Environmental Inequality

By Abby Fry, Mauricio Moreno, Dawn Pham

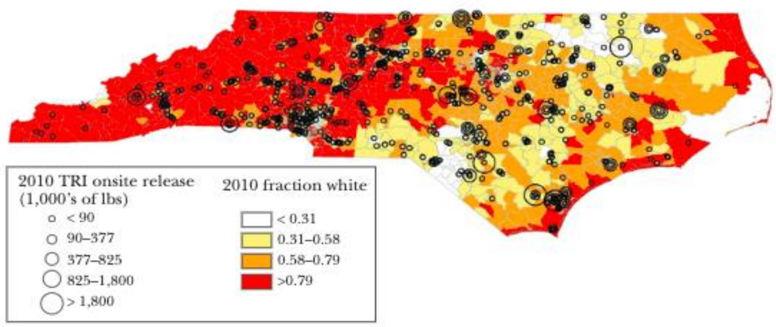
Abstract

This paper studies whether communities of color face greater pollution. Existing studies establish a strong correlation between race and pollution distribution, but do not account for economic factors or other factors that may affect the pollution distribution variable, which is represented by carcinogen released per capita in the county in this study. We show that including such factors depicts a statistically significant association between carcinogen release per capita and percent of minorities using county-level data. Naïve regressions using an OLS model makes it possible to include such omitted variables and identify the strength of the relationships between these variables. Results suggest that in 2018, an increase in carcinogen released per capita has a significant negative relationship with a higher percentage of minorities within a county when economic factors are included in the regression. This relationship is highly robust across samples, variable definitions, and model specifications.

Introduction

In the 1980s, activists and researchers concerned over the distributional impacts of pollution and began using a term to describe the correlation between environmental outcomes and race: *environmental racism.* (Hamilton 1995) Over the year's studies have convincingly shown that poor people and minorities are more likely than other groups to live in polluted neighborhoods. Many research studies suggest that there is a correlation between race and pollution. According to Banzhaf et al. (2019), discriminatory politics and enforcement explains the patterns between polluters and residents. The correlation between race and pollution is depicted in Figure 1, which suggests that minority communities in North Carolina face more pollution exposure and environmental risks compared to other groups in the year 2010. (Banzhaf et al. 2019) The question as to whether high pollution levels in minority communities stems from discrimination has attracted increasing judicial, executive, and legislative attention. (Hamilton 1995) Disadvantaged groups may be living in more polluted areas because to be poor means not having the resources to afford the good things in life—including a clean environment.

Figure 1
Emissions from Large Polluters and Fraction Non-Hispanic White for North Carolina, 2010



Source: Authors using data from the Toxic Release Inventory and US Census.
Note: Using data from the Toxic Release Inventory (TRI), Figure 1 plots emissions from large polluters in North Carolina in the year 2010 with circles, against a heat map of the percentage of the population that is non-Hispanic white at the census tract level.

Source: *Banzhaf, Spencer, Lala Ma, and Christopher Timmins. "Environmental Justice: The Economics of Race, Place, and Pollution." The Journal of Economic Perspectives 33, no. 1 (2019): 185-208.*

In the United states, people of lower income and minority backgrounds experience overall poorer health than their affluent and white counterparts, in part, by a higher concentration of environmental harm and hazardous facilities in their communities. According to a study by the Forum of International Respiratory Societies' Environmental Committee, it is estimated that about "500,000 lung cancer deaths and 1.6 million COPD deaths can be attributed to air pollution, but air pollution may also account for 19% of all cardiovascular deaths and 21% of all stroke deaths". (Schraufnagel et al. 2018, 417) The study also concludes that air pollution yields a host of indirect effects on the development of other serious health conditions. (Schraufnagel et al. 2018) Environmental justice asserts these health disparities. The US Environmental Protection Agency defines "environmental justice" as a fair treatment where "no group of people should bear a disproportionate share of the negative environmental consequences resulting from industrial, governmental and commercial operations or policies," regardless of "race, color, national origin, or income with respect to the development, implementation and enforcement of environmental laws, regulations and policies." (Banzhaf et al. 2019, 186) However, there are still many cases in the U.S. where communities of color and of low income experience a disproportionate share of hazardous exposure and pollution.

