financing (PACE), and the Solarize campaign. Using a unifying descriptive model, the study tests the effects of these interventions on the income bias of solar adopters. The adoption income bias was calculated to be the difference between adopter incomes and county median incomes. The primary data used in this study was the Lawrence Berkeley National Laboratory's Tracking the Sun dataset which includes data for more than 70% of all residential PV systems installed in the United States. The study found that LMI incentives were the most effective in increasing income equity, followed by leasing and then PACE. The Solarize campaign and traditional financial incentives did not increase income equity in this study. As shown in Figure 4.1, implementing a LMI incentive resulted in an increase in LMI solar installations in low-income zipcodes and a decrease in LMI solar installations in high-income zipcodes. This means that LMI incentives were effective in shifting solar deployment towards low-income communities.

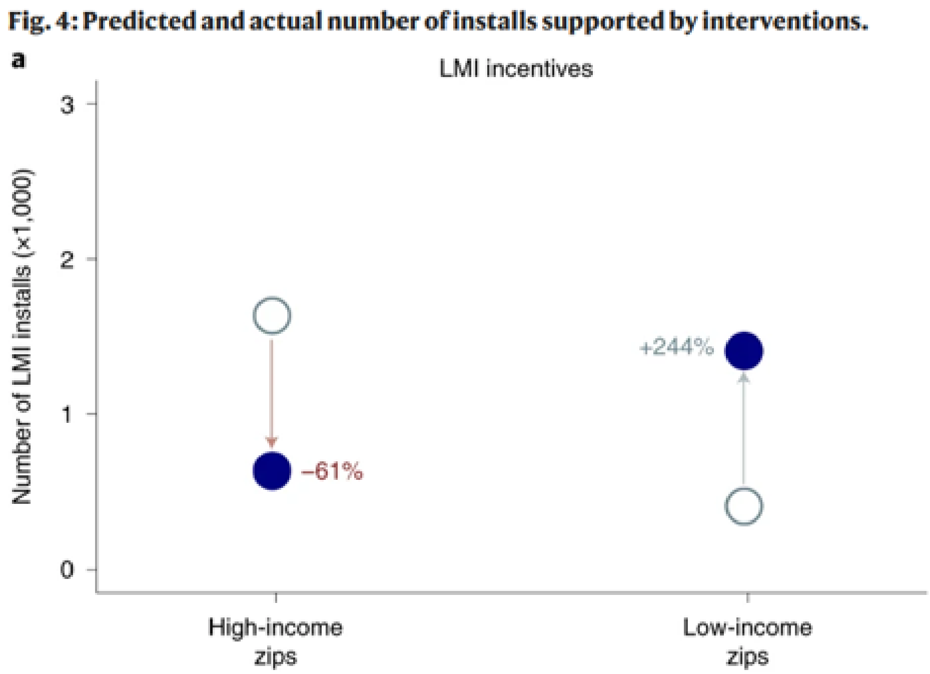

Figure 4.1: Image from [5]. Graph shows predicted (open circle) and actual (solid circle) solar installations in LMI households in high-income communities and low-income communities.

The LMI policies modeled in this study were based off of existing programs in California, Connecticut, and New York. The LMI incentive was based off of a program in California called the Single-Family Affordable Solar Housing (SASH) program. The SASH program provides a $3 per watt capacity-based incentive to LMI households that qualify. While this program has successfully installed close to 10,000 PV systems, inequities between installation in disadvantaged communities and non-disadvantaged communities still exist. As mentioned in [15], disadvantaged communities are the top 25% of census tracts that face exposure to pollution. By evaluating the SASH program

4.2 Evaluating Effectiveness of LMI Incentives

Lukanov and Kreiger tracked the number of solar installations in DAC and non-DAC communities from 1998 to 2017 and found that a significantly greater amount of solar capacity was installed outside of DACs than inside [6]. This paper also opened a discussion on the effectiveness of the Single-family Affordable Solar Housing (SASH) program in California. They looked at the difference in solar installations between disadvantaged communities (DAC) and non-disadvantaged communities (non-DAC). Figure 4.2 shows the annual rate of solar additions in DACs vs non-DACs in California as their relative weight of the total amount of kWs installed per capita. An equal rate of deployment in DACs and non-DACs would mean that both points are at the 50% mark. However, in 2017 we see that the deployment rate in DACs was half as much as in non-DACs. From this yearly analysis, the limitations of the SASH program can be seen.

