Climate Vulnerability in Boston:

Through the Lens of Infrastructure

Executive Summary

This paper details the development, findings, and ancillary analysis of a Climate Vulnerability Metric created using Building Permit and adjacent Tax Assessment data for the City of Boston. The metric aims to identify communities where public and private infrastructure is potentially insufficient to shelter occupants as weather and heat events are made more severe by climate change.

Main Findings

- High Vulnerability census tracts are clustered in **Dorchester, Hyde Park, and East Boston.** Low Vulnerability census tracts are clustered in **Back Bay, the South End, and Downtown.**
- There is a significant relationship between the racial demographic of a census tract and the Vulnerability measure.
    - Moderate correlation ($R = +0.58$) between proportion POC and Vulnerability ($n = 176$).
    - Significant difference in mean Vulnerability between majority white census tracts ($n=94$) and majority POC census tracts ($n=82$): **$t = -8.367$, $p=1.9e-14$**
    - The 13 least vulnerable census tracts are all majority white.
    - **The 34 most vulnerable census tracts are all majority People of Color.**
- Since these results are based solely on the built environment, it indicates that **climate injustice is built into the physical infrastructure of the city.**

Implications on Policy

Organizations such as Climate Ready Boston and the Metropolitan Area Policy Council (MAPC) are already aggregating Climate Vulnerability Metrics on the local level. These metrics incorporate a variety of social, economic and geographic factors. The Building Permit approach in this paper constrains this research to solely infrastructure, resulting in a description of the climate inequities that are physically ingrained into Boston's built environment. It is important for policymakers to be proactive on this issue because of the time, money and public commitment it takes to improve infrastructure at this scale.

Introduction

In December 2016, Climate Ready Boston released an executive summary (Arcadis Et al., 2016) outlining detailed projections for the climate crisis' impact on the city. Unavoidably, Boston is going to experience more extreme temperatures and heavier precipitation events in the coming decades. The average summer temperature, historically 69 °F, could increase to 76 °F by 2050 and **84 °F by the end of the century**. From 1958 to 2010, there was a **70% increase in the amount of precipitation** that fell in the heaviest events. This trend is expected to continue, leading to more dangerous blizzard and flooding events.

For Climate Ready Boston and other organizations dedicated to researching these impacts, a common tool is to model vulnerability geographically. The Climate Ready Boston summary (above) includes maps detailing the changing flood threats over time. MAPC takes a wider approach to a vulnerability map, detailing the risk of both Heat and Flood threats in their model for the Greater Boston Area (Flingai, 2019). There are also resources detailing the social vulnerabilities to the climate crisis. Climate Ready Boston's Social Vulnerability Index (Boston Maps, 2017) contains population counts of marginalized and vulnerable groups around the city. This data will be used later in the paper to compare to an infrastructure-based Vulnerability Metric.

In tune with many public crises, the effects of climate change disproportionately impact minority groups. In the Northeast Region of the US, **minority groups are projected to see 10% more new asthma diagnoses** than whites due to a global Temperature increase of 2 °C. They are also **15% more likely to experience property damage or loss** under the same Temperature increase (EPA, 2021).

A major factor in these discrepancies is the historical impact of **spatial segregation**. Even today, **two-thirds of Boston's Black population reside in 3 neighborhoods**: Dorchester, Roxbury and Mattapan (Data: Analyze Boston - Demographics, 2021). These and other neighborhoods with a majority of People of Color tend to have disadvantaged economic and physical infrastructure as a result of historic marginalization. Roxbury and Dorchester have some of the oldest building stocks in the city, ranking 5th and 6th out of all neighborhoods with a **median age of around 105 years** each (Jtpollak, 2015).

This built environment has significant implications on community health as it is the first line of defense against the elements. Both public and private infrastructure in an area provide emergency and everyday shelter for severe heat and weather events. There is also an argument that general disorder and dilapidation in an area impacts overall safety and quality of life (Kelling, 1999).

Regardless, infrastructure resilience is a crucial factor in mitigating the impacts of climate related disasters. In this paper, a Climate Vulnerability Metric based solely on Boston's built environment will be created. The metric will be aggregated using both Building Permit data and

Property Assessment data to incorporate existing infrastructure and active renovation patterns, creating a nuanced representation to help inform policymakers of climate injustice in the city.

Data & Methods

Data: Building Permits

The Building Permits Dataset is an aggregate of 480,000+ Permit Requests collected over the last 15 years. Sourced from BARI via the Harvard Dataverse, the data contains information on the applicant, construction type and estimated valuation. The data is marked geographically by Parcel, Block and Census IDs. The rest of the data is mostly text-based, with the estimated valuation and the permit fee being the only quantitative values included. As such, a significant amount of preprocessing is required for any numerical analysis.

