12-4-2019

Air quality and environmental justice: An analysis of county-level data in the United States

Alyssa Tittle
Western Michigan University, alyssa.tittle@gmail.com

Air quality and environmental justice:

An analysis of county-level data in the United States

Alyssa N. Tittle

Western Michigan University

Abstract

This paper analyzes the relationship between air pollution and poverty and race/ethnicity at the county level in the United States. Using air pollution data from the Environmental Protection Agency and population characteristics data from the Census Bureau to model relationships in RStudio, results show that race/ethnicity, population, and median value of owner-occupied housing units are the most statistically significant variables associated with air quality.

1. Introduction

Environmental justice is defined by the United States Environmental Protection Agency (EPA) as a societal goal in which everyone, regardless of “race, color, national origin, or income,” enjoys equal protection from threats to their environment and health as well as equal access to the decision-making processes that affect protection in all areas of daily life (2019). Environmental injustice might then be defined as the disproportionate exposure of the poor and certain racial/ethnic communities to environmental hazards as well as disproportionate representation in policymaking. Those disproportionately burdened by environmental hazards may experience lack of access to clean water, healthy food, adequate housing, and clean air, for example. This paper focuses on the air pollution aspect of environmental hazards, analyzing the relationship between median Air Quality Index (AQI) values and county-level characteristics in the United States.

The paper is organized as follows: The literature review summarizes the understanding of environmental justice in various studies, especially in relation to air pollution, poverty, and race/ethnicity. The following section describes the population and air pollution data, with a brief discussion on sampling bias. The findings of the models are discussed in the results section, followed by a conclusion.

2. Literature Review

The Clean Air Act of 1970 (and its amendments) mandates that the EPA set National Ambient Air Quality Standards (NAAQS) for six criteria air pollutants (EPA, 2016). Included in these criteria air pollutants are particulate matter 2.5 (airborne particles that have a diameter of less than 2.5 micrometers) and ozone (a highly reactive gas), two measures of primary interest to researchers to gauge air quality across the United States. Criteria air pollutants also include carbon monoxide, lead, sulfur dioxide, and nitrogen oxide. Studies on environmental justice rely on data collected by the EPA as well as population characteristics from the Census Bureau, such as race and income, to explore inconsistencies across different groups in the application of environmental protection.

The environmental justice movement in the United States gained momentum in 1987 when the "Toxic Wastes and Race in the United States" report was released by the United Church of Christ Commission for Racial Justice (CRJ, 1987). The CRJ reported on two cross-sectional studies on the relationship between population demographics and commercial hazardous waste facilities and uncontrolled hazardous waste sites. The CRJ reported communities with the highest percentages of black and Hispanic residents to also have the greatest numbers of waste facilities across the nation. Low-income communities showed a similar relationship but the correlation for proximity to waste sites was not quite as high as for race/ethnicity. The CRJ also revealed that "three out of every five black and Hispanic Americans lived in communities with uncontrolled toxic waste sites" (CRJ, 1987). This report showed clear relationships between hazardous environmental conditions and minority communities.

Location is also important to consider in regard to present-day EPA air monitoring sites. Miranda, Edwards, Keating, and Paul (2011) find that the EPA concentrates its monitoring in densely populated areas, typically near highways or industrialized centers. This study concludes that observed differences in population characteristics are likely linked to the EPA's method of data collection in which air pollution monitors are located near significant masses of people. Miranda et al. finds, as a result, that the EPA fails to gather information in many rural areas of the United states, leaving gaps in pollution data for the older, non-Hispanic white population.

Studies by Benzhaf et al. (2019), Bento et al. (2015), and Miranda et al. (2011) find that exposure to air pollutants varies across groups according to income and race/ethnicity, although there are differences in findings. Benzhaf et al. (2019) suggests three key explanations for environmental injustice in the United States: issues related to Coasean bargaining, siting, and sorting. The Coase Theorem states that private settlements over property rights will lead to socially optimal outcomes where all parties are better off in the end. However, parties disadvantaged due to race, language barriers, and/or poverty may be unable to adequately express their value for a cleaner environment, thus leading to inefficiency. In addition, disparities in environmental conditions could be addressed by reshaping zoning rules within communities, although long-run housing prices may bring low income households to polluted areas through resorting.