One way California is attempting to increase the effectiveness of their LMI incentive is through their new DAC-SASH program. In 2018, the DAC-SASH program was created for low-income households in disadvantaged communities to receive no-cost solar installations. This program is more specific than the earlier implemented SASH program in how it requires a household to be located in the top 25% of disadvantaged census tracts in California. By targeting disadvantaged communities, LMI incentives move beyond only considering income disparities. Disadvantaged communities are not only areas that are exposed to higher levels of pollution; they are areas that also face a higher degree of socioeconomic disparities.

4.2.1 LMI Solar Program Evaluation

It is only through effective program evaluation that California was able to identify limitations in the SASH program and implement a better program that targets DACs specifically. Lawrence Berkeley National Lab recently released a report assessing the evaluation practices of low-to-moderate income solar programs [16]. The report assess 41 different LMI solar

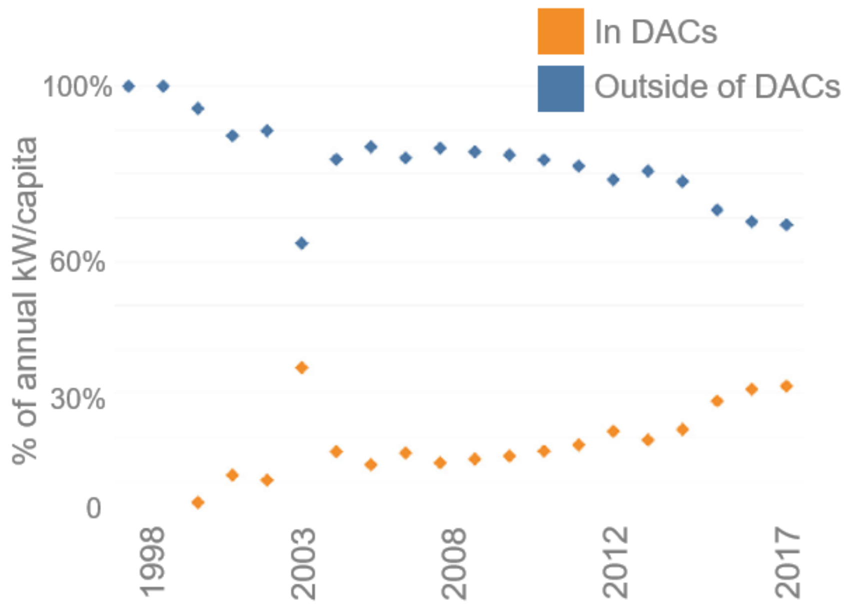

Figure 4.2: Image from [6]. The graph shows the relative weight of annual residential solar deployments per capita as percent of the total installed kW per capita for that year. An equal rate of deployment would be at the 50% line.

programs. Out of all the programs, the report found that the California's single-family and multi-family (SASH and MASH) programs were notable examples of program evaluation.

The report also identified six best practices of program evaluation.

- Continuous evaluation: This means ongoing feedback and data collection to improve the program.

- Applying best practices from other fields: This means consulting Energy Efficiency programs with well defined program evaluations.

- Collecting multi-dimensional data: This means collecting data about participants to identify correlations between adopters and socioeconomic disparities.

- Standardizing LMI solar program evaluation: This means standardizing the evaluation metrics so that they can be implemented at a smaller scale.

- Integrating non-energy benefits: This means including improved wealth, health, and safety into the evaluation metrics.

- Process evaluations: This means examining program efficacy and stakeholder feedback. This is different from impact evaluations that measures if a goal was achieved.