To simplify the creation of the Vulnerability Metric, the Building Permit data is trimmed to only include records marked as a “Renovation”. The other three types (New Construction, Addition and Demolition) all have different scopes and implications which add unnecessary variance. This action does not sacrifice sample size as most Permits are marked as Renovation; the resulting subset still has 400,000+ records.

Data: Property Assessment Data

The Property Assessment is a wide-ranging description of 175,000 properties across the city of Boston last updated in 2019. Also from BARI via the Harvard Dataverse. the data contains many property details such as Living Area, building age and Air Conditioning type. The data is marked geographically by Parcel, Block and Census IDs. As such, it can be easily matched with the Building Permit Data, resulting in **150,100 total records** of combined data. To satisfy the tradeoff between sample size and noise, the data is aggregated at the Census Tract level (n=179).

Supporting Data: Climate Ready Boston Social Vulnerability

The Climate Ready Boston Social Vulnerability data (Boston Maps, 2017) is an index of 7 social metrics known to impact community resilience. These include Older Adults, Children, People of Color, Limited English Proficiency (LEP), Low to No Income, People with Disabilities and Medical Illness. The data is aggregated on the Census Tract Level and is available as a Shapefile. This information will be used to compare the calculated vulnerability metric with other social factors present in the City of Boston.

Supporting Data: US Disaster Fatality Data

The United States Disaster Fatality Data is used in creating weights for the Vulnerability Metric. The data is a cumulative table and is shown in detail in the **Methodology Appendix 3**.

Metric Creation: Theory & Components

The first step in creating the Climate Vulnerability Metric was to identify characteristics of vulnerable infrastructure. The original list of potential factors is displayed in the **Methodology Appendix 1**. The 4 factors that were pursued in the final project are as follows:

1. Value of New Investments in the Property
2. Presence of Air Conditioning
3. Age of Property
4. Owner Attention → Frequency of Upkeep / Repairs

The first factor can be found directly in the **declared_valuation** variable in the Permit dataset. This records the monetary estimation of the value added by the project according to the applicant. While this value is subjective and has some empty values, that uncertainty is accounted for when adding weights to the components. For ease of development, it is renamed **val**.

The second factor can be found directly in the **R_AC** variable in the Assessments data. This describes the type of Air Conditioning in the property as one of four categories: 'central_ac', 'ductless_ac', 'none' or 'yes'. The 'central_ac', 'ductless_ac', and 'yes' were set to True and the 'none' was set to False to create a Boolean representation.

The third factor can be derived from the **YR_BUILT** variable, also in the Assessments data. The variable **age** is calculated by subtracting the year from 2021.

The final factor requires a string-based comparison to identify repair frequency. This is aggregated into a single variable, **repair**, which is detailed in the **Methodology Appendix 2.**

Metric Creation: Aggregation

To create a representation on a geographical level, these 4 variables are aggregated by census tract. The quantitative variables (**val, age, repairs**) are aggregated by mean. The **R_AC** is used to create **no_ac**: the proportion of the Census Tract without any sort of Air Conditioning infrastructure. Sample size concerns for these aggregations are described in **Metric Creation: Validation**. Once the values for each variable are aggregated, each vector is linearly scaled to a range of [0,1] for simplicity.

Metric Creation: Weighting

Once all 4 components are created, they can be aggregated into a single metric. To do so, it is important to appropriately adjust them by their individual impact on Climate Vulnerability. Since there isn't any training or reference data relevant enough to learn weights statistically, the weights are calculated using a rule-based system described in depth in the **Methodology Appendix 3.** The finalized component weights are shown in Table 1 below.

Table 1. Component Weights

Component	Weight
no_ac	0.367
repair	0.334
age	0.207
val	0.092

Metric Creation: Validation

The following chart shows the distribution of the final vulnerability metric.