In this paper, we revisit the relationship between race and pollution distribution, but focusing on a more recent year, 2018. We investigate the question as to whether communities of color face greater pollution distribution by observing nationwide and individual state data. First, we will investigate counties across the U.S., with a focus on Texas, North Carolina, and Georgia —looking at carcinogen per capita released in the county and seeing whether counties with a higher percent of minorities face more pollution. Second, and more important, there is the potential for omitted variable bias. Some other factors may determine the inequality of pollution distribution.

Our strategy is to include control variables and identify the strength of the relationships between these variables: Median household income, GINI index, a county's political leaning, and how urban the county is. The data on facilities that released toxic materials was taken from the 2018 Toxic Release Index and the demographic data is provided by the 2014-2018 American Community Survey 5-Year Estimates. Ultimately, we decided to represent pollution levels of each county by log of grams of carcinogen released per capita. This would measure the level of carcinogens released by facilities within the county per person. From the demographic data, we generated the percent of the population of the county that are minorities, in which represented communities of color. While naïve regressions are not a panacea for controlling time-invariant unobserved individual characteristics, they are well suited to the investigation of the relationship between race and pollution. The major source of potential bias in a regression of carcinogen released per capita on percent minorities is county-specific, economic factors influencing the amount of toxins released by facilities within a county. If these omitted variables are to be included in the regression, they will be removed as a source of bias. Consider, for example, median household income. According to the standard interpretation of the Coase Theorem, polluting firms will locate where it does the least environmental damage because this is where the cost for externalities is the least. Some externalities that firms consider when deciding on a location includes: “the number of people affected, incomes, property values, and residents’ willingness to pay for environmental amenities.” (Hamilton 1995, 110) Firms choose to locate their polluting facilities in low-income areas where compensation demands and expected liabilities from operation are lower. Unfortunately, residents that remain or move in these areas are mainly low-income minorities who lack the resources to pay for environmental amenities. The idea of including omitted variables, such as median household income, is to consider other factors that may affect the amount of toxins released by facilities within a county and reduce some omitted variable bias.

However, there are some potential threats to validity that does exist in our regression because not every hazardous waste site handles an equal amount of waste, nor does every TRI site emit an equal amount of pollution. Carcinogens also vary in their toxicity. For example, beryllium (released from the burning of coal) is over three million times more hazardous as an air toxic than the same amount of dichlorotetrafluoroethane (often used as a refrigerant). We also found that some of our variables were not statistically significant, which is a potential threat to external validity. There may also be a violation of heteroscedasticity, which violates the OLS assumptions—lowering precision intervals.

Our first result, before omitted variables are introduced, the positive relationship between carcinogen released per capita and percent minorities disappears. In terms of nationwide data, we found that the percent of the population who were minority did not have a statistically significant impact on the grams of carcinogen toxins released per capita. However, when we estimated a regression that included all our variables, this association became positive, yet remained not statistically significant. While the percent minority population was not statistically significant, our economic indicator variables were statistically significant. Our second result, when looking at individual states data, the percent minority population is not statistically significant for any of the individual states.

This paper proceeds as follows. In Section I we describe the data and econometric model. Section II presents the results for the nationwide sample and individual state samples. Section III presents our figures and heat maps. Section IV concludes.