Benzhaf et al. (2019) also makes the case that disparities in environmental conditions could be addressed by reshaping zoning rules within communities, although housing decisions for individuals with constrained budgets may lead them to neighborhoods with lower environmental quality anyway. Housing market prices tend to reflect neighborhood features and those with less desirable features (e.g., pollution) tend to be less expensive. Zoning rule changes may initially improve issues related to disproportionate exposure to pollution, but long-run housing prices may bring low income households to more polluted areas through resorting.

Benzhaf et al. (2019) states that environmental injustice issues related to siting and sorting require addressing the larger issue of income inequality, pointing to income as a major driving force behind uneven exposure to air quality.

Another study related to income inequality is research on the 1990 Clean Air Act Amendments (CAAA) by Bento et al. (2015). Bento et al. (2015) finds that the CAAA created benefits for those in the lowest quintile of the income distribution that amounted to over double the benefits received by those in the highest quintile. These benefits amount to roughly 0.3% of the lowest quintile's yearly income, though benefits primarily include short-term clean-up efforts and may reflect the higher exposure of poor communities to begin with.

Contrary to the conclusions drawn by Benzhaf et al. (2019) and Bento et al. (2015) is a study conducted by Miranda et al. (2011) that suggests environmental injustices are more consistent with race/ethnicity differences than measures of income. Using two criteria air pollutants as determined by the EPA, particulate matter and ozone, they use multivariate analysis to show that counties with the worst incidences of particulate matter have higher rates of poverty. However, this study finds that counties with worse incidences of ozone pollution have lower rates of poverty. Though findings for particulate matter and ozone in relation to income are conflicting in this study, results regarding race and air quality tend to be more consistent. Miranda et al. (2011) finds that the non-Hispanic black population is more likely to live in areas with higher exposure to daily particulate matter and ozone. The Hispanic population is also more likely to live in areas with higher exposure to daily particulate matter, but less likely to experience ozone pollution. Though poverty and pollution present an inconsistent relationship here, this research consistently finds that the non-Hispanic black population is subject to worse air quality than other groups.

3. Data

The ‘countyComplete’ dataset from the RStudio ‘openintro’ package contains 2010 Census Bureau data by county that provides population characteristics for this paper. Characteristics include ‘poverty’, or the percent of the population below the poverty line from 2006 to 2010, as well as the median home value of owner-occupied units over the same period. Also included is the population in 2010 and population density, or persons per square mile. Finally, I include the percent of the population that is black and the percent of the population that is Hispanic. To be more specific, the 2010 census does not list ‘Hispanic’ as an option when asking about a person’s race. This particular census classifies ‘Hispanic’ as an origin, or ethnicity, which is why this distinction is made throughout this paper (Humes, Jones, & Ramirez, 2011).

