Katie Lee
12/12/20

Analysis of Participation in Providence Water's 0% Interest Loan Program

Background

Starting in 2007, lead levels in Providence drinking water have routinely exceed the federal limit set by the Lead and Copper Rule of 1991.[1] According to the rule, if levels exceed 15 parts per billion (ppb), the system must take action to control corrosion. From 2009 to 2013, lead levels spiked to double the federal action limit to 30ppb. Recently, in May of 2020, Providence Water announced another spike in lead levels, though the areas affected were unspecified.[2] Decreasing lead levels is particularly critical for child development, infant and maternal health, and particularly the neurodevelopment of infants. According to the Rhode Island Department of Health, about 10% of children in Providence with blood lead levels high enough to qualify for a required home inspection also had high lead levels in their tap water.[3] Historically, lead paint was the primary cause of exposure. However, as regulation of lead paint becomes more robust, the proportion of exposure attributed to lead in water only grows each year. Houses built before the 1940s are at a high risk for having lead lines across the country. In Providence, a city with the fourth oldest housing stock in the nation, about 43% of the houses were built before 1940.[4]

Providence Water, which delivers water to homes in Providence and surrounding towns, has taken measures to mitigate the harms of lead. They've added corrosion inhibitors to water to decrease lead solubility, replaced public lead lines, and suggested flushing pipes at home. The problem of lead in water will nonetheless persist as long as private side lead lines remain. While residents are not compelled to replace their own lead pipes, Providence Water has offered financial assistance to homeowners to lower financial barriers to private lead line replacements. The 3-year 0% Loan Program, funded by the RI Infrastructure Bank, offers an incentive for residents to hire private contractors to replace their own lead lines. The average loan is about $3,000.[5] In order to ensure that this program is effective, it is important to boost awareness and participation amongst all homeowners. So far, 363 residents have participated in the program. This study will be looking for a correlation between distribution of participation among the 24 zip codes served by the Providence Water Utility, and variables of education, income, proportion of households with children, and proportion of households that are occupied by renters that correspond with each zip code.

Expectations

The question is centered on the idea that the listed variables may have a relationship with program awareness or participation. Conventional expectations would suggest that more years of education may increase awareness and procedural knowledge that make loan applications easier. In the study "What is education's impact on civic and social engagement?", David Campbell used the European Social Survey, and found that increased education within nations correlated with an increase in civic engagement. While this is a much larger scale study, with only tentative

---

[1] Frank Carini, "Latest Biannual Test Finds Lead in Water," EcoRI News, https://www.ecori.org/public-safety/2018/8/31/latest-biannual-test-finds-lead-in-providence-water

[2] Madeline List, "High Lead Levels found in water," Providence Journal, https://www.providencejournal.com/news/20200518/high-lead-levels-found-in-water-in-some-providence-homes

[3] Rebecca Renner, "Exposure on Tap," Environ Health Perspect, *US National Library of Medicine*, (2010) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2831942/

[4] EPA, "Lead Poisoning Prevention," https://www3.epa.gov/region1/eco/uep/pdfs/5yr_rpt_pdfs/caseprovid.pdf

[5] Providence Water, Lead free is the way to be, https://www.provwater.com/water_quality/lead-center/loan

results, Campbell expresses the difficulty of determining a causal relationship, writing "Perhaps education only appears to have an effect, when the real causal mechanism lies elsewhere." For this reason, it is important to account for the potential spurious results that failed to account for other intervening or interacting variables, such as race and age. Nevertheless, according to Campbell's conclusion, "Two independent studies have examined natural experiments… and found that formal education does appear to have a truly causal relationship on civic and social engagement."[6] In this study we are looking to find whether education and loan program participation are related. Since loan program participation requires knowledge about government programs and resources, we expect to find that more years of education will be related to high participation in the program.

Additionally, low-income households with children may be more motivated to participate as well. The Center for Social Development published a study conducted by Amanda McBride, "Civic Engagement among Low-Income and Low-Wealth Families," wherein two potential factors for this relationship were determined. First, "Forms of community involvement such as coaching sports, leading Girl Scouts, and volunteering to help in the classroom or on field trips represent important civic roles, impacting the functioning of vital civic institutions." Essentially, parents may receive more exposure to community and its resources through the connections formed through their kids. This may be particularly true if programs that could benefit children, such as this loan program, are promoted through channels that specifically target parents. The study also stated that parents may have increased interest in applying for the loan because, "when work and family demands are at their height, it may be that there is not a quantitative decrease in civic activity among parents but a shifting of the context in which they engage."[7] The context being added concerns about potential harm caused by lead pipes.

