Examining Injustice in Environmental Toxicity

A Thesis

Presented to

Department of Statistical Science

Duke University

Anne Driscoll

May 2018

Approved for the
Bachelor of Science in Statistical Science

_____

David Banks

_____

Committeemember O. Name

_____

Committeemember T. Name

_____

Mine Cetinkaya Rundel, DUS

Acknowledgements

I want to thank a few people.

Preface

This is an example of a thesis setup to use the reed thesis document class (for LaTeX) and the R bookdown package, in general.

Table of Contents

List of Tables

List of Figures

Abstract

The preface pretty much says it all.

Second paragraph of abstract starts here.

Dedication

You can have a dedication here if you wish.

Abstract

Environmental justice literature has been significantly limited by data available on toxicity. Through neighboorhood to state level research, it's been shown that toxicity is steadily dropping for everybody as production methods are improved and regulations are tightened. Environmental justice literature has followed the same vein of local study, working hand in hand with organizers to assess and understand micro-level toxicity disparity patterns in communities. As an injustice inherent to the societal structure of our nation, this is a question that needs to be addressed at a national scale to see where we have been and where we are going. The Risk Screening Environmental Indicator model (RSEI) data was released for public use in 2015, opening a window to work with fine grain data available nationally since 1988. Combining RSEI toxicity data with Census demographic data allows the creation of experienced toxicity distributions for any subset of the American public.

Using better measures of toxicity, over the 1988-2013 period, we are able to broach more systematic questions. Specifically, we look at: how the 5th, 50th, and 95th percentiles of toxicity experienced by each race group varies over time; how protective higher incomes are for differing race groups; and focusing on an unexamined group, look at changes in toxicity over time for the wealthiest. Through creating toxicity distributions for minority and low-income groups over the time period, we are able to see how toxicity burdens change over that time period, and how changes were distributed over different groups. Results show that toxicity is decreasing dramatically for all, and though the gap is closing, there still remains a large difference in toxicity experienced between white and minority groups.

Keywords

Chapter 1

Environmental Justice

1.1   Environmental Justice Theories

This growing body of environmental literature has shown with increasing certainty that on average minorities in America experience a higher poullution level than the white population. The literature has focused primarily on air pollution and industrial facilities, the two areas with the most consistently accessible data. Many authors have focused on point source pollution, measuring individual pollution as a binary: the presence or absence of a polluting facility in the geographic level of analysis (SOURCE). An alternate measure of pollution exposure has been the count of polluting facilities within various buffer zones.

Pollution has historically been constantly improving, Approaches to the idea of environmental justice are not ????, many approach the issue from the context of point source pollution (SOURCES), some use data on transportation emissions, and others use broader environmental data, like levels of NO2 and PM2.5. Results are not unanimous. Many studies conclude that minorities do experience a disproportionate toxicity or pollution burden (SOURCES), others find that economic class accounts for most of the difference in experienced toxicity (SOURCES), yet others find no racial differences (Anderson).

Environmental justice research is unique in that research and advocacy have gone hand in hand, leading to questions guided by the needs of communities. This trend has led to unusually focused work, evaluating specific communities for the presence of specfic toxins at specific times (SOURCE). Though some work has had a broader net, few authors have broached national data (SOURCES) and even fewer have used data spanning significant time (SOURCES). Due to the definitions of pollution being used, uncertainty in accessible pollution data, and differing geographic units of analysis, work spanning broader time periods as well as a wider geographic range is neccessary to try to understand processes that operate on a national scale that lead to environmental inequality.

Despite differing conclusions from the collective body of research, how we define and think about environmental justice is evolving. Alongside descriptive studies that investigate specific pollution trends, theories of how environmental injustices are created have arisen. The main theorized explanations for environmental injustice are economic, sociopolitical and discriminatory explanations.

- what is the overall trend in toxicity, do we expect to see that same trend for everyone?
- those studies are not unanimous, but many show a different in group toxicity

1.1.1   Economic Explanations

Theories focusing on economic explanations for environmental inequality rely on individual rational choice as an explanation. Especially in the case of point source pollution, companies must make economically rational decisions when siting facilities. Those decisions can be multifaceted: if the location has easy access to transportation, or is close to raw materials, but a primary concern is cost of land. Given companies want cheap land for polluting facilities, and minority communities tend to be near cheap land (SOURCE), it's rational that there would be more facility siting in minority areas.

In the same vein, it would be economically rational for those with the means to escape newly sited pollution to do so. Under this theory, cheaper land is more likely to be near minority communities, those who can afford to escape do so, devaluing the land further, and making it all the more attractive for facility siting. This theory should be testable, given the direct cause and effect theorized, but results have been mixed. Several studies found that more facilities were sited in minority communities (SOURCES), while others didn't (SOURCES).

- low income is not a protected class, so there is no legal recourse to this.

1.1.2   Sociopolitical Explanations

An alternate explanation still relies upon siting decisions, but more explicitly exploitative. Sociopolitical explanations theorize that companies rationaly chose to site in minority communities, as they will provide less resistance. At risk communities would show a lower voter turnout, would have a large older population, may be disproportionately renters, and may be poorly politically represented. In theory, communities with lower organizing potential are more attractive for facility siting, as they pose a low risk of creating political or economic resistence to the proposed facility.

1.1.3 Racial Discrimination

Racial discrimination theories are divided in to overt and insitutional racism theories. In discussion of these theories the language used tends to change, using the term 'environmental racism' rather than 'environmental justice'.

The overt discrimination theory speculates on the intentions of the individuals making the final siting decisions. In this case, overt racism refers to the siting decisions as choices made intentionally to cause strife to minority communities. On the flip side of the coin, but the same outcoem, it can refer to the siting choices as deicisons made in order to protect wealthy white communities.

The intstitutional discrimination theory relies primarily on housing dynamics. Based on the knowledge that the number of minority individuals living near a sited facility has risen since first measured (SOURCE), the theory states that minorities move towards polluting facilities due to reduced choice in the housing market. Reduced choice can arise in several ways: the choices realtors make in showing homes or the availability of mortgages or loans. Local governments can compound the problems with housing availability for minorities, as minority areas are more likely to be rezoned for industrial purposes by local government (SOURCE).

These three theories all boil down to the same outcome. Despite using different language and justifications all represent different orders of instutional discrimination. Economic explanations rely on the segregation of minority communites created by redlining. Redlining destroyed investment in communities, prevented black families from investing in homes, and enforced segregated communities. Redlining created the geography and culture that enables economic and sociopolitical explanation, making it just a consequence of previous overt discrimination.

Though the current theories of siting choices would all create the same behaviour, the different reasonings can help to explain (and perhaps change) the trends that lead to environmental injustice.

1.2 Goals

- address the first question of ej literature by (with this new+improved data, over a longer period of time than analyzed in any of the literature) by looking at how the groups differ, and how that changes over time. -improve by using accurate measures of location (not people within range of TRI, not amount of pollution, but the toxicity of the pollution experienced by ppl)

- what is the relative toxicity experienced by minority groups over this time period

- how protective are finances?

- Fisher J. B., Kelly M. & Romm J. (2006). Scales of environmental justice:

Combining GIS and spatial analysis for air toxics in West Oakland, California. *Health & Place, 12*, pp. 701-714.

- Uses data from the ISCST3 air dispersion model, and local knowledge of non-point soureces of diesel emissions in West Oakland.
- Does a disparate impact analysis based on the EPA's
- Proves that the point sources are not randomly dispersed accros the Bay area, let alone the Oakland area, and shows statistically significant clusters of facilities.

- Kravitz-Wirtz N., Crowder K., Hajat A. & Sass V. (2016). The Long-term Dynamics of Racial/Ethnic Inequality in Neighborhood Air Pollution Exposure, 1990-2009. *Du Bois Review, 13*(2), pp. 237-259.

  - Uses data from the Panel Study of Income Dynamics (PSID), and data on nitrogen dioxide (NO2), and particulate matter (PM2.5 and PM10) from the Toxic Resource Inventory (TRI) and the Air Quality System (AQS).
  - Find that "Blacks and Latinos are, on average, more likely to be exposed to higher levels of NO2, PM2.5, and PM10 than Whites."
  - More specifically, examines the "correspondence between neighborhood racial composition and pollution levels" using the household level data in PSID.
  - Additionally uses measures of population density and residential segregation (measured using multigroup entropy index) of the areas each of the households are located in.
  - Uses a multilevel repeated measures model. Level 1 is individual intercepts and slopes, Level 2 is the average of all intercepts and slopes, with coefficients for race and time, Level 3 accounts for metropolitan areas.
  - Conclusions: There are substantial racial differneces in exposure to several kinds pollution. This persists even for households in similar circumstances. Declines in pollution over time have gone predominantly to minority groups.
  - Future work: Disentangle if differences come from "racially distinct patterns of individual mobility", "neighborhood change", or industrial siting.

- Gilbert A. & Chakraborty, J. (2011). Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida. *Social Science Research, 40*, pp. 273-286.

- Ard K. (2015). Trends in exposure to industrial air toxins for different racial and socioeconomic groups: A spatial and temporal examination of environmental inequality in the U.S. from 1995 to 2004. *Social Science Research, 53*, pp. 375-390.

- Burwell-Naney K. et al. (2013). Spatial disparity in the distribution of superfund sites in South Carolina: an ecological study. *Environmental Health, 12*(1).

- Moghadam A. K. & Kayahan B. (2017). What influences the pattern of pollutant releases? An investigation of firms' siting and households' sorting decisions in

Ontario, Canada. *Journal of Environmental Planning and Management, 60*(4). pp. 743-754.

- Pais J., Crowder K. & Downey L. (2013). Unequal Trajectories: Racial and Class Differences in Residential Exposure to Industrial Hazard. *Social Forces, 92*(3). pp. 1189-1215

Chapter 2

Data

2.1 Raw Data

The Risk Screening Environmental Indicators (RSEI) Model is a geographically detailed data set produced by the Environmental Protection Agency (EPA). RSEI data is based upon the Toxic Release Inventory (TRI), which (for about 30 years) has been collecting data on all toxic releases in the US. The TRI program manages the regulations, policies and facilities that ultimately are mandated to be reported.

