The Effect of Income Distribution on Solar PV Installation by Zip Code

James Kurtak[a,*]

[a]*Department of Economics, 7146 College Hill Rd, Clinton, NY 13323*

Abstract

I use income, demographic, and solar PV installation data to test whether wealth inequality has an effect on Solar PV usage within the State of Arizona. I find that there is no statistically significant evidence that a higher proportion of low income households within a given zip code results in a decreased installation rate of solar PV. Rather, as the proportion of individuals earning between $50,000-75,000 dollars a year or $100,000-$200,000 dollars per year rises, the number of solar PV installations increase. Finally, I find there is a positive correlation between the median age of a zip code and solar PV installations.

JEL codes: Q20, Q29, Q42

Keywords: Environmental Economics, Solar PV, Residential Solar, Income,

Introduction

Since 1977, the cost of solar panel installation and maintenance within the US has gone down from $76.67/watt to $0.698/watt today. This cost reduction is a direct result of technological advances and a plethora of government back subsidies, tax exemptions, and rebate programs. Unsurprisingly, with the increased accessibility of this renewable technology, more households and industrial firms have begun utilizing solar PV. Despite this, many economists believe that even with this marked decrease in the price of solar PV technology, the idea of using solar energy in a domestic setting can only be realized by higher earning Americans. Interestingly, though Arizona has the second most installations within the United States, the majority of literature on this subject has only examined the income installation relationships within California or New York. Because of this, I seek to test the degree of the income-solar relationship within the state of Arizona.

In order to do so, I compare the cumulative installation data by zipcode for Solar PV's within Arizona with demographic data from the census and income data from the IRS- also available at the zipcode level. Ultimately I find that the proportion of low income individuals ($0-$25,000 USD per year) within a given zipcode has little

*Thesis

Email address: jkurtak@hamilton.edu (James Kurtak)

to no perceivable effect on the Solar PV installation rate. Moreover, I find that there is a positive correlation between the annual income levels of $50,000-$75,000 and $100,000-$200,000, higher levels in median age, and the number of Solar PV installations per zipcode.

Intuitively, one would expect the lower income households to have a lower installation rate compared to all other income brackets. Because of this, in the final section of my analysis, I seek to explain why my results differ from previous literature, either through possible flaws within my data and analysis, as well as possible differences which might exist within the varying income brackets to create these differences.

Literature Review

Boreinstein (2014)

Boreinstein explains the benefits of domestic solar energy in California compared to traditional grid based electricity, given the higher price of electricity during peak hours, and the ample government systems in place to encourage solar consumption. As income data per household installing PV was not available for Borenstein he relied upon a five bracket income system, where he grouped census blocks of roughly 600 households into income brackets dependant on their energy usage; these brackets are $0-$20,000, $20,000-$40,000, $40,000-$60,000, $60,000-$100,000, and above $100,000. He then compared the CSI census block income with the adoption rate of solar PV. Through his analysis, he found that solar adoption was skewed to wealthier households, though has been declining between 2011 and 2014. While overall, I did not use his methodology in my analysis, I found the bracketed income system worked very well as a means identifying wealth distribution within a zip code. Additionally, his work was invaluable in describing the economic incentives of using solar energy- a factor which is extremely relevant when describing why one might see different solar adoption rates when using solar.

Josp? (2014)

In 2014, a group of ten Columbia students at the school of International and Public affairs worked with the non-profit organization GRID-Alternatives in order to determine whether Solar PV was reaching low income communities within New York as well as what policies could be implemented to make solar energy more accessible to these communities. By combining zip code based residential solar PV installation data with ZCTA data from the US census, they created a scatter plot comparison of the residential solar installations per 10,000 residents against the median household income by county.Ultimately, I mimic this methodology for my basic graphic analysis of the income relationship with PV installation rates, although my method of measuring income is less specific, as I will discuss later in the Data section.

Rothfield (2010)

Rothfield's primary focus was to better understand the policy options the California government could adopt in order to increase solar PV installation rates among residential households. To do so, they created a cost analysis model which measured energy incentives and energy prices against installation rates, with each observation being a California zipcode. While doing so, they controlled for demographic information such as income level, age, race, and education. Ultimately, they found a positive correlation between education and high-income levels against the rate of Solar PV usage.

It is important to note that Rothfield's income variable 'rich' was a boolean value which was true given the household income was $125,000 or more; though they found high-income levels are correlated with solar PV installations, the principal objective of their analysis was clearly not an in depth analysis of the income/solarPV relationship. Otherwise, we would see several unadulterated incomes per household or at the very least income bracket variables such as we saw in Borenstein's work. Additionally, Rothfield's education variable differs from my own in that it includes only college education and up- whereas I include highschool education and up. In any case, I found their basic regression model would serve as the basis for my own- with the exception of the Electricity Price, Electricity Demanded, Incentive variables.

Data Sources

For my regression, I use data from the US Census provided through Factfinder.com, data from the IRS, and data from The Open PV Project.A table of my summary statistics can been seen in figure no. 1 in my last section.

From the US Census, I obtained all demographic information including the variables zipcode/ZCTA, Total.Pop, Median.Age, White, Urban, UrbanCluster, Rural, and HighschoolPlus. Rothfield (2010) used education as a "proxy for awareness and perceived seriousness of global warming". I intend to do the same, with HighschoolPlus serving as a control an individual's care for the environment. It is worth mentioning that the variable "zipcode" is actually an estimation of where the census data was retrieved from. Because zip codes are small areas used by the US postal service and change with frequency, the US census has chosen another method of classification, where zip codes are "lumped together" to form Zip Code Tabulation Areas (ZCTA's). Factfinder.com takes this into consideration, assigning a zip code to a given ZCTA based on the most predominant zip code within the geographic area. Though this will inevitably lead to some inaccuracy in my analysis, it is a necessary evil- demographic provided through the census is simply unavailable at the true zipcode, household, or individual level.

From the IRS, I acquired tax report information for the years 2011, 2012, and 2013. From this data I created

a set of variables which represent the percentage of individuals who filed a tax return of one six income brackets within a given zip code. This is expressed in the variable AVGretX where X takes the value of either $1-$25,000, $25,000-$50,000, $50,000-$75,000, $75,000-$100,000, $100,000-$200,000 or $200,000+ dollars per year. All data is "self-reported", by those who choose to file taxes, and is available at the State, county and zip code level.

Finally, from The Open PV Project, I obtained the cumulative number of solar PV installations per zip code between the years 1977 and 2013. The Open PV Project is a collaborative initiative between the government, industry, and the public in order to estimate the total number, size, date of installation and cost of all solar PV installations within the United States. All data is available at the zip code level and is submitted voluntarily by the installer of the panel, though the Open PV project encourages the installers to submit their data. This discrepancy is potentially problematic when interpretting my data and later regression results. I discuss why this is in the final portion of my analysis.

Data Analysis

Overall the standard deviation of my variables fits within the bounds I expected, though there seems to be a signifigantly higher variance within the variables AVGret1to25 and AVGret100to200 when compared to the other income brackets. Moreover, we observe an interesting distribution within the income brackets; the zip code average for the percentage of income returns of $1 to $25,000 dollars is 40%! Oddly, the number of individuals making between $100,000-$200,000 thousand is two percent higher than those of the $75,000-$100,000 income bracket.

In figure no.2-3 we see a graphical representation of income distributions on the solar panel installations within a given zipcode. On the scatter plot side, one can see that the majority of zipcodes contain less than 2,500 installations per 10,000 individuals regardless of income distributions.When viewing the smooth-line graph side it appears to be that the income brackets of $50,000-$75,000 and $100,000-$200,000 have a large positive impact on the Solar installation rate while the $1-$25,000 dollar bracket has a strong negative cooralation. However, I must note that the grey area one sees surrounding the line is the variance one could expect given the data within 95% accuracy. This means that we need to control for some variables before we can make an accurate analysis!

Additionally, both Median Age and Highschool education variables appear to be coorelated with solar PV installation rates. Again, because of the possible variance within the line given the data, one can not draw conclusions until a proper regression has been conducted.None-the-less the relationship appears to be rather strong.