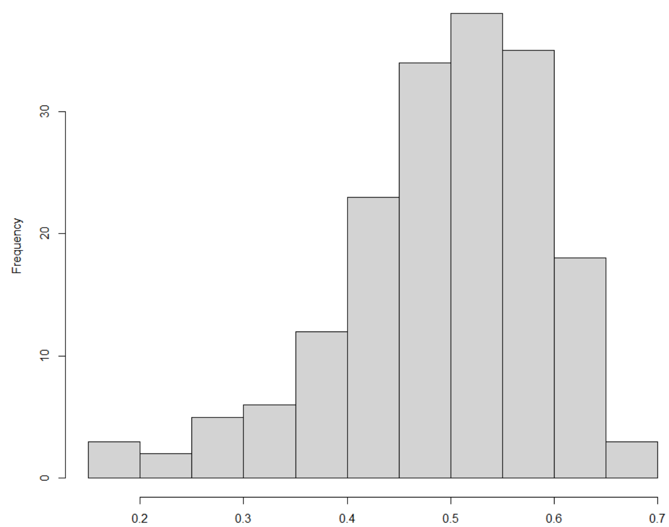

Figure 1. Frequency Histogram

Figure 1 reveals a semi-normal distribution with left skew and an overall range of [0.1, 0.7]. The reason that there is no data at the extreme end of the possible range [0,1] is due to the cumulation and relative scaling of the metric. For a Census Tract to have a Vulnerability of 1, for example, it would have to have the most severe value for each of the four components. Since this is not expected of any single part of the city, the lack of extreme values is not a reason for concern.

The more serious worry is sample size. Even though there are over 150,000 total values for just 179 census tracts, tracts with low population can experience a lack of data. The chart below shows the Histogram of Sample Counts that went into each Tract’s Vulnerability Calculation.

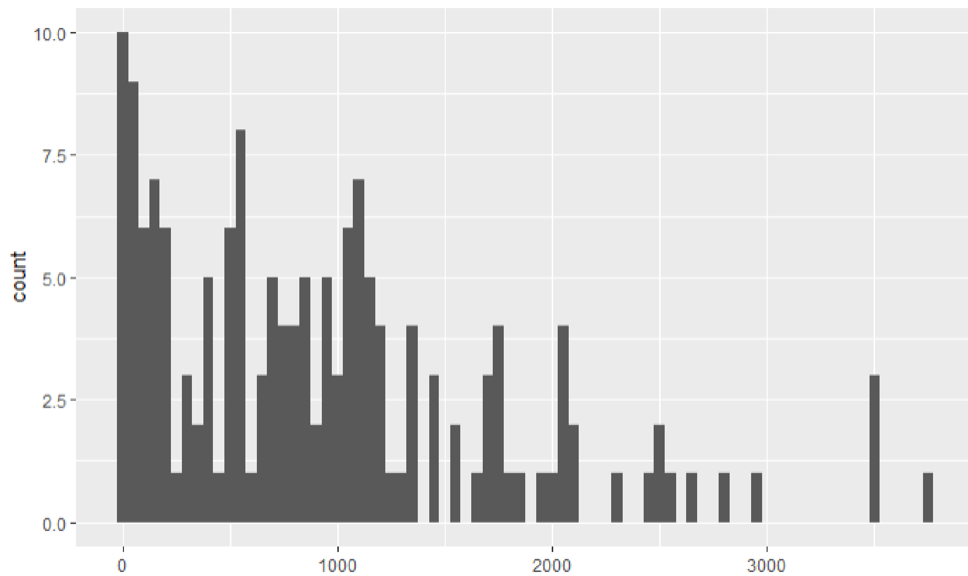

Figure 2. Sample Counts in each Census Tract

The graph reveals that there are 10 Census Tracts (5.6% of sample) with less than 30 records. Recall that each record is a single building permit. As such, the data points may be subject to noise if there is an extreme value being factored into the aggregation. For example, if an applicant here added an extra digit to their declared valuation, it would spike the mean for the overall **val** variable much more than in a district with 1000+ records.

In addition to these 10 tracts, there are 3 tracts (1.7% of sample) with no records at all. Counterintuitively, this is not as much of an issue as the tracts can be ignored in any mapping or analysis. This is an appropriate dismissal as all of these areas are census tracts with no built environment.

Overall, the vulnerability metric has no extreme issues that would invalidate any analysis. While sample size does have the potential to cause noise, it is a small enough proportion of the city to not impact any overall trends. However, it does have some quantitative limitations. It is important to remember that the Vulnerability is an abstract metric. Due to its cumulative nature, Vulnerability is only relevant when compared to another Census Tract. This lack of tangibility may cause confusion when viewing any regression or numerical analysis.

Results & Discussion

Geographic

The following figure displays the heat map of vulnerability. Note that darker shades of red indicate higher Vulnerability; since Vulnerability is abstract the scale is omitted.