A report by the Clean Energy States Alliance [17] emphasizes these same points on how policymakers can expand access to solar in LMI communities. The report additionally stresses that different submarkets need different strategies. The efficacy of the DAC-SASH program in California might differ in Washington. In addition, it is unlikely that a single program will eliminate all barrier to solar access. What is important is that Washington diligently evaluates the effectiveness of any LMI program they propose and track data throughout the process.

4.3 Recommendations for the Washington Department of Commerce

Currently, Washington state does not have any solar incentives like the SASH or DAC-SASH program in California that targets LMI households. A similar SASH program could help address existing solar income disparities in Washington state while a similar DAC-SASH program could help benefit disadvantaged communities who are often marginalized in multiple ways. Based on my analysis, there are solar disparities that I identified that need to be examined further to determine what type of solar equity policy will be the most effective in Washington. In addition, no policy should be implemented before consulting community leaders who are most attuned to the barrier to solar access that are unique to each community.

In this section, I propose the following recommendations to address solar equity based on the findings from my study:

- Explore why rural census tracts have a positive correlation between the percent of people of color in the tract and the solar per capita. As my analysis found, when only examining rural tracts for socioeconomic correlations to the solar per capita in a tract, the strongest correlation was a positive correlation with the amount of diversity in a census tract. This is the opposite of what I examined in urban tracts.

- Explore why there is a strong correlation between census tracts with a higher percentage of college educated people and solar per capita. As my results from the zero-inflated Poisson regression model showed, the strongest socioeconomic influence on the solar per capita in a census tract was the percent of the population with a college degree. Is the way solar is being marketed accessible to those with less education?

- Identify disadvantaged communities based on a curated list of socioeconomic factors that should receive equitable attention in any solar policy making. As my analysis found, disparities inside and outside DAC communities are not as distinct in Washington as in California. However, if disadvantaged communities in Washington are defined by the Sensitive Populations and Socioeconomic Factors categories of the Washington Environmental Health Disparities map, a clear solar disparity in highly ranked census tracts is visible.

- Continually track and monitor changes in solar disparities in Washington state. My analysis uses the most recent census-level count of solar installation available. As these open-source datasets get updated, changes in solar disparities should be tracked.

- If possible, collect socioeconomic information of solar adopters to gain more granular data on solar disparities than a census-tract analysis can provide. As Appendix B states, there are limitations in identifying solar disparities on the tract level. Collecting socioeconomic information from new solar adopters would allow for a more accurate understanding of existing barriers to solar access.

- Continue building partnerships between state energy agencies and community-based organizations. As the Washington 2021 State Energy Strategy recognizes, an environmental justice framework that uplifts the voices of community members is essential for an equitable deployment of residential rooftop solar.

BIBLIOGRAPHY

[1] T. G. Reames, "Distributional disparities in residential rooftop solar potential and penetration in four cities in the united states," *Energy Research Social Science*, vol. 69, p. 101612, 2020.

[2] Washington State Department of Commerce, "Fuel mix disclosure," Available at https://www.commerce.wa.gov/growing-the-economy/energy/fuel-mix-disclosure (2019).

[3] G. Barbose, S. Forrester, N. Darghouth, and B. Hoen , "Income trends among u.s. residential rooftop solar adopters."

[4] D. A. Sunter, S. Castellanos, and D. M. Kammen, "Disparities in rooftop photovoltaics deployment in the united states by race and ethnicity," *Nature Sustainability*, vol. 2, no. 1, p. 71–76, 2019.

[5] E. O'Shaughnessy, G. Barbose, R. Wiser, S. Forrester, and N. Darghouth, "The impact of policies and business models on income equity in rooftop solar adoption," *Nature Energy*, vol. 6, no. 1, p. 84–91, 2020.

[6] B. R. Lukanov and E. M. Krieger, "Distributed solar and environmental justice: Exploring the demographic and socio-economic trends of residential pv adoption in california," *Energy Policy*, vol. 134, p. 110935, 2019.