Methodology

In my analysis I compare the effect of wealth distribution of six income brackets within a given zip code on the number of solar PV installations within that zip code. These income brackets are manifest in the AVGret variables seen below. In addition, I control for the total population, the median age, the percentage of individuals who've received high school education or greater, and whether or not the zip code is considered a rural or urbans area. SolarPanelper10k serves as my y variable and is indicative of the number of solar panels per 10,000 individuals which exist within a given zip code. This model is loosely based on both Borenstein, with the use of income brackets as an indication of wealth, and Rothfield, through the demographic control variables described above. My regressional formula is as follows:

$$\begin{aligned} SolarPanelper10k &= \beta_0 + \beta_1 AVGret1to25 + \beta_2 AVGret25to50 + \beta_3 AVGret50to75 \\ &+ \beta_4 AVGret75to100 + \beta_5 AVGret100to200 + \beta_6 Total.Pop + \beta_7 Median.Age \\ &+ \beta_8 Urban + \beta_9 Rural + \beta_{10} HighschoolPlus + \varepsilon \end{aligned}$$

Results and Inferance

In figure 7 one can see my regression results.

Surprisingly, I found Solar PV usage was equal to average in the lowest income communities and only above average in the middle class bracket of $50-$75,000 & $100-$200,000 dollars - while Borenstein and Josp? found a large disparity between lower and upper income levels. This is even more curious given Arizona solar panels are even more accessible, factoring price and solar saving programs within the state, than either California or New York. This indicates that this variance is not due to some difference in solar panel policy between the states, but some other reason altogether. I can guess several explanations as to why this may be.

Firstly, solar PV energy lowers long-term cost while dramatically increasing short term cost; while poorer individuals might not consume solar because of the wealth barrier, more affluent individuals might be just as unlikely to utilize the technology given they are indifferent to energy costs and the environmental impact of energy usage. Assuming altruism takes second place to economic benefit, the middle class would be the prime candidates for the cost savings, while simultaneously having enough economic capital to realize them. In essence, low income and high income individuals are both installing less Solar PV than expected, though for entirely different reasons. This would indicate that my results are differing from those of Borenstein and Rothfield simply due to behavioural differences between the affluent in California or New York and Arizona, where the wealthy in Arizona are simply more indifferent to the environmental implications of solar PV technology. However, this an unlikely scenario, at least when comparing Arizona to New York, given that Arizona has a much higher number of installations, lower cost per kilowatt for installation, and a higher solar capacity.

An alternative possibility is that my original Solar PV data was flawed in that the installers belonging to the middle income brackets were the most likely to report solar panel installations. While poorer individuals had few solar panels to report, wealthier individuals may not have been as willing to share private information about their property or their finances. Likewise, it is quite possible that wealthy individuals within California and New York were simply more inclined to share information about their installations, which would account for the perceived differences between my analysis and previous studies.

Third, it is possible that the findings of Rothfield and Borenstein differed from my own because of their methods of classification of the poor or their models themselves. Borentein's model used energy consumption per household to estimate income; on one hand, his observations were at a much more detailed than my own (household vs zip codes). On the other, he estimated income rather than using IRS data and his income brackets were different than my own. His method of estimation could be flawed in some way, or were could simply be observing differences due to different cut-off points for our income brackets. Furthermore, Rothfield determined low income as a household income less than $44,050 USD a year. My own model assumes low income only when a household is earning between $1 and $25,000 USD.

Another important difference I found with my results and those of previous literature is the non-correlation between education and Solar PV installation rates. However, again I must note, my variable for measuring education (and thus concern for environmental matters) was dependant on a highschool education or higher, whereas Rothfield used college education or higher.

Unsurprisingly, I did not find any evidence to support a correlation with race solar PV installation rates, though I did find a positive relationship between age and solar PV usage. Unfortunately, the latter could be correlated to income, so I am hesitant to perscribe any true relationship between age and solar PV installation rates, even given a p-value$<$0.01. With that in mind, it was rather unexpected to find no relationship between the Rural and Urban variables given one might expect there to be more solar installations where there is more available land and sunlight, i.e. outside of the city. However, the cities within Arizona are rather 'spread out' rather than 'built up' like we might see in California or New York. This in turn could result in little to no difference between urban and rural solar PV installation rates, while controlling for population and income levels.

References

Josp?, C. (2014). Ensuring New York Solar Programs Reach Low Income Residents. Retrieved from http://mpaenvironment.ei.columbia.edu/files/2014/06/GRIDAlternativesProject.Final_.pdf

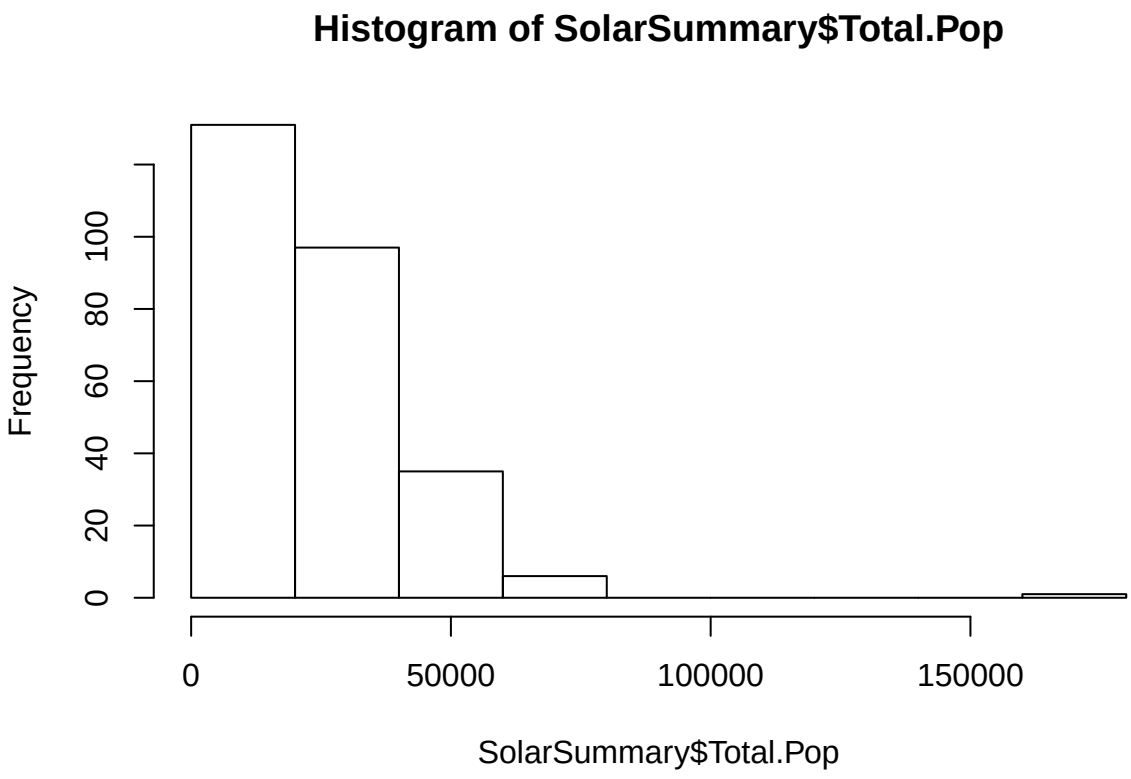
Rothfield, E. (2010). Solar Photovoltaic Installation in California: Understanding the Likelihood of Adoption Given Incentives, Electricity Pricing and Consumer Characteristics. Retrieved from https://sites.duke.edu/djepapers/files/2016/10/Rothfield_DJE.pdf

Borenstein, S. (2015). A Cost comparison of Solar PV with Grid Electricty. Nber Working Paper Series. Retrieved from http://www.nber.org/papers/w21342.pdf

Cost of Solar Panels - 10 Charts Tell You Everything. (2014, September 17). Retrieved May 07, 2017, from http://costofsolar.com/cost-of-solar-panels-10-charts-tell-you-everything/