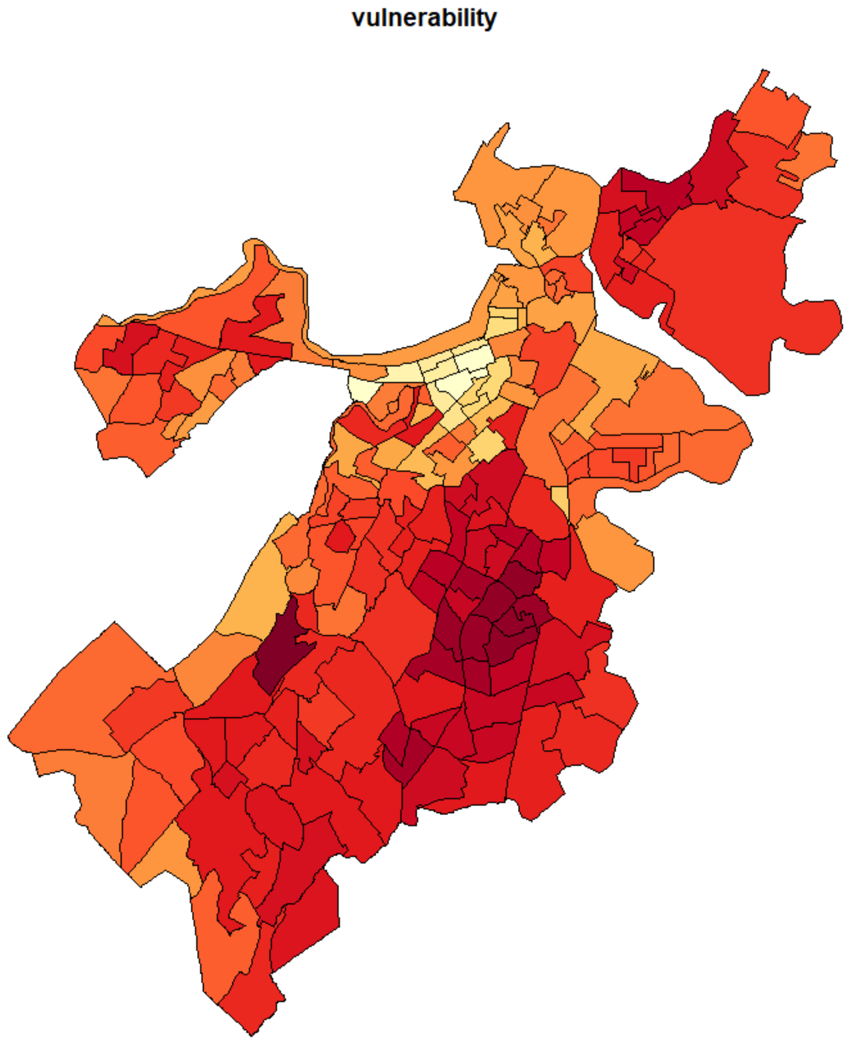

Figure 3. Climate Vulnerability Map

The most vulnerable cluster of Census Tracts is located in Dorchester. Mattapan, Hyde Park and East Boston also have high vulnerability rankings. On the other hand, Back Bay, the South End and Downtown have the lowest vulnerability rankings.

Note that there is some noise as anticipated in **Metric Creation: Validation**. The highest vulnerability in the city (see Jamaica Plain) belongs to a tract mostly occupied by the Arnold Arboretum. Since there isn't much built environment here, the individual permits filed have an extreme impact on the tract's overall Vulnerability.

Trends: Exploratory Analysis

To explore overall trends in the distribution of Vulnerability, the census level data is merged with the Climate Ready Boston Social Vulnerability Index. A scatterplot matrix is created to identify features of interest.