[7] C. L. Kwan, "Influence of local environmental, social, economic and political variables on the spatial distribution of residential solar pv arrays across the united states," *Energy Policy*, vol. 47, p. 332–344, 2012.

[8] J. Yu, Z. Wang, A. Majumdar, and R. Rajagopal, "Deepsolar: A machine learning framework to efficiently construct a solar deployment database in the united states," *Joule*, vol. 2, no. 12, p. 2605–2617, 2018.

[9] Google, "Project sunroof," Availble at https://www.google.com/get/sunroof.

[10] G. Barbose, N. Darghouth, S. Elmallah, S. Forrester, K. K. S. H., D. Millstein, J. Rand, W. Cotton, S. Sherwood, E. Oshaughnessy, and et al., "Tracking the sun: Pricing and design trends for distributed photovoltaic systems in the united states - 2019 edition," 2019.

[11] B. O. Sigrin and M. E. Mooney, “Rooftop solar technical potential for low-to-moderate income households in the united states,” 2018.

[12] King County, “King county 2020 strategic climate action plan,” Available at https://your.kingcounty.gov/dnrp/climate/documents/2020-SCAP-Full-Plan.pdf (2020).

[13] A. B. Downey, *Think stats probability and statistics for programmers.* O’Reilly, 2011.

[14] D. McFadden, “Conditional logit analysis of qualitative choice behavior,” Available at https://cowles.yale.edu/sites/default/files/files/pub/d04/d0474.pdf (1974).

[15] Office of Environmental Health Hazard Assessment, “Calenviroscreen 3.0,” Available at https://oehha.ca.gov/media/downloads/calenviroscreen/report/ces3report.pdf (June 2018).

[16] B. Paulos, S. Forrester, E. O’Shaughnessy, C. Dyson, G. Barbose, and R. Wiser, “An assessment of evaluation practices of low- and moderate-income solar programs,” 2021.

[17] “Expanding access to solar for low-to-moderate income households and communities: Lessons learned for state agencies,” Available at https://www.cesa.org/resource-library/resource/solar-lessons-learned-for-state-agencies/, May 2021.

[18] Washington State Department of Commerce, “Washington 2021 state energy strategy,” Available at https://www.commerce.wa.gov/wp-content/uploads/2020/12/Washington-2021-State-Energy-Strategy-December-2020.pdf (2020).

[19] B. Paulos, “Bringing the benefits of solar energy to low-income customers,” *Clean Energy States Alliance, 2017.*

[20] B. Sigrin, M. Gleason, R. Preus, I. Baring-Gould, and R. Margolis, “The distributed generation market demand model (dgen): Documentation,” 2016.

[21] S. A. Klugman, *Loss models : from data to decisions*, fifth edition. ed., ser. Wiley series in probability and statistics. Hoboken, NJ: John Wiley Sons, Inc., 2019.

[22] W. H. Greene, “Accounting for excess zeros and sample selection in poisson and negative binomial regression models,” 1994.

Appendix A
CHECKING MULTIPLE REGRESSION ASSUMPTIONS

As stated above, the assumptions for a multiple regression model as shown below.

1. Linear Relationship: There is a linear relationship between the dependent variable and independent variables.

2. Independence: The values of the residuals are independent.

3. Homoscedasticity: The variance of the residuals is constant

4. Normality: The values of the residuals are normally distributed.

I check that there is a linear relationship between the dependent variable and independent variables by graphing a scatterplot for each independent variable versus the dependent variable. Figure A.1 shows the results for all six independent variables. In these plots as well as our regression model, the dependent variable is logged. As shown in the plots, even though there is high variance in some of the relationships, I can assume a linear relationship between all independent variables and the dependent variable.

I check that the values of the residuals are independent by ensuring that the independent variables are not highly correlated. A correlation matrix of Pearson correlation coefficients can be used to check that none of the coefficients are above 0.7 or below -0.7. As shown in Figure 3.3, there is high collinearity between some of the independent variables of interest. For this reason, percentage of renters, percentage of the population that is white, and the daily solar radiation are excluded from the model. Figure A.2 shows the Spearman correlation coefficient for all the independent variables included in the Poisson regression model.

