Investigating the Impact of the Change in Flood Risk Level in Lower Brooklyn: An Analysis of Housing Affordability and Socioeconomic Factors

Master Thesis
June 2023

Written by Jurgis Jančauskas, supervised by Leon Bremer

Master in Spatial, Transport, and Environmental Economics, Vrije Universiteit Amsterdam

Student Number: 2778966

Table of Contents

i. Abstract

Flooding is a major source of hardships for private citizens and businesses in many of New York's boroughs. Following the Hurricane Sandy catastrophe of 2012, many issues were raised about the disadvantages of implementing flood insurance policies and other flood regulations. The federal government, through Federal Emergency Management Agency (FEMA), carried out many regulatory changes, with one of them being a change in the special flood hazard areas (SFHA). The study focuses on Lower Brooklyn, where many vulnerable communities are concentrated and seeks to assess the impacts of the SFHA change on house prices, changes in the share of income groups, and earnings. Because many policies, such as insurance rates, depend on flood risk maps, the 'treatment' described above presents a starting point in evaluating the effects. Through the model, it is discovered that when the change in the SFHAs is considered as a dummy indicator, where 1 is a census tract treated, and 0 when the census tract is not treated, no significant changes in the composition of the census tract is recorded, when comparing it to the control census tracts. However, because the flood risk consists of 6 individual levels, the model is again tested on each level. This is where variation arises. Some specific levels, usually the lowest and the highest ones, record massive changes in house prices, earnings, and the shares of different incomes. Most importantly, house prices and earnings, both decrease by around 11% in risk level 1, but house prices grow massively, by 32% in risk level 5, while earnings continue to decrease. Another discovery is that the share of the poor grows in the lower risk levels but decreases in the higher ones.

1. Introduction

Flooding is a major source of hardships for private citizens and businesses in many of New York's boroughs (NYC Mayor's Office of Climate Change & Environmental Justice, 2022). The easiest way of alleviating these hardships is acquiring flood insurance. The National Flood Insurance Program (NFIP), introduced by Congress in 1968, paved the way for enforcing new floodplain management programs to ease the consequences of the unavailability of private insurers (Dixon et al., 2006). At the center of these programs lies the requirement of constructing elevated residential homes in Special Flood Hazard Areas (SFHAs). These are the sectors/boroughs where the expected annual chance of a flood is 1% or below. Additionally, for already constructed buildings, that are below the base flood elevation (BFE), it is required to be raised if more than 50% of the building has been damaged by a flood. These requirements also apply to new non-residential buildings, and older buildings that do not meet these requirements, must be upgraded (Dixon et al., 2006).

According to the literature, a vast majority of homeowners are uninsured against risk-associated flood damage (Kunreuther, Watchter, Kousky, & LaCour-Little, 2020; Sanjay, Garrahan, & Raghuveer, 2020). The floodplains of SFHAs, as identified by Federal Emergency Management Agency (FEMA), require flood insurance on loans issued by federally regulated lenders (Kousky, Kunreuther, LaCour-Little, & Wachter, 2020). Note that this does not apply to homeowners with no outstanding mortgage payments. Flood Insurance Rate Maps (FIRMs) set the insurance rates, but they are flawed and incomplete (Koslov, 2019; Pralle, 2019). A major reason for their flawed design is the use of past events as a base for creating future projections of floodplains. When Hurricane Sandy hit the Eastern United States back in October of 2012, $19 billion of damages were caused, with more than 88,000 buildings being overrun (Herreros-Cantis, P., Grabowski, & & McPhearson, 2020). The real estate market and people's lives were severely impacted. Insurers and lawmakers did and do play a direct role in risk management, by assessing and pricing potential risks caused by floods. The city and the Federal Government have received a substantial amount of criticism from the public, who insist that inadequate preparations for an event of such magnitude were made, and dejectedly, for not being inclusive to the marginalized communities (Herreros-Cantis, P., Grabowski, & & McPhearson, 2020).

After the devastating events, New York implemented new measures and regulatory changes, one of them being the revision of SFHA regulations. Still, flood insurance requirements and policies pose significant implications for low-income communities, who happen to live in flood-prone areas, whether due to cheaper housing as these areas are usually further away from the city center[1]. The government must protect vulnerable households, or any household for that matter, from flood risk by above mentioned regulations, but on the contrary, strict regulations may increase the costs of living in these areas, leading to displacement and gentrification in the communities. This "trade-off" is especially relevant to densely populated urban areas like New York City, where marginalized communities are disproportionately concentrated in flood-prone areas (Maantay, Maroko, & Culp, 2009). This thesis seeks to assess some of the impacts caused by a change in the flood risk level in Lower Brooklyn, where a lot of the vulnerable communities are concentrated.

The study focuses on the periods before and after Hurricane Sandy. An analysis of the regulatory changes is made, particularly focusing on the changes in the median house price, the median income, the share of the poor (making below $25,000), and other economic indicators. The motive is to identify the potential trade-offs between flood risk management policies and housing affordability. The study uses a difference-in-differences (DiD) approach to compare the consequences from before and after changes in the SFHA areas. Moreover, the effect of regulatory changes and the disproportionate impact, if any, on low-income communities is explored. Additionally, possible implications for future policy interventions will also be investigated. The initial null hypothesis states that house prices and income levels are unaffected by the SFHA change. Alternatively, the results may suggest that there is a significant difference in housing prices and income levels, with higher prices and income levels being in the treated areas as compared to the controlled ones. However, after a deeper insight into the analysis, some intriguing new patterns emerge, challenging the first assumptions. The hypothesis is then adapted to consider and explore the data nuances. It is revealed that individual risk levels, ranging from 0 to 5, influence the outcome. Where lower risk levels, even though being part of the treatment variable have negative effects on house prices and earnings, decreasing by 11.3% and 11.9% respectively. However, at the highest risk

1 *See: https://rebuildbydesign.org/who-lives-in-nycs-floodplain/*

level, the results indicate a massive 32.97% growth in house prices. Additionally, the share of poor people grows in the lower risk levels and then drops in the highest level.

The thesis structure consists of 6 sections. Section 1 contains the Introduction to the topic, Literature Review can be found in Section 2, the Methodology in Section 3, Empirical results, and their analysis in Section 4, and finally, in Section 5, the thesis is concluded. For full regression results and other miscellaneous data representations, check the Appendix in section 6.

2. Literature Review

Flood risk maps and the accompanying insurance rate maps are critical in identifying high flood-risk areas and informing about flood mitigation efforts. The literature review provides a clear perspective of the process of mapping and why it is underdeveloped. Also, the regulations of flood insurance, and more importantly the National Flood Insurance Program (NFIP) is reviewed. Lastly, existing literature on Hurricane Sandy and its consequences are assessed, highlighting the importance of these topics. However, the review leaves us with a crucial gap – specifically regarding a comprehensive assessment of flood risk mapping on lower-income communities, just like the ones in Lower Brooklyn and beyond. This gap serves as a bridge to the methodology and regression analyses.

2.1 Flood Insurance and Risk Mapping

As introduced above, FEMAs maps and FIRMs are incomplete. First, it must be explained that once FEMA publishes these maps, the regulation maps, like FIRMs also come into effect, as they are closely linked and layered. An analysis from Lis Koslov (2019), who examines the contested case of New York flood risk mapping, presents a conflict between risk representation and uncertainty. FIRMs and their way of representing risk do not diminish the feeling of uncertainty but rather redistributed it. The collective action, from governmental institutions, is reworked in such a way that it can obstruct the progress leading toward a more flood-adaptable future (ibid). Looking into a real-world example, documented by The New York Times in 1977, a business owner, finding his business in a newly classified high flood risk zone was left dumbfounded. For him, the government stepping in and publicly releasing information via maps, was comparable to slapping a label on someone's back. The same could have been said about thousands of others, who found their properties and businesses in these maps. The government must be fully certain of its facts, before publishing highly sensitive information for everyone to see (White, 1977). It is not to say that the risk and insurance maps are wrong per se, however, their implementation process does raise questions.

Wanyun Shao et al. talk about individuals' perceptions of flood risk, which have significant impacts on voluntary purchases of flood insurance. The behavior is said to be influenced by the individual's trust in the local government, his income level, and of course, education (Shao, et al., 2017). The authors conclude that flood maps should be updated frequently to reflect accurate information and new policies that increase homeowners' trust in the local government, should be designed (easier said than done…). Most importantly though, the affordability question for low-income homeowners should be addressed heads on, as well as designing educational programs for those who need it (ibid).

2.1.1. Effects on Low-Income Communities

Flood mapping does not only influence New York but the whole country too, affecting low-income communities especially hard (Rhode, 2012; Faber, 2015). The marginalized citizens are very often not able to afford the heightened flood-risk premiums (Tesselaar, et al., 2020) also, not only that but as research from the First Street Foundation has shown, the number of American homeowners who are at risk of getting flooded could be as much as 70% higher than FEMAs estimates (First Street Foundation, 2019). The additional problems of the latter can be easily inferred. Households of high and low incomes, who may be able to afford the risk-premium insurance, but continue without the coverage, have a concealed chance to be devastated by flood-associated costs (National Flood Services, 2021).

Herreros-Cantis et al. (2020) analyze the differential vulnerability of coastal flooding in New York. In their study areas, containing six districts of marginalized communities, the authors observe a 45.7% increase in the floodplain after the SFHA updates, a 10.5% increase in the exposed population, and a 7.5% increase in the exposed population living in vulnerable communities (Herreros-Cantis, Olivotto, Grabowski, & McPhearson, 2020). The prices of houses located on the floodplain do also. According to the writers, the variability is high between the districts, where in some communities, potentially, floodplain gentrification may be occurring (i.e., decreasing vulnerability due to shifting racial demographics and increasing income levels). Others experience a 'double jeopardy', where exposure and vulnerability increase.

As intended, in the Risk Mapping section above, it was not mentioned that FEMA's preliminary flood maps are open for public consultation by citizens, government agencies, etc., who can appeal or demand a revision. Some communities in New York City indeed made appeals against the boundaries of risk zones, however many communities considered to be low-income did not have the resources to file such appeals (Pralle, 2019). Thus, raising nationwide concerns relating to barriers of who has the time, wealth, and power of authority to file these appeals (ibid).

Lastly, Wil Lieberman-Cribbin et al., describe the unequal burden of flooding in a cohort of New York's residents. The authors seek to assess whether the effects of flooding were distributed equally according to socioeconomic demographics. Residents completed self-assessment questionnaires 1.5-4.0 years after the hurricane. Using multivariate logistic regressions, it was possible to determine the relationship between sociodemographic characteristics and flood exposure (Lieberman-Cribbin, Gillezeau, Schwartz, & Taioli, 2021). The results suggest that older participants were more likely to live in a flood-exposed household, and those living in high-income areas had a decreased chance of flooding (ibid). The authors conclude that future preparedness for natural disasters must understand flooding from an environmental justice perspective, so it could minimize disproportionate exposure and its outcomes.