TRI data is self reported by facilities, with each observation being a release of a reporting chemical at a reporting facility. For each observation, data is collected on which chemical was released, how much of it was released, and the facility from which it was released.

The detailed location and chemical data that is collected through TRI (as well as detailed weather data from NOAA) is reformatted through a fate and transport model to create RSEI. The RSEI data shows where each release has traveled on an 800m grid across the USA. The ultimate data we have access to through the RSEI data is an observation for each release, for each square in the grid that the release hits. This gives us an idea of how the chemicals spread from the release locations, and enables us to create a map across the entire nation for where TRI chemicals accumulate in any year between 1988 and 2014.

The initial data from RSEI is in the form of one observation per release per block on the 800m grid that it reached. Given that the 800m grid across the United States has on the order of 10 billion squares, each release hits many squares, we have many releases, and the data covers 27 years, this is an enormous data set. Overall, the disaggregated microdata is approximately 4 terabytes of data.

The RSEI reformat of the TRI provides some additional information computed from the release information, and additional geographic information. X and Y are the geographic identifiers for the square on the grid across the US, with (0, 0) in the

center of the US. Release number tells us which release that row is associated with. Chemical number through media are all release specific data, but all data beyond that is release and grid number specific. Conc (concentration) is the raw concentration of the amount of the chemical that reached that grid cell. ToxConc is the toxicity weighted concentration of that release in that grid cell. 'Score' variables are meant to be used as hazard created by that release in that grid cell, as they are weighted by the population in that grid cell.

Below see an example of the disaggregated microdata:

X	Y	Release	Chem	Facility	Media	Conc	ToxConc	Score
-185	51	2050156	317	3	1	4.6e-4	2.28e-3	0
-184	41	2050156	317	3	1	3.3e-4	1.65e-3	0
-184	42	2050156	317	3	1	3.3e-4	1.67e-3	0
-184	43	2050156	317	3	1	3.3e-4	1.66e-3	0
-184	44	2050156	317	3	1	3.4e-4	1.68e-3	0
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .

2.2   Important Caveats

Due to the nature of the data, there are a few interesting caveats to consider.

- The RSEI data is self reported, and has been thought to contain some severe underreporting.

- The data is entirely based off a black box fate and transport model. The model has uncertainty that we are not addressing.

- The data only captures releases for certain facility types within certain industries, for certain chemical types within those facilities. Not all chemicals are mandated reporting, and any analysis that is done based off the data can't be extrapolated to discuss toxicity more generally.

- Not only does the data not capture all chemicals, it also doesn't adress many common nuicances. Because of this, it is difficult to relate the RSEI scores to health or quality of life outcomes in an area. In discussion of environmental toxicity more generally, it is worth noting that certain toxicity types may be more or less likely to cause adcerse effects based on how they reach people in the area (eg. they may or may not reach local water systems). There are also more obvious environmental hazards, that are likely to have strong influence on public health and living conditions that TRI doesn't address, eg: brownfields, solid waste disposal, animal farming, hazardous waste, etc.

- RSEI gives the weight of the release and the chemical of the release, but chemicals have very different levels of toxicity. A small amount of mercury released is

much more problematic than a small amount of CO2. To that end, the EPA assigned each chemical a 'toxicity' weight, by which the amount of the chemical released is multiplied. This means that we can aggregate all the chemicals in an area, and compare the overall toxicity over space and time. The toxicity weights allow us to compare to other chemicals' toxicity weight, and overall toxicity of a given chemical release, but also means that the values only have meaning in comparison.

Despite the limitations that the data presents, it provides an incredibly detailed and complex view of toxicity in America that is worth delving in to.

2.3 Consistency in Reporting (detailed data description)

2.3.1 Census Comparability

For the time period that RSEI data is available for (1988 - 2014), Census geographies have experienced considerable overhaul. Many of our questions of interest involve demographic characteristics, and the changes we see in environmental toxicity over time for those demographics. As such, we need to be able to aggregate the toxicity to block group or tract level to be able to merge with Census data. To do so, we caluclate the toxicity values for each geographic unit for the closest temporal census geography (1990, 2000 or 2010). If we want to use Census areas as the unit of analysis, we also need to consider consistency of the units definition over time. To do so we would need to create crosswalks that help us transform past Census geographies to the current form, allocating the population appropriately to the new geographies.

2.3.2 Details of Chemical Consistency

Comparability across time, space, and chemicals is a consistent topic through this section. The EPA is incredibly detailed in their data collection, and creation of metrics to make the data meaningful on a broad scale. Unfortunately the EPA is subject to the changing scientific consensus of the times, and therefore hasn't been able to provide entirely consistent data.

Over time, as chemicals have been found to be toxic, they have been added to the list of TRI mandated reporting chemicals. There are also chemicals that through renewed understanding have been removed from the mandated reporting list. Because the list of chemicals reported changes over time, aggregating all the data would cause us to see artificial jumps in toxicity. These jumps wouldn't be reflective of an actual increase in toxicity, but rather that the toxicity was beginning to be measured. These jumps may change what areas appear as toxic i the data, as industries that don't have

to report at some point in the time frame will be entirely removed from the dataset. Several of those industries are very highly polluting, areas who focus strongly on those industries will show inaccurate low scores.

2.3.3   Details of Industry Reporting Consistency

TRI regulates who needs to self report using the North American Industry Classification System (NAICS), and before NAICS was available used its predecessor, the Standard Industrial Classification (SIC) system. Just as we see changes in the regulations for chemicals, we see changes in the regulations of various industries. NAICS codes that need to report are regulated independently of chemical codes, and NAICS codes that are not consistently reported across the time period of interest must be removed to maintain continuity. For a toy example, a textiles facility releasing mercury might have to report it, but the neighboring mining facility might not have to report their mercury emissions. If that changes over the time period, and suddenly mining needs to report, we will see an artificial huge jump in the mercury present in that area if we don't remove by industry.

2.4   Ultimate Data Form

The disaggregated microdata from RSEI is one observation per release per block on the 800m grid that it reached. The munging for these 4 terabytes of data filters each of the billions of observations to check that it is 1) from a chemical that is consistent across the relevant years 2) from an released that is linked to an industry that is consistent across years, and 3) allocates the observation to the appropriate geography.

This data cleaning is done through R, using the DBI and SQLite packages. Since there is so much data, it's not feasible to process it using typical R function, so after loading the data in to a database, it becomes queryable using SQL. This significantly speeds up the processes detailed below.

To accomplish the chemical consistency, we use a data table (provided by Rich Puchalsky) that contains a row for each chemical with the data of regulation, and the date of deregulation. Using this information, we can select the chemicals that are relevant to any set of years of interest. Chemicals are found by selecting the subset of chemicals where the year of initial regulation is before the interest period, and the year of deregulation is after the interest period, while also excluding delisted chemicals, and all observations are checked to be in this range.

Filtering out observations whose releases are not under a regulated NAICS category for the entire time is more complex. Using a similar table that contains the regulation and deregulation dates of NAICS codes we can find the consistent industry categories. However, the only reference to NAICS or SIC codes are in the facility table that the releases reference. This table provides 6 NAICS codes that are the most common

NAICS codes associated with the facility. However, NAICS codes are release specific, not facility specific, meaning that for each emission reported a NAICS code is reported. Removing by facility is not accurate, since facilities might have different types of NAICS emissions. The textiles facility we used as an example earlier might make both shoes and jackets, with different industry codes and different releases that have different reporting requirements. To get data on the NAICS codes by submission, data must be taken from the original TRI data, and linked to the disaggregated microdata by the document control number.

In the data cleansing process, we filter the data to remove industries and chemicals that were not consistent over the period of inquiry, and aggregate the data to the relevant Census geographies. The data, then forms a data set for each year, with an observation for each block group, and just one measure, an aggregated toxicity level.

block	concentration	area
010010201001	627.3050138	6.520168
010010201002	499.6297799	8.48669
010010202001	578.8311689	3.137173
010010202002	756.3733114	1.962949
010010203001	637.7356488	5.907125
. . .	. . .	. . .

The 'concentration' estimate must be interpreted in the context of the consistency adjustments. For example, towns with extremely high mining emissions may not show as exceptionally high pollution, as mining is one of the industries that had different regulations over the 1990-2010 time period, and therefore the reported mining emissions have been removed entirely. Essentially, the pollution estimates are comparable across time and location, but only in the context of continuous EPA regulation, and can not be interpreted independently of one another.

Chapter 3

Approaches and Methods

3.1 Geographic Level of Analysis

The questions we would like to pose - how the distributions of toxicity that individuals experience over time are predicted by their complex, multidimensional identities - is inherently intended to be a person level analysis. That intent may not be achievable given the available data.

This analysis depends on two data sources, the disaggregated RSEI toxic release data (as compiled to contain only releases that are consistently reported between 1990 and 2010) as well as relevant demographic information from the Census.

RSEI toxicity data can be obtained at extremely fine level (the 800 meter grid across the United States,) but the finest grain Census data is available at is the block level, which contains between 0 and a few hundred people. At such low geographic levels, few variables are available for demographics due to identifiability concerns. At low levels cross tabulations are not available due to concerns of identifiability. Using a low level of geography (like census blocks or block groups) is important for the environmental aspect of this analysis, since environmental hazards can be very localized, especially along neighborhood lines in urban areas.

Unfortunately, the availability of cross tabulations is equally important to the goal of this work in examining inequality of environmental burden held by minority groups in America. The intersection of social identies, especially those steeped in systems of oppression, is extremely important for identifying unequal burdens. For example, low income populations across the board may be more likely to experience environmental hazards, but low income minority populations could be much more likely than low income white populations to experience extreme hazard. The intersections of demographic charictersitics, such as race and income or race and education are likely to be important in teasing out differences true inequality burden.

We combine the computed aggregated toxicity for each block group and the demo-

graphic data. Now for each census geography, we have toxicity information as well as demographic data.

block	concentration	area	total_pop	white	black
010010201001	627.3050	6.520168	530	447	83
010010201002	499.6298	8.486690	1282	1099	126
010010202001	578.8312	3.137173	1274	363	824
010010202002	756.3733	1.962949	944	458	477
010010203001	637.7356	5.907125	2538	2152	384
. . .	. . .	. . .	. . .	. . .	. . .