I. Data and Descriptive Statistics

Table 1-Summary Statistics							
Variable	Definition	Source	Obs	Mean	Std. Dev.	Min	Max
logCarcino~n	Log of grams of carcinogen released per capita in a county	Toxin Release Inventory	2,428	2.73106	3.817751	-16.4989	13.86471
PCT_Minority	Percent of the population of the county that are minorities	American Community Survey	2,428	17.7101	16.35819	0.305236	88.99113
Income	Median household income of a county	American Community Survey	2,428	52213.6	14424.98	13262	136268
GINI	Gini index of a county (measures income inequality)	American Community Survey	2,428	0.44654	0.034795	0.3513	0.6179
Republican	If the county party was Republican, then the county had a value of 1. Otherwise, the county had a value of 0.	Agadajanian (MIT)	2,428	0.47158	0.499295	0	1
Urban	The percent of the population living in an urban area	Decennial Census	2,428	47.8345	29.85097	0	100
segregation	The summation of the absolute value of the differences between the percentage minority population of the census tracts and the percentage minority population of the county divided by the number of census tracts in the county.	Decennial Census	2,391	7.40480	6.277465	0.021866	40.87672

Notes: Observations were taken from the year 2018

The data on facilities that released toxic materials was taken from the 2018 Toxic Release Index, an index provided by the Environmental Protection Agency listing all toxic release facilities and the data. The demographic data was provided by the 2014-2018 American Community Survey 5-Year Estimates. This demographic data was provided for both the county level and the census tract level. From this demographic data, we generated the percent of the population of the county that are minorities. To generate the percent of minorities in a county, we subtracted the percent of the population who identified as non-Hispanic white from 100. To compute the toxicity faced by individuals living in a county, we took the data of all the toxic release facilities within the county and combined their grams of carcinogen released. We then divided the total grams of carcinogen released in the county by the total population within the county. From there, we took the log of the grams of carcinogen released per capita. We did this as the grams of carcinogen per capita had multiple observations that could be considered outliers, and thus could be violations of OLS assumptions. By taking the log of the grams of carcinogen per capita, we aimed to reduce the influence of these outliers. This limited our observations to counties that had more than 0 grams of Carcinogen being released of which there were 2428.

Since firms choose to place their polluting facilities in areas populated with lower-income individuals, economic factors such as income need to be controlled to remove omitted variable bias. Income would have a negative bias on the minority coefficient as income has a negative relationship with grams of carcinogen per capita but has a positive covariance with percent minority. Minority populations, on average, have lower incomes than white populations. According to Pew Research, as of 2016, white median household income was still almost twice as much as African American and Hispanic median household incomes (NW 2016). However, since firms place polluting plants in areas where they face few liabilities, counties with a lower median income but with greater wealth inequality would have less grams of carcinogen released per capita. Counties with larger minority populations also likely have greater wealth inequality because people of color make between 58% and 73% of their white counterparts' wage. (Patten 2016) To control for wealth distribution, we used the county Gini coefficient, a tool used to measure economic inequality. The Gini coefficient would have a negative bias on the minority coefficient as Gini has a negative relationship with grams of carcinogen per capita and a positive relationship with the percent of the population who identify as minorities. Both data for median household income and the Gini coefficient for the county were provided by the 2014-2018 ACS survey.

Beyond economic factors, we aimed to control two other factors that could cause an omitted variable bias in our regression: political leanings of the county and how urban the county is. We included a dummy variable denoting whether the county leaned Republican as Republicans have adopted an anti-environmentalist stance in the past decades since the Reagan administration. The anti-environmentalist stance undertaken by the party could result in less environmental protection in a county and fewer controls on the building and operation of toxic release facilities. Republican counties also have a lower percentage minority population. Therefore, whether a county is Republican or not would result in a negative omitted variable bias if not controlled. This variable was generated from election data provided by the Massachusetts Institute of Technology (Agadajanian 2018). Using the lowest election results that were available, which for most counties were 2016 Governor election results, we found the party that received the most votes. If the party was Republican, then the county had a value of 1. Otherwise, the county had a value of 0. To control for size of urban areas in the county, we generated the percent of the population living in an urban area by dividing the Urban population by the total population then multiplying that value by 100. The urban population data was provided from the 2010 Decennial Census (Manson et al. 2010). We predict that percent of the population living in an urban area in a county would have a negative relationship with grams of carcinogen released as urban areas typically have a higher population density and higher property values than rural areas both factors that would disincentives the firms to build toxic release facilities. In the United States, urban areas typically have a higher percent minority population than rural areas. This would result in the percentage of the population living in urban areas having a negative omitted variable bias as there is a positive covariance between urban areas and minority populations but negative relationship between urban areas and grams of carcinogens released per capita.