Figure 1. U.S. black population by county in 2010

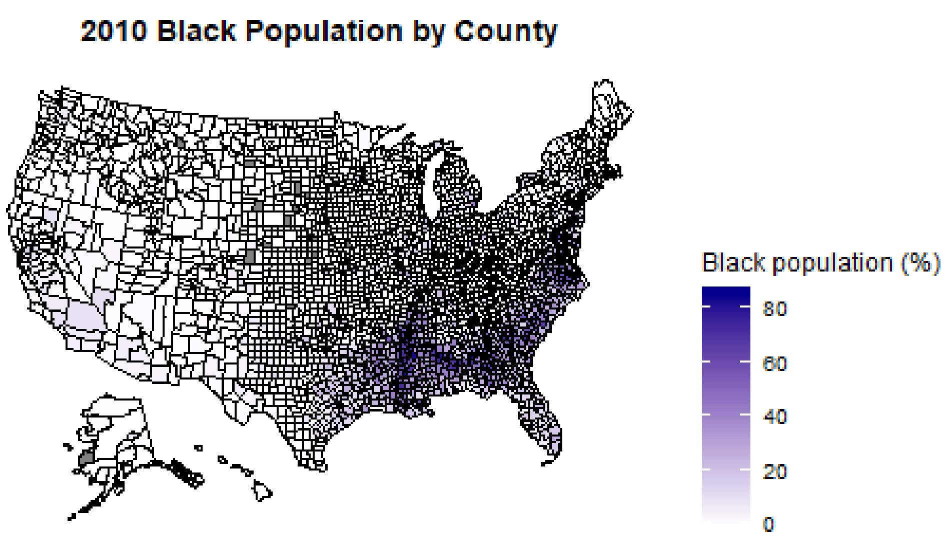
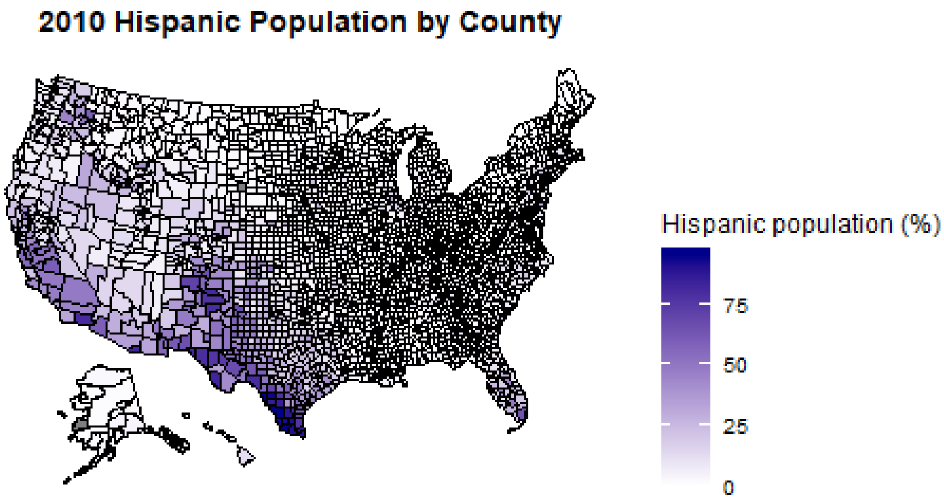


Figure 2. U.S. Hispanic population by county in 2010



Pollution data is taken from the EPA's AirData Air Quality Index Summary Report that provides summaries for AQI values by county in 2012. AQI values, gauges of overall air quality, are assigned to each county based on the concentrations of the six criteria air pollutants, ground-level ozone, particulate matter (2.5), carbon monoxide, lead, sulfur dioxide, and nitrogen dioxide. A higher AQI value implies more air pollution. The US EPA states that most AQI values range 0 to 200, with only about 0.3% of counties with AQI values 201 or above (2018).

The main dependent variable modeled, median AQI, represents the median AQI value for each county where half of the values in 2012 were less than or equal to the median and half of the values were equal to or greater than it. I also use unhealthy AQI days for sensitive groups as the dependent variable in one of my models which is defined as the "number of days in the year having an AQI value 101 through 150." Finally, unhealthy days, or "the number of days in the year having an AQI value 151 through 200," is also included as a dependent variable in one model. (US EPA, 2018).