To move to a person level analysis, we can assign each of the people the toxicity for the geography they originated in. By assigning each person the toxicity of their block group, we can aggregate nationally to find the distribution of toxicity that each group experiences. This approach is restricted in ability to approach the problem in an intersectional manner, since we can only build a distribution for each of the crosstabs we have available. For higher levels of geography (where we might, for example, have race by income) we would be able to build national distributions for each income by race group.

In the case of the table above, to build a distribution for the white population, we would assing 447 people a toxicity of ~627, 1099 people a toxicity of ~499 and so on until we have the full distribution of toxicities experienced by the white population.

In choosing the level of aggregation at which to assign toxicity, in order to balance the needs of accuracy of toxicity and availability of cross tabulations, we create the overall toxicity distribution for Americans at each of the levels of geography. The process described above can be execcuted with the data shown above, or at a cruder level of geography, such as state. Using block group as the smallest form of geography, and state as the largest (including tract and county in between) we see how the distribution changes at each level of aggregation.

Distribution of Toxicity for US Population
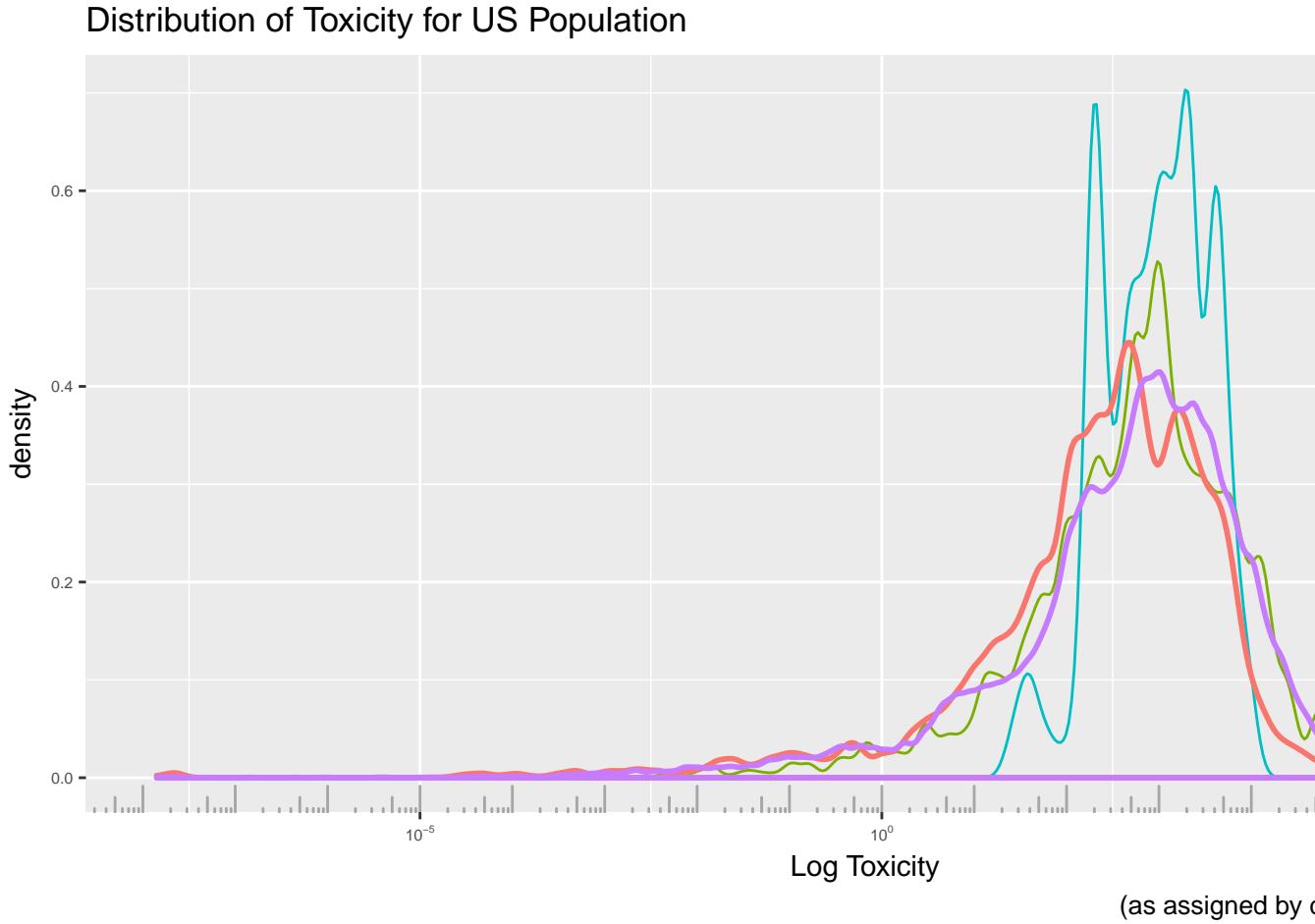
As expected, the state level assignment is a poor approximation of the lower level assignments. Given that we are assigning each individual the mean toxicity in their entire state, we are eliminating most of the variation from the data. Interestingly tract and county data seems to build a distribution quite similar to the block group level assignment. Initial results were replicated using all 3 levels of assignment, and no different conclusions were reached. This may be because the block group level is aggregating a large enough group of our fine grain toxicity data that it lost the street block by street block variation that we had deemed so crucial, meaning aggregating several block groups gives us a conceptual equivalent 'neighborhood' level of aggregation.

3.2 Temporal Level of Analysis

Given the layout of this non-parametric method, we can find the changes in positional distribution for any two given years. Though we are most interested in the complete change from the starting year of 1988 to 2014, as that is the data we have, the gradual changes and the speed of change from year to year is also of interest. While we

have the full range of data from 1988 to 2014 for the toxicity data, we only have available Census data from the decennial census and the ACS. That means that we have snapshots of data from 1990, 2000, and continuous data for 2010 on.

In order to do a continuous analysis from 1990 until the most recent toxicity data (2014), we chose to use linear interpolation between the decenial census. For current data, we use the 5 year ACS sample.

3.3   Simulation

To examine the change of environmental burden over time for groups we first use simulation to tease apart the forces in play as each group's distribution changes over time. For any given value, the percentile it holds in a minority distribution is likely to be different from the percentile it holds in the overall distribution. For example, in 2010 the 75th percentile of the black distribution is 2528.45, while that same value is the 81.44th percentile in the overall distribution.

We expect the mean of minority distributions to reduce over time for two reasons: first we assume the entire distribution will slowly be shifting right as we see improvements in environmentally friendly production technology and more comprehensive envirnmental regulation. Secondly, we hope that with Title VI protetections, and the work of civil rights advocates, minority communities will better protected against the economic power frequently held by polluters.

In order to track how minority distribution and the overall distribution have changed over the period of study, we use the positions that minority groups held in the overall distribution at the start of the period of study to simulate how each group's distribution would have progressed through time assuming a static position in society at large. This simulation proceeds as follows:

- Build an emipirical distribution of toxicity experienced for the entire population and for groups of interest in the starting year.

- Sample individuals from the empirical distributions for the entire population and the groups of interest.

- For each sampled value, find the percentile in the empirical distribution for the entire population in the starting year.

- Create an empirical distribution for the entire population in the ending year.

- For each sampled percentile, find the corresponding value in the full empirical distribution of the ending year and assign to the appropriate group of interest.

Using this method we can hold constant the place each individual (and more broadly each group) holds in the overall distribution, but follow the changes in the distribution as a whole. The collection of values simulated now represents where each individual

or group would have been had there been no positional improvement for the group as a whole.

If there had been improvement for a group, we would expect the simulated distributions to paint a bleaker picture of the environmental burden borne than the true distribution of the ending year.

3.3.1 Simulation Accuracy

We use bootstrap sampling to approximate the standard errors of the simulated estimates of the 5th, 50th and 95th quantiles with different values for n. This is done by simulating the new distribution at time t repeatedly with varying n, and describing the distribution of quantiles produced. This sampling helps us to choose a size of the sample to get an accurate measure, and frames the final values we report.
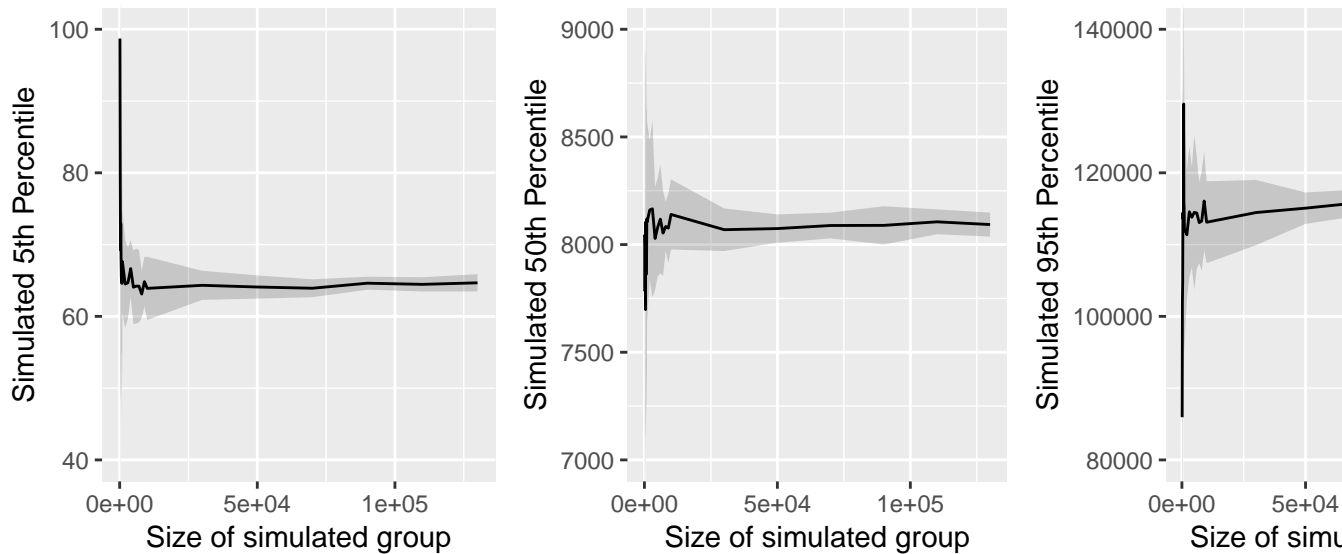
For the 5th quantile, a relatively low n produces a fairly stable result, sampling with n = 15,000 appears sufficient. Due to the extreme right skew of the data, the 95th percentile requires a larger sample. Still n = 50,000 is sufficient.

