Final Project Memo

Evan Bowman

2022-10-18

Introduction

A major discussion in today's politics is the consequences of climate change and what is to come if no meaningful attempt at mitigating atmospheric greenhouse gasses occurs. These consequences, on a meta level, range from higher mean temperatures, sea level rise, and changes in the ability to grow sustainable food. These potential consequences poses large public health risks such as increases in heat related deaths and illness. This is cause for concern as about 25 percent of natural hazard mortality in the country is caused by heat exposure (Benz and Burney 2021). Furthermore, heat-related mortality in the USA causes roughly 1,500 more deaths annually than other severe weather events, as well as other outcomes such as heat strokes, dehydration, loss of labor productivity, and decreased learning (Hsu et al. 2021). The potential increase in heat related health incidents requires a substantial need for policy to combat increases in mean temperatures.

Little discussion looks at the discrepancies these effects might have on urban populations. On the topic of increased global temperature averages, there is a need to look at the intersectionality of an already occurring phenomenon: urban heat island effects. This concept revolves around the key fact that urbanized city centers already create higher mean temperatures compared to their rural counterparts. Not only are these urban centers more likely to have higher mean temperatures than rural areas, there are clear sociodemographic discrepancies between populations within these cities. Few studies have been published analyzing these demographic discrepancies. However, in one report, research has found that roughly 77 percent of counties studied show that there are large differences in change of daytime temperature between census tracts with the highest and lowest Black population proportions. Furthermore, 81 percent of counties studied with lower difference in daytime temperature are majority non-Hispanic White. (Benz and Burney 2021). These findings are important and should be analyzed for the District of Columbia to see if similar trends exist.

There is also empirical evidence that shows discrepancies in urban heat island effects targeting communities that are more impoverished. For example, a 2020 study found that there is a negative correlation of .88 between cities' urban heat island effects and income (Chakraborty et al. 2020). This is further cause for concern when looking at potential correlations between poverty and race in city census tracts. This shows a clear need for social policy that protects majority minority census tracts from being disproportionately affected by increased heat island effects.

Research Questions

What are environmental metrics that are correlated with Urban Heat Island Effects? What environmental characteristics of certain census tracts create larger heat exposure sensitivity indexes? Understanding the environmental factors would help in identifying policy approaches that would mitigate the probability of heat-induced health events.

What are reoccurring sociodemographic trends that are associated with higher heat exposure sensitivity indexes? Who are the communities are at a larger risk of heat-induced health events? Identifying sociodemographic trends can assist policy makers in identifying the communities who would best benefit from policy implementation.

Heat Sensitivity Data from Open Data DC

The data for this study will be sourced from the D.C. government's open data website. This data is already clean and requires no specific packages for analysis.

The dataset contains some identification variables that are associated with the census tract that the data corresponds to, as well some administrative variable columns . Below is a modified data frame eliminating some of these variables for analysis and easier viewing.

```r
heat <- heat %>%
  select(TRACT, TOTALPOP:OBESITY, P_TREECOVER:AIRTEMP_MEAN, HSEI)
tibble(heat)
```

```
## # A tibble: 206 x 16
##    TRACT  TOTALPOP P_POC P_CHILD P_ELDERLY P_POVE~1 P_DIS~2 P_LIM~3 ASTHMA   CHD
##    <chr>     <dbl> <dbl>   <dbl>     <dbl>    <dbl>   <dbl> <chr>    <dbl> <dbl>
##  1 001600     5014  76.5     4.9      22.8        4     8.9 0.9        9.6   5.7
##  2 001500     6016  30.5     6.8      27.6        6     6.6 3.8        8.3   5.2
##  3 010300     4272  82       2.3       9.5       19     9.2 10.5      10.4   4.8
##  4 001702     3667  61       3.5      16.5       10     8.8 4.4        9.9   3.9
##  5 001803     4161  88.7     5.9       9.6       40    17.2 20.6      12     5.4
##  6 001804     5803  93.6     7.3      11.8       48    18   23.9      12.2   6.6
##  7 001401     3641  40.3    13.4      18.7        6     5.5 2.5        8.5   3.8
##  8 009505     3907  89.2     1.1      13.6       12    13.5 0         10.8   5.9
##  9 001902     2460  82       7        15.4       23    12.8 8.3       10.6   5.7
## 10 001901     4220  84.1     3.1      14.4       23     9.6 13.4      10.4   5.1
## # ... with 196 more rows, 6 more variables: OBESITY <dbl>, P_TREECOVER <dbl>,
## #   P_NOTREE <dbl>, P_IMPSURF <dbl>, AIRTEMP_MEAN <dbl>, HSEI <dbl>, and
## #   abbreviated variable names 1: P_POVERTY, 2: P_DISABILITY, 3: P_LIMENG
```