Data Appendix

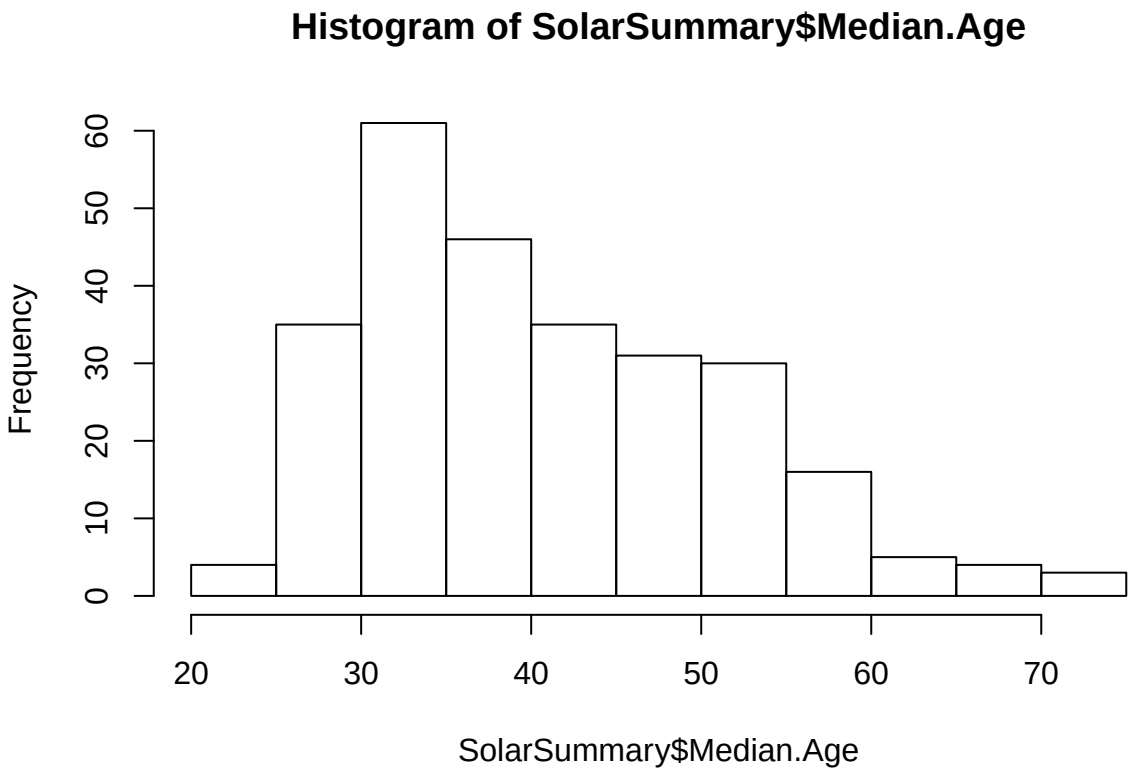
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     311    6705   21220   23180   35970  170000
```



- Median.Age
- US Census
- 0 observations missing
- Measures the median age within a given zipcode

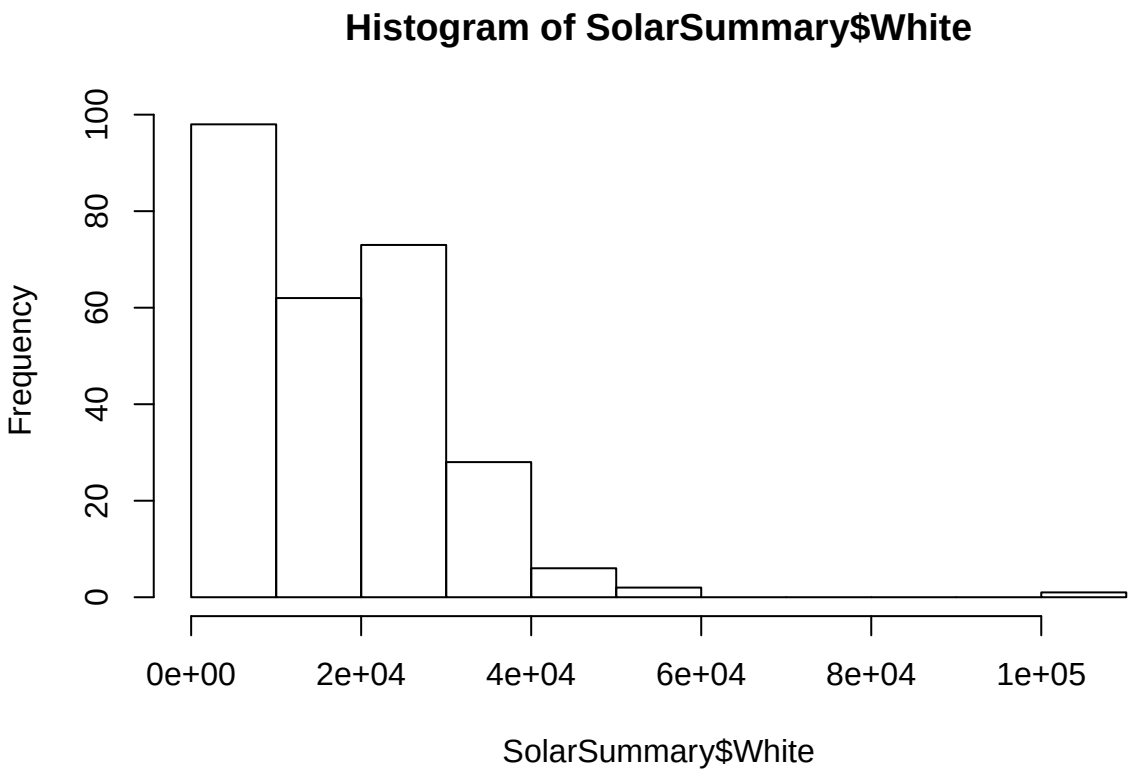
- Median.Age
- The unit of measurement is “years” while each unit of observation is zipcode

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   23.50   32.10   38.20   40.82   48.08   74.30
```



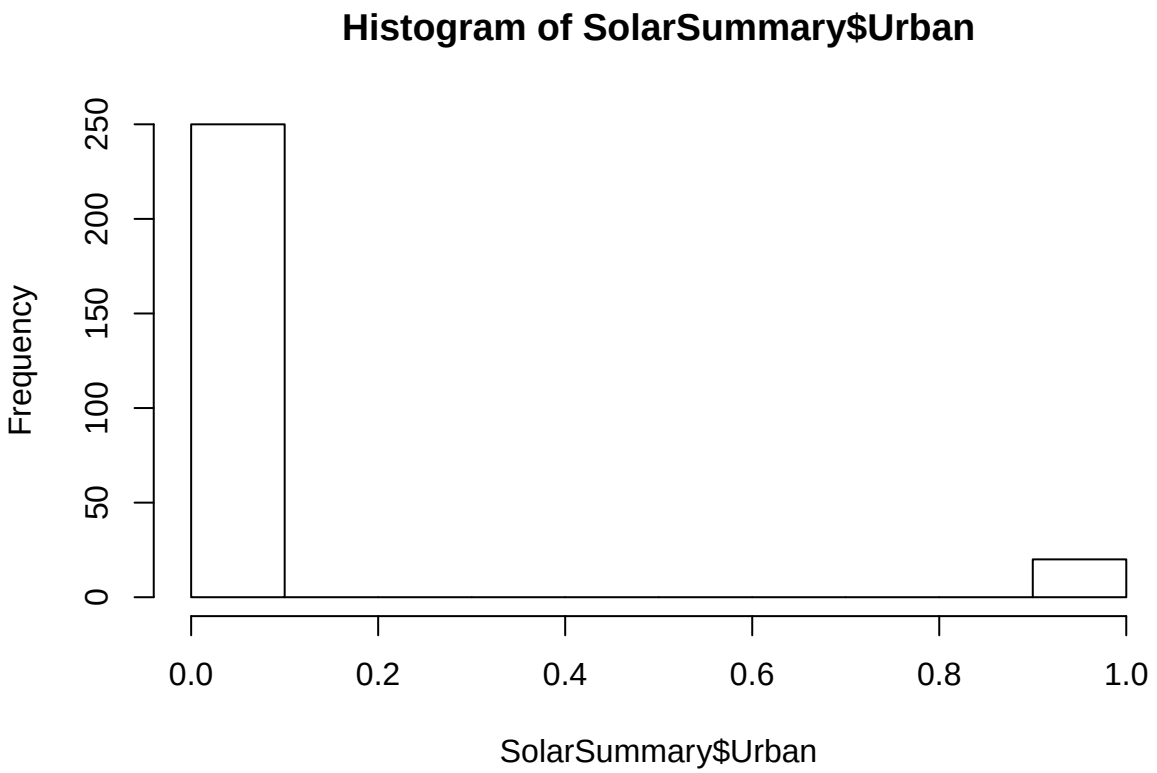
- White - US Census - 0 observations missing - Measures the total number of individuals who identify as white within a given zipcode - White - The unit of measurement is “race” while each unit of observation is zipcode