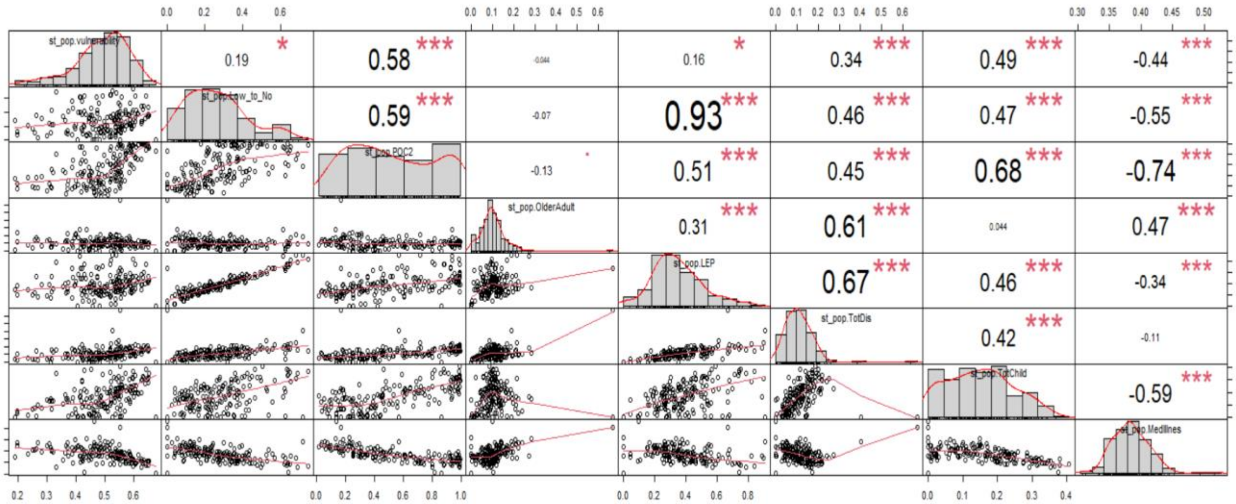

Figure 4. Exploratory Analysis between Climate & Social Vulnerability

From the figure above, the most strongly related feature is **POC2**, which represents the Proportion of People of Color in any given census tract. Some other features that display significant positive relationships are **TotDis** and **TotChild**, which represent the proportion of People with Disabilities and the Proportion of Children, respectively. Interestingly, there is one category with a significant negative relationship: **MedIllness** (medical illness). It is likely that this relationship is a result of response bias due to an underreporting of certain diseases in disadvantaged communities.

Because of the strength of the trend (and prior supporting research) the next few sections of the paper will dive deeper into the relationship between Proportion of People of Color and Vulnerability.

Trends: Racial Correlation

A simple geographic visualization provides some more context to this relationship.

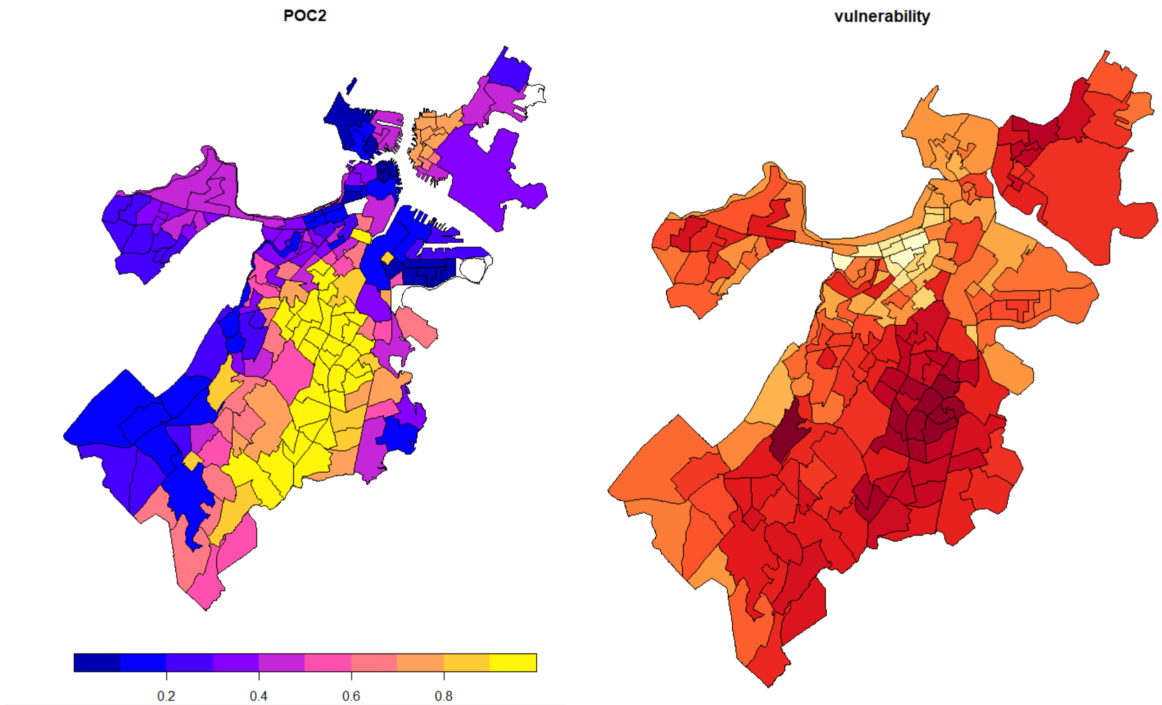

Figure 5. Vulnerability & People of Color Comparison

On the demographic map one can see the results of spatial segregation, most majority POC tracts are located in Dorchester, Roxbury and Hyde Park. One can also see a resemblance to vulnerability, with the most vulnerable areas lining up with yellow tracts (>90% POC). To quantify this relationship, a linear regression is created. The resulting visualization and model output is shown below.