Therefore, we expect counties with low-income households with children to participate at higher rates in the loan program. Education may also have a role, with more years of education increasing the likelihood of participation.

Data

In order to find whether or not these variables are correlated with participation in the program, this study will use two sources of data. First, Providence Water has provided the raw data, aggregated and anonymized by zip code, of loan program participants. Of the 24 zip codes served, only 14 have participants, with a total of 363 individual loans granted. The data was last updated on September 30, 2020.[8] Second, to determine the breakdown of education, income, race, family size, and age, the study will be using the 2018 American Community Survey data conducted by the US Census Bureau. This data is comprehensive and reliable, often cited by businesses, journalists, advocacy groups, and government agencies in both research and reporting. It provides four categories of data, all of which will be useful for this study. Social Characteristics provide "Family Households" and "Educational Attainment," Economic Characteristics provide "Income and Benefits", and Demographic Characteristics provide "Age"

---

[6] David Campbell, "What education's impact on civic and social engagement," OECD, (2006) http://www.oecd.org/education/innovation-education/37425694.pdf

[7] Amanda McBride, "Civic engagement among Low-Income Families," Center for Social Development, (2004) https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=1494&context=csd_research

[8] "Private Side Lead Service Replacements," Providence Water, 9/30/20

and "Race."[9] These can all be tabulated by zip code, which makes it compatible with the Providence Water data.

The data are estimates based on a sample of 5900 household respondents, and 750 individual respondents. The sampling took place in two stages. In the first stage, selection used systematic assignments of new addresses to sub-frames and identified appropriate sub-frames associated with the year's samples. The second stage sample selected the sample from the selected sub-frames systematically. Group quarters, facilities that house groups of people such as dorms, nursing facilities, and group homes, were also assigned to sub-frames. People housed in GQs were also eligible to be interviewed, with the same probability of selection.[10] According to the coverage rates published alongside the 2018 ACS, 99.3% of housing units were covered, along with 91% of the general population.[11] The overall sample of zip codes provided are all 90 zip codes in Rhode Island. Our analysis sample will be the 24 zip codes served by Providence Water.

The dependent variable we are studying is rates of participation in Providence's lead pipe replacement loan program. This participation is divided by area zip code, and the data offers the exact number of households in each zip code that have utilized this loan program. In order to make the values comparable, I calculated the rate of participation from the raw participation numbers. In order to do this, I pulled data on the number of households in each zip code from the 2018 ACS, and divided the raw participation number by the number of households. The mean rate of participation was .26% with a standard deviation of .533.

variable	observations	mean	standard deviation	minimum	maximum
rate of participation	19	0.00263	0.00533	0	0.01767

Table 1: Description of dependent variable statistics

[9] US Census Bureau, ACS Information Guide, https://www.census.gov/content/dam/Census/programs-surveys/acs/about/ACS_Information_Guide.pdf

[10] US Census Bureau, ACS Design and Methodology Chapter 4: Sample Design and Selection, (2014) https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_ch04_2014.pdf

[11] US Census Bureau, ACS Coverage rates, (2018), https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/coverage-rates/index.php

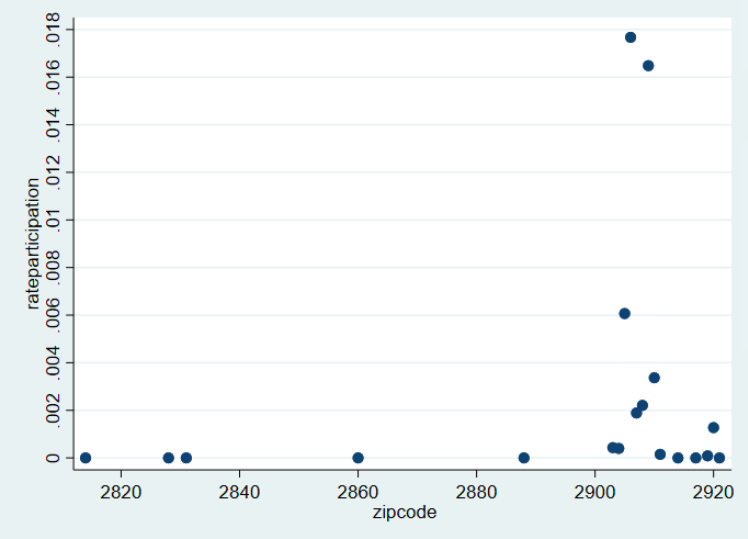

Figure 1: Distribution of participation rates across zip codes

The sample for independent variable data variables was taken for the 2018 American Community Survey by the US Census Bureau. 5,900 housing units were interviewed in Rhode Island alongside 750 individual interviews. According to the reports of coverage, this sample should represent 99.3% of households and 91.4% of individuals.[12] The sample of zip codes covers the areas served by Providence Water Utility, with only 20 observations. There are four independent variables that are used in this research project. Two key variables of interest are education and income.