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     129    5527   16350   17190   25840  102300
```



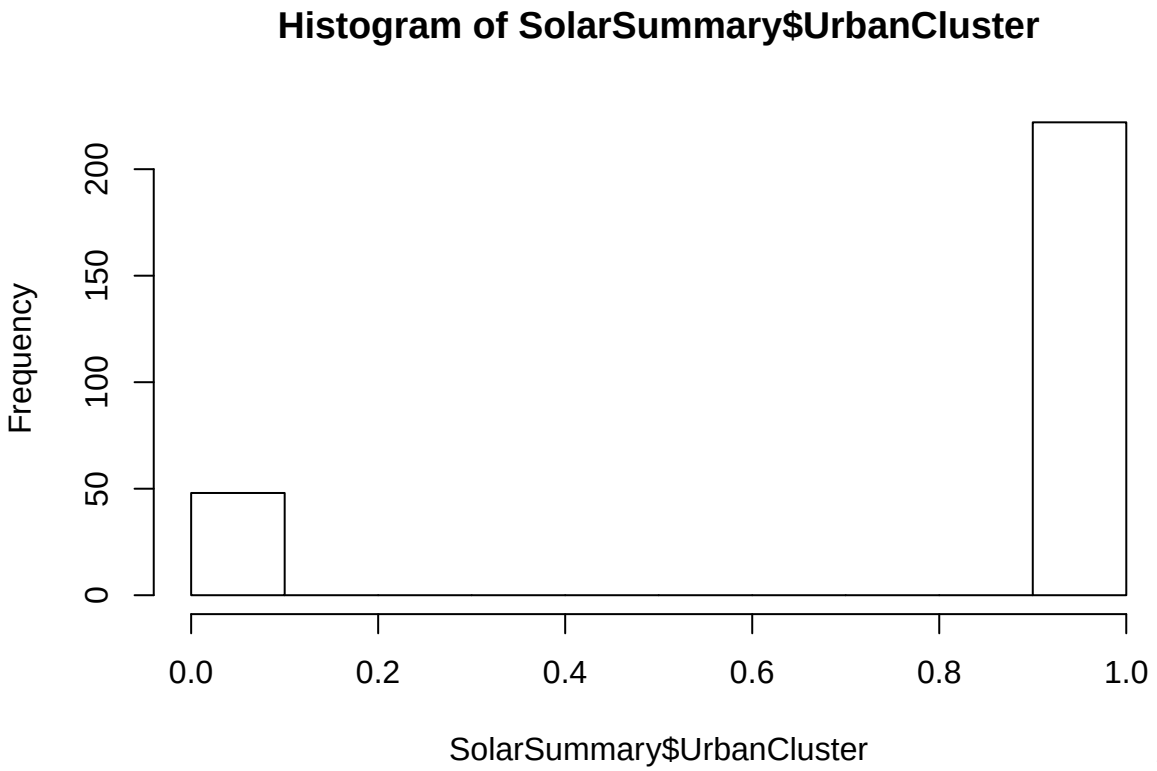
- Urban - US Census - 0 observations missing - Determines whether a zipcode is Urban or not. - Total.Pop -The unit of measurement is a boolean value of 1 or 0. The variable takes a 1 when popluation is greater than 50,000.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.00000 0.00000 0.07407 0.00000 1.00000
```



- UrbanCluster - US Census - 0 observations missing - Determines whether a zipcode is an Urban Cluster. - Total.Pop -The unit of measurement is a boolean value of 1 or 0. The variable takes a 1 when popluation is between 2,500 and 50,000.

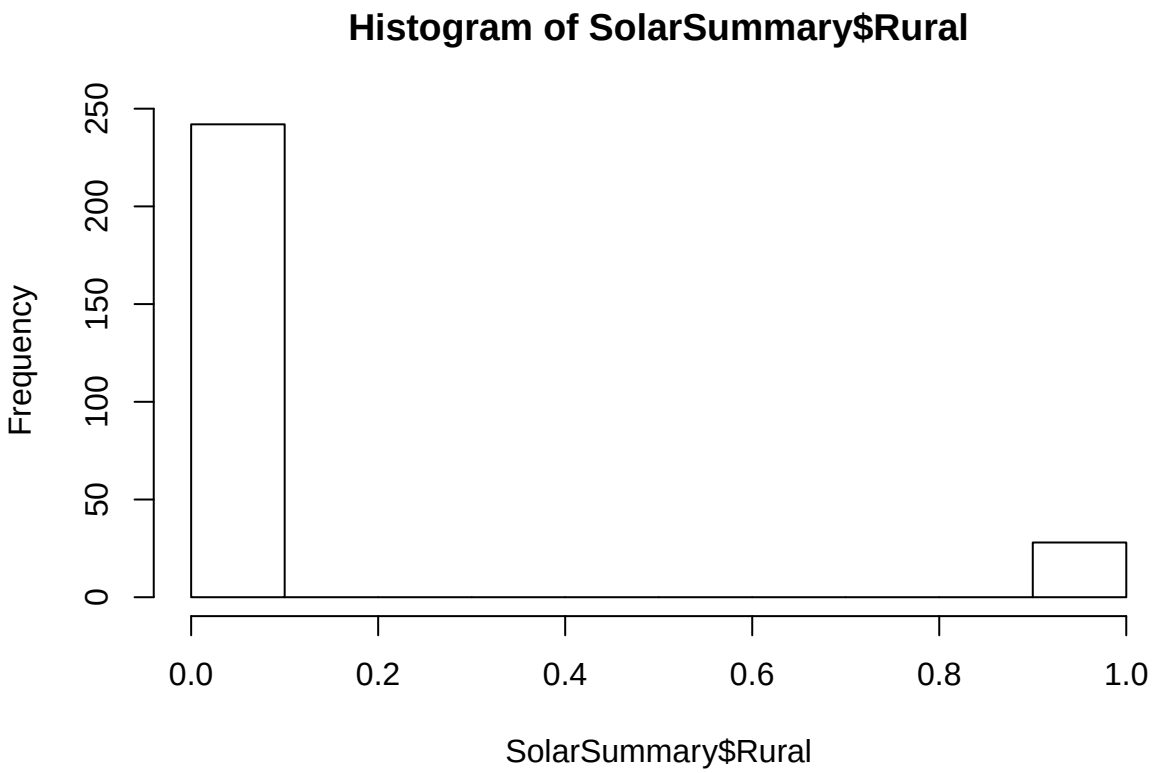
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  1.0000  1.0000  0.8222  1.0000  1.0000
```



- Rural

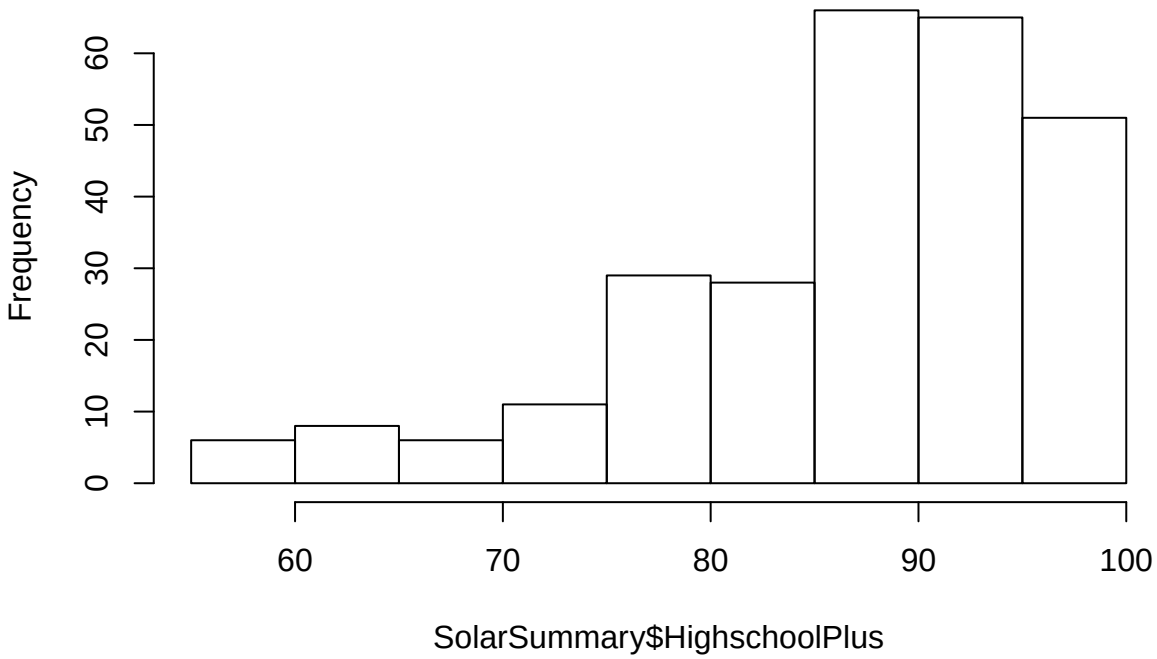
- US Census - 0 observations missing - Determines whether a zipcode is Rural. - Rural -The unit of measurement is a boolean value of 1 or 0. The variable takes a 1 when popluation is less than 2,500.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.1037  0.0000  1.0000
```