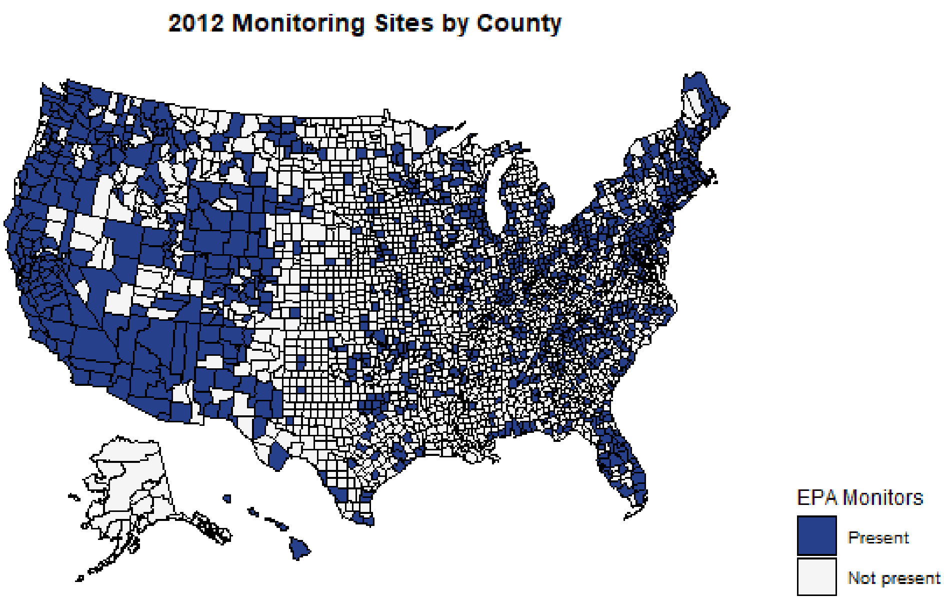
Table 1. Descriptive statistics for each variable used

Descriptive Statistics

Statistic	N	Mean	Min	Max
Median AQI	1,071	38.286	0.000	171.000
Unhealthy AQI Days: Sensitive Groups	1,071	8.533	0.000	134.000
Unhealthy AQI Days	1,071	1.505	0.000	221.000
Poverty	3,143	15.499	0.000	53.500
Median home value	3,143	132,544.900	0.000	1,000,001.000
Population 2010	3,143	98,232.750	82.000	9,818,605.000
Density	3,143	259.322	0.000	69,467.500
Black	3,126	8.931	0.000	85.700
Hispanic	3,143	8.284	0.000	95.700

Table 1 shows descriptive statistics for each variable. The population characteristic variables have observations for all 3,143 counties recorded in the census, except for the 'black' variable, which is missing 17 observations. We see that air pollution variables have only 1,071 observations, which makes sense, as the EPA does not have air pollution monitoring sites in every county in the United States. As found by Miranda et al. (2011), the EPA tends to have air monitors in densely populated, urban areas, and so this data likely exhibits bias. The distribution of air pollution monitoring sites is shown below in figure 3.

Figure 3. 2012 EPA air monitoring sites by county



The pollution data comes from non-randomly selected sampling points, so the issue of biasness in this data is corrected using the Heckman correction method. The Heckman correction package in RStudio works through two stages, utilizing a selection (probit) equation and an outcome equation in a two-step method. The results from the selection equation are used to estimate an inverse Mills ratio that is factored into the outcome equation as an additional explanatory variable that corrects for bias. The inverse Mills ratios in each of my regressions have negative coefficients which implies that, without the sampling correction, my estimations would be downward-biased (Setzler, 2014).

4. Results

The EPA's monitoring sites tend to be located in densely populated, urban, and often poor areas, and initially, this data may not have been representative of the population as a whole. Simply put, to solve the issue of sampling bias, we need better information and a more random means of collecting data. To correct for bias issues, I use Heckman's standard sample selection model with each of my regressions and output their results in table 4. Estimations before the Heckman correction can be seen in table 2 below, and the probit model that estimates where the monitors are located in model 1 is shown in table 3.