As seen from the tibble above, there are **16 variables of interest** for the **206 DC census tracts**. The data set consists of sociodemographic variables such as proportions of minority, child, elderly, poverty, disability, and limited English for each census tract that will be key for our second research question. The data also contains a few health variables such as prevalence of asthma, congenital heart disease, and obesity that could help with the analysis of certain demographics and heat index values.

There are also environmental variables found within the data frame including percent tree cover, percent of tract with no trees, percent of tract with impervious surfaces, and mean ambient air temperature (degrees Celsius). There are three index variables, heat sensitivity index (HSI), heat exposure index (HEI), and heat sensitivity exposure index (HSEI). HEI is calculated by 50% Air Temp, 25% lack of tree canopy, 25% impervious surface and HSEI is calculated by 50% HSI, 50% HEI. We will not include HSI or HEI due to the covariation with HSEI. For our first research question, we will use the HSEI as the dependent variable.

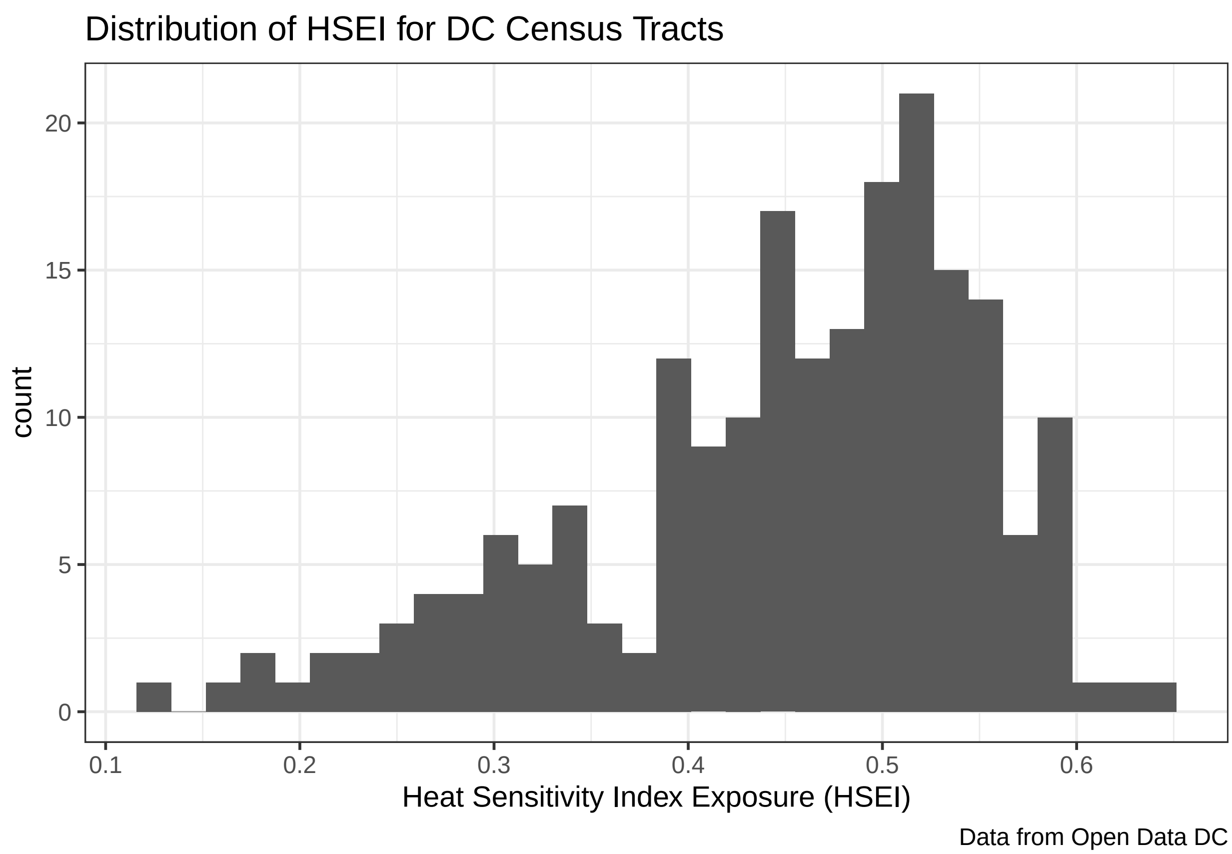
Initial Exploratory Data Analysis