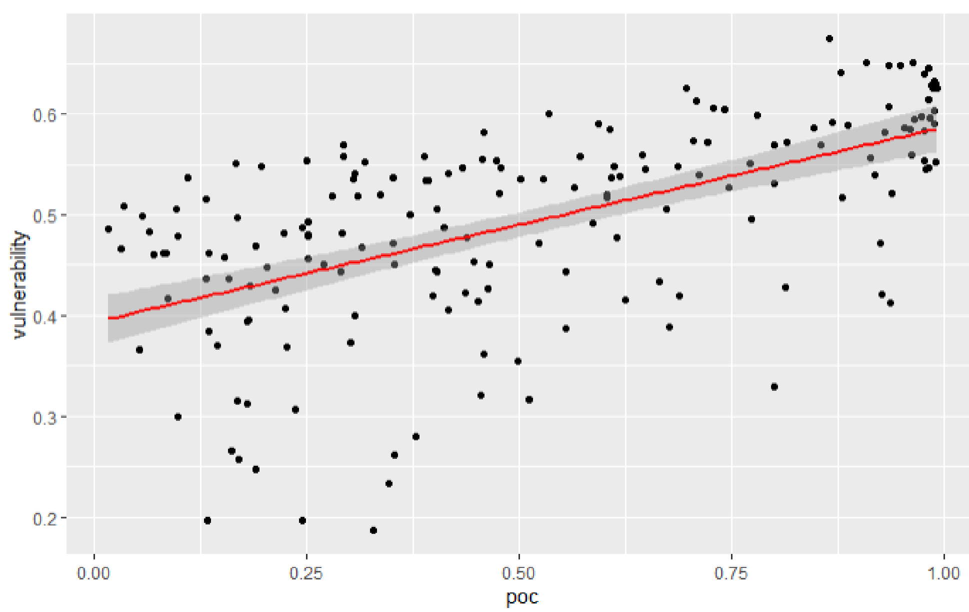

Figure 6. POC and Vulnerability Regression

```
Call:
lm(formula = df)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.26974 -0.03481  0.01656  0.06135  0.12556 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.39338    0.01235  31.852   <2e-16 ***
poc          0.19372    0.02068   9.367   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08273 on 174 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.3352,	Adjusted R-squared:  0.3314 
F-statistic: 87.74 on 1 and 174 DF,  p-value: < 2.2e-16
```

Figure 7. POC and Vulnerability Output Parameters

As seen in the exploratory analysis, there is a moderate positive correlation (R=+0.58) between the two groups (DF = 174). While race is not a sole indicator, it has a definite relationship with vulnerability.

Trends: Majority POC vs. Majority White

To create a more tangible comparison, the data is split into two groups: majority POC (N=82) and majority white (N=94). A T-Test for Difference in Means yields a significant result (**t = 8.367, p=1.9e-14**) with a 95% Confidence Interval of [0.082, 0.132]. Note that this interval isn't very insightful given the abstract nature of the Vulnerability Metric. A more conceptual understanding of these distributions can be gained from the violin chart shown on the next page.

Figure 8. Distribution of Majority POC & Majority White Groups

This visualization is useful in depicting the disparities towards the extreme end for each group. The exact numbers are as follows: the **13 *least* vulnerable** census tracts are **majority white** while the **34 *most* vulnerable** census tracts are **majority POC**. Recall that the Vulnerability metric does not include any sort of social or race-related factors, this discrepancy is purely based on the built environment.

Discussion: Implications

The fact that **41%** of all POC majority census tracts are ranked as more vulnerable than the **most** vulnerable white majority tract has significant implications. While this number may seem impossibly high, the relationship can be explained by the geographical clustering of both POC populations and vulnerability. These clusters are remnants of a long history of spatial segregation. To combat the inequity, the City of Boston needs to focus on these marginalized areas with targeted infrastructure policy.

For a guiding example, Boston should look towards existing programs. Climate Ready Boston is clearly going in the right direction in the context of climate vulnerability. As for the Built Environment, Go Boston 30 is a Green Infrastructure Initiative working to improve safety and reduce emissions. For affordable infrastructure, the Boston Housing Authority is the largest housing provider in the city working at the intersection of the public/private sectors to provide safe and accessible shelter. These types of programs have historically secured funding to proactively ensure citizen’s shelter and develop future-ready infrastructure. Policymakers should consider using these existing groups to identify and mitigate widespread infrastructure vulnerabilities in the most at risk census tracts.

Conclusion

Climate Vulnerability Metrics (and similar measures) can play a significant role when it comes to creating equitable policy. With the proliferation of open data and the new urban science, aggregated indicators are playing an increasingly important role in identifying disparity in cities. Previous research has shown these types of Metrics to correlate with social factors, making them capable of identifying injustice on a city-wide scale.