2.2 Effects of the Flood

Ortega and Taṣpınar (2018), measured the impacts of Hurricane Sandy by constructing a large parcel dataset with geographic data from 2007-2013, containing property values, along with data from FEMA on the same properties. The authors divided flood effects into direct, like damages related to flooding, and indirect, like changes in property prices of the undamaged buildings that were constructed in high flood-risk zones. The empirical results of the research indicated that Hurricane Sandy had inflicted a 9% decrease in housing prices in flood-prone areas, in comparison to similar properties outside of these areas (Ortega & Taşpınar, 2018). Rather more interestingly, a 17% to 22% price drop in damaged properties was discovered right after the disaster, however, five years later, the price discount had converged with the price level of non-damaged properties. Still, this price level was 8% lower than before, having no gestures towards recovery (ibid).

To describe this phenomenon, the term price penalty can be applied. The explanations for this may include neighborhood deterioration or the expectations associated with increasing flood insurance costs. As Kozlowski et al. (2015) theorize, and as hypothesized by the previous authors, the phenomenon may be due to the belief updating process, i.e., the current knowledge and the perceived risk of extreme events in flood-prone areas. The statements above tie in closely with the propositions put out by Kahneman and Tversky (1982). They suggest that individuals frequently make irrational judgments that involve uncertainty and risk, due to their cognitive biases, or, "rules of thumb", many times subjecting themselves to systematic biases.

Several studies have shown that the negative effects of a flood tend to be short-lived, disappearing over the following 5 years, which indicates that humans have a temporal adjustment in their perceptions (Hallstrom & Smith, 2005; Atreya, Ferreira, & Kriesel, 2013; Chandra-Putra & Andrews, 2019). This is contrary to Ortega's and Taşpınar's paper, where the estimations convey more persistent negative effects, suggesting more systems at work. In a paper written by Chandra-Putra and C. Andrews, where the authors use an integrated agent-based hedonic pricing modeling system, evidence of price capitalization in property value due to flood risk is found, although, the price discount diminishing over time.

As it turns out, it is also discovered that there is a distinct equity-to-efficiency trade-off when public policies are designed to reduce the cost to society. While it is widely accepted that direct regulations in real estate practices can be very effective in reducing costs associated with flood damage, low-income homeowners and especially homebuyers will often be driven away (Chandra-Putra & Andrews, 2019). Again, this is where mapping and flood regulations play a significant role in vulnerable communities. The above authors suggest a managed retreat in case of a flood occurrence, so low-income and marginalized citizens could remain in the flood risk zones, although still vulnerable to flooding. Nonetheless, many experts say otherwise. Projects of this magnitude are full of fairness issues, like who decides and who relocates (Yarina, Mazereeuw, & Ovalles, 2020).

Additionally, in a study done by Okmyung Bin et al., the authors employ a hedonic property price method to examine the effects of coastal flood hazard areas on the property values in Carteret Country, North Carolina. The regression results suggest that a location within a floodplain leads to a lower average property value by around 7.3% or $11,598 (Bin, Kruse, & Landry, 2008). The study uses NFIP zones as risk measures and can differentiate between 2 risk levels – 100-year floodplain (1.0 percent annual chance of flooding), or 500-year floodplain (0.2 percent annual chance of flooding). The results, as expected, also show that implicit prices are sensitive to the presence of coastal amenities (ibid). The common finding of the study suggests that the location within the floodplain lowers property values from 3 to 12 percent.

Soon after Hurricane Sandy had inflicted billions of dollars in damages to New York City, the city initiated large-scale investments in recovery and improved resiliency infrastructure. As of June 2022, 73% of the $15 billion of federal grants was spent by the city. Of this number, $10 billion was in FEMA Sandy grants, of which $6.62 billion have been spent. The rest was in other development grants, of which 92.4% were spent (NYC Comptroller Brad Lander, 2023). However, some of the uncompleted Coastal Resiliency Projects will not be completed until at least 2030, which is almost 20 years after the hurricane!

The contrary to already discussed literature about insufficient FEMA's flood risk mapping, we now look at an independent study done by Doglian Sun et al., who use the Suomi National Polar-Orbiting Partnership (Suomi NPP) spacecraft's Visible Infrared Imaging Radiometer Suite (VIIRS), as well as Advanced Technology Microwave Sounder (ATMS) data to map the floods caused by Sandy. Through mixed-pixel linear decomposition, their method calculates the water fraction from coarse-resolution VIIRS and ATMS data (Sun, et al., 2015). The flood map that was derived from the coarse-resolution VIIRS and AMTS measurements was then extrapolated to higher spatial resolution using topographic data. At first, the flood map derived from this process showed a much less overrun area than FEMA's flood map. However, there is a bias of time difference in observations, as VIIRS can only detect the hazards under clear weather conditions. After developing a new method that can derive the flood maps from ATMS microwave observations, the authors present an agreement between their flood map and FEMA's flood map, with a correlation of 95%. This, however, is another example of using historical data to produce flood maps. Scientists argue that the redrawn maps are too conservative (Bagley, 2013). As mentioned earlier, it can be argued that the updated maps do not include the future vulnerability of climate change.

3. Methodology

3.1 Data

This section presents the data sources and collection methods used for the study, as well as introduces and describes the treatment. Also, it provides descriptive statistics, like summaries and corresponding plots. The thesis draws on the American Community Survey datasets on economic and housing indicators as well as FEMAs coastal flooding risk maps, namely SFHAs, and NFHLs.

The American Community Survey (ACS) is collected using mailed-in questionnaires, phone interviews, and visits from the bureau's representatives (U.S. Bureau of Labor Statistics, 2019). These data profiles have the most frequently requested housing, socioeconomic, and demographic data. Specifically, the data used was accessed and initialized through the United States Census Bureau's website, which offers 5-year estimates on individual census tracts, ZIP codes, and other subdivisions. Note, five-year estimates were used, as historical data on single years were not available. To narrow down the analysis, a set of census tracts was selected in Lower Brooklyn and Brighton Beach, as this area is known to house lower-income individuals and is coastal, but more on this later. Selected census tracts are geographically defined regions, containing multiple neighborhoods in lower Brooklyn.

The data contains the 5-year estimates for every year from 2010 to 2021, where the post-treatment period starts from 2014 onwards. For the regression, 44 census tracts from the borough of Brooklyn in Kings County were selected as the treatment variables. This area had experienced a revised version of the FIRM maps, on which the SFHAs are based, as well as where the median income is comparatively low. Additionally, 45 census tracts that saw no regulation changes were also selected for comparison as the control variables in the same year. This area borders the treatment area and shares most of the public and private amenities present. In short, the treatment area is the changed/updated SFHA area, which took place after the occurrence of Hurricane Sandy, and the control area is an area neighboring the treatment area, where the

SFHAs were not updated (i.e., No Rating). Note that preliminary SFHAs were introduced in early 2013, and the 5-year ACS estimates usually come around December. The time gap may not be significant enough to include 2013, so we include 2014 in the treatment period instead. Also, 2014 data may not show the full picture, as the 5-year estimates take the average of the summed 2011, 2012, 2013, and 2014 data. The year in which the estimate was fully calculated after the initial treatment period is 2017 (2014, 2015, 2016, 2017).

These map changes can be found on the FEMAs National Risk Index[2] (also from Figure 2 to Figure 3), where the coastal flooding risk data can be layered on top of census tracts to show maps from 2020 and 2011 and present where the changes occurred. See Appendix A for the census tracts map. The tracts chosen, overlap with the areas of Brooklyn where the median household income was lower than $20,000 per year in 2013. However, this is where we encounter our first issue, as it is possible for the income to be above $100,000 and below $20,000 in two bordering blocks, and the same census tract. See the representation in Figure 1[3], from Business Insider, below. Additionally, as with every survey, some biases may potentially be present in the data. Namely, sampling bias, where certain groups may be under or overrepresented, reporting bias, and of course, selection bias, as only certain tracts were selected.

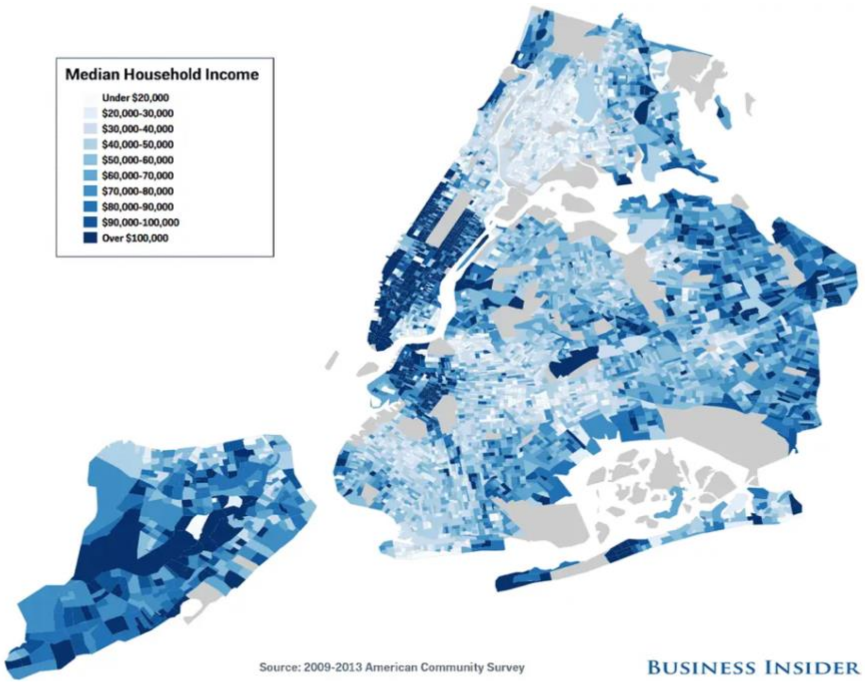


Figure 1: Median Income

2 *Source:* https://hazards.fema.gov/nri/map
3 *Source:* https://www.businessinsider.com/new-york-city-income-maps-2014-12?r=US&IR=T

Finally, the whole data set is comprised of 1140 observations on census tracts. It is important to mention that the data used for the study is aggregated at the census tract level and does not include individual-level observations. For the record, individual-level ACS data exists as a part of the Public Use Microdata Sample (PUMS), where each observation represents one correspondent. However, to ensure the confidentiality of the respondents, the location of their residence is not recorded, meaning there is no possibility of knowing whether the person lives in a High Flood Risk area or Low.