Chapter 4

Results

4.1 Toxicity Trends

Spurred by fear created by an disaster release of toxic chemicals in India, the Community Right to Know Act was created in 1986 to require emergency planning and enforce reporting on hazardous chemicals being realeased in to the environment by individual facilities. TRI data, beginning in 1988, is the public release of the data created through the Community Right to Know Act.

The time after TRI began collection included a lot of environmental regulation, including the Montreal Protol, the Clean Air Act Amendments of 1990, the Oil Pollution Act of 1990 and the 1994 Executive Order directing 'federal agiencies to identify and address the disproportionately high ... environmental effects ... on minority populations.' (SOURCE) As more and more environmental protections went in to place, and collective environmental attitudes became greener, we expect to see large drops in toxicity experienced by all groups. Especially in the early 1990's with mandated phase outs of certain CFC's and requirements for facilities to switch to Best Available Control Technology (BACT), there should be large drops in toxicity.

TRI releases by weight over time
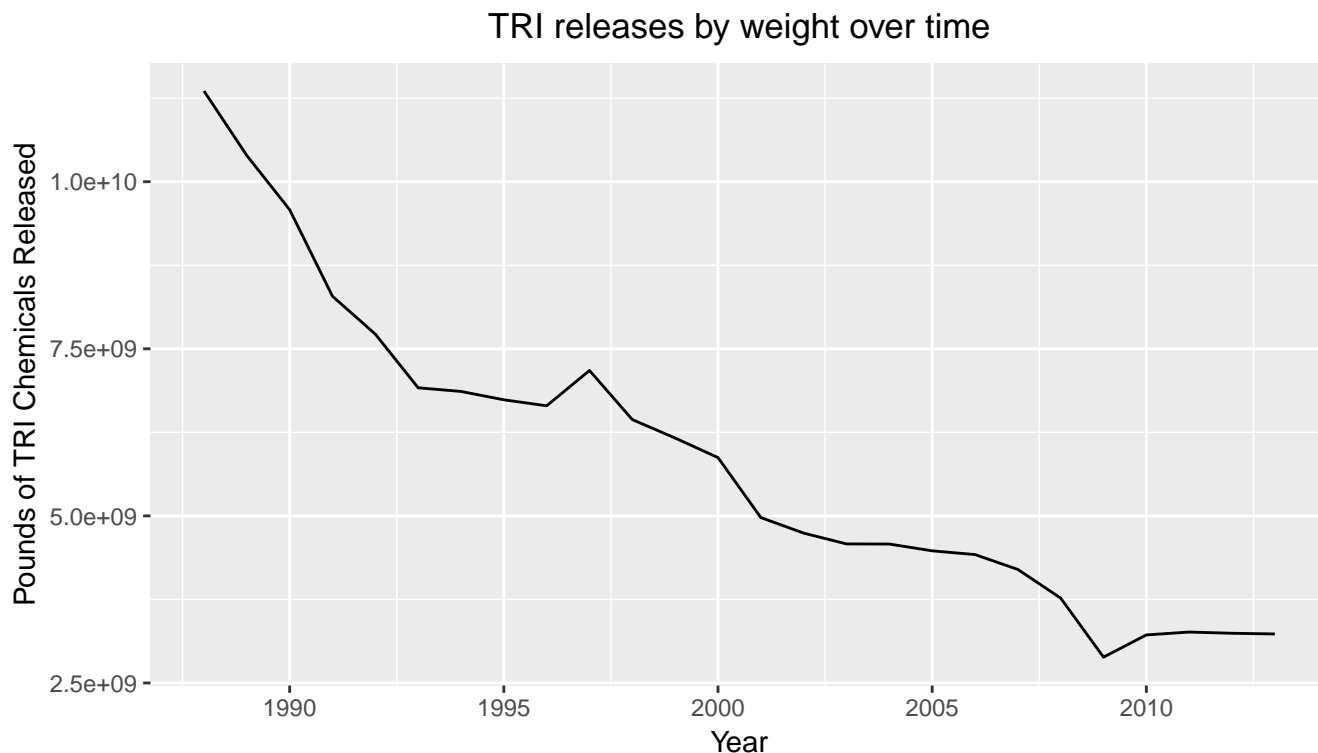
Looking at the net volume of TRI releases, we do see a fairly linear drop in the amount of TRI chemicals being released nationally between 1988 and 2013. Despite the common use of volume of TRI releases as a study measure (SOURCES) the rest of this work proceeds to use TRI toxicity as the measure of interest. TRI toxicity refers to the weight of each release multiplied by the toxicity weight assigned to the chemical by the EPA. Given that TRI chemicals vary significantly in their potential human effects, toxicity experienced is a better measure of the potential human impact than the net weight of toxic chemicals experienced. Given that this seeks to find differences in toxicity experienced by minority groups, toxicicty is even more important, as there is potential that net releases are less likely to differ across communities than the toxicity of those releases.

Toxicity trends look very different than the net release trends shown above. This is partially attributable to the strongly skewed distribution of toxicity. Some releases are not especially large in regards to weight, but are of extremely toxic chemicals, meaning the distribution of toxicities experienced is extremely right skewed by the uncommon, extremely toxic dumps.

Toxicities experienced in 2000
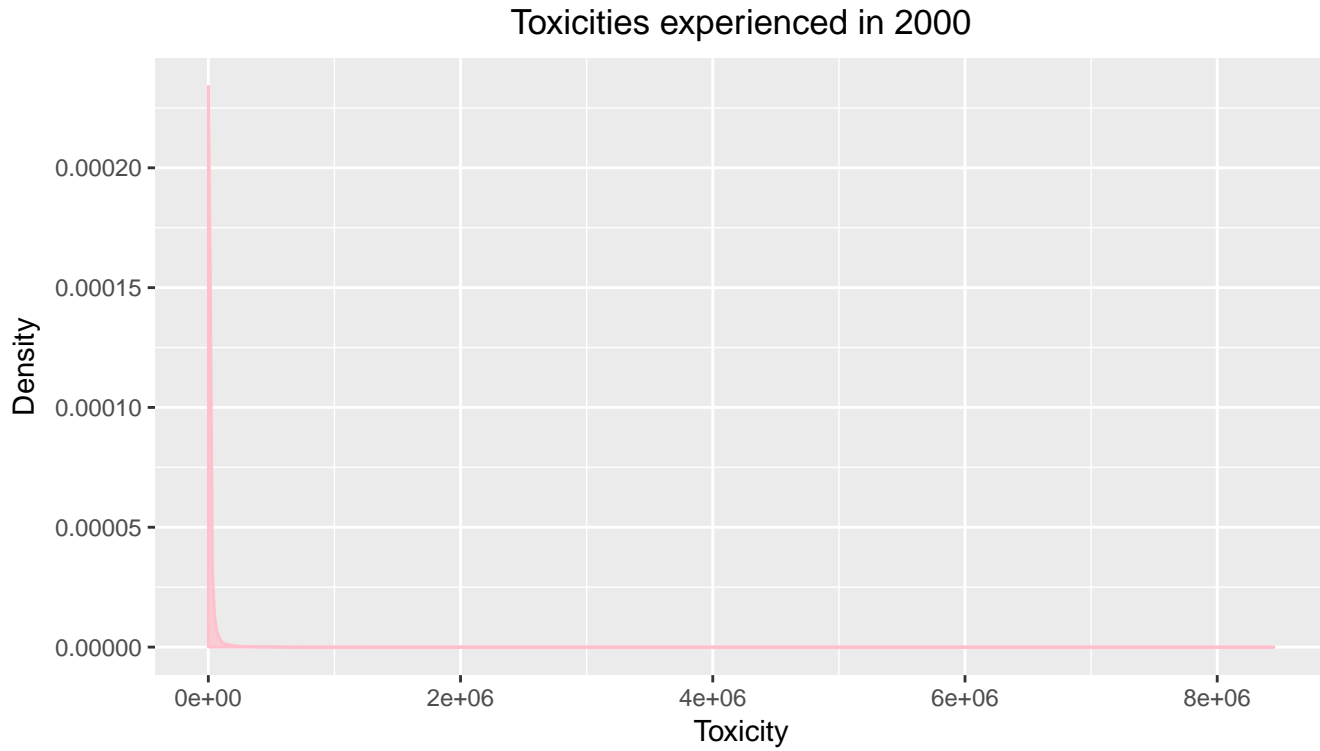
Given the extremely right skewed nature of this distribution, for ease of analysis and interpretation, from here on toxicicity is reported as log toxicity. The distributions of log toxicity are much more interpretable, but care must be used when examining the results to contextualize the results.