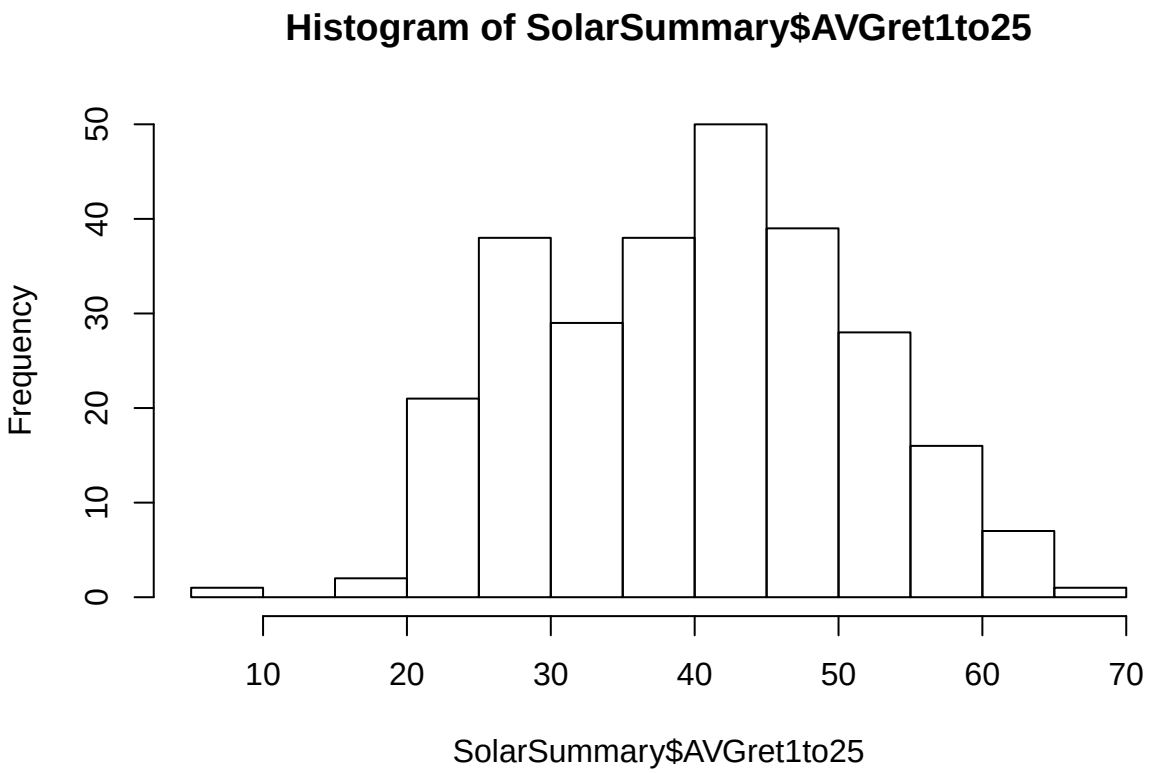
- HighschoolPlus - US Census - 0 observations missing - Indicates what percentage of individuals attended highschool education or higher within a given zipcode. - HighschoolPlus -The unit of measurement is an individual. The unit of observation is a zipcode.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   55.50   81.58   88.90   86.50   93.60   98.80
```



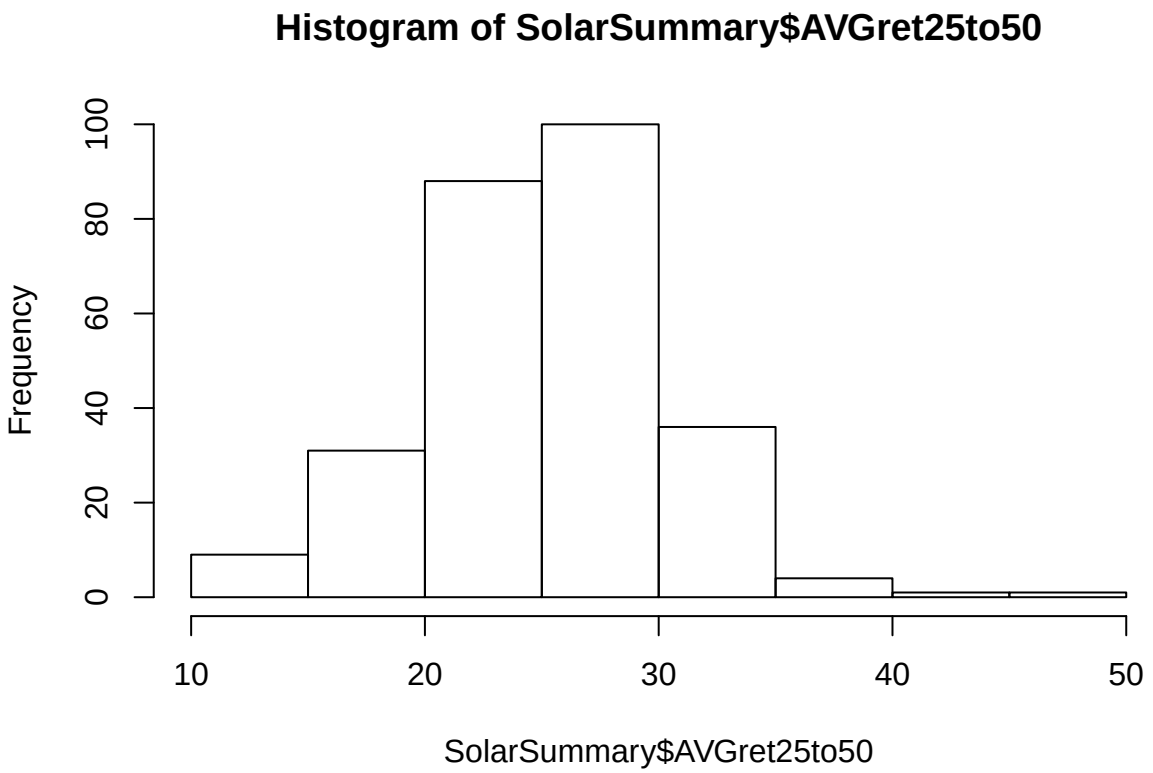
- AV-Gret1to25 - IRS - 0 observations missing - Indicates the average percentage of individuals of the income bracket $1-$25,000 which existed in a given zipcode between 2011 & 2013 - $1 under $25,000 -The unit of measurement is a percentage. The unit of observation is a zipcode.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   9.921  31.090  41.300  40.120  48.190  65.520
```