One of the main limiting factors faced in this paper is the usage of county-based data. While county-based data is more freely available than census-tract data, using county-based data requires us to make more assumptions about the population. Counties are often made up of smaller population clusters and those population clusters could vary widely. For example, Harris County, overall, has a 31.1 percent white population. Yet, that population is not evenly spread. One city, Tomball, has a 64.2 percent white population. Another, Channelview, has a white population of 18.3 percent. Using census tract data would provide greater accuracy for our results. However, as the Toxic Release Index data lacks census tract information, we included a variable to account for the differences in racial spread across counties[1]. This is necessary as if certain areas have widely different percentages of minority population and the percent of the population that identify as minority does affect the grams of carcinogen released, the areas with high minority populations could have a greater mass of carcinogens released per capita. The segregation variable took the summation of the absolute value of the differences between the percentage minority population of the census tracts and the percentage minority population of the county then divided that by the number of census tracts in the county. The introduction of the segregation variable did reduce the observations used in the final regression as there were 39 counties that were missing census tract level racial demographics data.

Figure 2

Equation to Generate Segregation Variable

$$\boldsymbol{Segregation} = \frac{\Sigma \left| \frac{\boldsymbol{Minority\ Population_{Census\ Tract}}}{\boldsymbol{Total\ Population_{Census\ Tract}}} - \frac{\boldsymbol{Minority\ Population_{County}}}{\boldsymbol{Total\ Population_{County}}} \right|}{\boldsymbol{N}}$$

II. Results/Discussion

In terms of the nationwide data, shown in Table 2, we found that the percent of the population who were minority did not have a statistically significant impact on the grams of carcinogen toxins released per capita. (see in appendix table 2) We had hypothesized that an increase in the percentage of the population who were minority would also lead to an increase in the grams of carcinogen released per capita. However,

[1] Theoretically, the addresses provided by TRI can be converted to census tract information but there is no accessible way to do so as it requires finances and coding ability outside the scope of this course.

our naïve regression estimated a non-statistically significant negative relationship, with a coefficient of -0.0065. This coefficient means that, according to our naïve regression estimate, for every increase of one percentage points in percent of population that identify as minorities grams of carcinogen released decreases by 0.648 percent. However, this decrease was not statistically significant. When we estimated a regression that included all our variables, this association remained negative but became statistically significant. Our regression estimated that an increase of one percent in the minority population was associated with grams of carcinogen released per capita decreasing by 1.69 percent, a statistically significant difference. This negative relationship could be due to the inclusion of the segregation variable which has a statistically significant positive coefficient.

Our economic indicator variables were statistically significant. GINI and Median Household Income in 10000 USD were both statistically significant at the 1% level. An increase of 10000 dollars in Median Household Income was associated with a –47.06 percent decrease in grams of carcinogen released per capita in a county. Median Income was not the only economic factor that was estimated during this regression. The GINI coefficient increasing by 0.1 was associated with a –57.26 percent decrease in grams of carcinogen released per capita. This relationship is statistically significant to the 1% level. According to the regression, between two counties with the same median income, the county with greater income inequality will have less grams of carcinogen released per capita in a county. This finding suggests that having a wealthy population in an otherwise low-income county does reduce grams of carcinogen being released.