Log Toxicities experienced in 2000
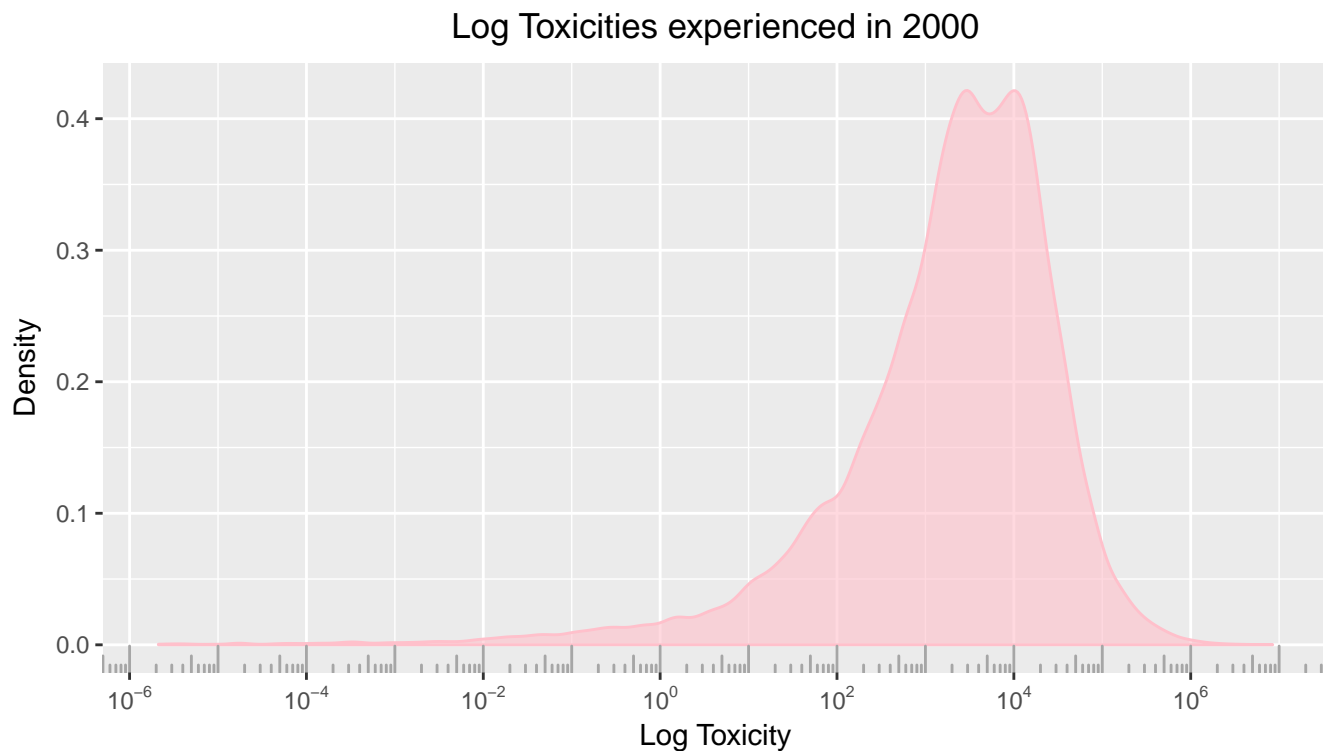
In this application, there are locations that are orders of magnitude more toxic than the bulk of the United States. These are sites to be aware of, where toxicity is more likely to have an effect on people's lives or health. As discussed in the {#data} section, toxicity measures don't always align with the locations that we know are truly hazardous.

Many of the locations that arise as especially toxic in the data are assuredly so. High Point, NC is consistently one of the most toxic tracts in the early 2000's, as confirmed by EPA reports identifying it as a location prime for testing toxicity abatement measures. Tracts outside of Salt Lake City commonly known for acting as dumping grounds are consistently ranked as toxic in the early 2000's. Mobile, AL, whose 3 superfund sites were proposed in the early 1990's, routinely ranks as toxic over that time period.

On the other side of the coin, there are many known toxic events and locations that fail to ever reach the top 5% of toxic tracts, like Flint, MI, or the Houston major superfund site.

4.2   Trends in Minority Toxicity Burden

Using data at the tract level, we assign the tract concentration to the counts of each group in the tract. By assigning toxicity for each tract, we build four distributions: toxicity for the white population, toxicity for the black population, toxicity for the

hispanic population, and toxicity for all other groups. In this case, hispanic is not a mutually exclusive group, and the race groups may contain ethnically hispanic individuals.

For each of the distributions, the 5th, 50th, and 95th percentiles are calculated for each time period. In the case of toxicity, the 5th and 9th percentiles are especially of interest, as they represent the non-polluted neighborhoods that each group has access to and the level of pollution trap that each group falls in to. Clearly, for all variables, lower toxicity is better.

The values for each race group are shown for each quantile below. On the left is the log toxicity for the quantile over time. On the right is the toxicity minorities experience relative to the white distribution. These values are not computed from the log, but are the net toxoicity of the given quantile for a group divided by the net toxicity of the quantile for the white distribution.
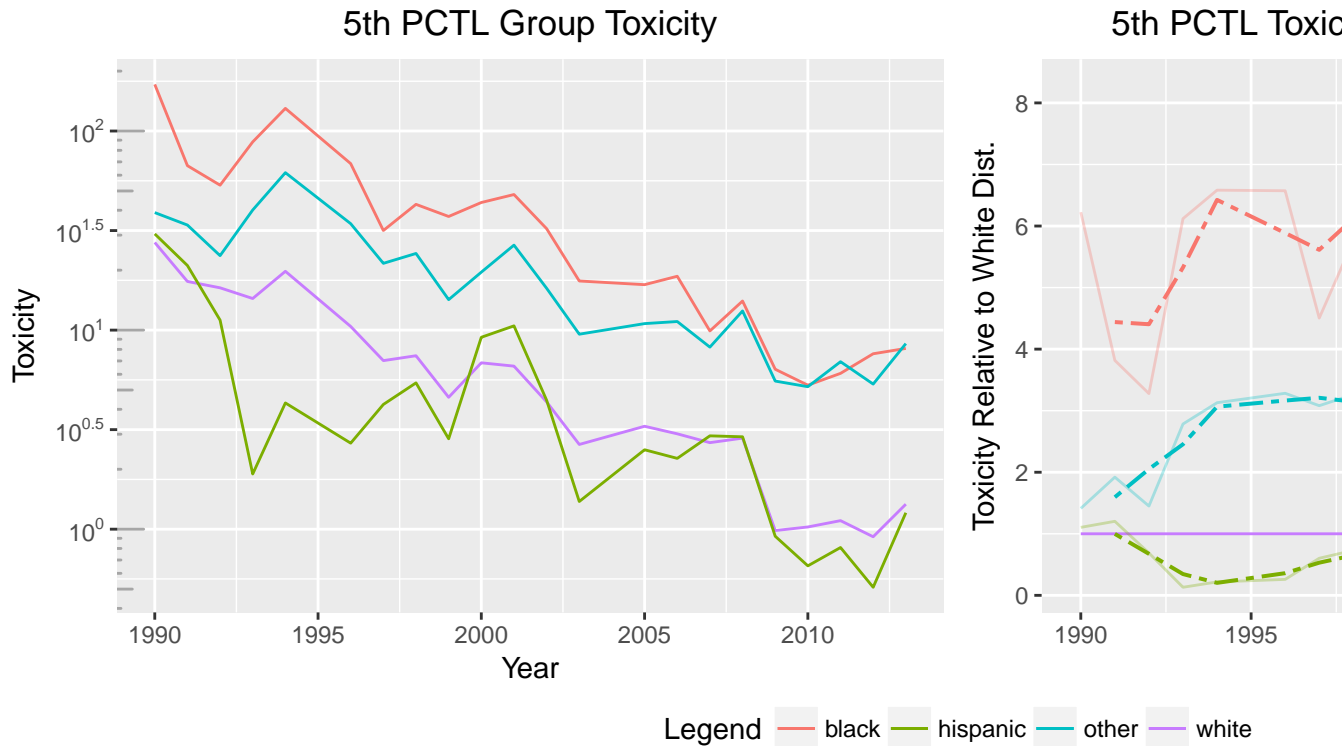

As we see in the 5th percentile plots, toxicity is steadily decreasing. Given the y axis has been logged, and we see a linear trend, we understand that toxicity has seen an exponential drop over the years. There appear to be two distinct groups, with the black and other groups significantly higher than the white and hispanic groups in 1988, and distinctly seperated by 2013. Though we see incredible improvements in toxicity for all groups, the relative improvement shows a different story. The difference for the group containing all other minorities is most impressive, with minorities originally having 5th percentile toxicity at approximately 1.5 times the rate of the white 5th percentile, but ending at over 6 times the white 5th percentile.

From this plot we see that though minorities have had the largest net gains in 5th percentile toxicity, those gains are an artifact of how much higher their toxicity began at. When framing improvements relative to the white population, there has been no improvement, in some cases relative toxicity has regressed.
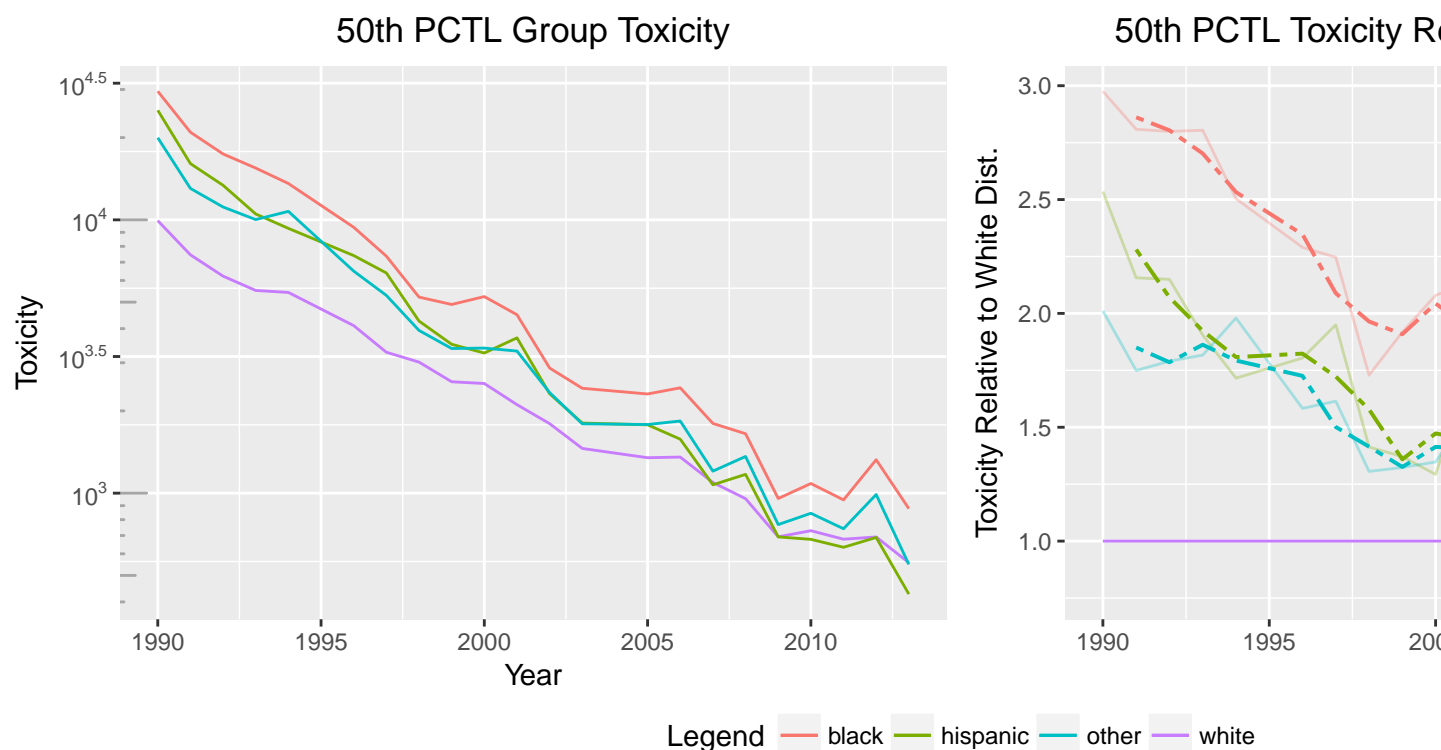
The toxicity changes for the 50th percentile tell a very different story. Again, we see linear improvement in the logged toxicity plot, indicating an exponential decrease in the toxicity experienced by the 50th percentile of the population. The difference at the 50th percentile is that minority groups have improved faster than their white counterparts, meaning that the relative toxicity of all groups is converging towards the 50th percentile of the white distribution.