In order to measure education, there are two variables taken into account. First is percentage of zip code population with at least a high school degree. Second, we include a variable for percentage of zip code population with at least a college degree. By splitting the factor of education into two different variables will lead to a more nuanced understanding of the relationship between the dependent variable, rate of participation, and education. This is an important relationship because if zip codes that have lower rates of educational attainment have lower rates of participation, this could inform how the water utility promotes the loan program and forms its accessibility. The mean percentage of residents in a zip code with at least a high school degree is 86.68% with a standard deviation of 7.69%. The mean percentage of residents in a zip code with at least a college degree is 31.49% with a standard deviation of 14.79%.

The second independent variable we are interested in is median income. Once again, this information was obtained by the 2018 ACS data. The mean of median income in each zip code is $84,498.42 with a standard deviation of $96,445.01. Income varied greatly among zip codes. Understanding the relationship between income and loan program participation will reveal the demographics of those who are and are not taking advantage of the program. If certain

[12] https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/coverage-rates/index.php

populations are left out, the utility can more specifically target those populations and increase accessibility of the program and thus pipe replacements in general.

The third independent variable is percentage of housing units occupied by renters in a zip code. Providence Water has previously expressed concern that renters are a demographic that know very little about the utility and the loan program opportunity. Outreach about the program is primarily done through mailers attached to water bills which are sometimes not received by renters. With this expectation, we will be looking at the relationship between percentage of renters and rates of participation. The median percentage of housing units occupied by renters was 42.36% with a standard deviation of 21.82%.

Our final independent variable is potentially a control variable. One of the expectations we held in the proposal was that households with children are more likely to apply for a loan. First, the motivation to do so may be higher as children are an at-risk population for lead exposure. Second, having children can expose guardians to more channels of communication that deliver information about resources such as the loan program by Providence Water. The variable measured here is percentage of households with one or more children under the age of 18. The mean percentage was 29.07% with a standard deviation of 8.04%. By understanding the influence this variable can have on relationships between other independent variables and the dependent variable will ensure we are taking into account potential spurious relationships.

variable	observations	mean	standard deviation	minimum	maximum
highschool	19	0.8668	0.0769	0.709	0.98
college	19	0.3149	0.1479	0.137	768
medianinc	19	84498.42	96445.01	25425	471777
renters	19	0.4236	0.2182	0.113	0.819
children	19	0.2907	0.0804	0.081	0.477

Table 2: Description of independent variable statistics

Regression:

With no p-values under the alpha of .05, there are no significant relationships between rates of participation and the four independent variables. We were challenged with a small sample and limited data points. Of these insignificant relationships, some had much larger p-values than others. Having children and a college education had the smallest p-values and could be further explored.

Katie Lee
12/12/20

Source	SS	df	MS	Number of obs	=	21
				F(6, 14)	=	1.00
Model	122.683526	6	20.4472543	Prob > F	=	0.4615
Residual	285.602189	14	20.4001563	R-squared	=	0.3005
				Adj R-squared	=	0.0007
Total	408.285714	20	20.4142857	Root MSE	=	4.5167

participationnum	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
incnum	-.3318193	.2004863	-1.66	0.120	-.7618197	.0981811
owneroccupiedpercentnum	.4861614	.2983619	1.63	0.126	-.1537612	1.126084
renteroccupiedpercentnum	.2736954	.3773939	0.73	0.480	-.5357341	1.083125
highschoolnum	-.2662151	.2747369	-0.97	0.349	-.8554672	.323037
collegenum	-.2450839	.2360693	-1.04	0.317	-.7514022	.2612344
childrennum	-.0394722	.2500041	-0.16	0.877	-.5756776	.4967332
_cons	9.26723	3.494608	2.65	0.019	1.77204	16.76242

Table 3: Regression of participation with all our independent variables

Regression Diagnostic

The first modified model after the regression is one without the outlier of zip code 02912. This zip code was dropped as an outlier because it is the zone of Brown University, which is primarily Brown University students. As the student body does not participate in the loan program, it does not belong in the dataset. Having dropped the outlier of zip code 02912, there is a significant relationship between income and participation rates. For income, P = .038 which is less than the alpha of .05.