Among other control variables, while the political association of the county were not statistically significant, the percent of the population who lived in an urban area and the segregation of a county were both significant. Contrary to the negative relationship we had hypothesized, an increase of one percentage point of urban population was associated with grams of carcinogen increasing by 1.69 percent. The reason for this could be that most of America's manufacturing takes place in urban areas. The manufacturing sector is a major source of pollution and the increased presence of factories in urban areas, as compared to rural areas, would increase the number of carcinogens in an area.

When looking at individual states data, shown in Table 3 (see in appendix table 3), the percent minority population is not statistically significant for any of the individual states. The control variables no longer have a statistically significant relationship with the grams of carcinogen released per capita. This is likely due to the shrinking of the sample size. While in the nationwide data there were over 2000 observations, for each of the states there are only roughly 100 observations. The smaller sample size means that the coefficient intervals for statistical significance are smaller and thus the relationship between the control variables and grams of carcinogen released per capita need a larger magnitude coefficient to be statistically significant.

The largest threat to internal validity in this regression is correlation between the control variables. Most of the control variables had a statistically significant relationship with each other. The GINI coefficient, in particular, was correlated with all other proxy variables. However, we were unable to develop an instrumental variable for GINI, as finding a variable that correlated with GINI coefficient but did not also correlate with Income proved difficult.

III. Heat Maps

To produce these heat maps we utilized the 2018 Toxic Release Index and demographic data provided by the 2014-2018 American Community Survey 5-Year Estimates. To present data at a county-level, we summed up all zip code specific data across each individual county. Of the 2429 total county observations, 51 counties were unrecognized by the mapping program and were dropped from our data observations. These counties included all of Puerto Rico (42 counties total) and several smaller population areas across the United Staes. Because Puerto Rico was not a particular area of interest in this research, dropping those observations did not seem detrimental to our data figures. Moving forward, we aim to utilize different programming software to include Puerto Rico in the general U.S. heat map.

One issue that we encountered once starting to generate our heat maps for carcinogen grams per capita, was that there were a few counties with absurdly high reported levels which threw off the distribution of color intensity. We calculated the mean and standard deviation of the county values to be $\overline{x} = 1449$ and $\sigma\overline{x} = 23024$, which left 158 county values as outliers. To adjust for this, we selected for 2000 to be the upper limit of carcinogen grams per capita as we saw that the large spike in carcinogen releases occurred above 2000; any amount greater than 2000 was manually adjusted to now be 2000. After this, the new mean and standard deviation were calculated to be $\overline{x} = 308$ and $\sigma\overline{x} = 570$, and no outliers exist within the new range of values. The heat maps for Carcinogen Grams per Capita and BIPOC (Black, Indigenous, and People of Color) Population Densities as of 2018 for the U.S., Texas, North Carolina, and Georgia are provided below in figures 2-9.

We noticed that counties with a greater proportion of BIPOC individuals tended to also have greater carcinogen output in grams per capita, a trend that we expected to see but this did not quite match up with our statistical findings; our percent minority variable had a statistically significant **negative** coefficient. Although the relationship may be calculated to be negative, having the provided heat map visuals allows us to identify a general trend.

IV. Conclusion

The conventional wisdom in economic literature is that there exists a positive relationship between race and pollution distribution. In this paper, we investigated the question as of whether this is true. Omitted variables—most probably economic factors and other controlling variables—appear to have shaped a statistically significant association between carcinogen release per capita and percent of minorities, leading to a negative association between the two when looking at nationwide U.S. data. Our economic indicator variables were also negatively statistically significant. While the segregation of a county and the political association of the county were not found to be not statistically significant, the percent of the population who lived in an urban area was significant. Regressions that did not account for economic factors and controlling variables show that percent of the population who were minority did not have a statistically significant impact on the grams of carcinogen toxins released per capita. Consequently, when looking at individual state data, the percent minority population is not statistically significant for any of the individual states. These results shed considerable doubt on the conventional wisdom both in academic literature and in the popular press that race, income, and pollution distribution are positively correlated.