95th PCTL Group Toxicity
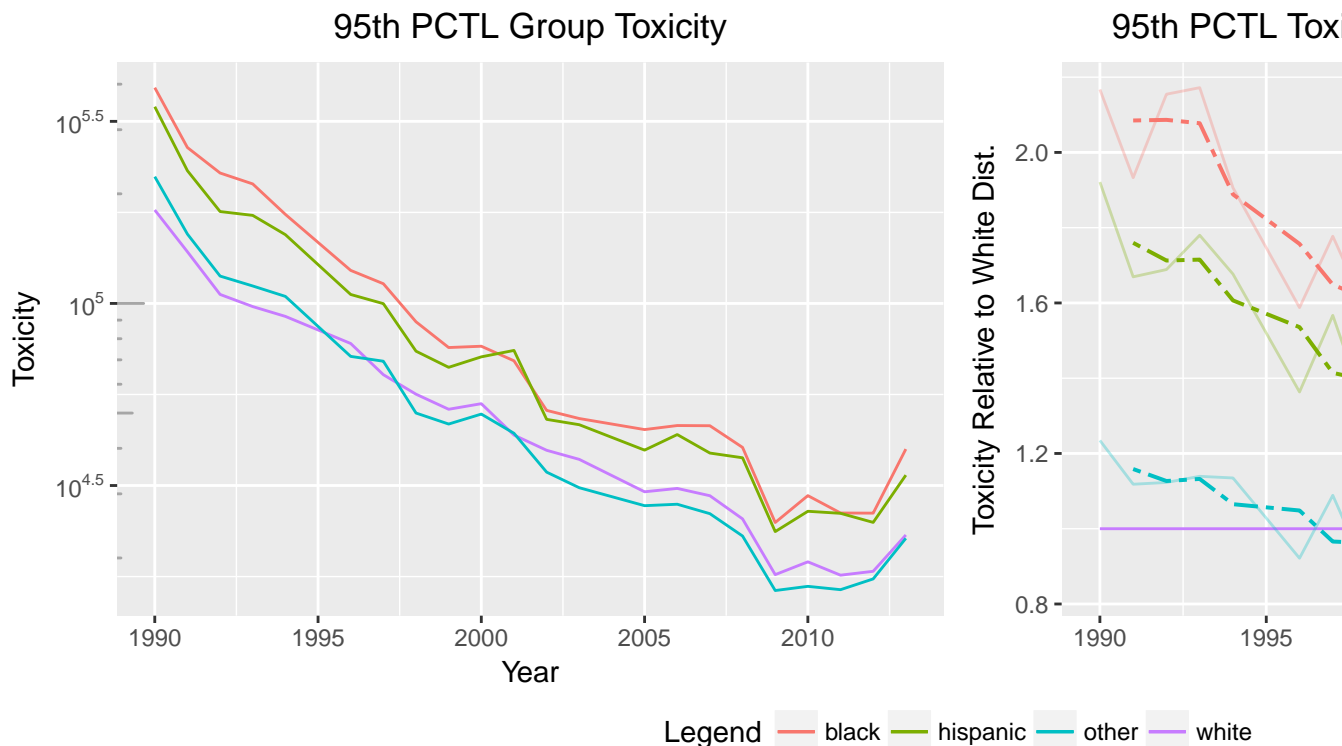
Ninety-fifth percentile toxicity is in some ways the most of interest. Though the difference between estimates for white and minorities is still approximately an order of magnitude, an order of magnitude at this level is much more likely to be harmful than an order of magnitude in the 5th percentile. An interesting difference is that at the high end of the distribution, we see the other group has a relative toxicity close to 1, where in previous plots the other group had significantly worse relative toxicity. Similarly, though the hispanic group previously performed closest to the white group, at the 95th percentile the hispanic group performs more similarly to the black distribution.

Examining the changes in each of these plots independently gives us a glimpse at how the distributions as a whole are shifting. The lack of change, or even worsening, of relative toxicity for the black and other distributions at the 5th percentile may be indicative of barriers minorities face at access to clean communities. As wealthier communities are getting ever cleaner, minorities are still unable to gain access to them, meaning less improvement in the low tail of the minority distributions. This hints towards the institutional racism explanation of environmental justice.

Comparatively, the changes at the 50th percentile show that minorities have all consistently improved faster than the white 50th percentile, meaning that the bulk of the distribution has been sucessfully shifting faster than the white distribution.

In the 95th percentile, we see the hispanic group performing similarly to the white distribution, while the black and hispanic distributions improving until the early 2000's, and then flattening out or rising relative to the white distribution. With the

widening gap in the 5th percentile, the closing gap in the 50th percentile, and the stagnation of change in the 95th, we see that the distribution is shifting left and condensing as the left tail stagnates, and the right tail slowly moves in.

In essense we are seeing the "rich getting richer" phenomenon in the 5th percentile of the white distribution. The hispanic group appears to have remained constant

4.3   Simulating Changes

Being able to build the true distributions that each group experiences over time allows us to see how relative toxicity burden has changed over time. Much has changed in how we think about equality and environmental issues in the time period being studied, leading to many legal changes in the way environmental justice is defined and the way environmental infractions are managed.

The changes presented in the analysis above are solely the net changes experienced by each group. Given minorities are more likely to be in the right tail of the distribution, any compression of that tail would result in both a net improvement for minority groups and a relative improvement to the white distribution. Those changes, though positive, don't represent a true change in the position minorities hold in the distribution, just that the distribution has shifted favorably.

To find how much of the change is attributable to changes of minority position, we use the simulation method discussed in section 3. By holding the percentile of each individual constant, but applying it to the new distribution, we can see what the minority distribution would like if the minority position had not changed. Effectively, this means that the initial situation, where minorities were disproportionately represented in the upper quantiles would no longer be true, or would at least be minimized.

Improvement will only represent true environmental justice when all groups are equally likely to be represented at any given quantile, not when the distribution is compressed enough that differences are negligible.

## 5th PCTL Group Toxicity

5th PCTL Toxi
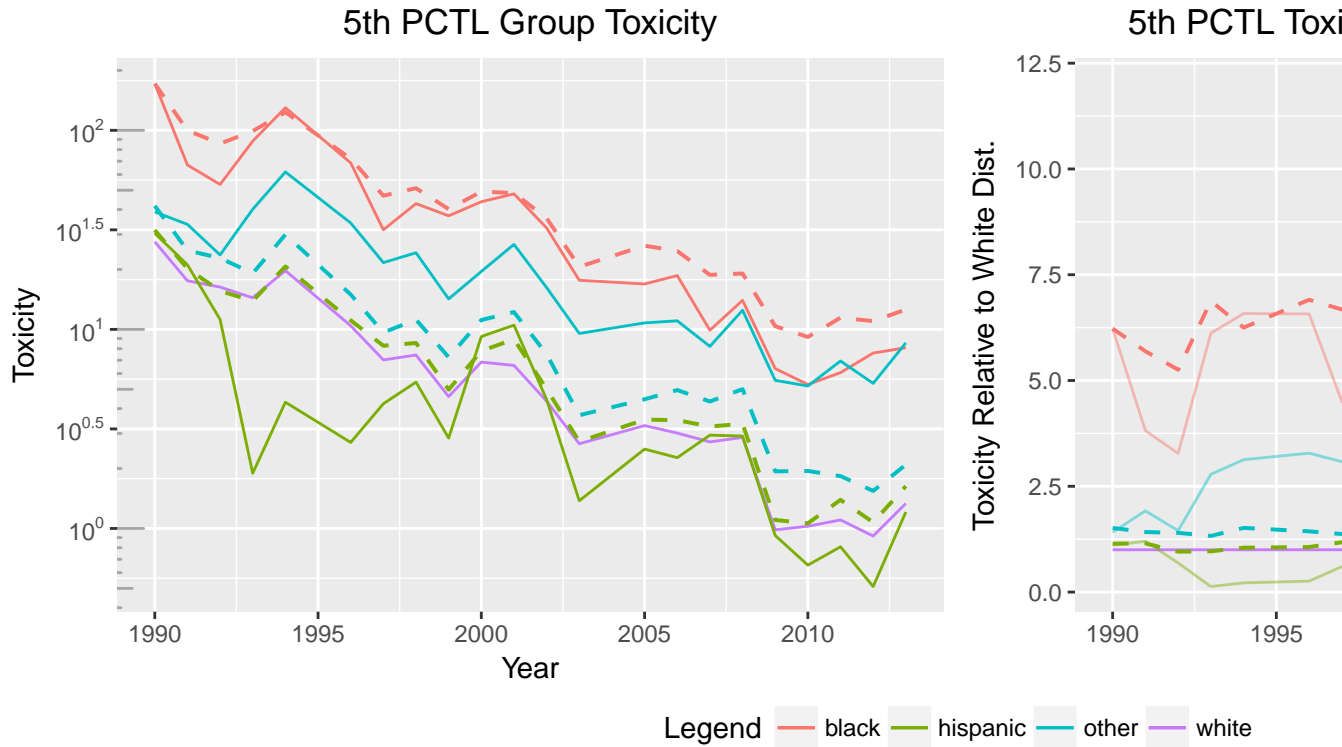
For the 5th percentile, we see a few interesting things. The 5th percentile of the simulated black distribution is significantly higher than that of the true black distribution, rising to approximately 12 times that of the white distribution. This means that more black individuals were able to move to locations with lower toxicity, as if they had maintained their location in the overall distribution, they would have experienced significantly higher toxicities. Interestingly, the opposite is true for the 'other' distribution. Their true 5th percentile has significantly worse toxicity than their simulated 5th percentile. This indicates that non-black minorities at the low end of the distribution have worse places in the overall distribution than they did in 1990.