Table 2. Regression results before sampling correction

Regressions Before Heckman Correction

	Dependent variable:				
	Median AQI			Sensitive	Unhealthy
	All variables	Without race	Without income/wealth	All variables	All variables
	(1)	(2)	(3)	(4)	(5)
Poverty	-0.2498***	-0.0623		-0.1410	0.0871
	(0.0902)	(0.0801)		(0.0919)	(0.0594)
Hispanic	0.0986***		0.0624**	0.1077***	0.0212
	(0.0323)		(0.0293)	(0.0329)	(0.0213)
Black	0.1495***		0.1153***	-0.0621*	-0.0331
	(0.0370)		(0.0348)	(0.0377)	(0.0244)
Median home value	-0.000004	-0.000001		-0.00001	0.00001*
	(0.000004)	(0.000004)		(0.000004)	(0.000003)
Density	-0.0001	-0.0001	-0.0002	-0.0003**	-0.0002**
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Population 2010	0.00001***	0.00001***	0.00001***	0.00001***	0.000003***
	(0.000001)	(0.000001)	(0.000001)	(0.000001)	(0.000001)
Constant	39.2855***	37.9922***	35.6136***	8.9601***	-1.1988
	(1.6502)	(1.6106)	(0.5495)	(1.6797)	(1.0871)
Observations	1,000	1,000	1,000	1,000	1,000
R^2	0.1155	0.0963	0.1085	0.1629	0.0467
Adjusted R^2	0.1102	0.0927	0.1049	0.1579	0.0409
Residual Std. Error	12.0040	12.1212	12.0392	12.2188	7.9084
F Statistic	21.6148***	26.5187***	30.2832***	32.2157***	8.1077***
Note:					*p < 0.1, **p < 0.05, ***p < 0.01

Population size consistently shows a positive correlation and statistical significance in my regressions. This trend is unsurprising given that a larger population size might be related to more human activity that results in more polluting behaviors, such as more people driving or having their energy supplied by non-renewable, polluting sources. More densely populated areas may also be more likely than sparsely populated towns to have manufacturing plants and power plants.

Table 3. Model 1 probit selection equation

Probit Selection Model for Model 1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.5010000	0.1253000	-11.972	< 2e-16 ***
Coal capacity	0.0001727	0.0000824	2.095	0.0363 *
Petroleum capacity	0.0008573	0.0011550	0.742	0.4579
Gas capacity	0.0000532	0.0000652	0.816	0.4145
Poverty	0.0107500	0.0058560	1.835	0.0666 .
Hispanic	-0.0042490	0.0025980	-1.636	0.1021
Black	-0.0162700	0.0026900	-6.049	1.67e-09 ***
Median home value	0.0000040	0.0000005	8.263	2.24e-16 ***
Density	-0.0001847	0.0000200	-9.216	< 2e-16 ***
Population 2010	0.0000068	0.0000004	15.400	< 2e-16 ***

Note: Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 4. Regression results after sampling correction

Regressions After Heckman Correction

	Dependent variable:				
	Median AQI			Sensitive	Unhealthy
	All variables	Without race	Without income	All variables	All variables
	(1)	(2)	(3)	(4)	(5)
Poverty	-0.2002**	-0.0366		-0.1134	0.0235
	(0.0987)	(0.0876)		(0.0987)	(0.0302)
Hispanic	0.0472		0.0123	0.0786**	0.0304***
	(0.0387)		(0.0350)	(0.0386)	(0.0117)
Black	0.1589***		0.1451***	-0.0509	-0.0166
	(0.0401)		(0.0371)	(0.0400)	(0.0122)
Median home value	-0.00002***	-0.00002***		-0.00002***	-0.000001
	(0.00001)	(0.00001)		(0.00001)	(0.000002)
Density	0.00004	0.0001	-0.0002	-0.0001	-0.0001***
	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.00004)
Population 2010	0.000004***	0.000005***	0.000004***	0.00001***	0.000003***
	(0.000001)	(0.000001)	(0.000001)	(0.000001)	(0.000000)
Constant	47.3634***	48.7478***	41.3909***	16.4414***	0.5491
	(2.2347)	(2.3119)	(1.1770)	(2.2188)	(0.6651)
Observations	2,610	2,626	2,610	2,610	2,610
rho	-0.5159	-0.6268	-0.4762	-0.4552	-0.0940
Inverse Mills Ratio	-6.2701***	-7.9970***	-5.7450***	-5.4463***	-0.3278
	(1.1069)	(1.1482)	(0.9610)	(1.0993)	(0.3297)
Note:					*p < 0.1, **p < 0.05, ***p < 0.01

Poverty is only statistically significant in model 1 and exhibits a negative relationship with median AQI. This coefficient suggests that as we increase the percentage of the population under the poverty level, we might find a decrease in median AQI, or an improvement in air quality at the 95% confidence interval. We can also refer to the coefficients on the median value of owner-occupied housing units variable for more information about wealth/income.