Initial EDA will mainly revolve around our second research question regarding demographic trends. Analysis regarding our first question, environmental factors, will come at a later date.

Below is the distribution of HSEI values for 203 census tracts. Three tracts were filtered out because of no reported HSEI values. The distribution is skewed heavily to the left, with a mean appearing to be hovering around the .45 mark. Further EDA below will explore the correlations between different demographic factors and HSEI values.

```
heat %>%
  filter(HSEI > 0 & HSEI < 1) %>%
  ggplot(aes(HSEI)) +
  geom_histogram() +
  ggtitle("Distribution of HSEI for DC Census Tracts") +
  xlab("Heat Sensitivity Index Exposure (HSEI)") +
  labs(caption = "Data from Open Data DC") +
  theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



As shown by the graph below, there seems to be a clear positive correlation the percentage minority population in a given census tract and its HSEI value. This is further supported by the boxplot of HSEI values comparing majority minority tracts versus majority White tracts.

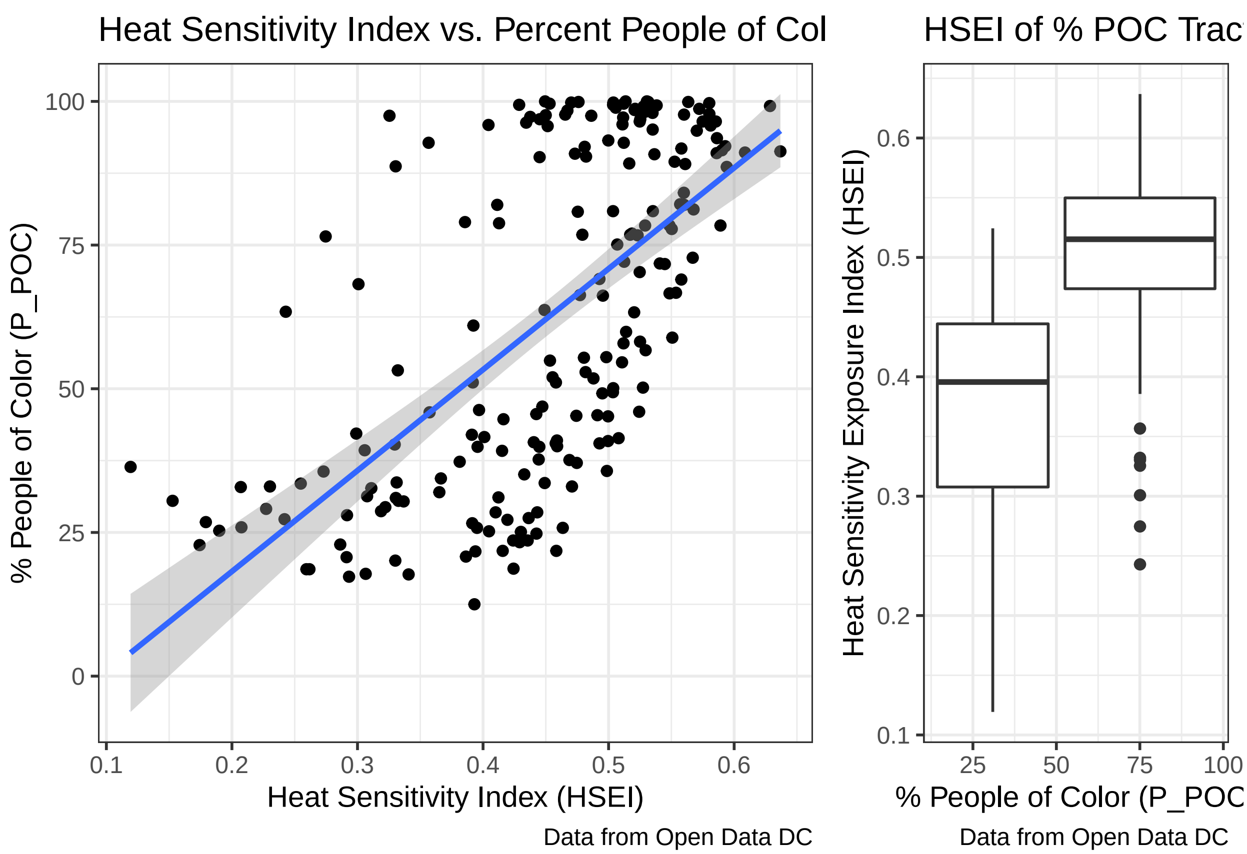
```
scatter <- heat %>%
  filter(HSEI > 0 & HSEI <1) %>%
  ggplot(aes(HSEI, P_POC)) +
  geom_point() +
  geom_smooth(method = "lm") +
```