In this paper, a Climate Vulnerability Metric was created using only Building Permit & Property Assessment data. By cross-examining these results with a social vulnerability index, the underlying trends connecting minority populations with increased climate vulnerability were confirmed. Because the metric did not account for any traditional socioeconomic factors, the resulting trend indicates that Climate Vulnerability is inequitably distributed in Boston's physical infrastructure.

As a result, it is important for policymakers to be proactive in mitigating the impacts of climate change. Large-scale infrastructure improvements require a significant amount of time, money and commitment. Boston should take advantage of existing climate and infrastructure initiatives to direct attention and funding in order to improve the built environment of marginalized regions.

Methodology Appendix

Appendix 1: *Original Factor Ideas*

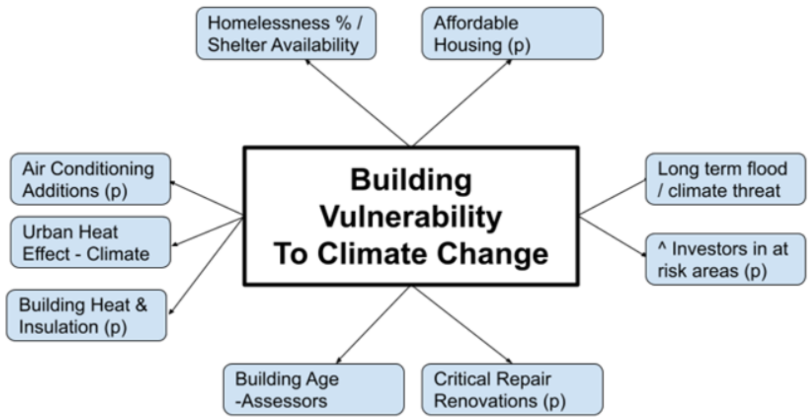

Figure 9. Brainstormed Chart of Possible Climate Vulnerability Indicators

Appendix 2: *Variable Construction: repair*

The fourth component variable repair will loop through a list of common 'repair' keywords and synonyms and assign each row True or False. The targets for the String search are: "repair", "fix" (excluding fixture), "mend", "restore", "overhaul", "service", and "rebuild". While these aren't comprehensive, it should be enough to mark a significant portion of permit data for each neighborhood grouping. Once the Boolean list is created, I will aggregate it by census tract in the same method used for the other variables.

Appendix 3: *Process and Assignment for Weighting Components*

Process

Reweighting the existing (and new) variables precisely would require significant background knowledge and focused research that is out of the scope of the project. So, the weights will be based roughly off of their implications in the context of the national death count from natural disasters (1970-2004). I acknowledge that these do not perfectly represent the fatality proportions in Boston's climate, but any regional data was far too sparse to use. I will keep it in mind during the weight assignment. The Disaster Data Table is shown below (Borden, 2008).

Deaths From Natural Disasters by Type, 1970-2004

	NUMBER	PERCENT
All Deaths From Natural Disasters	19,959	100.0
Heat or Drought	3,906	19.6
Severe Weather (severe storm/thunderstorm, fog, hail, wind)	3,762	18.8
Winter Weather	3,612	18.1
Flooding	2,788	14.0
Tornado	2,314	11.6
Lightning	2,261	11.3
Coastal (storm surge, rip current)	456	2.3
Hurricane or Tropical Storm	304	1.5
Geophysical (earthquake, tsunami, volcano)	302	1.5
Mass Movement (avalanche, landslide)	170	0.9
Wildfire	84	0.4

Source: Kevin A. Borden and Susan L. Cutter, "Spatial Patterns of Natural Hazards Mortality in the United States," *International Journal of Health Geographics* 7, no. 64 (2008): figure 1 and table 3, accessed at www.ij-healthgeographics.com/content/7/1/64, on Dec. 13, 2010.

Figure 10. Natural Disaster Deaths US (1970-2004)

Assignment

To assign weights, use the following process.

1. For each variable, estimate the relevant proportion for its implications on each disaster. For example, not having insulation and heat may be assigned 90% to "Winter Weather", and the remaining 10% to the more temperate "Severe Weather" category.
2. For each variable, estimate the certainty of the impact from 0 to 1. For example, I am sure that not having AC would impact the building's vulnerability (x=1) while the 'repair' variable may only have an 80% chance of being structural and making the building more resilient (x=0.8). Then, reweight these values so the impact proportions for all variables sum to 1.
3. For the final component, take the proportion of deaths from the disaster fatality chart. To roughly account for New England's weather, add all of the percentage points from the "Tornado" section to "Winter Weather".
4. Linearly combine the implication matrix with the impact and disaster proportion vectors by summing the rows and completing the following element-wise multiplication.