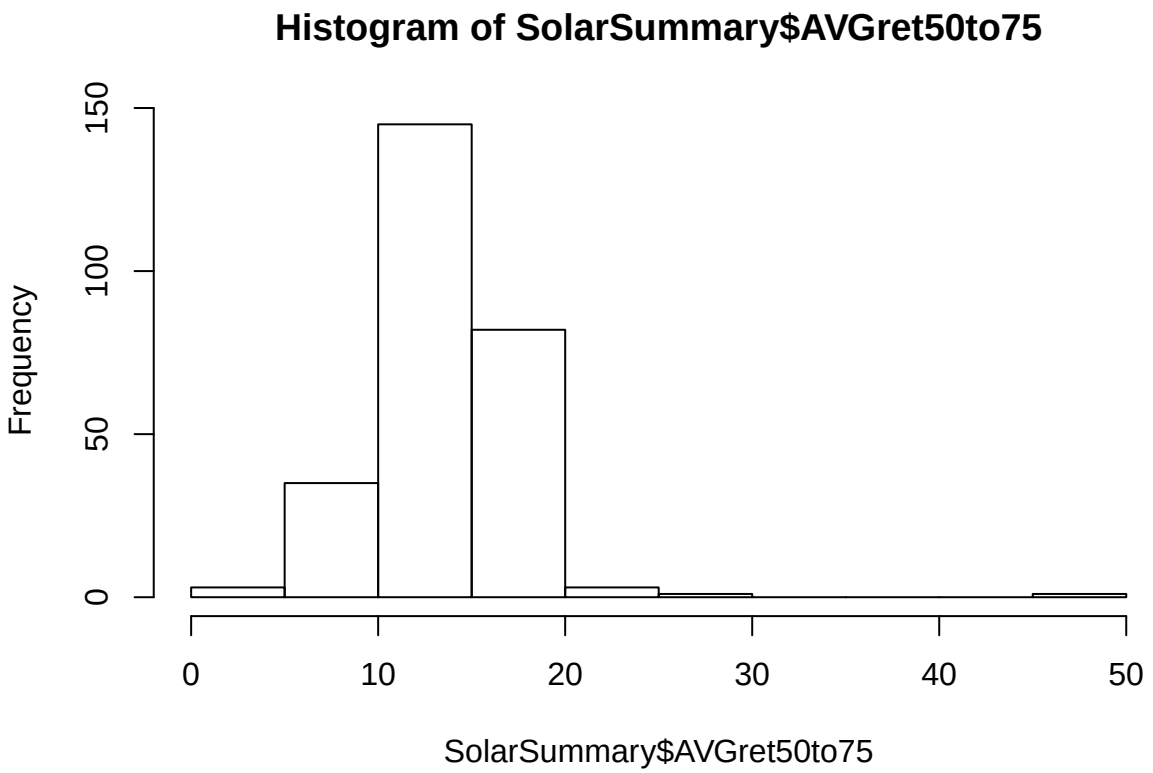
- AV-Gret25to50 - IRS - 0 observations missing - Indicates the average percentage of individuals of the income bracket $25-$50,000 which existed in a given zipcode between 2011 & 2013 - $25,000 under $50,000 -The unit of measurement is a percentage. The unit of observation is a zipcode.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.96   22.33   25.31   25.15   28.53   46.88
```



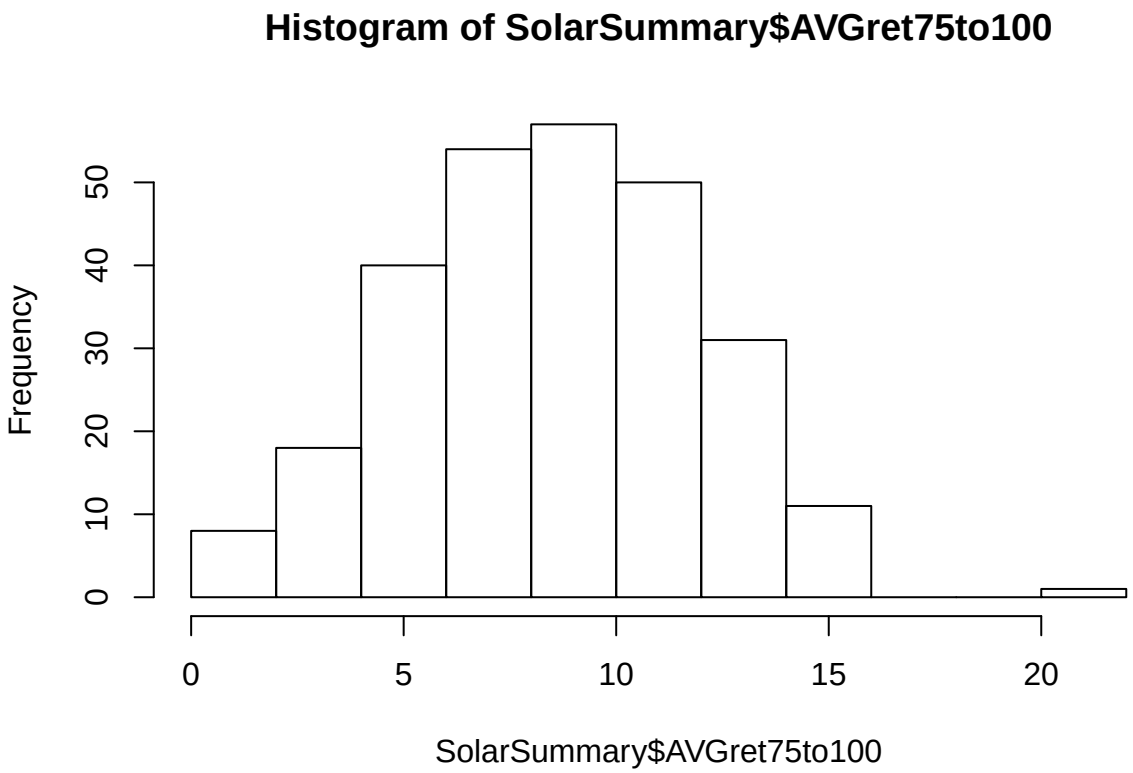
- AV-Gret50to75 - IRS - 0 observations missing - Indicates the average percentage of individuals of the income bracket $50,000-$75,000 which existed in a given zipcode between 2011 & 2013 - $50,000 under $75,000 -The unit of measurement is a percentage. The unit of observation is a zipcode.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   11.71   13.71   13.59   15.58   48.58
```



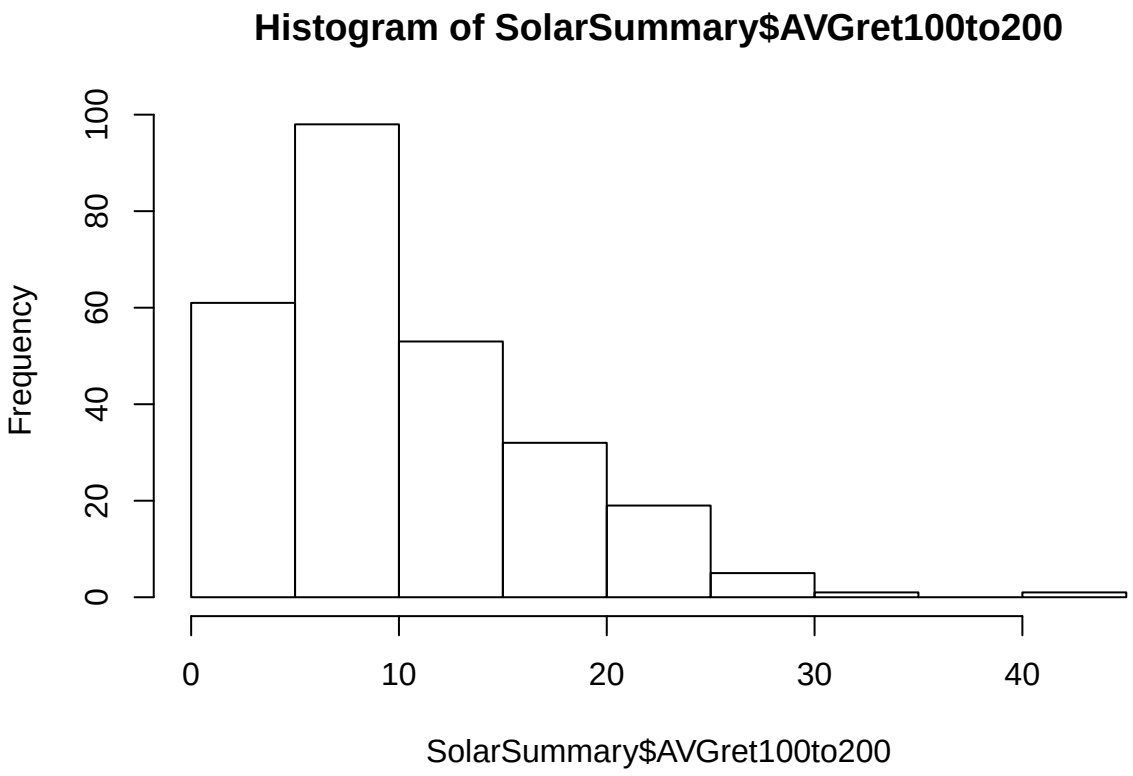
- AV-Gret75to100 - IRS - 0 observations missing - Indicates the average percentage of individuals of the income bracket $75,000-$100,000 which existed in a given zipcode between 2011 & 2013 - $75,000 under $100,000 -The unit of measurement is a percentage. The unit of observation is a zipcode.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   6.143   8.289   8.470  11.020  21.760
```



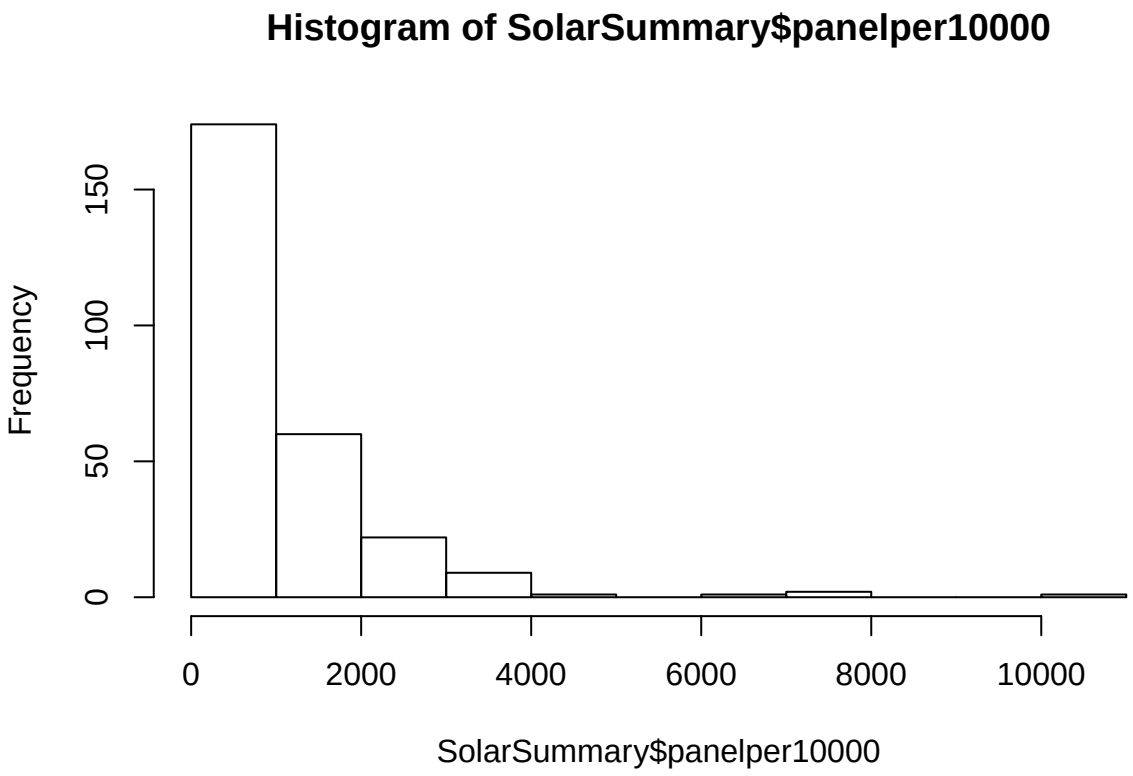
- AV-Gret100to200 - IRS - 0 observations missing - Indicates the average percentage of individuals of the income bracket $100,000-$200,000 which existed in a given zipcode between 2011 & 2013 - $100,000 under $200,000 -The unit of measurement is a percentage. The unit of observation is a zipcode.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   5.321   8.247  10.080  14.060  43.040
```