Before the analytical framework of the model is presented, the census data of ACS on housing and ethnic composition should be noted:

Treatment Area	2010	2020	Percentage Change
Population	177,004	186,919	5.6
Housing Units	77,586	77,276	-0.4
Vacant Units	6,336	4,123	-34.9

Control Area	2010	2020	Percentage Change
Population	139,686	150,691	7.9
Housing Units	66,944	64,534	-3.6
Vacant Units	6,481	3,293	-49.2

Table 1: Housing Composition

Additionally, the population of whites decreased by 6.8%. Every other race saw growth, with the group of 'nonhispanic with two or more races' having grown by 161.8% (NYC Planning, 2010; 2020)[4]. However, the white population was the largest at 53% of the total. In the *Control Area,* the number of whites decreased by 14.9%, while the number of every other race grew. Again, the 'nonhispanic with two or more races' grew the most, by 175.6%. However, the white population remained the largest group at 47.2% of the total.

From the data above we can see that our comparison *Control Area* experienced a higher population but a much higher drop in total housing units when comparing data from 2010-2020. However, the data availability for the Decennial Census on the selected census tracts only started in 2010, meaning that it is difficult to determine a consistent trend over time from the data provided by NYC's municipality. Based on the

[4] PUMA selected by selecting the corresponding census tracts. Deviations from the stated numbers are possible.

results above, it is only a guessing game of why such trends are occurring. For this reason, the results presented next use a DiD regression to estimate the causal effect of the SFHA change on the outcome variables in the *Treatment Area*, relative to the unaffected *Area*.

3.2 Analytical Framework and the Regression Model

To estimate the causal effect of the SFHA policy change on housing outcomes, a Difference-in-Difference (DiD) approach is used. The changes in outcomes over time between the treated group, which experienced SFHA change, and the control group, which did not, are compared by the method. To better understand to what extent the SFHA was changed, see Figure 2 below. Inside the red square are all the treatment census tracts that were selected for the regression. They number 45 (see Appendix A for the official numbers of census tracts selected). The colors ranging from red to blue explain the coastal flooding risk, with red being very high, dark blue being very low, and white having no rating (rated as having no rating in terms of flood risk) in the year 2020 (maps changed in 2013).

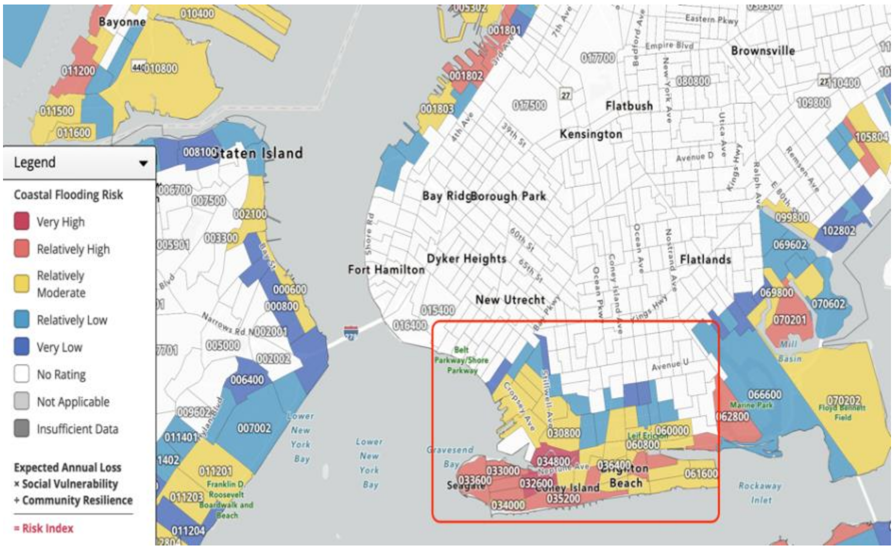


Figure 2: SFHAs from 2013 and Census Tracts

Figure 3 shows how these shaded zones (SFHAs) used to look before Superstorm Sandy. The risk zones have changed massively, and the northern selected census tracts were almost unaffected by regulations before the change occurred. Further search is then redirected to FEMA National Flood Hazard Layer data (NFHL), as historical flood risk data by census tract is unavailable on the NRI website (FEMA, 2021).

Therefore, to analyze the changes in flood risk levels, an alternative approach has been adopted. A map of census tracts was layered on top of the NFHL map of the flood zones from 2007.

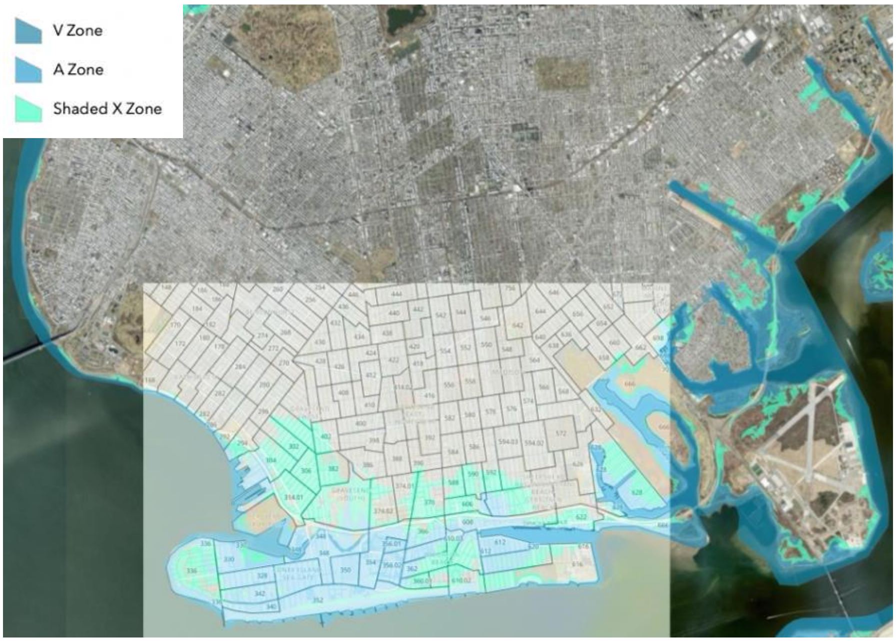


Figure 3: SFHAs from 2007

Additionally, the above map only provides 3 risk zones, with Shaded X Zone being the least risky, and to V Zone being the riskiest. Due to the unavailability of more detailed historical flood risk zone data by census tract[5], a visual interpretation approach was employed. o account for the different number of risk zone categorizations in the 2007 and 2013 maps, the visual interpretation involved applying the closest corresponding risk zone based on the relative proportion of the census tract being occupied by each of the 3 risk levels. If less than 40% of the census tract was in the X Zone, it got assigned a Very Low-risk zone. If the census tract was fully encompassed by the X zone, it got assigned the Relatively Low-risk zone. A zone got the same treatment, where levels went from Relatively Moderate to Moderate. Lastly, the V zone also got the same treatment as the X zone, where if 40% of the census tract or below was in the zone, it got assigned the High-risk zone, if above, then the Very High-risk zone. While efforts were made to ensure accuracy, by doing this, a level of subjectivity is introduced and may potentially be biased, again, due to individual judgments. In fact, after the review and introduction of newly assigned risk levels, it may be said most of the 2007 risk zones went up a level, and some two. Not one census tract went down in flood risk level. Because we cannot use the string data of the

[5] Note, the 2015 map shows 5 different risk zones.

flood risk variable, the 6 zones are assigned a corresponding level number. So, the "No Rating" zone is matched with a 0, and the "High Risk" zone is matched with a 5.

In general, both treatment and control census tracts, which will sometimes be referred to as *Treatment Area* and *Control Area*, are at the lower spectrum of the median household income. The two do share a border and are similar in their makeup factors, which will be discussed in section 4. They are also located at a similar distance from downtown Manhattan and other business centers. Both are close to public transport stations and share the same bus lanes[6] as well as metro lines[7]. What the study aims to find are discrepancies between the affected and unaffected *Areas*, before and after SFHA changes. The treated and control areas were matched based on individual judgments, as well as their characteristics.

To reiterate, *treatment* and *control areas* are compared, using the data from 2010 to 2021. Any differences persisting in the outcome variable from 2014 can be controlled with the DiD analysis. DiD combines time series difference and cross-sectional difference, comparing outcomes between treatment and control groups.

Furthermore, the dependent variables used are median house price, income level, and the share of poor (individuals making below $25,000). The independent variables used for the study range from economic to housing characteristics, including income and earnings by census tract, multiple housing variables, as well as house prices, information on vacancy rates, mortgages, and rent, people with social security, healthcare, and others.

The following is the DiD estimation equation for the methodology introduced above. Note, additional analyses of this estimation will be encountered in the Empirical Results section, as the section builds upon the methodology, exploring alternative specifications of the regression model. The rationale for each identification will be discussed in detail later.

$$Y_{it} = \beta_0 + \beta_1 X + \beta_2 X + \beta_n \ldots + \beta_{n+1} Treat_i + \beta_{n+2} Post_t + \beta_{n+3}(Treat_i * Post_t) + \varepsilon_{it}$$

[6] *Source:* *https://new.mta.info/project/brooklyn-bus-network-redesign*
[7] *Source:* *https://www.nycsubwayguide.com/subway/subway_map.aspx*

Where Y_{it} is the outcome variable for tract i in year t, $Treat_i$ is the binary indicator variable that equals to 1 when a tract is treated or 0 when it is controlled. $Post_t$ is the post-treatment binary indicator variable equal 1 for the post-treatment period (2014-2020) and 0 for the pre-treatment period (2006-2011). β_0 is the intercept that represents the average outcome of the control group in the pre-treatment period. β_1, β_2 and β_n are the effects of independent control variables. β_{n+1} is the average treatment effect, which indicates the differences between the treatment tracts and control tracts in the pre-treatment period. β_{n+2} is the average time effect, which represents the change in outcome for the control tracts over time. β_{n+3} is the average treatment effect over time, which indicates the differential change between the outcome of control and treatment tracts. $Treat_i * Post_t$ is the interaction term between the treatment and post-treatment dummies, and ε_{it} is the error term.

The below is an extended specification model, with fixed effects and more independent variables to control for unobserved variables. Adding many control variables may pose a threat of multicollinearity, which would make it hard to isolate their individual effects. Thus, the variables are selected carefully, based on their theoretical relevance and availability, and will be noted below. The specification now looks like this:

$$(\log)Y_{ixt} = \alpha_0 + \sum_{j=1}^{J} \alpha_i X_{ixtj} + \beta_1 T_t C_x + \gamma_x + \delta_t + \epsilon_{ixt}$$

Where $T=1$ if the observation is in the post-treatment period, and 0 if it is before. $C=1$ if the census tract is in the treated area, and 0 if it is not. Y_{ixt} denotes earnings; house price; share of poor of observation i in a specific census tract x and year t. X_{ixtj} will denote the characteristics j of observation i, γ_x the census tract fixed effect, δ_t the yearly fixed effect, and ϵ_{ixt} the error term.

