Investigating the Urban Heat Island Effect and Risk of Heat-Related Mortality in Liver and Kidney Transplant Patients

by

Nathan Davis

A Senior Thesis Submitted to the Department of Earth and Environmental Science, University of Pennsylvania

In Partial Fulfillment of the Requirements for the Degree of Bachelor of Arts

University of Pennsylvania, Philadelphia

Advisor: Irina Marinov

May 2023

**Acknowledgements**

I would like to thank the following people for offering their time and guidance to support me in completing this project.

Dr. Irina Marinov – For guidance in background research, model creation, writing clearly and descriptively, and connecting with other researchers.

Dr. Therese Bittermann, MD – For assistance in identifying potential data sources and accessing UNOS data.

Dr. Jane Dmochowski – For feedback on all stages of the research and writing process, and for her patience and adaptability.

Professor Maria Antonia Andrews – For guiding the formulation of a research question and preparation of a research proposal.

**Abstract**

Climate change is the “single biggest health threat facing humanity,” according to the World Health Organization.[[1]](#endnote-1) Increased exposure to heat, including rising ambient temperatures and more severe and frequent heatwaves, has been observed to have profound impacts on human health. Past research has sought to quantify the mortality impacts of heat exposure and how heat exposure will increase as a result of climate change. However, few studies have specifically investigated how these effects vary from rural to urban areas. It has been hypothesized that urban areas experience greater mortality from heat as a result of the urban heat island effect.

We create two Poisson generalized additive models (GAMs) to examine the association of all cause-mortality in liver and kidney transplant recipients with temperature, population density, and greenspace in the United States from 2005 to January 2023. We find that mortality is higher on days with higher average heat index, higher in zip codes with higher population density, and higher in zip codes with less greenspace. Additionally, we find that zip codes with low amounts of green space (< 25%) are more sensitive to increasing average heat index overall. The relationship between mortality and greenspace was stronger than the relationship with population density, suggesting that the known risks from the urban heat island effect can potentially be mitigated by expanding greenspaces in cities. Previous work has shown that areas which most severely experience the urban heat island effect tend to have lower income and less access to healthcare; climate change is thus likely to magnify existing social inequities in urban places with low greenspace areas. Further work needs to be done to link health implications of climate-driven heat increases with social inequities.

1. **Introduction and Background**

Climate change, which includes to the long-term changes in temperature and humidity patterns, is the ‘single biggest health threat facing humanity’, according to the World Health Organization.[[2]](#endnote-2) Increased exposure to heat, including rising ambient temperatures and more severe and frequent heatwaves, has been observed to have profound impacts on human health.[[3]](#endnote-3) Additionally, the World Health Organization has found that increased exposure to extreme heat will be one of the most significant impacts of climate change. It is estimated that under mid-range emissions scenarios increased exposure to heat as a result of climate change could result in 94,000 annual excess deaths globally.[[4]](#endnote-4)

It is expected that heat risks are greater in urban areas due to the heat island effect.[[5]](#endnote-5) However, it has not yet been definitively shown whether the interplay of heatwaves and the urban heat island effect results in significantly higher mortality in urban areas.[[6]](#endnote-6) This thesis will investigate this question by comparing how the risk of heat varies between zip codes across the US according to population density and greenspace. This research will improve our understanding of who is most likely to be impacted by climate change and where adaptation is most critical.

***I.1 Quantifying exposure to heat in health-related studies***

It has been well documented that extreme heat is harmful to human health. A global study examined the relationship between heat and mortality in 384 locations and consistently found an increase in mortality with higher mean daily temperatures.[[7]](#endnote-7) This has been corroborated by studies other studies conducted at the global, national, or regional level and studies examining individual or collections of cities.