- panelper10000 - The Open PV Project - 0 observations missing - Indicates the number of solar panels per 10,000 individuals within a given zip. - $100,000 under $200,000 -The unit of measurement is solar panel per 10,000. The unit of observation is a zipcode. -This variable "technically" did not exist in the original data set. It is derived by adding all installations of Solar PV installations within a given zip between the years 1988 & 2013. This composite number is then divided by the population over 10,000.

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##     9.91   433.50   670.50  1094.00  1449.00 10370.00
```


Histogram of SolarSummary$panelper10000
Frequency
0
50
100
150
0
2000
4000
6000
8000
10000
SolarSummary$panelper10000

List of Figures

1 Summary Stats 21
2 Income1to25 vs Solar PV 22
3 Income50to75 vs Solar PV 23
4 Income100to200 vs Solar PV 24
5 Education Effect 25
6 Median Age Effect 26
7 Regression Results 27

Statistic	N	Mean	St. Dev.	Min	Max
Total.Pop	270	23,175.800	19,046.560	311	170,047
Median.Age	270	40.818	10.599	23.500	74.300
White	270	17,190.520	13,115.730	129	102,315
Urban	270	0.074	0.262	0	1
UrbanCluster	270	0.822	0.383	0	1
Rural	270	0.104	0.305	0	1