In the DiD framework, β_{n+3} is of particular interest, as it estimates the treatment causal effect. Coefficients can be interpreted as follows[8]:

	Treatment Group ($T_i = 1$) (1)	Control Group ($T_i = 0$) (2)	Difference (1) – (2)
Post-Treatment Period ($P_t = 1$) (a)	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_0 + \beta_1$	$\beta_2 + \beta_3$
Pre-Treatment Period ($P_t = 0$) (b)	$\beta_0 + \beta_2$	β_0	β_2
Difference (a) – (b)	$\beta_1 + \beta_3$	β_1	$\boldsymbol{\beta_3}$

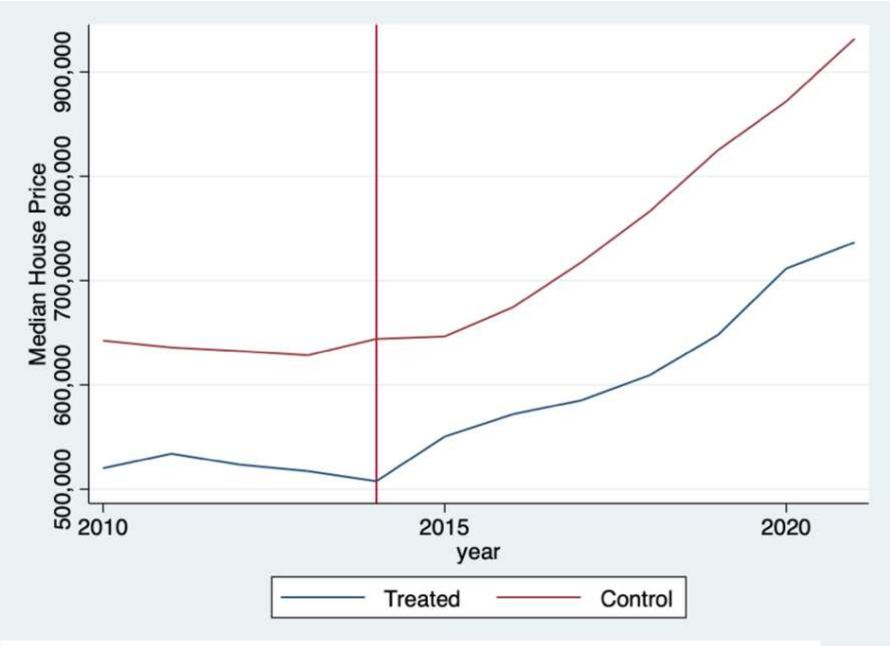
Figure 4: Interpretation of DiD coefficients

Lastly, for later regressions, the following variables should be noted: the *share of poor* represents the income groups that make up to \$25,000, divided by the total household number; similarly, the *share of lower income* represents the share of income groups that make up to \$50,000; *share of middle income* represents the share of income groups that make up to \$100,000; *share of high income* represents the share of income groups that make up to \$200,000; lastly, the *share of rich* is the share of income groups that make above \$200,000.

3.3. Data Visualization

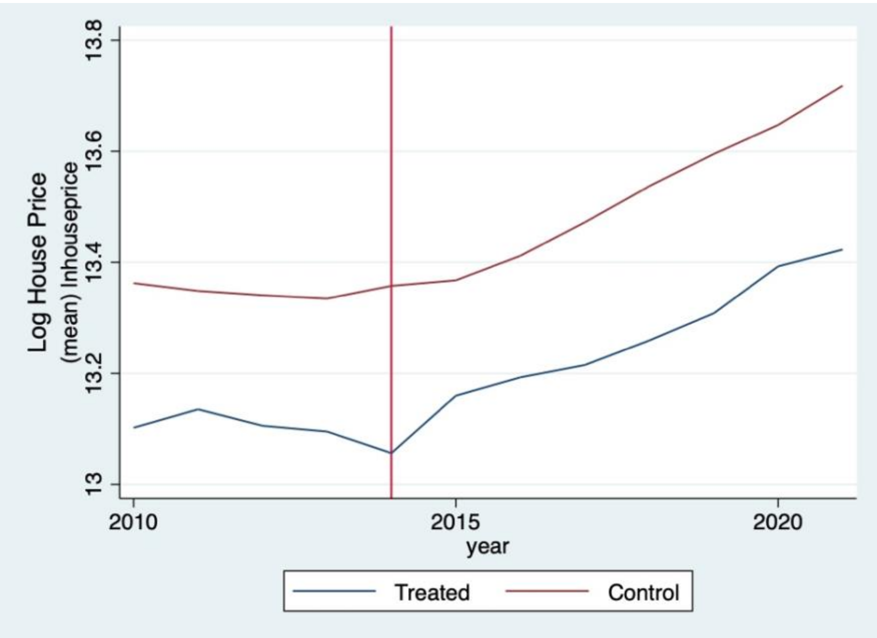
Before the regression results are presented, the overall trends of the data should be discussed. In the below plot, we can see the treatment effect which takes place in 2014. The areas in which SFHAs are changed, experience a slightly smaller growth in house prices, meaning that SFHA indeed has an impact. Because real estate can be better understood in the longer run, from 2020 onwards, noticeable changes can be seen in house prices when compared to the beginning of the treatment. The plot also reveals that the parallel trend assumption cannot be disregarded, as both the control and treatment areas experienced a similar decrease in median house prices before the treatment period. More on assumptions in Section 4.2.1.

[8] *Source: https://medium.com/eatpredlove/regression-difference-in-differences-208c2e787fd2*



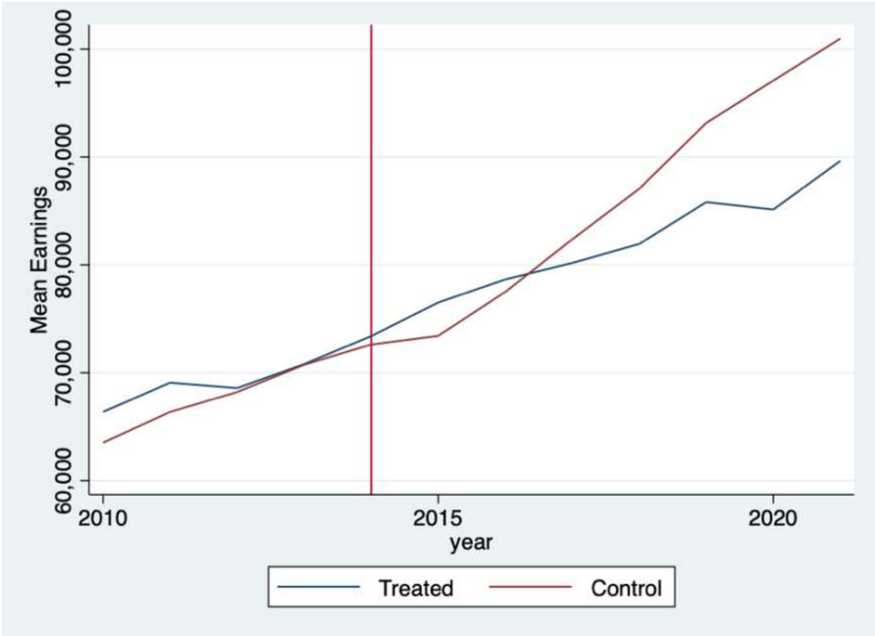
Plot 1: Treatment Effect Visualized in Median House Prices

And below is the plot of log house prices:

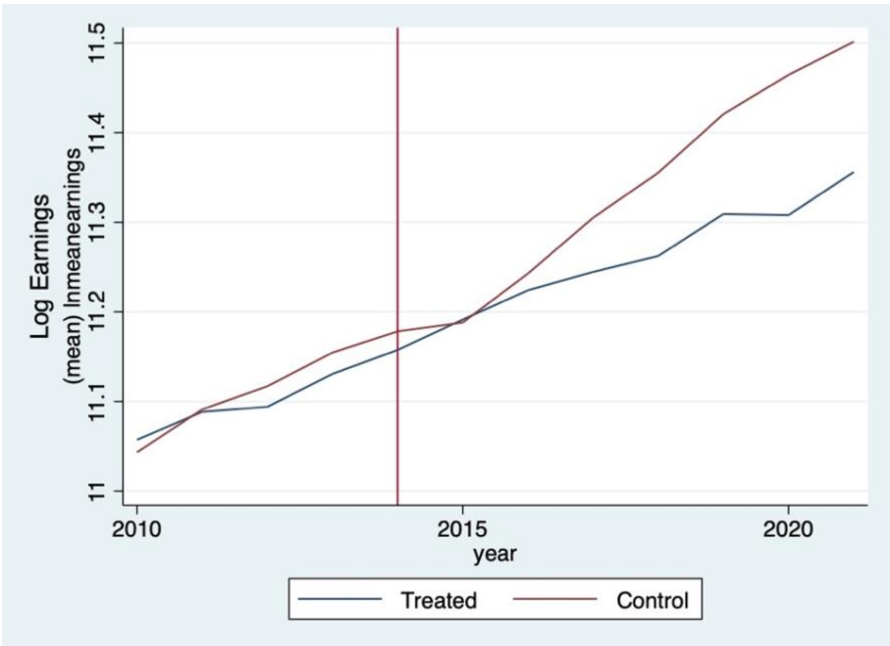


Plot 2: Treatment Effect Log House Price

Similarly, if we graph a two-way plot for mean earnings (not per household), parallel trends are noticeable before the treatment period, while after 2014, people in the control groups start earning significantly more when compared to treatment groups, where people's earnings stagnate. This may be explained by higher-earning people moving out or people earning less. The log plot is quite similar, where the control group is making just above the treated.



Plot 4: Treatment Effect Visualized in Mean Earnings



Plot 3: Treatment Effect Log Earnings

Also, the below table presents summary statistics of the flood risk, showing how many observations/census tracts are in each risk level:

Flood Risk Level	Frequency	Percent	Cumulative
0	584	51.23	51.23
1	88	7.72	58.95
2	100	8.77	67.72
3	252	22.11	89.82
4	96	8.42	98.25
5	20	1.75	100
Total	1,140	100	

Table 2: Summary Statistics of Flood Risk Levels

Flood risks 1 and 5 have the least observations, while risk level 0 has more than 50% of all observations, as it is the risk level of all the control census tracts. If flood risk 0 is not accounted for, then flood risk level 2 has the most observations with 252 or 22.11%, making it the most likely risk level for individuals residing in SFHAs.