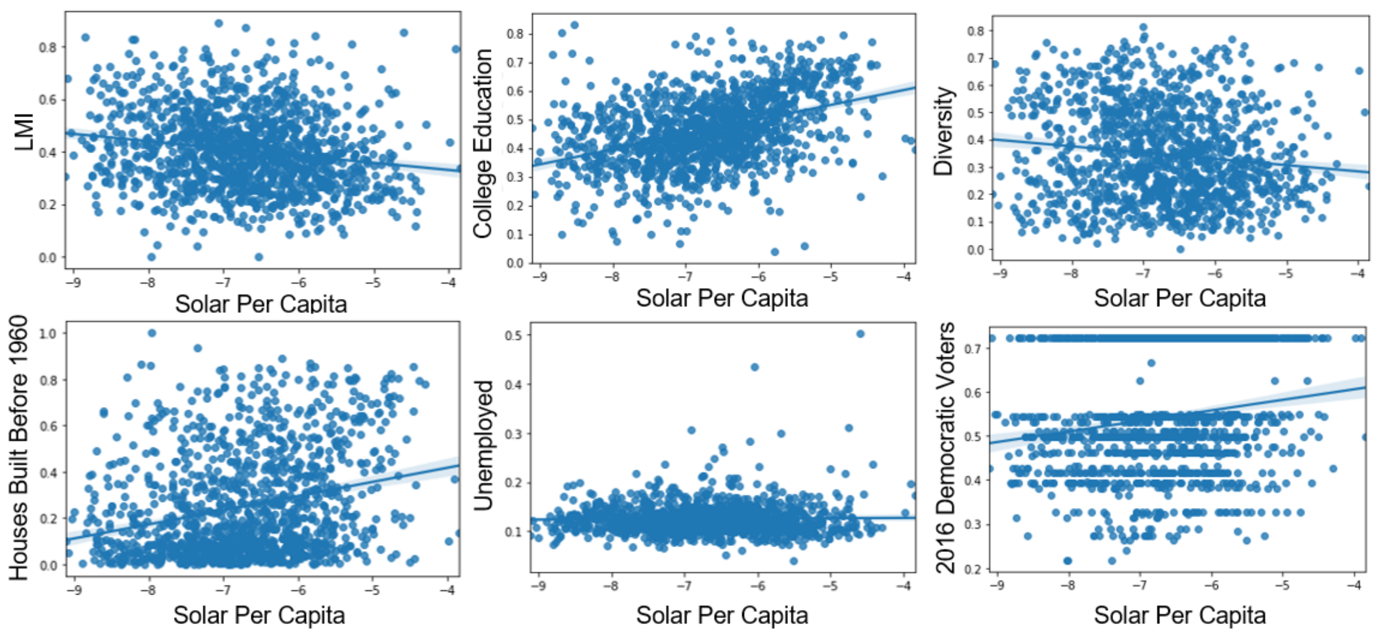

Figure A.1: Six scatterplots of the solar per capita in each census tract and six socioeconomic factors: LMI, college education, diversity, houses built before 1960, unemployment, and 2016 democratic voters

I check that the variance of the residuals is constant by graphing the residuals and the predicted y values to see if the size of the error term changes with different values of the predicted y values. As Figure A.3 shows, the average residual value does not change.

I check that the values of the residuals are normally distributed using a Normal QQ Plot. A Normal QQ Plot is a visual tool that shows if a set of data comes from a normal distribution. If the scatterplot forms a straight line at a 45 degree angle, that means that the data is close to normal. As Figure A.4 shows, the residuals form a straight line, indicating the residuals are normally distributed.