Regression without Zip Code 02912

Source	SS	df	MS	Number of obs	=	20
				F(6, 13)	=	2.80
Model	194.87477	6	32.4791283	Prob > F	=	0.0566
Residual	150.92523	13	11.6096331	R-squared	=	0.5635
				Adj R-squared	=	0.3621
Total	345.8	19	18.2	Root MSE	=	3.4073

participationnum	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
incnum	-.3486929	.1513248	-2.30	0.038	-.6756103	-.0217755
owneroccupiedpercentnum	.1288142	.2483318	0.52	0.613	-.4076741	.6653025
renteroccupiedpercentnum	-.0942765	.30451	-0.31	0.762	-.7521302	.5635773
highschoolnum	-.2452893	.2073482	-1.18	0.258	-.6932379	.2026593
collegenum	-.2507536	.1780947	-1.41	0.183	-.6355038	.1339966
childrennum	-.0307787	.1886164	-0.16	0.873	-.4382596	.3767022
_cons	18.37692	3.755488	4.89	0.000	10.26369	26.49016

Table 5: Regression of participation with all our independent variables having dropped zip code 02912

Katie Lee
12/12/20

Next we ran individual regressions for each independent variable. This was in order to determine whether any individual variable had a significant relationship with participation. Therefore, we produced four more models that demonstrate the regression between participation and income, percentage occupied by owners, high school education and college education individually. While the relationship between education and participation rates remained insignificant, income once again had a significant relationship with participation rates when its regression was modified. Additionally, percentage occupied by owners had a newly significant relationship with participation rates when regressed individually. For income, the P = .014 and for owner occupied the P = .033.These are both below the alpha level of .05 and thus we can reject the null hypothesis that they have no significant impact on participation rates.

Income

Source	SS	df	MS	Number of obs	=	20
				F(1, 18)	=	7.41
Model	100.873684	1	100.873684	Prob > F	=	0.0140
Residual	244.926316	18	13.6070175	R-squared	=	0.2917
				Adj R-squared	=	0.2524
Total	345.8	19	18.2	Root MSE	=	3.6888

participat~m	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
incnum	-.3894737	.1430443	-2.72	0.014	-.6899987	-.0889487
_cons	13.57895	1.84022	7.38	0.000	9.712789	17.44511

Table 5: Regression of participation with income

Owner

Source	SS	df	MS	Number of obs	=	20
				F(1, 18)	=	5.35
Model	79.247648	1	79.247648	Prob > F	=	0.0327
Residual	266.552352	18	14.808464	R-squared	=	0.2292
				Adj R-squared	=	0.1863
Total	345.8	19	18.2	Root MSE	=	3.8482

participati~m	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
highschoolnum	-.3877086	.1675974	-2.31	0.033	-.7398177	-.0355996
_cons	13.05463	1.913842	6.82	0.000	9.033795	17.07546

Table 6: Regression of participation with percent occupied by owners

Highschool

Katie Lee
12/12/20

Source	SS	df	MS	Number of obs	=	20
				F(1, 18)	=	0.46
Model	8.70600414	1	8.70600414	Prob > F	=	0.5040
Residual	337.093996	18	18.7274442	R-squared	=	0.0252
				Adj R-squared	=	-0.0290
Total	345.8	19	18.2	Root MSE	=	4.3275

participat~m	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
collegenum	-.1200828	.1761208	-0.68	0.504	-.4900989	.2499333
_cons	10.45093	2.20503	4.74	0.000	5.818336	15.08353

Table 7: Regression of participation with percent of population with at least high school education

College

Source	SS	df	MS	Number of obs	=	20
				F(1, 18)	=	0.30
Model	5.65852456	1	5.65852456	Prob > F	=	0.5910
Residual	340.141475	18	18.8967486	R-squared	=	0.0164
				Adj R-squared	=	-0.0383
Total	345.8	19	18.2	Root MSE	=	4.347

participationnum	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
owneroccupiedpercentnum	.0845818	.1545678	0.55	0.591	-.2401531	.4093167
_cons	8.080789	2.100928	3.85	0.001	3.666902	12.49468

Table 8: Regression of participation with percent of population with at least college education

Discussion

Our sample has limited statistical power because of the small scale of the sample size. However, by regressing participation rates multiple times with each individual independent variable, we have thoroughly observed potential relationships between income, owner occupation rates, and education with participation. Additionally, each of these individual independent variables has sufficient range to provide for a statistically diverse analysis. For instance, the standard deviation for income is 96,445.01, while standard deviation for rate of home occupied by renters is .21. (See Table 2)

We also analyzed the kernel density of the residuals of our regression, which is displayed against a normal distribution below. Evidently, the distribution of residuals from our regression is very similar to a normal distribution. This satisfies the assumption that residuals are not skewed nor indicative of an uneven distribution of the residuals.