These results raise the question of why there is a statistically significant negative association between pollution distribution and race in 2018 in the U.S. We provided evidence that this can be due to the inclusion of the segregation of a county. However, there could also be concerns of endogeneity, where percent minorities, income, Gini index, political leanings, and a county's urban identification, is the cause of the amount of carcinogen released per capita. Here, there is a risk of reverse causality.

Nevertheless, some caution is necessary in interrupting our results. Firstly, although the coefficient percent minorities and other economic variables such as income and GINI index, were negative and statistically significant, this association might be conditional on some other characteristics that were not included in our regression. Because of the early nature of this research, we were unable to take as many variables as we wished into account.

Appendix

Table 2-OLS Estimates of Pollution Distribution on Race using Nationwide data

	(1)	(2)	(3)
Estimation Method	OLS	OLS	OLS
PCT_Minority	-0.00645	-0.00475	-0.0175**
	(0.00486)	(0.00573)	(0.00798)
Income		-3.42e-05***	-3.86e-05***
		(7.15e-06)	(7.15e-06)
GINI		-12.95***	-14.51***
		(2.925)	(3.004)
Republican		0.0398	-0.0467
		(0.156)	(0.156)
Urban		0.0176***	0.0168***
		(0.00315)	(0.00339)
segregation			0.0558***
			(0.0193)
Constant	2.845***	9.525***	10.32***
	(0.115)	(1.425)	(1.460)
Observations	2,428	2,428	2,391
R-squared	0.001	0.020	0.026

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Dependent variable is log carcinogen released per capita in the county. Used data from the 2018 Toxic Release Index and 2014-2018 American Community survey 5-year estimates.

Table 3-OLS Estimates of Pollution Distribution on Race using Individual State data

	Texas	North Carolina	Georgia
	(1)	(2)	(3)
Estimation Model	OLS	cluster	cluster
PCT_Minority	-0.0125	-0.0122	-0.0354
	(0.0640)	(0.0313)	(0.0253)
Income	5.01e-05	-8.01e-05	-2.96e-05
	(3.28e-05)	(7.13e-05)	(3.97e-05)
GINI	2.207	35.72	12.78
	(11.97)	(22.14)	(10.58)
Republican	-1.029	-0.0759	-0.960
	(0.703)	(0.845)	(0.669)
Urban	0.0136	0.0285	0.0243*
	(0.0108)	(0.0183)	(0.0135)
segregation	0.105	-0.0994	0.0262
	(0.117)	(0.144)	(0.0677)
Constant	-1.776	-9.889	-2.434
	(6.463)	(11.77)	(5.708)
Observations	153	85	120
R-squared	0.065	0.087	0.070

Robust standard errors in parentheses

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

Notes: Dependent variable is log carcinogen released per capita in the county. Used data from the 2018 Toxic Release Index and 2014-2018 American Community survey 5-year estimates