Figure A.2: Spearman correlation coefficient matrix of the solar per capita in each census tract and the six socioeconomic factors included in the regression model

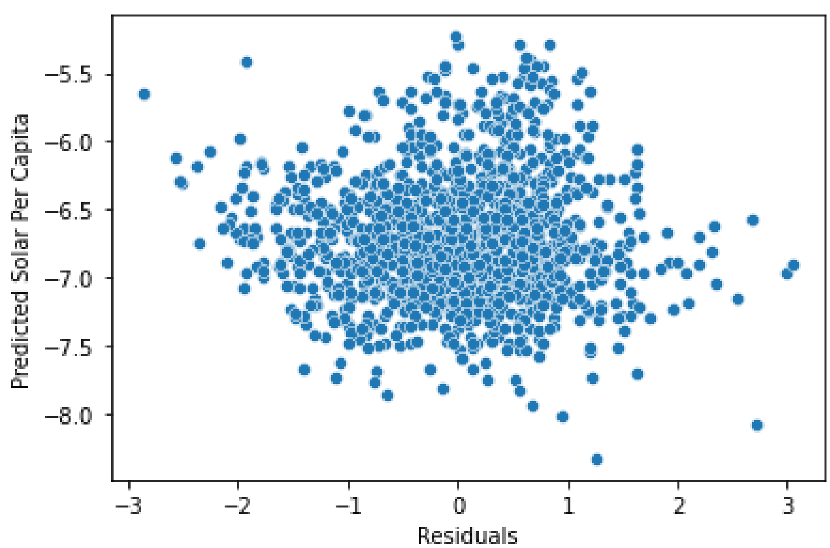

Figure A.3: A scatterplot of the predicted solar per capita in each tract based on the results of the linear regression model and the residuals from the model

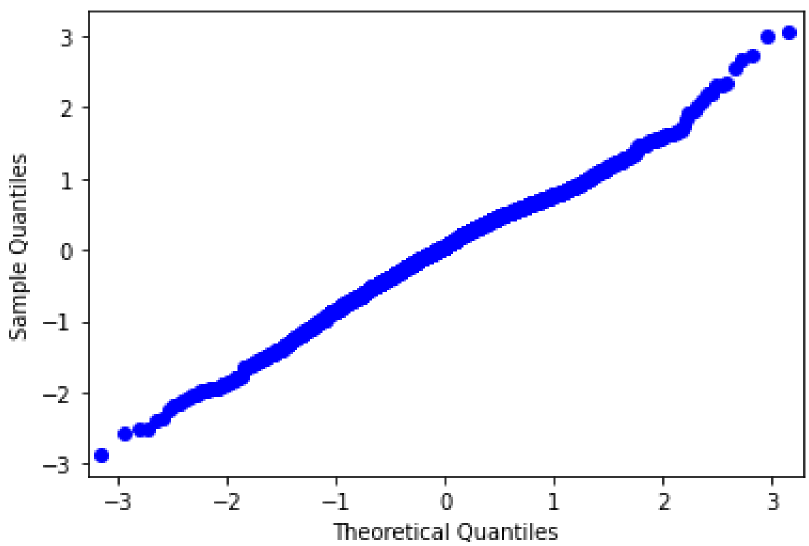

Figure A.4: A Normal QQ Plot of the residuals from the multiple regression model

Appendix B

LIMITATIONS OF CENSUS TRACT LEVEL DATA

Individual socioeconomic factors are not publicly available for each solar adopter. For this reason, the most granular analysis I could conduct in Washington state is at the census-tract level. While I identified existing disparities at the census-tract level, the true magnitude of these disparities are unknown. One way I can examine the limitation of a census-tract level analysis is looking at income disparities. As my analysis found, there was only a positive correlation with households with a 120% AMI or more and the amount of solar per capita. In addition, census tracts with a higher percentage of LMI households were likely to have less solar. I can now compare these findings with the income disparities identified by the Lawrence Berkeley National Lab, where each solar adopter was classified into an income bracket.