50th PCTL Group Toxicity
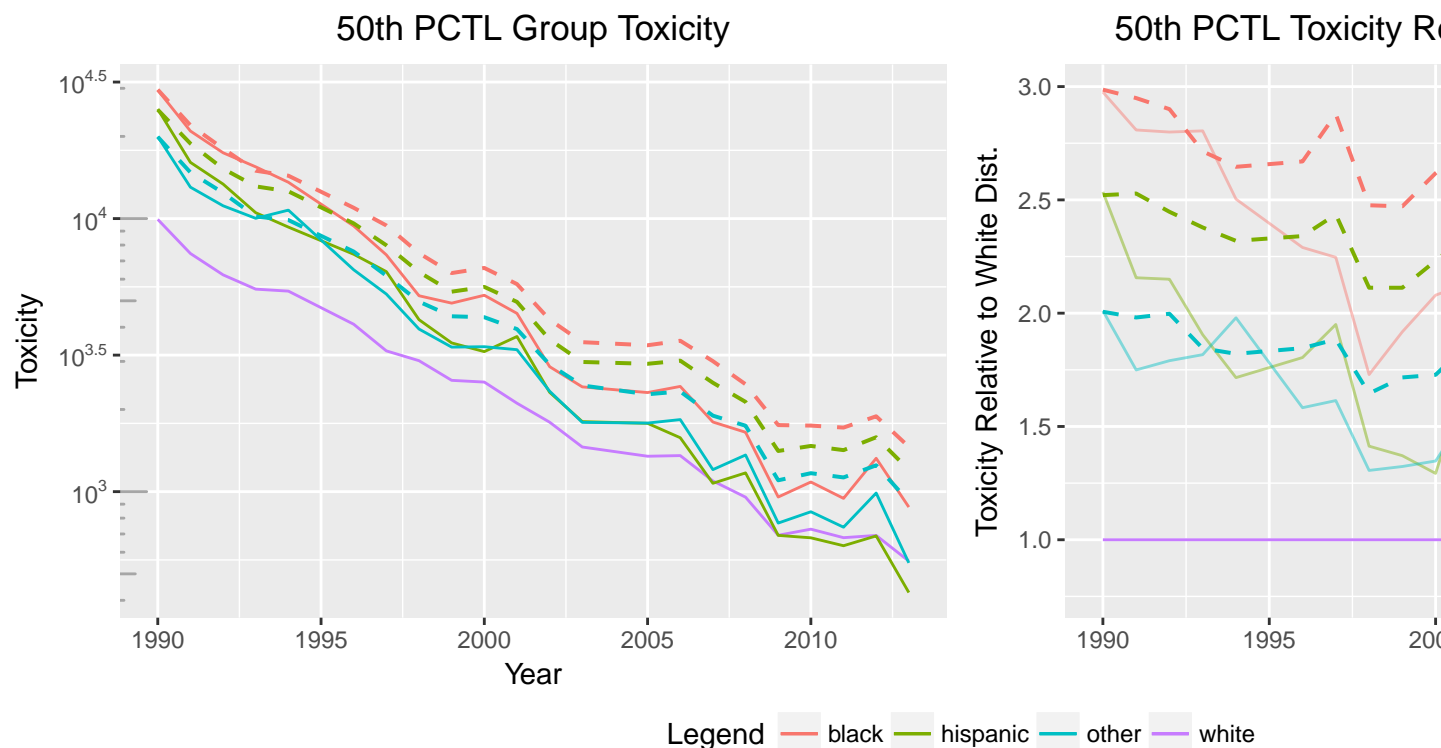
Changes at the 50th percentile were much more impressive in the initial analysis, and remain that way. The simulated distributions still show improvement, but very minor compared to the observed improvement. The simulated black 50th percentile drops from 3 times the white 50th percentile to 2.75 times, where the true distribution drops from 3 times to 1.5 times the white distribution. All the simulations show the same pattern, with the hispanic group showing the largest difference between true and simulated 50th percentile, the simulated group ending at 2.25 times the white distribution, and the true group ending at 0.75 times the white distribution.

Interpreting these changes as the portion of change attributable to true change in where each group stands in the distribution, it seems that there has been significant change in the positions minorities hold in the overall toxicity distribution.

95th PCTL Group Toxicity
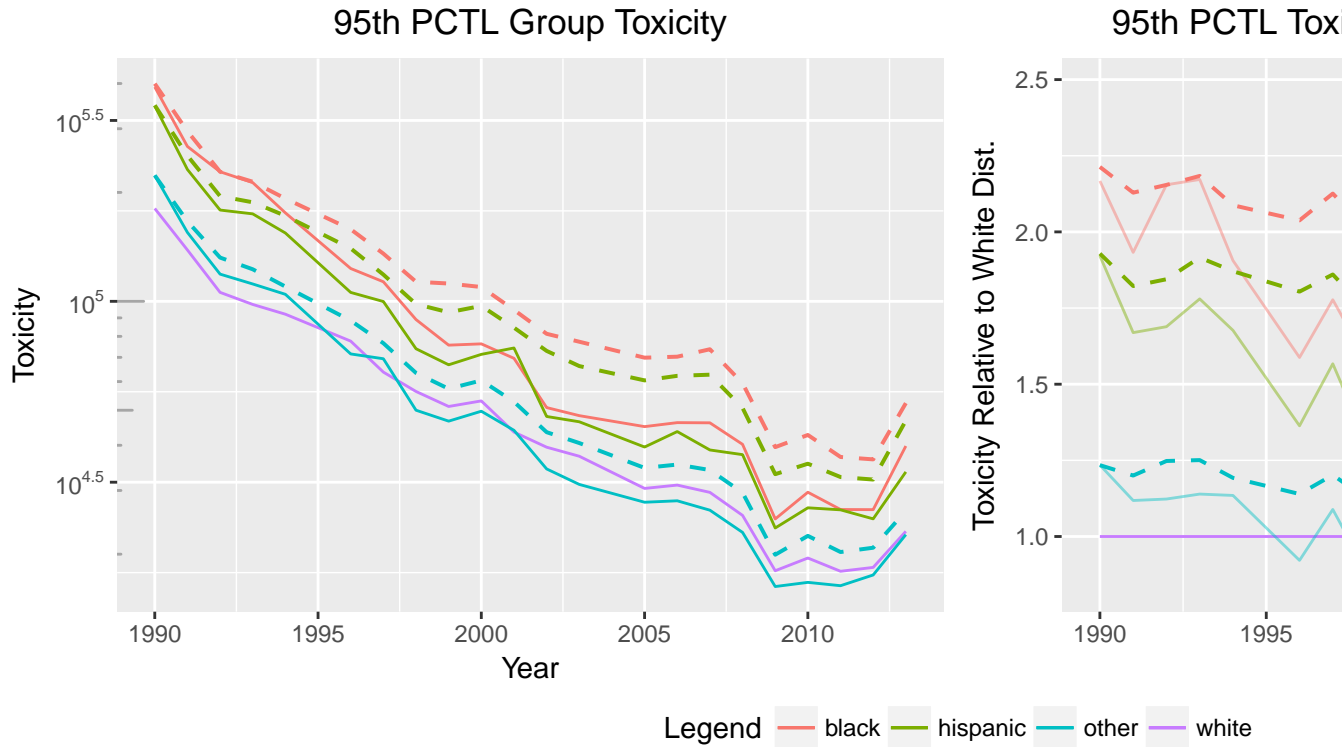
Simulations of the 95th percentile show even less improvement, meaning nearly all the change minorities experience is from moving towards a lower quantile in the overall distribution. At the 95th percentile the black distribution shows the largest difference to it's simulated distribution. The simulated black 95th percentile ends at 2.25 times that of the white distribution, while the 95th percentile of the true distribution ends at 1.75 times that of the white distribution.