4. Empirical Results

This section presents the empirics and results extrapolated from multiple OLS and DiD regressions. The extensive results can be found in the Appendix. To first see the significance of our treatment and post-treatment period variables, in a DiD analysis, we run an ordinary least squares (OLS) regression with no additional independent variables. The function takes a simplified form as described in Section 3.2. The coefficients are presented below:

Log House Price		
F(3, 1095)	=	91.17
Prob > F	=	0
R-squared	=	0.1775
Root MSE	=	0.316

Log Earnings		
F(3, 1136)	=	76.24
Prob > F	=	0
R-squared	=	0.119
Root MSE	=	0.26571

Regression Results	*Log House Price*	*Log Earnings*	*Share of Poor*
Treated	*-0.237**** *(-8.13)*	*-0.0089* *(-0.36)*	*0.0642**** *(5.39)*
Post Treatment	*0.167**** *(9.28)*	*0.230**** *(3.71)*	*-0.0330**** *(-4.89)*
Treated#Post	*-0.0241* *(-0.63)*	*-0.0665** *(-2.10)*	*0.00961* *(0.65)*
Constant	*13.35**** *(999.56)*	*11.10**** *(901.26)*	*0.281**** *(49.85)*
Observations	*1099*	*1140*	*1140*

Find the t statistic in parentheses

*Statistical significance is indicated by * p<0.05, **p<0.01, *** p<0.001*

Table 3: Initial Regression for Treated and Post

The F-statistic for log house price as the dependent variable is 91.17 and the P-value associated with the F-statistic is 0.0000, considering the rule of thumb (F>10), the model shows statistical significance. Besides, the P-value is below the conventional significance level of 0.05, concluding that the model indeed has some significance. The R-squared, with the current independent variables, indicates that 17.75% of the variation in the dependent variable is explained by the treatment and post-treatment period variables.

The coefficient of the treated areas dummy variable (-0.237***) indicates that on average, being in the treatment group can be associated with lower (log) median house price, compared to the control group, by around 21.1%. The three asterisks indicate that the variable has a high level of significance. The coefficient of the post-treatment period dummy variable (0.167***) indicates that, on average, the post-treatment period is associated with higher (log) median house prices, by around 18.17%. This is also statistically significant. The interaction variable between the treatment and post-treatment period does not show any significance, but it may just be because no independent and relevant variables were included in the regression.

Running the same regression, but with log mean earnings as our dependent variable gives us an F-statistic of 76.24, P-value of 0.0000, and R-squared of 0.119 indicating strong statistical significance. Treated, in this case, shows no strong significance, the post-treatment dummy variable indicates that earnings, on average, have risen in the post-treatment period. The interaction term between the two dummy variables (-0.0665*) indicates slight significance, where the census tracts that were treated, show smaller earnings, by around 6.4%, when compared to the controlled census tracts. Additionally, a variable for the share of poor, where it is the share of households who earn below $24,999 per year when considering all households, was generated. In this regression, the coefficients for treated (0.0642***) and post (-0.0330***) also show significant results, where the share of poor in treated regions is higher than in control. The dummy variable interaction (0.00961), like others, does not depict significance, but more about this later.

4.1.1 Fixed Effects DiD Regression with Control Variables

From here on, the relationship between the treatment and outcome variables will further be explored. Relevant independent variables, that are expected to capture other factors, will be added to the regression model. Moreover, to account for time-invariant characteristics (time and year), fixed effects will also be introduced into the model. This will allow for the control of unobserved heterogeneity. The results below show the selected variables from the regression:

Regression Results	(1) Log House Price	(2) Log Earnings	(3) Share of Poor
Post treatment	0.324*** (11.13)	0.374*** (7.22)	-0.03 (-1.67)
Treated#post	-0.0327 (-0.95)	-0.0505 (-1.82)	0.0045 (0.4)
Vacant Households	-0.0000557 (-0.25)		
With Social Security	-0.00000706 (0.05)	-0.00018 (-1.81)	-0.00000977 (-0.25)
Rental Vacancy Rate	-0.00192 (-0.67)		
Median number of rooms	0.0691 (0.14)		
Unemployed	-0.000233 (-1.18)		
No bedroom	-0.0000219 (-0.10)		
1 bedroom	0.0000791 (0.63)		
2 bedrooms	0.000138 (1.06)		
3 bedrooms	0.000302 (1.90)		
4 bedrooms	0.000701* (2.01)		
5 or more bedrooms	0.00115*** (4.01)		
Public Assistance			0.0000473 (0.49)
2010	0 (.)	0 (.)	0 (.)
2011	0.0137 (1.01)	0.0371** (2.95)	-0.00364 (-0.76)
2012	-0.0101 (-0.62)	0.0532** (3.28)	-0.00413 (-0.61)
2013	-0.0191 (-0.99)	0.0839*** (4.06)	0.0041 (0.48)
2014	-0.338*** (-13.24)	-0.241*** (-6.56)	0.0412*** (3.64)
2015	-0.307*** (-13.86)	-0.224*** (-5.81)	0.0395** (3.36)
2016	-0.267*** (-12.34)	-0.185*** (-5.57)	0.0276* (2.4)
2017	-0.222*** (-9.35)	-0.148*** (-5.35)	0.0219* (2.2)
2018	-0.173*** (-7.45)	-0.113*** (-4.75)	0.0173* (2.11)
2019	-0.122*** (-5.95)	-0.0588* (-2.54)	0.00596 -0.82
2020	-0.0483*** (-3.41)	-0.0344** (-2.89)	0.00522 (1.3)
2021	0 (.)	0 (.)	0 (.)
Earners		-0.0000999 (-1.08)	
Business, science, management, or arts		0.000280*** (6.23)	-0.0000321 (-1.54)
Service occupations		-0.000154* (-1.99)	0.0000123 (0.57)
Sales and office		0.0000385 (0.55)	-0.0000263 (-1.09)
Natural resources		-0.000129 (-1.17)	-0.000112** (-2.95)
Production		-0.000082 (-0.78)	-0.0000322 (-1.00)
Median Rent		0.0000612 (0.88)	-0.0000683** (-2.97)
Constant	12.96*** -66.26	11.07*** -98.5	0.417*** -13.8
Observations	1099	1136	1120

Find the t statistic in parentheses
*Statistical significance is indicated by: * $p<0.05$, ** $p<0.01$, *** $p<0.001$*

Table 4: Regression Results

After initially running the regression with the share of poor as the dependent variable, all time variables for fixed effects showed significant coefficients (*$p<0.05$, **$p<0.01$, ***$p<0.001$), with 2014 and on showing the significance of $p<0.001$. After carefully reviewing the variable, the outlier tracts, where the share of poor was more than 0.69 were removed (only for this regression), a total of 16 observations. The above results reflect that change. It is clear that the share of poor in treated tracts, after the initial treatment period, have comparatively more low-earning individuals than in non-treated tracts. This effect lasts from 2014 to 2019 when the coefficients again become insignificant. The result is as it should be expected due to the time trend of post-treatment. The interaction term, on the other hand, does not seem to indicate significance. In the post-treated tracts, SFHA change has a slightly negative effect on log earnings, a slightly negative effect on log house prices, and a slight positive effect on the share of the poor.

For log earnings, the control variables selected include socio-economic factors, like their occupation. First off, there are hundreds if not thousands of different factors that may affect a person's earnings (including occupation). Occupational coefficients, in this case, provide meaningful insights, as working in business, science, management, or arts has the biggest impact on an individual's earnings (0.000280***) increasing the mean earnings by 0.028% for each additional individual working in the field. Working in the services industry decreases the mean earnings by 0.0154% for each additional individual working in the industry.

The house prices are affected by housing factors like the number of bedrooms (see Appendix C for a full variable list and their coefficients), where having 4, 5, or more bedrooms increase the house price by 0.07% and 0.115% respectively. Again, the interaction term does not seem to indicate significance. Therefore, we cannot reject the null hypothesis yet. While the results provide satisfactory insights on the policy change, where 1 represents SFHA change and 0 does not, the findings also highlight the need for further sensitivity analysis and possible explorations of alternative model specifications. Possibly, the individual risk level, as there are 6 of them, may play a larger role in impacting the outcomes of the study. We can refine our hypothesis (H1) by proposing that it is the individual risk levels that may result in a more pronounced effect on house prices, the share of the poor, and the individual's earnings, when compared to lower risk levels. For instance, an individual, who gets 'treated' and transitions from risk level 0 to risk level 1 may experience substantially different after-effects than an individual who transitions from risk level 0 to risk level 5 or 6. In the next section, the model is re-specified to meet the newly proposed hypothesis.

4.1.2 Accounting for 6 Levels of Flood Risk

The treatment dummy only considers if a tract was treated or was not treated, but there are a total of 6 risk levels. To analyze the effect of each risk level, a new interaction variable floodrisk_numeric#post is created. See the results below.

Regression Results	(1)	(2)	(3)
	Log House Price	*Log Earnings*	*Share of Poor*
Flood risk 1	-0.0411 (-0.65)	0.0666 (1.78)	0.0467 (-1.19)
Flood risk 2	0.053 (0.54)	0.0735 (1.40)	0.0303 (0.65)
Flood risk 3	0.0958 (1.04)	0.126* (2.08)	0.0157 (0.34)
Flood risk 4	0.0571 (0.43)	0.140 (1.78)	0.00327 (0.07)
Flood risk 5	-007 (-0.49)	0.161* (2.05)	0.0491 (-0.99)
Post Treatment	0.315*** (9.95)	0.380*** (7.37)	-0.0324 (-1.67)
Floodrisk#post 1	0.0327 (0.39)	-0.0596 (-0.61)	-0.0878* (-2.01)
Floodrisk#post 2	-0.137 (-1.36)	-0.127* (-2.17)	0.0311 (-0.87)
Floodrisk#post 3	-0.120* (-2.06)	-0.102* (-2.16)	0.0154 (-1.16)
Floodrisk#post 4	-0.0026 (-0.03)	0.00124 (0.02)	0.0086 (-0.51)
Floodrisk#post 5	0.285*** (9.39)	-0.0467 (-1.75)	-0.0328** (-3.16)
Constant	12.92*** (57.14)	11.04*** (100.38)	0.414*** (-11.95)
Observations	1099	1136	1136

Table 5: Flood Risk#Post regression results

Note that the control variables were kept the same as specified earlier in the section. The new findings suggest that certain flood risk levels indeed change our dependent variables. For risk level 3, in the pos-treatment period, the coefficient (-0.120*) suggests:

$$(e^{-0.12} - 1) * 100 = -11.3\%$$

Meaning an 11.3% decrease in house prices when moving from risk level 0 to 3. On the other hand, when moving from risk level 0 to 5, the coefficient (0.285***) indicates a draconian 32.97% increase in house prices. However, in our data set, there are only 20 observations where the flood risk level is "very high", so this conclusion is only justifiably true for Lower Brooklyn and Brighton Beach. While the available data may not be all-encompassing to view the underlying truth, it nevertheless presents some interesting findings. Although, it is important to acknowledge the potential for the refinement of the analysis with an expansion of the dataset. When considering the log earnings as the dependent variable, the coefficient for 2nd and 3rd risk levels show significance. For risk level 2, the coefficient (-0.127*) indicates an 11.9% drop in earnings when a household moves from the base category of 0 to risk level 2. With a p-value of 0.032, we can state that these results are statistically significant. Similarly, in risk level 3, the coefficient (-0.102*) suggests that when compared to risk level 0, the post-treated areas, earn 9.7% less. Risk levels 4 and 5 do not show significant results.