Katie Lee
12/12/20

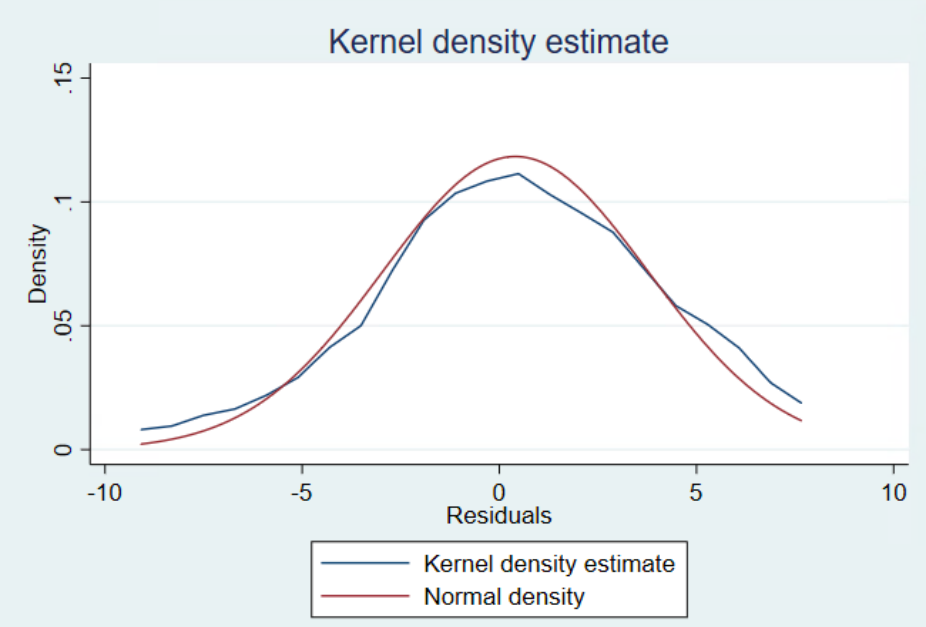

Figure 2: Kernel Density Distribution

Additionally, we measured collinearity between the independent variables. This was conducted by measuring a VIF, or variance influence factor. If the VIF is too high, the standard errors in the kernel density measurement may be inflated or unstable. However, the VIFs for the independent variables in the regression for our model are relatively low. The mean VIF is a 2.92, which is far below the threshold of 10 that would warrant further investigation. Therefore, collinearity did not impact the regression, meaning independent variables do not relate to each other to affect each other's impact on the dependent variable.

Variable	VIF	1/VIF
rentero~tnum	6.31	0.158525
owneroc~tnum	4.20	0.238015
childrennum	2.10	0.476432
highschool~m	1.95	0.512203
collegenum	1.65	0.606261
incnum	1.31	0.762389
Mean VIF	2.92	

Table 9: VIF statistic for all independent variables except percent occupied by renters

Some of these statistics support our expected findings. The significant relationships that came from regressing income and rate of owner occupation after deleting the outlier of our observation of zip code 02912, and regressing individual independent variables confirm our hypothesis that income and renter rates affect participation rates. However, contrary to our expectations, education did not have a significant impact on participation rates. The k density

and the VIF show that the regression is still stable despite a lack of power due to small sample size.

Conclusion:

Through this study, we attempted to determine relationships between various demographic factors and participation in the Providence Water 0% Loan Program in order to better understand how to promote the program and bolster participation. While a geographic distribution is readily available, understanding the populations within these zip codes is crucial for tailoring outreach initiatives.

Going forward there are multiple areas of potential study to be pursued. These independent variables are very generalized and could be individually explored in relation to participation rates. Some qualitative research could bolster understanding of how best to reach residents of Providence, particularly in targeting specific income groups or education levels. Using the initial findings of this study, it would be possible to pursue a well-informed focus-group study that covers the range of income and education that were not explored in depth by this study. Additionally, while the original regression did not produce substantial relationships, the reason behind the change after individual regressions is an area of potential study.

Given the results, it is possible to leverage the power of statistical analysis to inform outreach and engagement initiatives for organizations such as Providence Water. In order to ensure that individuals receive an equal opportunity to apply to the program, thus promoting the replacement of private side lead pipes, residents require accessibility to valuable resources such as the loan program. In addition to the current successful promotions of the program, this analysis could inform future promotion as well. Rather than upscaling the magnitude of outreach, this analysis would inform direction and methods, particularly looking into target demographics who are likely to miss promotions. This may be due to renter status or income level.

Overall, the analysis conducted by the regressions in this study provided preliminary insight into the demographics of program participants, as well as the seeds for future studies.