Median home value is used in these models as a proxy for wealth since owning a more valuable house implies the ability to afford it. The study by Benzhaf et al. (2019) finds that areas with better environmental features (such as clean air) tend to have homes with a higher value. Though model 5 exhibits insignificance, the highly significant and negative coefficients of regressions 1, 2, and 4 suggest that if we increase the value of a home, we would expect to see a decrease in the median AQI value, or an improvement in air quality, as previously suggested by Benzhaf et al. (2019).

Race is highly significant in models 1 and 3 and carries a positive relationship in both instances, though it is statistically insignificant in models 4 and 5. For example, in reference to model 1, if we increase the percentage of the population that is black by 10, we would expect to see an increase in median AQI value by 1.59, all else constant. These coefficients are consistent with the findings of the study by Miranda et al. (2011) that non-Hispanic black communities are more likely to be exposed to higher rates of pollution.

The Hispanic population and air pollution relationship is also worth considering, although it does not have quite as significant of a relationship as the black community shows. In each regression there is a positive relationship between air pollution and the Hispanic population, though it is only significant at the 95% confidence interval in model 4 and at the 99% confidence interval in model 5. In general, the higher percentage of the population that is Hispanic, the greater we would expect the median AQI value to be.

5. Conclusion

One way the EPA aims to provide protection from environmental hazards is through regulation of ambient air quality. Air quality regulations are not perfect, however, and many people in the United States are subject to poor environmental conditions. Though the United States monitors and regulates air pollution, it has failed to attain environmental justice nearly 50 years since the implementation of the Clean Air Act. It is important to identify patterns in disproportionate pollution exposure so we may concentrate efforts to improve the health and quality of life for groups saddled with environmental burdens through fair clean-up efforts, enforcement of air pollution standards, and effective environmental policy.

These results suggest race/ethnicity, population, and median value of owner-occupied housing units to be the most significant variables associated with air quality in the United States, although these results vary slightly between models. Future studies might account for geographic variation, differences in industrial development, traffic congestion, types of energy production, and variations in enforcement of air pollution regulations, for example. Further modification of the above models is needed to draw more definitive conclusions to bring about awareness and policy change for the complex relationships that exist between race/ethnicity, poverty, and air pollution in the United States.

References

Banzhaf, H. S., Ma, L., & Timmins, C. (2019). Environmental Justice: Establishing Causal Relationships. *Annual Review of Resource Economics*, *11*(1), 377–398. https://doi.org/10.1146/annurev-resource-100518-094131

Bento, A., Freedman, M., & Lang, C. (2015). Who Benefits from Environmental Regulation? Evidence from the Clean Air Act Amendments. *Review of Economics and Statistics*, *97*(3), 610–622. https://doi.org/10.1162/REST_a_00493

Commission for Racial Justice. (1987). Toxic Wastes and Race in the United States [PDF]. Retrieved from https://www.nrc.gov/docs/ML1310/ML13109A339.pdf

Humes, K. R., Jones, N. A., & Ramirez, R. R. (2011). Overview of Race and Hispanic Origin: 2010 [PDF]. Retrieved from https://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf.

Miranda, M. L., Edwards, S. E., Keating, M. H., & Paul, C. J. (2011). Making the Environmental Justice Grade: The Relative Burden of Air Pollution Exposure in the United States. *International Journal of Environmental Research and Public Health*, *8*(6), 1755–1771. https://doi.org/10.3390/ijerph8061755

Setzler, B. J. (2014). Selection Bias Corrections in Julia, Part 1. Retrieved from https://juliaeconomics.com/tag/inverse-probability/

US EPA. (2018). About Air Data Reports. Retrieved November 3, 2019, from https://www.epa.gov/outdoor-air-quality-data/about-air-data-reports