Lastly, taking the share of the poor as our Y, the coefficients for risk levels 2 to 4 indicate raising numbers of poorer individuals in these risk zones, but the significance can be debated. Moving from risk level 0 to risk level 1, the share of poor drops by 8.4%. At least in lower Brooklyn, the share of poor also drops in risk zone 5, by 3.2%. Both level 1 and level 5 show statistical significance. The other levels have positive coefficients, with p-values being insignificant.

4.1.3 Heterogeneity Analysis on Income Groups

Previously, only individuals making up to $25,000 were included in the analysis. Here, all the different income groups are tested in a heterogeneity analysis, including the share of poor, the share of low income, the share of middle income, the share of high income, and the share of rich. The treatment group is again floodrisk_numeric#post. The aim is to test the effect of the change in SFHA on the different sub-groups. Omitted years and risk levels were removed and control variables were left the same as specified by the share of poor from the regression results above. Additionally, the coefficients for year-fixed effects are also not shown here. As the occupational control variables are not of interest, they can be found in Appendix C along with the year fixed effect coefficients.

Regression Results	(1) *Share of Poor*	(2) *Share of Mid. Income*	(3) *Share of High Earners*	(4) *Share of Rich*
Floof Risk 1	*0.0467* (*1.19*)	*-0.0232* (*-1.35*)	*0.0204* (*1*)	*0.0192* (*1.9*)
Flood Risk 2	*0.0303* (*0.65*)	*-0.0367* (*-1.45*)	*0.0416* *-(.71)*	*0.025* *-(.91)*
Flood Risk 3	*0.0157* (*0.34*)	*-0.0348* (*-1.55*)	*0.0102* (*0.43*)	*0.0371** (*2.49*)
Food risk 4	*0.00327* (*0.07*)	*-0.0348* (*-1.18*)	*0.0218* (*0.85*)	*0.0375** (*2.4*)
Flood Risk 5	*0.0491* (*0.99*)	*-0.158**** (*-5.27*)	*0.0661** (*2.48*)	*0.0443*** (*2.69*)
Post Treatmeent	*-0.0324* (*-1.67*)	*-0.0301* (*-1.88*)	*0.0194* (*1.7*)	*0.0493**** (*5.7*)
Flood Risk#Post 1	*-0.0878** (*-2.01*)	*0.0490** (*2.35*)	*-0.0355* (*-1.37*)	*-0.0136* (*-1.18*)
Flood Risk#Post 2	*0.0311* (*0.87*)	*0.0131* (*0.45*)	*-0.0298* (*-1.15*)	*-0.0272** (*-2.18*)
Flood Risk#Post 3	*0.0154* (*1.16*)	*0.018* (*1.22*)	*-0.00481* (*-0.36*)	*-0.0174* (*-1.75*)
Flood Risk#Post 4	*0.0086* (*0.51*)	*0.00067* (*0.03*)	*-0.0259* (*-1.78*)	*-0.0104* (*-1.18*)
Flood Risk#Post 5	*-0.0328*** (*-3.16*)	*0.0631**** (*7*)	*-0.0543**** (*-7.71*)	*-0.0202**** (*-4.63*)
2015	*0.0399*** (*3.2*)	*0.0175* (*1.76*)	*-0.00979* (*-1.26*)	*-0.0346**** (*-6.35*)
2016	*0.0274** (*2.3*)	*0.0154* (*1.7*)	*-0.00457* (*-0.61*)	*-0.0306**** (*-6.04*)
2017	*0.0219** (*2.08*)	*0.0109* (*1.37*)	*-0.00153* (*-0.23*)	*-0.0255**** (*-5.40*)
2018	*0.0178** (*2.2*)	*0.00973* (*1.38*)	*0.00135* (*0.21*)	*-0.0205**** (*-4.81*)
2019	*0.00644* (*0.89*)	*0.00614* (*1.11*)	*0.00466* (*0.82*)	*-0.0139**** (*-4.04*)
2020	*0.0051* (*1.22*)	*0.00212* (*0.52*)	*0.00314* (*0.83*)	*-0.00636** (*-2.34*)
Constant	*0.414**** (*11.95*)	*0.177**** (*6.95*)	*0.0498** (*2.44*)	*-0.00579* (*-0.42*)
Observations	*1136*	*1136*	*1136*	*1136*

Find the t statistic in parentheses
*Statistical significance is indicated by: * p<0.05, ** p<0.01, *** p<0.001*

Table 6: Heterogeneity Analysis on Different Income Groups

After an initial look at the findings, it can be said that the share of poor, as already discussed earlier, is decreasing in the post-treatment period, with the biggest jumps being from risk level 0 to 1 and to risk level 5. The share of middle-income earners is increasing in every flood risk level, however, similarly to the share of poor levels 1 and 5 are affected the most. By going from risk level 0 to 1, the share of middle-income earners increases by 5%, and by going from risk level 0 to 5, the share increases by 6.5%. These results are of .05 and 0.001 % significance levels respectively. The share of high earners is constantly decreasing in each risk level, with risk level 5 showing the highest statistical significance, where the share decreases by 5.28%. The share of high-income earners, interestingly, is also decreasing in each flood risk level. The highest decreases in the share can be found in risk levels 2 and 5, with 2.68% and 1.99% respectively.

4.2 Analysis of Key Findings

So far, multiple regression results have been reported and certain trends uncovered. In the sub-sections below these trends are further explored and discussed, with each dependent variable/topic (house prices, mean earnings, shares of income groups) having its sub-section. Furthermore, by examining the trends in the outcome variables in the pre-change period, the model assumptions can be assessed.

Again, as mentioned in the introduction, the initial null hypothesis (H0) states that house prices and income levels are unaffected by the SFHA change. Alternatively (H1), the results may suggest that there is a significant difference in housing prices and income levels, with higher prices and income levels being in the treated areas as compared to the controlled ones. However, the hypothesis has been altered to accommodate the changes made in the model. The new hypothesis states that each flood risk level may influence the outcome variable.

4.2.1 Assumptions

From the initial OLS regression model, For the difference-in-difference method, the following assumptions were made:

1. Parallel Trends

The parallel trends assumptions states that in the pre-policy year 2011, before the Hurricane Sandy catastrophe, the trends presented by the outcome variable should be similar between the *treatment* and *control areas*. Any noticeable differences that arise in the outcomes between the two *areas* prior to the implementation of SFHA changes, can be attributed to factors other than the changes themself. There are no statistical tests for this assumption, however visual inspection can be used when there are many time points (Columbia University, 2023). This is represented in *Plot 1* and *Plot 2*, where similar trends are seen before the treatment period. Additionally, endogeneity issues may have arisen due to the non-random assignment of the treatment areas. Thus, a common strategy was employed - control areas that were selected were close in proximity, sharing many characteristics (explained in the Methodology Section). Also, as suggested by Neumark & Simpson (2014), who argued that geographical proximity does not necessarily mean similar trends in socioeconomics, the control area selected was shown to have faced similar vulnerability levels, ethnic composition, and average income.

2. SUTVA

The Stable Unit Treatment Value Assumption assumes that there are no spillover effects between the treatment and control groups (Gerber & Green, 2010). In other words, the changes in SFHA in the treatment group do not affect the outcomes of the control group. As in the case of this study, it is assumed that the change in SFHAs in the treated tracts does not directly affect the outcomes in the control tracts, that are not subject to the same flood risks. A possible spillover may be the relocation of individuals into the control area due rising house prices, however as *Plot* 1 depicts, house prices are overall higher in the control region before and after treatment. So, even if people are moving out from the treated area, the chances of them moving to the control area are very slim. Individuals moving from the treatment to the control area due to unrelated reasons cannot be accounted for. It should also be acknowledged that some indirect factors could pose an effect on both the treated and control tracts.

4.2.2 Analysis and Discussion of House Prices

The initial visual examination of the data in a way supported the null hypothesis. The parallel trends were holding before the implementation of the treatment, and after the SFHAs were corrected, the growth of house prices of the treated census tracts started to lag the control census tracts around 5 years after the implementation. However, even though the first simple OLS regression model showed some statistical significance, the interaction term of treated#post bestowed a coefficient of -2.4%, with an insignificant p-value. As this model was in the simplest of forms, it was decided to expand on it.

After adding control variables and year and census tract fixed effects, the results were slightly different. The addition of individual characteristic independent variables improved the overall model. As expected, the number of bedrooms in the house had a positive effect on house prices. However, the addition of more control variables such as local amenities, location, and distance to the metro or school would have further improved the significance of the model. Again, the interaction term between the treated tracts and the post-treatment period presented no significance, although now the coefficient indicated 3.2% lower house prices in the treated census tracts when compared to non-treated. These are the right trends to anticipate, yet based on previous literature, like Okmoyung Bin et al. (2008), it should be expected for the trends to be more noteworthy. Overall, the results from the improved regression suggested that the treatment did not have large consequential impacts on the outcome.

Perhaps, the treatment dummy variable is not enough to see how the house prices change due to the flood risk increase. For this reason, another regression model was tried, with each risk level present in the interaction term. This time it was indeed discovered that individual risk levels had significant impacts on the outcomes, especially in risk level 3, where house prices dropped by 11.7%, and risk level 5, where house prices grew by almost 33% when compared to risk level 0. One speculation may suggest that the houses in risk level 3 are more inland, hence the demand for the house being much lower than the demand for a house on the coast, i.e., risk level 5. With higher flood risk, comes the associated cost of flood insurance. Another speculation may suggest that the 11.3% decrease could be just a trade-off between the house price and the insurance cost. Besides, the drop in prices may just reflect the concerns of potential floods, with individuals having lived through Hurricane Sandy not too long ago. It should also be mentioned that his number is very close to the numbers presented by O. Bin et al., mentioned previously, and indeed, something that is expected to happen. On the other hand, the 33% jump in house prices in level 5 could align with the expectation that coastal buildings come with desirable features, like distance to the beach and surrounding amenities. A study done by a professor from Yale University suggests that prices are not falling in the coastal areas that are expected to be affected the most (Allen, 2021). This reason alone may not drive up the value as much when considering flood-associated risks. Hence, further research is required.

4.2.3 Analysis and Discussion of Mean Income

As for the earnings, the initial OLS regression, unlike the Log House Prices, did show significance in its interaction term between treated tracts and the post-treatment period. The coefficient indicated lower income by 6.4% in the treated tracts post-treatment period. The model was then expanded. With control variables not added, occupational characteristics were the biggest determinants of the log mean earnings. As expected, working in management, any business field, science field, or arts had the biggest positive impact on earnings. Working in the services industry, however, negatively affected the mean earnings. Heidi Shierholz (2014) explains that the highest-paid service industry workers are the managers, who make around $15.42 an hour, which is still lower than the overall median wage outside of the industry. The results do follow the expected trends. However, it seems that the economic background due to the increase in flood risk does not change significantly, and the insignificant interaction term may suggest the presence of time-specific effects and factors that influence earnings.