```
  ggtitle("Heat Sensitivity Index vs. Percent People of Color") +
  xlab("Heat Sensitivity Index (HSEI)") +
  ylab("% People of Color (P_POC)") +
  labs(caption = "Data from Open Data DC") +
  theme_bw()

box <- heat %>%
  filter(HSEI > 0) %>%
  ggplot(aes(P_POC, HSEI)) +
  geom_boxplot(aes(group = cut_interval(P_POC, length = 50))) +
  ggtitle("HSEI of % POC Tracts") +
  xlab("% People of Color (P_POC)") +
  ylab("Heat Sensitivity Exposure Index (HSEI)") +
  labs(caption = "Data from Open Data DC") +
  theme_bw()

gridExtra::grid.arrange(scatter, box, ncol = 2, widths = 2:1)
```

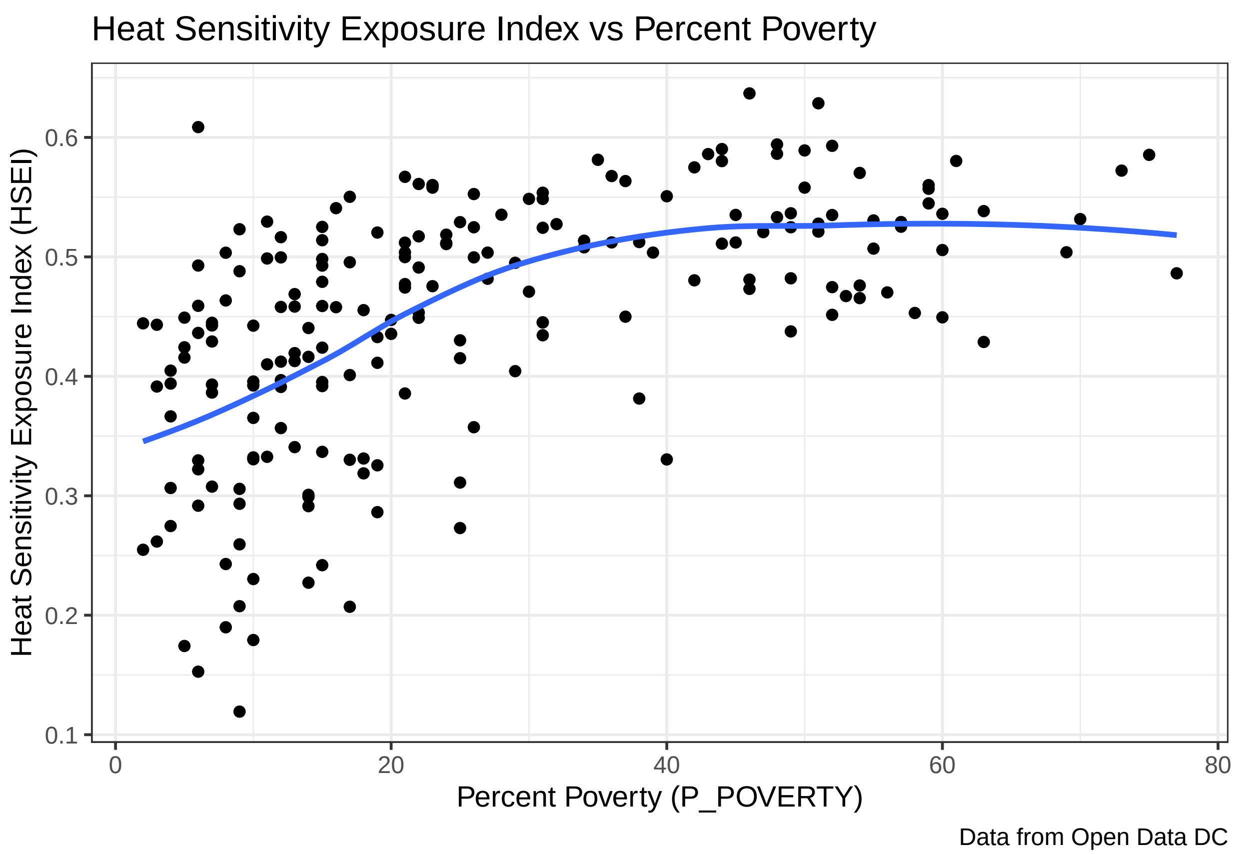
```
## ‘geom_smooth()‘ using formula ’y ~ x’
```



The next analysis to be conducted is looking at poverty and its correlation with HSEI values. The scatter plot shows a positive correlation between the two variables. There appears to be a linear increase that flattens off at about 35 percent poverty. Further analysis will need to be done to look at the exact correlation between the two variables.