HighschoolPlus	270	86.503	9.711	55.500	98.800
AVGret1to25	270	40.119	10.802	9.921	65.522
AVGret25to50	270	25.148	5.213	10.960	46.875
AVGret50to75	270	13.594	3.987	0.000	48.580
AVGret75to100	270	8.470	3.418	0.000	21.757
AVGret100to200	270	10.078	6.781	0.000	43.044

Figure 1: Summary Stats

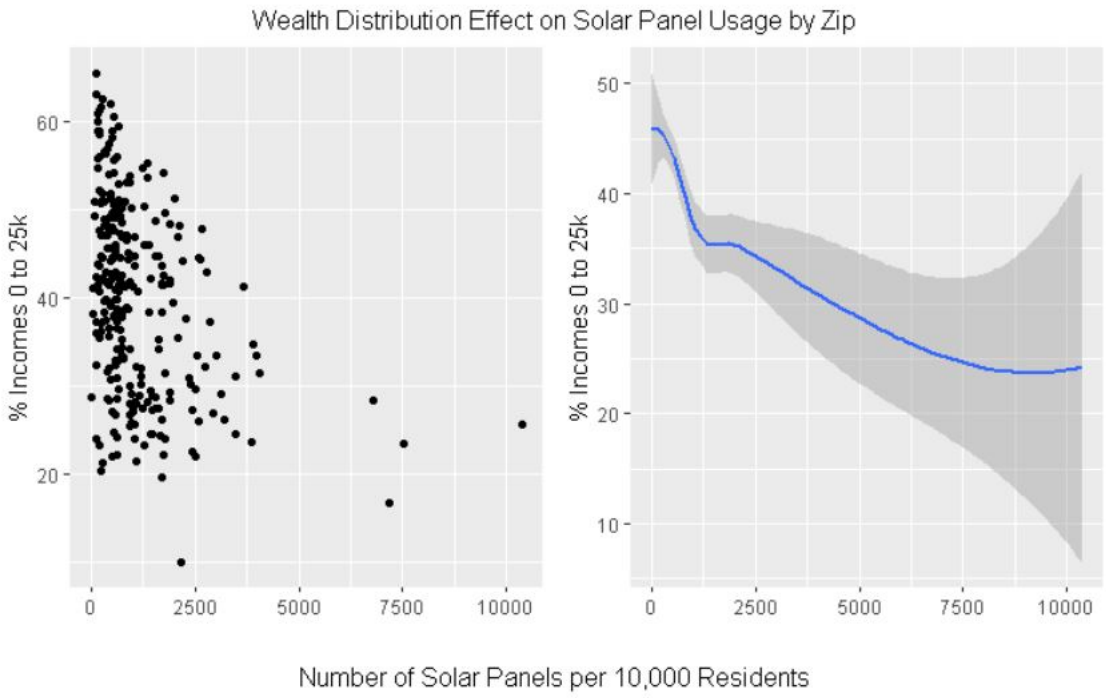


Figure 2: Income1to25 vs Solar PV

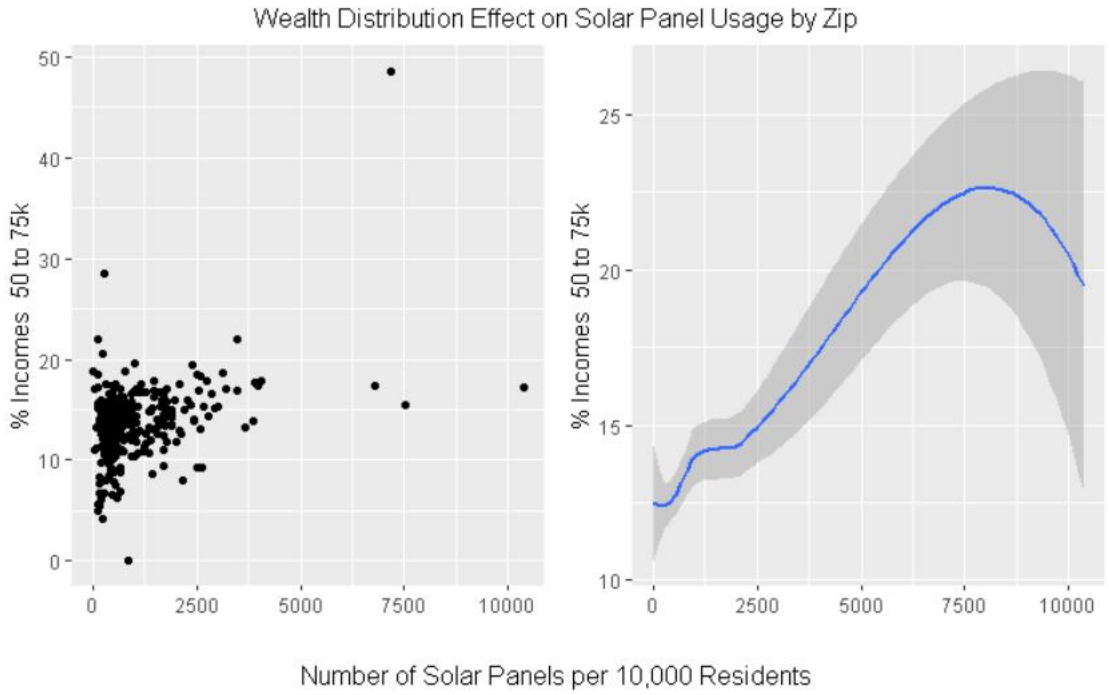


Figure 3: Income50to75 vs Solar PV

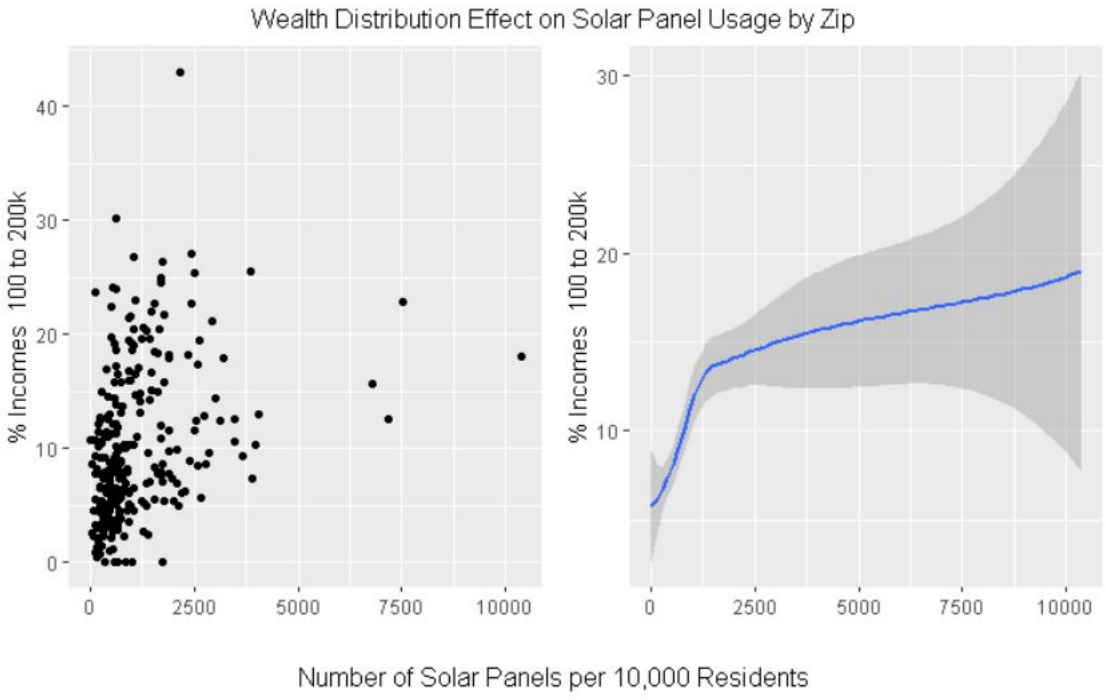


Figure 4: Income100to200 vs Solar PV

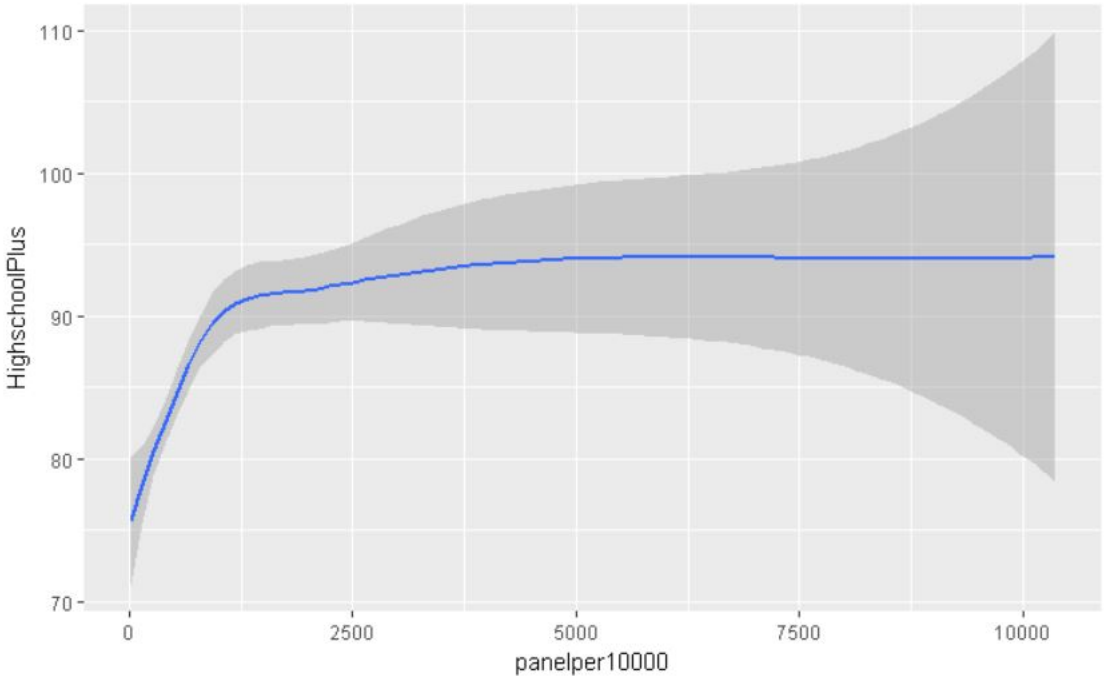


Figure 5: Education Effect

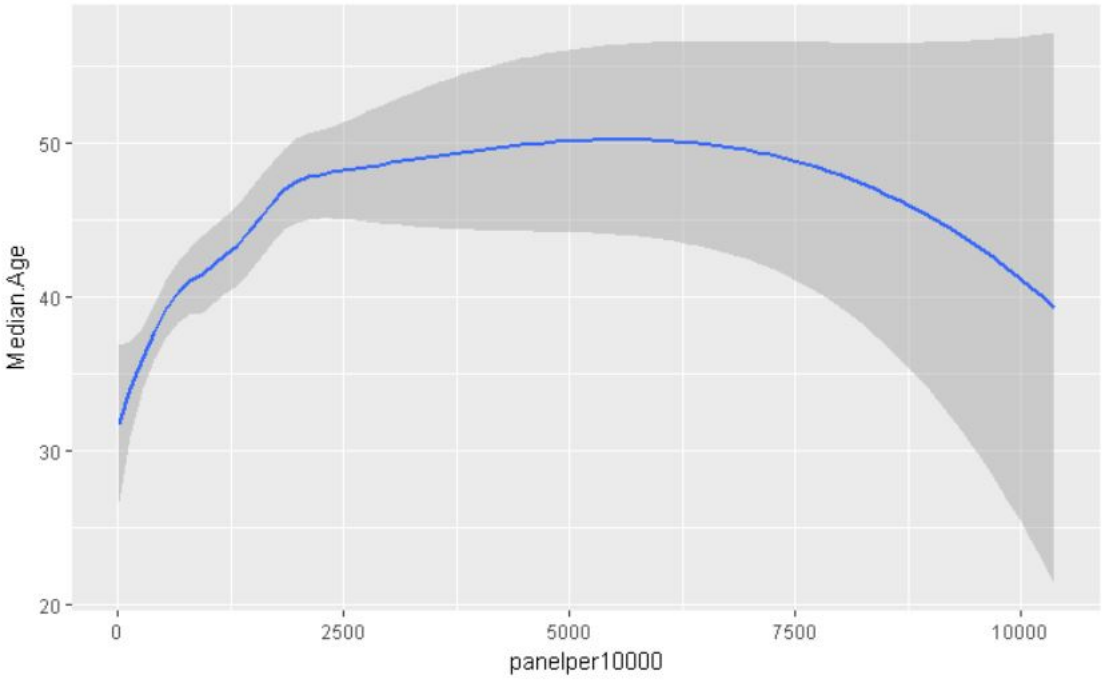


Figure 6: Median Age Effect

	(28.529)
AVGret25to50	35.530
	(29.609)
AVGret50to75	139.425***
	(33.427)
AVGret75to100	30.716
	(41.513)
AVGret100to200	111.908**
	(44.724)
White	-0.004
	(0.009)
Total.Pop	0.009
	(0.007)
Median.Age	28.222***
	(7.956)
Urban	19.160
	(323.255)
Rural	208.152
	(247.289)
HighschoolPlus	-14.708
	(11.848)
Constant	-4,893.952*
	(2,936.870)
Observations	270
R2	0.263
Adjusted R2	0.231
Residual Std. Error	1,051.195 (df = 258)
F Statistic	8.358*** (df = 11; 258)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 7: Regression Results