When considering the individual flood risk levels, risk levels 2 and 3 display a drop in mean earnings of 11.9% and 9.7% respectively, which may imply that these risk levels may have adverse effects on economic opportunities or job prospects. The drop in earnings is very close to the drop in house prices, and when two variables are taken together, the change in SFHAs highlights the financial burden of residing in these flood-prone areas. The findings emphasize the need for successful risk management strategies and mitigation programs to ensure the continued resilience of these communities. The drop may also represent a shrink in the total number of people living in a household, but this is not tested. Risk levels 4 and 5 do not show significance, either due to data availability or, maybe, due to these zones being very close to the coast, where richer people live, as is reflected in the house prices of the same risk zones. Either way, the results hint towards rejecting H0 and accepting the refined hypothesis, which states that the flood risk effect can be seen in individual risk levels. Nevertheless, for a stronger statement, more significant coefficients should be established, and other factors investigated.

4.2.4 Analysis and Discussion of Income Groups

Lastly, the initial model included only the share of the poor as the dependent variable and only later the heterogeneity analysis was performed. The first simple OLS regression did not reveal much, just like with the other variables. In the DiD regression with fixed effects and control variables, the interaction term again showed no significance, but other control variables did. Just like with log earnings, the occupation variables had a big effect on the total share of poor people living in the selected census tracts. All occupations have a positive effect on decreasing the share of the poor, except the service industry, which, negatively affects the share of the poor in a tract.

When considering all the different flood risk levels, some effects on the outcomes can be observed. Flood risk levels 1 and 5 show a decreasing number in the share of the poor. These findings could be viewed from two angles, where changes in SFHA positively affect the poor as there are fewer of them, or it negatively affects them as the reason may be them moving out. It could be assumed that they were forced to move out due to rising insurance prices and an increase in regulations, but then it would also be expected for the other flood risk levels to fall under similar trends. Again, this could be the size of the dataset issue. A report from FEMA itself has disclosed that the insurance affordability program for the poor could be reworked (Scata, 2018). Further studies should investigate the correlation between the drop in the number of households that make under $25,000 and the changes in flood risk levels. As mentioned previously, H0, at least in lower Brooklyn, could be rejected if no other factors are taken into consideration. As per the study, it could be presumed that individual risk levels do matter and pose an influence on the outcome variables.

Finally, heterogeneity analysis on all the different income groups was performed. Interestingly, the share of middle income could be named as an outlier, as in all flood risk levels, the share is growing, while all other income groups are contracting. Just like with the share of poor, middle-income share experiences its biggest changes in flood risk levels 1 and 5. Economically, this is debatable, as in the past decades, the middle class has fallen by around 11% in the USA, where increases were recorded in the upper- and lower-income groups (Kochhar & Sechopoulos, 2022). Could this mark a decrease in upper-income shares in Brooklyn, signal a transforming share of poor, from low-income to middle-income, or could it both? With the evidence presented by the research, the latter sounds plausible, as the share of high-earning households and households making above $200,000 has also been decreasing in each flood risk level, with level 5 presenting the most significant decreases in the shares. This finding, however, could interfere with the findings presented by regressing the housing prices. Flood risk level 5 has seen drastic increases in housing prices, but also a massive drop in the share of high-earning households. The results indeed show that flood risk level change affects the discussed topics, but it is not entirely certain if it is the increased regulations or cognitive grasp on the issues at hand. The study could be improved by conducting surveys in Lower Brooklyn to better understand the reasons for such undertakings.

5. Conclusion

To conclude, the study aimed to investigate and assess the impact of SFHA change in Lower Brooklyn post-Hurricane Sandy. An analysis was performed on housing affordability, the effect on shares of different income groups, and the mean earnings of selected census tracts. Through a comprehensive analysis and multiple regression models, including fixed effects and heterogeneity testing the impact of the change flood risk mapping. Literature reveals that flood risk maps developed by FEMA fail to consider the differences between all the subgroups in the flood 'hazard' areas. Map changes are responsible for flood insurance rates and additional regulations, with the belief being that more variables could alleviate some of the problems caused by the mapping processes. The data from the American Community Survey is used to test the hypothesis that the change in flood risk levels does not influence the selected variables. It is discovered that that effect on housing prices and the overall composition of the selected census tracts is noticeable. When the treated areas i.e., census tracts where the SFHAs were changed, are represented by a dummy variable (1 for treated, 0 for non-treated), the post-treatment period does not reproduce the results needed to reject the above state hypothesis. However, when each flood risk is considered, some risk levels have more significance in affecting the outcomes than others. The share of the middle-income group has grown at almost every level, while the poor and rich have shrunk. House prices have fallen in the lower risk levels while raising drastically in the highest risk level 5. The mean earnings of the households also shrink in some risk levels more than others, showing similar numbers to the decrease in house prices – by around 11%. Nonetheless, the study raises other concerns about data availability and the need for further studies, especially on individual responses themselves, which would allow us to see the individual migration (i.e., moving in or out) patterns more clearly. It should also be noted that the study was performed using survey data from the Lower Brooklyn area, meaning the transferability to other areas is uncertain.

6.Bibliography

Allen, S. (2021, April 20). *Study: Rising Seas Aren't Causing Coastal Property Values to Decline*. Retrieved from Yale Insights: https://insights.som.yale.edu/insights/study-rising-seas-arent-causing-coastal-property-values-to-decline

Atreya, A., Ferreira, S., & Kriesel, W. (2013). Forgetting the Flood? An Analysis of the Flood Risk Discount over Time. *Land Economics*, 577-596.

Bagley, K. (2013). *Climate Change Impacts Absent from FEMA's Redrawn NYC Flood Maps*. Retrieved from Inside Climate News: https://insideclimatenews.org/news/06022013/climate-change-global-warming-flood-zone-hurricane-sandy-new-york-city-fema-federal-maps-revised-sea-level-rise/

Bin, O., Kruse, J. B., & Landry, C. E. (2008). Flood Hazards, Insurance Rates, and Amenities: Evidence From the Coastal Housing Market. *Journal of Risk and Insurance*.

Chandra-Putra, H., & Andrews, C. J. (2019). An integrated model of real estate market responses to coastal flooding. *Journal of Industrial Ecology*.

Columbia University. (2023). *Columbia University Irving Medical Center*. Retrieved from Difference-in-Difference Estimation: https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation#:~:text=Parallel%20Trend%20Assumption&text=It%20requires%20that%20in%20the,observations%20over%20many%20time%20points.

Dixon , Lloyd; Clancy, Noreen; Seabury, Seth A.; Overton, Adrian;. (2006). *The National Flood Insurance Program's Market Penetration Rate: Estimates and Policy Implications*. Santa Monica: RAND Corporation. Retrieved from FEMA: https://www.fema.gov/sites/default/files/2020-07/fema_nfip_eval_market_penetration_rate.pdf

Faber, J. W. (2015). Superstorm Sandy and the Demographics of Flood Risk in New York City. *Human Ecology*, 363-378.

FEMA. (2021). *The National Risk Index*. Retrieved from FEMA: https://hazards.fema.gov/nri/

First Street Foundation. (2019). *Defining America's Past, Present, and Future Flood Risk*. Retrieved from First Street Foundation: https://firststreet.org/risk-factor/flood-factor/

Gerber, A. S., & Green, D. P. (2010). The Stable Unit Treatment Value Assumption (SUTVA) and Its Implications for Social Science RCTs. Yale University.

Hallstrom, D. G., & Smith, V. (2005). Market responsess to hurricanes. *Journal of Environmental Economics and Management*, 541-561.

Herreros-Cantis, P., O. V., Grabowski, Z. J., & & McPhearson, T. (2020). Shifting landscapes of coastal flood risk: environmental (in)justice of urban change, sea level rise, and differential vulnerability in New York City. *Urban Transform 2*.

Kahneman, D., & Tversky, A. (1982). *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kochhar, R., & Sechopoulos, S. (2022, April 20). *How the American middle class has changed in the past five decades*. Retrieved from Pew Research Center: https://www.pewresearch.org/short-reads/2022/04/20/how-the-american-middle-class-has-changed-in-the-past-five-decades/

Koslov, L. (2019). How maps make time. *City: Analysis of Urban Change, Theory, Action*.

Kousky, C., Kunreuther, H., LaCour-Little, M., & Wachter, S. (2020). Flood Risk and U.S. Housing Market. *Journal of Housing Research*.

Kozlowski, J., Veldkamp, L., & Venkateswaran, V. (2015). *THE TAIL THAT WAGS THE ECONOMY: BELIEF-DRIVEN BUSINESS CYCLES AND PERSISTENT STAGNATION*. Cambridge, Massachusetts, USA: National Bureau of Economic Research.

Lieberman-Cribbin, W., Gillezeau, C., Schwartz, R. M., & Taioli, E. (2021). Unequal social vulnerability to Hurricane Sandy flood exposure. *journal of exposure science & environmental epidemiology* .

Maantay, J., Maroko, A., & Culp, G. (2009). Using Geographic Information Science to Estimate Vulnerable Urban Populations for Flood Hazard and Risk Assessment in New York City. *Geospatial Techniques in Urban Hazard and Disaster Analysis*, 71-97.

National Flood Services. (2021). *Why an Incomplete Picture of Flood Risk Makes Marginalized Communities More Vulnerable to Flooding*. Retrieved from National Flood Services: https://nationalfloodservices.com/blog/why-an-incomplete-picture-of-flood-risk-makes-marginalized-communities-more-vulnerable-to-flooding/

Neumark, D., & Simpson, H. (2014). PLACE-BASED POLICIES. *NBER WORKING PAPER SERIES*.

NYC Comptroller. (2023). *Ten Years After Sandy Barriers to Resilience*. Retrieved from New York City Comptroller Brad Lander: https://comptroller.nyc.gov/reports/ten-years-after-sandy/#:~:text=Snapshot%20of%20Key%20Findings&text=Of%20the%20%2415%20billion%20of,%25%2C%20as%20of%20June%202022.

NYC Mayor's Office of Climate Change & Environmental Justice. (2022). *Chronic Tidal Flooding*. Retrieved from Climate City of New York: https://climate.cityofnewyork.us/challenges/chronic-tidal-flooding/

NYC Planning. (2010, 2020). *Population FactFinder*. Retrieved from Population FactFinder: https://popfactfinder.planning.nyc.gov/#14.72/40.59476/-73.97108

Ortega, F., & Taṣpınar, S. (2018). Rising sea levels and sinking property values: Hurricane Sandy and New York's housing market. *Journal of Urban Economics*, 81-100.

Pralle, S. (2019). Drawing lines: FEMA and the politics of mapping flood zones. *Climatic Change*.