Berkeley Lab's Solar Demographics Tool, as part of [3], displays graphs of solar adopter income distributions at a more granular level. Figure B.1 and Figure B.2 show the income distribution of solar adopters in Washington. Figure B.1 specifically shows the percent of solar adopters in Washington State that fall under six different income brackets. Figure B.2 re-configures this data and shows the income of solar adopters in relation to the AMI. Specifically, the bracket labeled "> 120%" includes the percent of solar adopters with an income that is greater than 120% of the area mean income. Meanwhile, the bracket labeled "< 60%" includes the percent of solar adopters with an income that is less than 60% of the area mean income.

As these result show, solar adopters are for the majority in the most affluent income bracket. This income disparity is stronger than I could identify at the census-tract level.

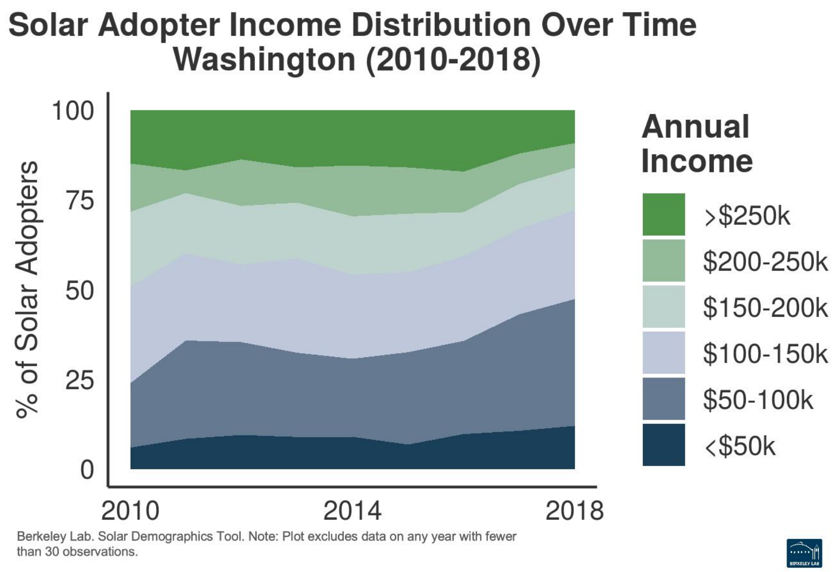

Figure B.1: Image from [3]. Area plot shows the income distribution of solar adopters over time in Washington from 2010-2018. Annual income brackets are broken into six groups.

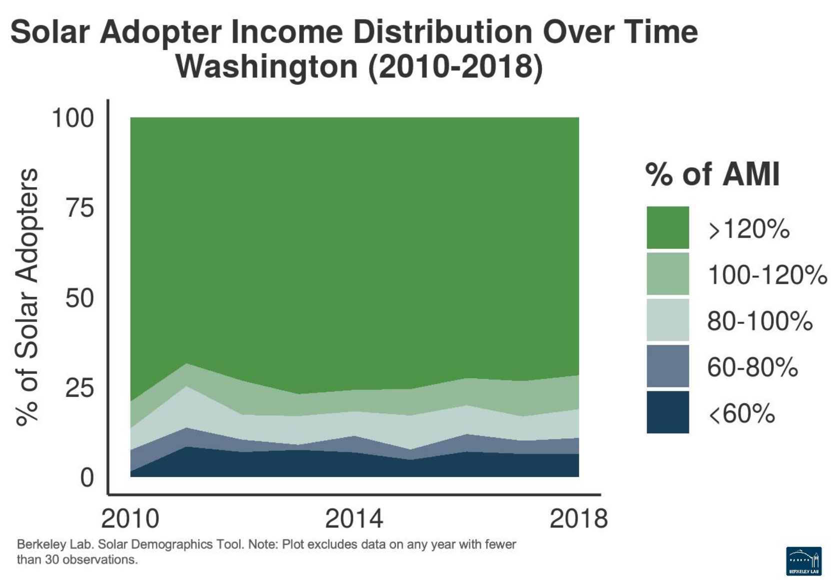

Figure B.2: Image from [3]. Area plot shows the income distribution as a percentage of AMI of solar adopters over time in Washington from 2010-2018. Income brackets as a percentage of AMI are broken into five categories.