Another way of thinking about it is that the distribution of quantiles occupied by minority groups was originally shifted towards the higher quantiles, but over time have shifted right. Plotting the percentiles of each group within their full distribution that year, we see that in 1990 all minority groups are strongly skewed towards the higher end of the distribution. As time goes on, more minorities exist in the lower percentiles, and more whites exist in the higher percentiles of the distribution.
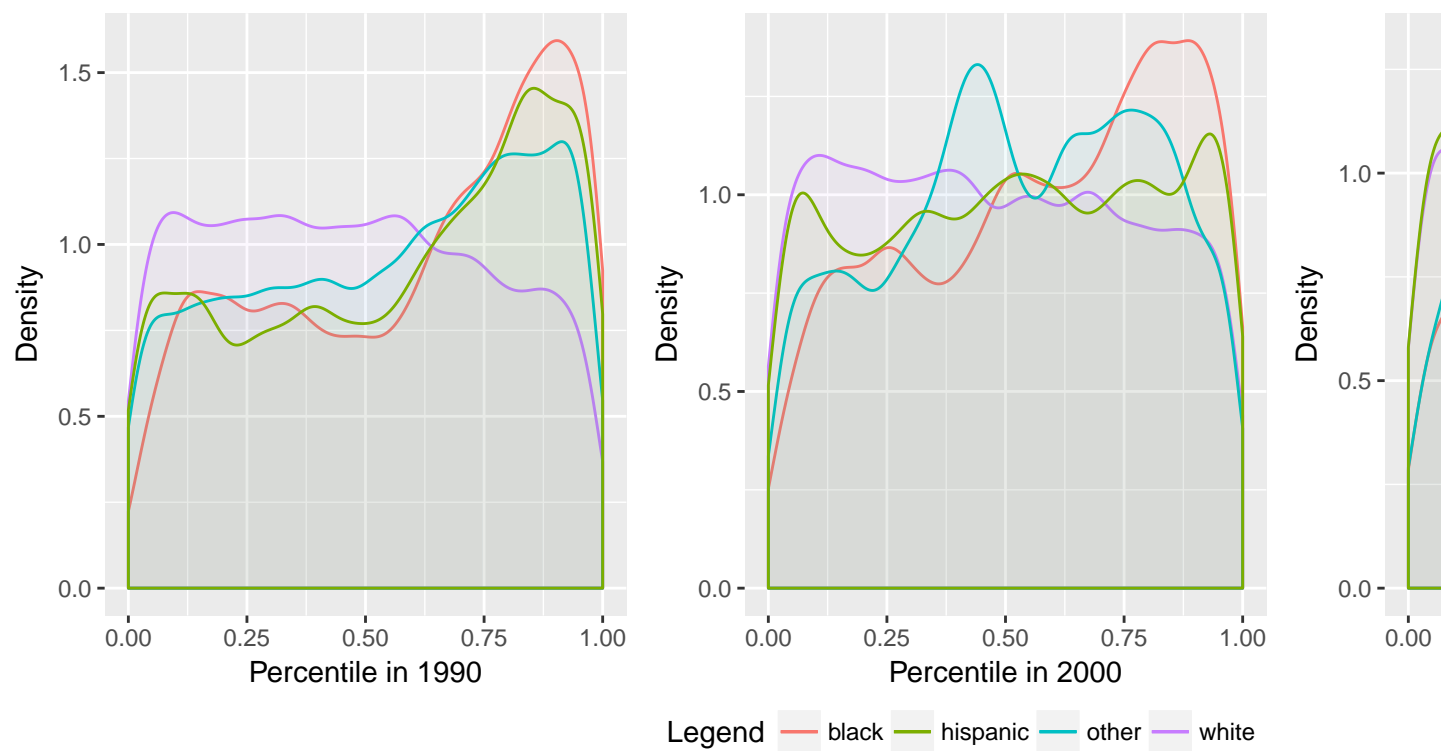
4.4   Interacting Risk Cases
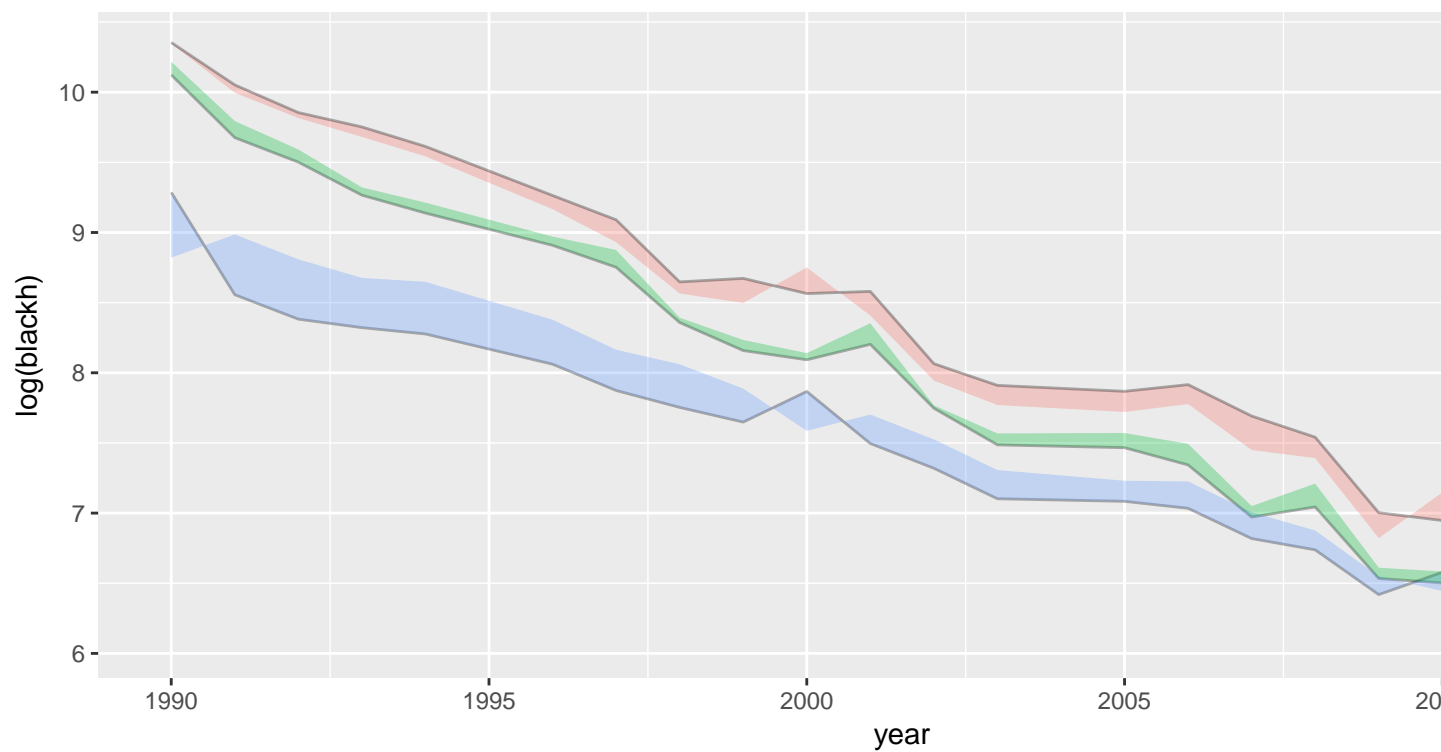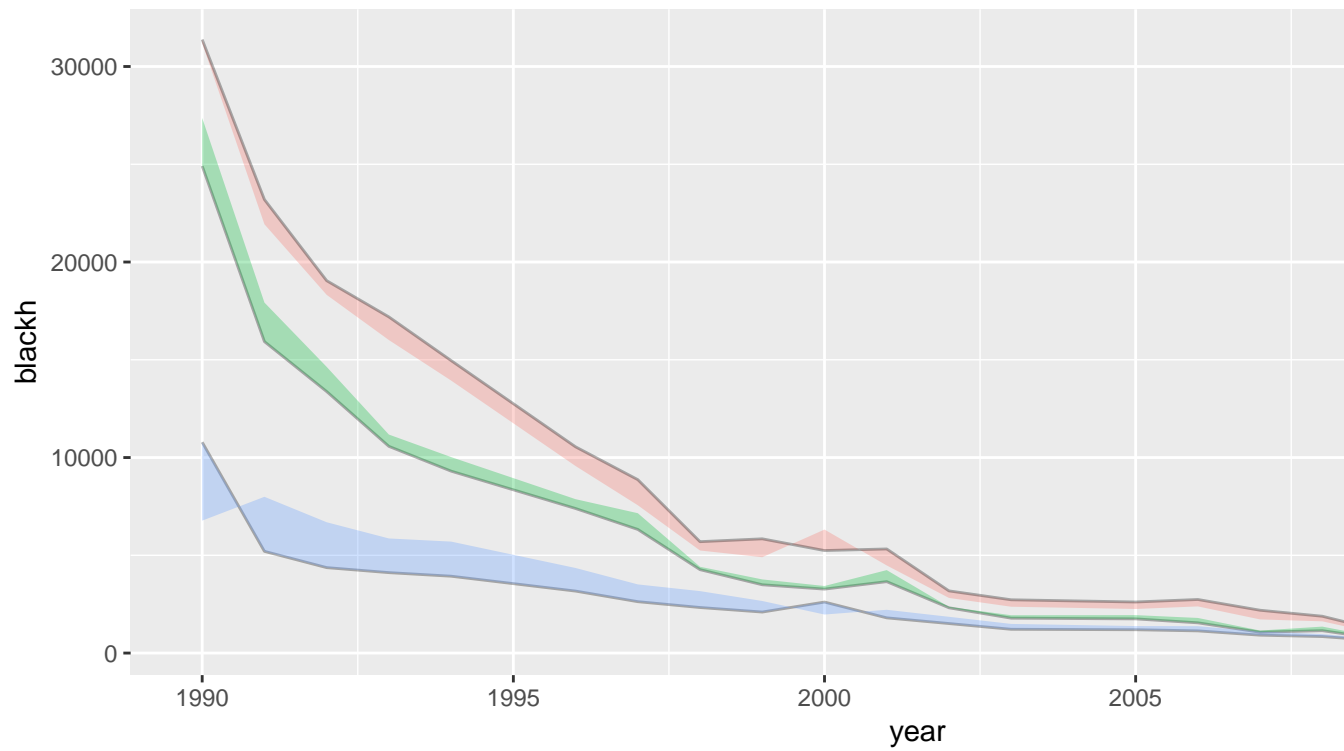
4.5 Geographic Changes in Exposure

4.6 Quantile Regression

Appendix A

The First Appendix

This first appendix includes all of the R chunks of code that were hidden throughout the document (using the `include = FALSE` chunk tag) to help with readability and/or setup.

In the main Rmd file

In Chapter 3:

References

Angel, E. (2000). *Interactive computer graphics : A top-down approach with opengl.* Boston, MA: Addison Wesley Longman.

Angel, E. (2001a). *Batch-file computer graphics : A bottom-up approach with quicktime.* Boston, MA: Wesley Addison Longman.

Angel, E. (2001b). *Test second book by angel.* Boston, MA: Wesley Addison Longman.