$$\boldsymbol{W}_{\boldsymbol{total}} = \sum_{j=1}^{4} \left(\boldsymbol{W}^{\boldsymbol{ij}}{}_{\boldsymbol{implication}}\right) \;.*\; \boldsymbol{W}_{\boldsymbol{impact}} \;.*\; \boldsymbol{D} \qquad \text{(Eq. 1)}$$

For step 1 (implications), the following proportions were assigned: **no_ac**: {“Heat or Drought” : 1.0} **age**: {“Severe Weather” : 0.5, “Winter Weather” : 0.3, “Flooding” : 0.1, Heat or Drought" : 0.1} **val**: {“Winter Weather” : 0.5, “Severe Weather” : 0.3, “Heat or Drought” : 0.2} **repair**: {“Severe Weather” : 0.5, “Winter Weather” : 0.3, “Heat or Drought” : 0.2}

$$\boldsymbol{W_{implication}} = \begin{matrix} \mathbf{1} & \mathbf{0.1} & \mathbf{0.2} & \mathbf{0.2} \\ \mathbf{0} & \mathbf{0.5} & \mathbf{0.3} & \mathbf{0.5} \\ \mathbf{0} & \mathbf{0.3} & \mathbf{0.5} & \mathbf{0.3} \\ \mathbf{0} & \mathbf{0.1} & \mathbf{0} & \mathbf{0} \end{matrix}$$

For step 2 (impact certainty), the following proportions were assigned: {**no_ac** : 1, **age**: 0.5, **val**: 0.2, **repair** : 0.8} When reweighted, this gives us the following proportions: {**no_ac**: 0.4, **age**: 0.2, **val** : 0.08, **repair** : 0.32}

$$\boldsymbol{W_{impact}} = \begin{matrix} \mathbf{1} \\ \mathbf{0.5} \\ \mathbf{0.2} \\ \mathbf{0.8} \end{matrix} \rightarrow \begin{matrix} \mathbf{0.4} \\ \mathbf{0.2} \\ \mathbf{0.08} \\ \mathbf{0.32} \end{matrix}$$

For step 3, these are the needed proportions (including Tornado -> Winter Weather). Any category not included in steps 1 or 2 is cancelled out in the matrix equation. Those percentage points were reassigned equally among all categories. {Winter Weather : 0.362 Heat or Drought : 0.239 Severe Weather : 0.229 Flooding : 0.170 }

$$\boldsymbol{D} = \begin{matrix} \mathbf{0.239} \\ \mathbf{0.229} \\ \mathbf{0.362} \\ \mathbf{0.170} \end{matrix}$$

Putting it all together, we get the following vector: { **no_ac** : .239, **age** : .135, **val** : .060, **repair** : .217 } Reweighting, this gives us our final weight vector: {**no_ac** : .367, **age** : .207, **val** : .092, **repairs** : .334 }

$$\boldsymbol{W_{total}} = \begin{matrix} \mathbf{0.239} \\ \mathbf{0.135} \\ \mathbf{0.060} \\ \mathbf{0.217} \end{matrix} \rightarrow \begin{matrix} \mathbf{0.367} \\ \mathbf{0.207} \\ \mathbf{0.092} \\ \mathbf{0.334} \end{matrix}$$

A quick sanity check confirms that this result aligns with an intuitive expectation of these different factors.

References

"Boston Neighborhood Demographics 1950-2010". Analyze Boston. 2019. Accessed December 16, 2021. https://data.boston.gov/dataset/neighborhood-demographics.

Arcadis, HR&A, University of Massachusetts Boston. 2016. *Climate Ready Boston (Executive Summary).*

Borden, Kevin A. and Cutter, Susan L. "Spatial Patterns of Natural Hazards Mortality in the United States" *International Journal of Health Geographics* 7, no.64. 2008.

"Climate Ready Boston Social Vulnerability Index" Boston Maps. 2017. Accessed October 17, 2021. https://data.boston.gov/dataset/climate-ready-boston-social-vulnerability.

EPA. *Climate Change and Social Vulnerability in the United States: A Focus on Six Impacts.* 2021. U.S. Environmental Protection Agency, EPA 430-R-21-003.

"Climate Vulnerability Analysis." Seleeke Flingai and Caitlin Spence. 2019. Accessed December 16, 2021. https://climate-vulnerability.mapc.org/.

Kelling, George L. 1999. Broken windows and police discretion. Washington, DC: U.S. Dept. of Justice, Office of Justice Programs, National Institute of Justice.

"Looking at the Age of Buildings in Boston with Statistical Tests". Jtpollak. 2015. Accessed October 17, 2021. https://sppua5262.wordpress.com/2015/12/01/looking-at-the-age-of-buildings-in-boston-with-statistical-tests/.