```r
heat %>%
  filter(HSEI > 0) %>%
  ggplot(aes(P_POVERTY, HSEI)) +
  geom_point() +
  geom_smooth(se = FALSE) +
  ggtitle("Heat Sensitivity Exposure Index vs Percent Poverty") +
  xlab("Percent Poverty (P_POVERTY)") +
  ylab("Heat Sensitivity Exposure Index (HSEI)") +
  labs(caption = "Data from Open Data DC") +
  theme_bw()
```

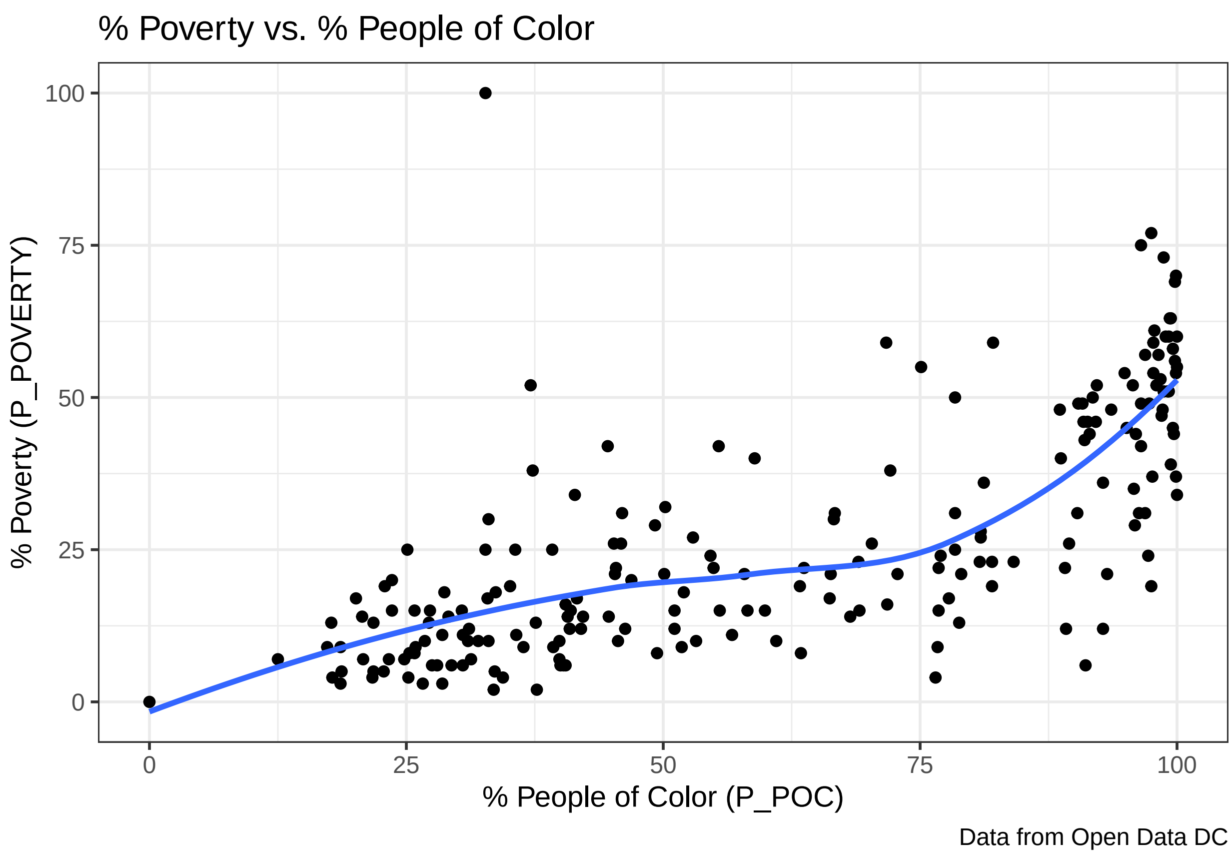
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Nonetheless, the next step is to look at the relationship between percent poverty and percent people of color to see if there is any correlation between the two variables.