Figure 2

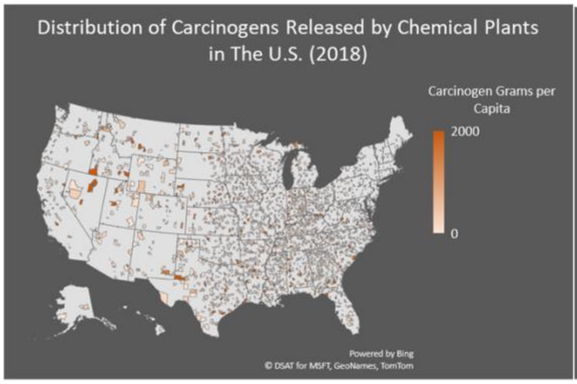


Figure 3

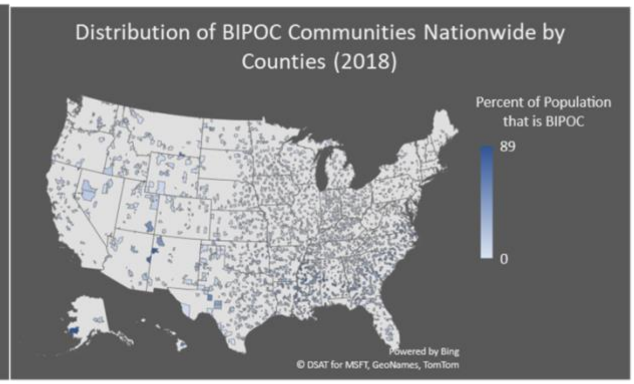


Figure 4

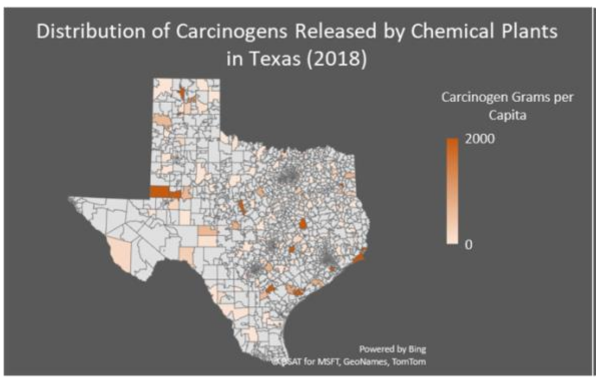


Figure 5

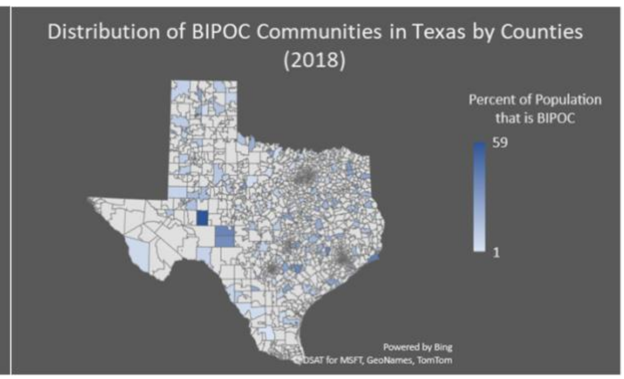


Figure 6

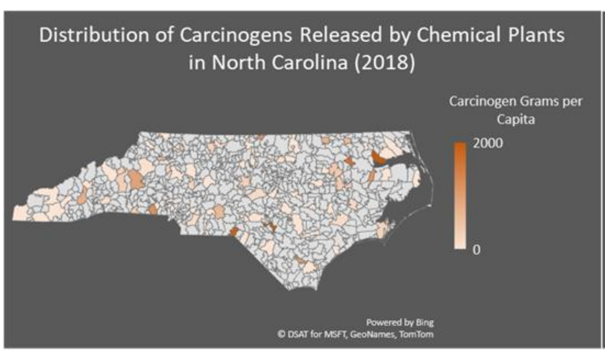


Figure 7

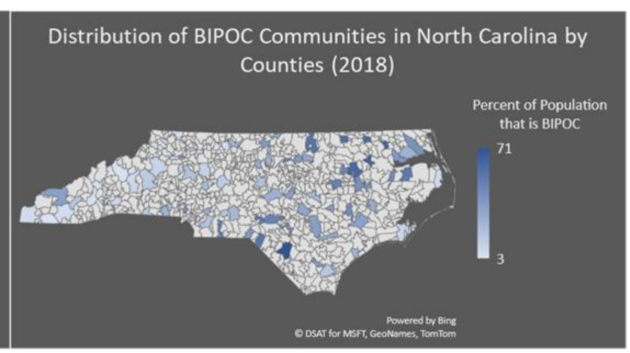


Figure 8

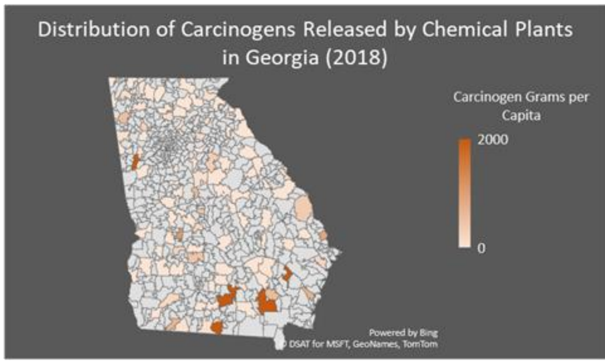
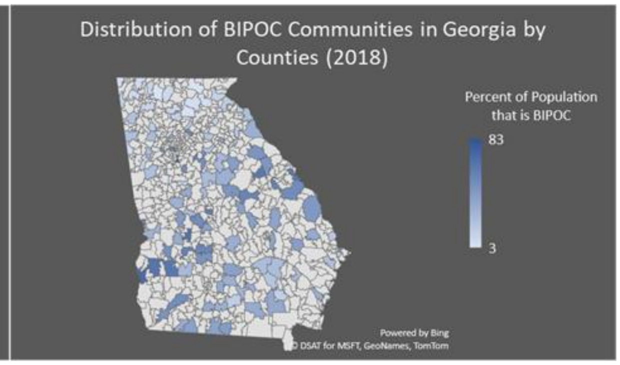


Figure 9



Notes: All above heat maps used data from the 2018 Toxic Release Index and 2014-2018 American Community survey 5-year estimates.

Bibliography

Agadjanian, Alexander. *MEDSL/2018-Elections-Unoffical*. 2018. MIT Election Data and Science Lab, 2020, https://github.com/MEDSL/2018-elections-unoffical.

Banzhaf, Spencer, Lala Ma, and Christopher Timmins. "Environmental Justice: The Economics of Race, Place, and Pollution." The Journal of Economic Perspectives 33, no. 1 (2019): 185-208. Accessed November 9, 2020. https://www.jstor.org/stable/26566983

Goodman, Chris. *Cbgoodman/Countyfips*. 2018. 2020, https://github.com/cbgoodman/countyfips.

Schpero, William. “Wschpero/Statastates.” *GitHub*, https://github.com/wschpero/statastates. Accessed 12 Dec. 2020.