Rhode, D. (2012, October 31). *The Hideous Inequality Exposed by Hurricane Sandy*. Retrieved from The Atlantic: https://www.theatlantic.com/business/archive/2012/10/the-hideous-inequality-exposed-by-hurricane-sandy/264337/

Sanjay, M., Garrahan, S., & Raghuveer, V. (2020). The flood insurance gap in the United States. *Climate Change and Natural Disasters*.

Scata, J. (2018, April 24). *Flood Insurance Subsidies Must Include Options to Lower Risk*. Retrieved from NRDC: https://www.nrdc.org/bio/joel-scata/flood-insurance-subsidies-must-include-options-lower-risk#:~:text=Poor%20Disproportionately%20Exposed%20to%20Flooding&text=FEMA%27s%20research%20found%20that%20low,in%20high%2Drisk%20flood%20zones.

Shao, W., Xian, S., Lin, N., Kunreuther, H., Jackson, N., & Goidel, K. (2017). Understanding the effects of past flood events and perceived and estimated flood risks on individuals' voluntary flood insurance purchase behavior. *Water Research*, 391-400.

Shierholz, H. (2014, August 21). *Low Wages and Few Benefits Mean Many Restaurant Workers Can't Make Ends Meet*. Retrieved from Economic Policy Institute: https://www.epi.org/publication/restaurant-workers/

Sun, D., Li, S., Zheng, W., Croitoru, A., Stefanidis, A., & Goldberg, M. (2015). Mapping floods due to Hurricane Sandy using NPP VIIRS and ATMS data and geotagged Flickr imagery. *International Journal of Digital Earth*.

Tesselaar, M., Botzen, W. J., Haer, T., Hudson, P., Tiggeloven, T., & Aerts, J. C. (2020). Regional Inequalities in Flood Insurance Affordability and Uptake under Climate Change. *sustainability*.

U.S. Bureau of Labor Statistics. (2019, December 6). *American Community Survey (ACS) Questions and Answers*. Retrieved from bls: https://www.bls.gov/lau/acsqa.htm#:~:text=The%20ACS%20is%20a%20large,3.5%20million%20household%20addresses%20annually.

White, J. R. (1977, June 13). Bergenfield Homeowners Protest Designation as Flood-Hazard Area. (I. T.N. Times, Interviewer)

Yarina, L., Mazereeuw, M., & Ovalles, L. (2020). A retreat critique: Deliberations on design and ethics in the flood zone. *Journal of Landscape Architecture*, 8-23.

7. Appendix

A.

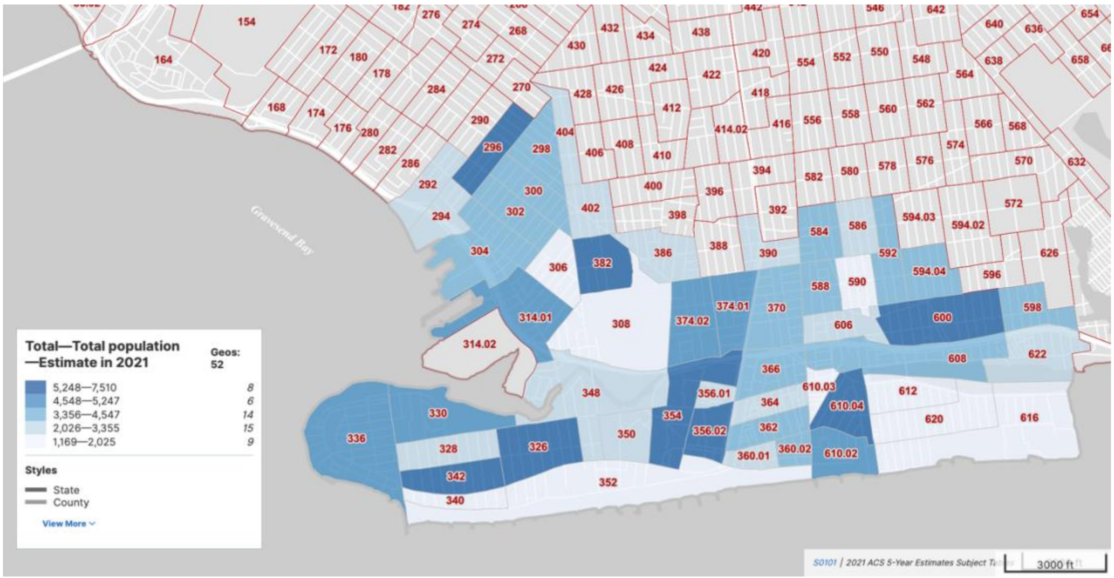


Figure 5: Treatment Census Tracts

The map above shows the treatment census tract numbers in lower Brooklyn, except 314.01, 314.02, 594.04 and 598 due to the lack of historical data. The treatment tracts that were selected are all shaded in. The map below shows all the selected control census tract numbers, except 594.03, 168 and 172 shaded in. Both *Areas* share a border.

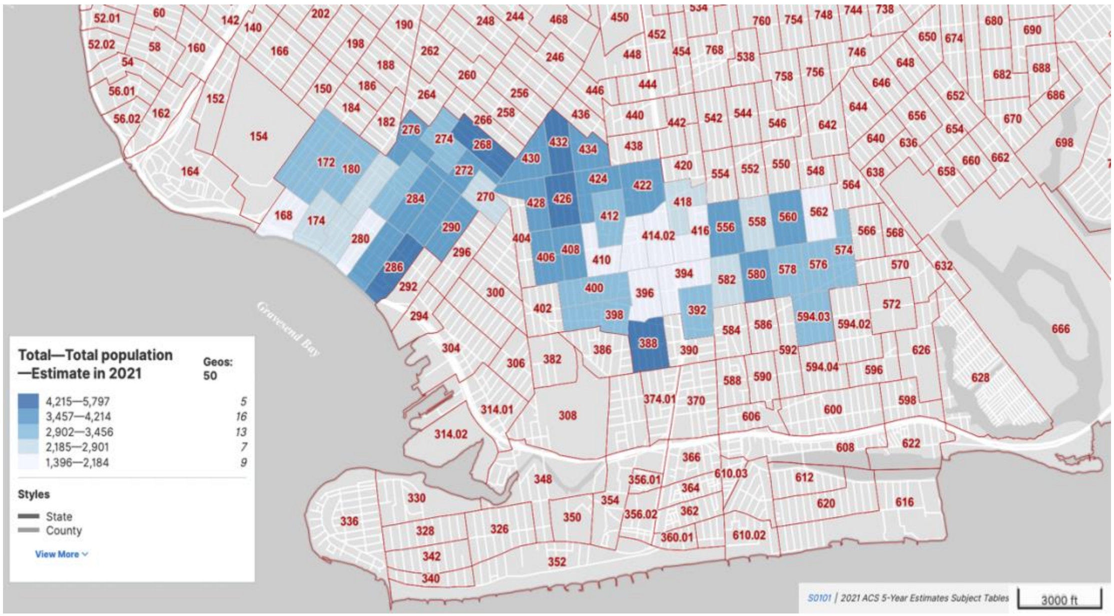


Figure 6: Control Census Tracts

B.

DiD Regression Results (OLS)

Log Earnings	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
1.treated	-0.0088991	0.0247327	-0.36	0.719	-0.0574259	0.0396277
1.post	0.2304935	0.0168077	13.71	0	0.1975158	0.2634711
treated#post						
1 1	-0.0664679	0.0316637	-2.1	0.036	-0.1285938	-0.0043421
_cons	11.10156	0.0123178	901.26	0	11.07739	11.12573

Table 7: Regression on Log Earnings

Log House Price	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
1.treated	-0.2370202	0.0291547	-8.13	0	-0.2942257	0.1798148
1.post	0.1665366	0.0179506	9.28	0	0.131315	0.2017581
treated#post						
1 1	-0.0241022	0.0381584	-0.63	0.528	-0.098974	0.0507696
_cons	13.34659	0.0133524	999.56	0	13.32039	13.37279

Table 8: Regression on Log House Prices

Share of Poor	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
1.treated	0.0642408	0.0119239	5.39	0	0.0408454	0.0876362
1.post	-0.0329599	0.0067387	-4.89	0	-0.0461816	-0.0197382
treated#post						
1 1	0.0096057	0.0146764	0.65	0.513	-0.0191902	0.0384016
_cons	0.2810647	0.0056384	49.85	0	0.2700019	0.2921276

Table 9: Regression on Share of Poor

Note that the share of poor is a generated variable, that adds up the estimates of households that make less than $10,000; $10,000 to $14,999; $15,000 to $24,999 and divides the number by the estimate for total households.

C.

	1 Share of Poor	2 Share of Mid. Income	3 Share of High Earners	4 Share of Rich
Business, management, or arts	-0.0000299 (-1.48)	-0.00000712 (-0.52)	0.0000211 -1.59	0.0000256* -2.49
Service occupations	0.0000162 -0.74	0.0000537** -3.11	0.0000331 -1.89	-0.0000388*** (-3.54)
Sales and office	-0.0000418 (-1.61)	0.0000166 -0.76	0.0000214 -1.24	0.00000357 -0.24
Natural resources	-0.000124** (-3.33)	0.000000222 -0.01	-0.0000158 (-0.51)	-0.0000181 (-1.01)
Production	-0.0000433 (-1.17)	0.0000579 -1.89	0.000000974 -0.04	-0.0000114 (-0.63)
Median Rent	-0.0000644* (-2.42)	0.0000131 -0.52	0.0000285 -1.61	0.0000207 -1.65
2011.year	-0.00453 (-0.93)	0.0000785 -0.02	-0.00264 (-0.76)	0.00168 -1.13
2012.year	-0.00543 (-0.79)	-0.00187 (-0.33)	-0.0027 (-0.62)	0.00321 -1.26
2013.year	0.00132 -0.15	-0.00702 (-0.96)	-0.00322 (-0.58)	0.00348 -1.05
2014.year	0.0413** -3.36	0.0156 -1.52	-0.0171* (-2.13)	-0.0367*** (-6.38)
2015.year	0.0399** -3.2	0.0175 -1.76	-0.00979 (-1.26)	-0.0346*** (-6.35)
2016.year	0.0274* -2.3	0.0154 -1.7	-0.00457 (-0.61)	-0.0306*** (-6.04)
2017.year	0.0219* -2.08	0.0109 -1.37	-0.00153 (-0.23)	-0.0255*** (-5.40)
2018.year	0.0178* -2.2	0.00973 -1.38	0.00135 -0.21	-0.0205*** (-4.81)
2019.year	0.00644 -0.89	0.00614 -1.11	0.00466 -0.82	-0.0139*** (-4.04)
2020.year	0.0051 -1.22	0.00212 -0.52	0.00314 -0.83	-0.00636* (-2.34)

Table 10: Control variable of the heterogeneity analysis