```r
heat %>%
  ggplot(aes(P_POC, P_POVERTY)) +
  geom_point() +
  geom_smooth(se = FALSE) +
  ggtitle("% Poverty vs. % People of Color") +
  xlab("% People of Color (P_POC)") +
  ylab("% Poverty (P_POVERTY)") +
  labs(caption = "Data from Open Data DC") +
  theme_bw()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



The scatter plot above highlights a positive correlation between percentage minority residents and percent poverty. The regression line shows an interesting trend. It appears the magnitude of correlation is relatively minimal until tracts with about 80 percent POC population. There is also an interesting potential outlier of 100 percent poverty. Upon investigation below, this tract only includes Catholic University students, which explains the high poverty percentage.

```
filter(heat, P_POVERTY > 90)
```

```
## # A tibble: 1 x 16
##   TRACT  TOTALPOP P_POC P_CHILD P_ELDERLY P_POVERTY P_DIS~1 P_LIM~2 ASTHMA   CHD
##   <chr>     <dbl> <dbl>   <dbl>     <dbl>     <dbl>   <dbl> <chr>    <dbl> <dbl>
## 1 009511     2156  32.7       0       2.8       100     4.5 NULL      12.8   4.5
## # ... with 6 more variables: OBESITY <dbl>, P_TREECOVER <dbl>, P_NOTREE <dbl>,
## #   P_IMPSURF <dbl>, AIRTEMP_MEAN <dbl>, HSEI <dbl>, and abbreviated variable
## #   names 1: P_DISABILITY, 2: P_LIMENG
```

Next Steps for Analysis

The next step in this analysis is to begin modeling the correlates of mean air temperature. This was not included in the initial analysis due to the need for more thoughtful mapping. There are very limited environmental variables found in the data set with two having high covariation (percent tree cover and percent no tree). This will lead to using different methods to achieve an accurate model. More likely than not, shrinkage methods (i.e. ridge regression or LASSO regression) will have to be used to mitigate the covariation.

The end goal of this project is to be able to make grounded policy recommendations on how to mitigate the public health risk of heat exposure. To do this, there are two distinct aspects that must occur for this to happen. First, we must identify strong correlates of mean air temp to better understand what environmental factors lead to higher mean air temperature for census tracts. Second, we must find demographic patterns that are associated with mean ambient temperatures and heat sensitivity exposure indexes. These identifications will allow us to recommend what vulnerable communities should be targeted for policy implementation.

References

Benz, Susanne Amelie, and Jennifer Anne Burney. 2021. "Widespread Race and Class Disparities in Surface Urban Heat Extremes Across the United States." *Earth's Future* 9 (7): e2021EF002016.

Chakraborty, T, A Hsu, D Manya, and G Sheriff. 2020. "A Spatially Explicit Surface Urban Heat Island Database for the United States: Characterization, Uncertainties, and Possible Applications." *ISPRS Journal of Photogrammetry and Remote Sensing* 168: 74–88.

Hsu, Angel, Glenn Sheriff, Tirthankar Chakraborty, and Diego Manya. 2021. "Disproportionate Exposure to Urban Heat Island Intensity Across Major US Cities." *Nature Communications* 12 (1): 1–11.