Hamilton, James T. 1995. “Testing for Environmental Racism: Prejudice, Profits, Political Power?” *Journal of Policy Analysis and Management* 14 (1): 107. https://doi.org/10.2307/3325435.

NW, 1615 L. St, et al. “Demographic Trends and Economic Well-Being.” Pew Research Center’s Social & Demographic Trends Project, 27 June 2016, https://www.pewsocialtrends.org/2016/06/27/1-demographic-trends-and-economic-well-being/.

"Racial, Gender Wage Gaps Persist in U.S. Despite Some Progress." Pew Research Center, 14 Aug. 2020, www.pewresearch.org/fact-tank/2016/07/01/racial-gender-wage-gaps-persist-in-u-s-despite-some-progress/.

Schraufnagel, Dean E., John R. Balmes, Clayton T. Cowl, Sara De Matteis, Soon-Hee Jung, Kevin Mortimer, Rogelio Perez-Padilla, et al. 2019. “Air Pollution and Noncommunicable Diseases.” Chest 155 (2): 417–26. https://doi.org/10.1016/j.chest.2018.10.041.

"Social Explorer." Social Explorer. Accessed December 12, 2020. https://www.socialexplorer.com/tables/ACS2019/R12682595.

Steven Manson, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 15.0

2010 Census. Minneapolis, MN: IPUMS. 2020. http://doi.org/10.18128/D050.V15.0

"TRI Basic Data Files: Calendar Years 1987-2019." US EPA. Last modified October 26, 2020. https://www.epa.gov/toxics-release-inventory-tri-program/tri-basic-data-files-calendar-years-1987-2019.