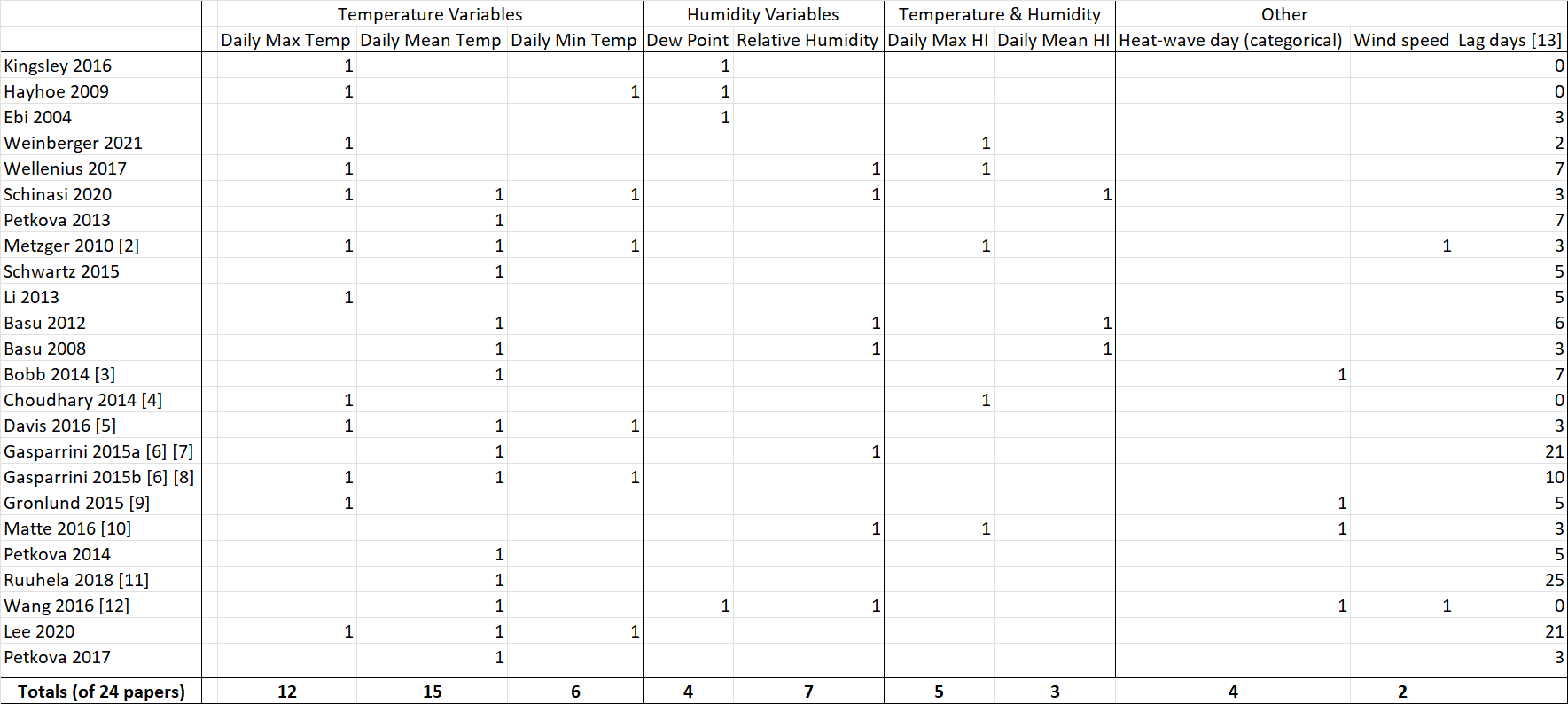
A partial review of existing research on the association between heat and mortality or morbidity was conducted to determine how researchers typically measure and classify exposure to heat. The results of this review are summarized in Table 1 and discussed here. The most common methodology in the studies reviewed for modelling the association between heat and mortality or hospital admissions was Poisson regression using splines to adjust for seasonal and long-term trends. Specifically, studies commonly used one or two splines to control for seasonality and a mixture of other unmeasured cofounders including infectious disease. It was common also to control for day of the week, and federal holidays (to account for the increased exposure on workdays of people who work outside). This type of regression is relatively common in environmental epidemiology and has been used in many studies evaluating the effects of air pollution. [[8]](#endnote-8)

In order to quantify exposure to heat, the vast majority of studies either considered daily maximum temperature[[9]](#endnote-9) or daily mean temperature.[[10]](#endnote-10) A few studies have considered daily minimum temperature in addition to one or both of these variables based on the theory that relief from extreme heat overnight can significantly reduce negative health effects.[[11]](#endnote-11) Humidity was also considered by many of these studies, since it affects the body’s ability to regulate its temperature through perspiration. This can be done by including a variable that represents humidity directly such as dew point,[[12]](#endnote-12) vapor-pressure deficit,[[13]](#endnote-13) or relative humidity.[[14]](#endnote-14) These variables essentially provide the same information as they can all be converted to each other when temperature is known.[[15]](#endnote-15) Alternatively, some papers examine heat index, which is a function of both temperature and humidity.[[16]](#endnote-16)

Occasionally papers also defined “heat-wave days” as a categorical variable. Bobb et al. (2014), and Wang et al. (2016) both defined heatwaves as at least 2 consecutive days with mean temperatures greater than the 97th percentile of all daily mean temperatures for a given location. Additionally, they both examined models with stricter definitions (4 consecutive days instead of 2, and 98th and 99th percentile instead of 97th). [[17]](#endnote-17) Gronlund et al. (2015) considered all days with a maximum (instead of mean) temperature greater than the 97th percentile of all daily maximum temperatures for a given location to be heatwave days (no requirement for consecutive days of heat).[[18]](#endnote-18) Finally, Matte et al. (2016) considered all days with heat index greater than 100F to be a heatwave day in addition to at least 2 consecutive days with heat index greater than 95F.[[19]](#endnote-19)

Studies also occasionally consider spatial synoptic classification, which essentially defines the type of air mass (e.g. dry tropical, moist polar, etc.). This is because of the theory that some types of air masses may be more harmful than others, referred to in Hayhoe et al. (2010) as “oppressive air masses.” Hayhoe argues specifically that the impact of oppressive air masses grows on each consecutive day when that type of air mass is present.[[20]](#endnote-20)

In studies of environmental epidemiology, it is common to consider the lagged effect of variables. In the context of exposure to heat, this means considering weather conditions on the day of an observed health outcome and on several days before. A substantial majority of studies reviewed consider the lagged effects of heat for 2-25 days.[[21]](#endnote-21) Studies commonly find that 3-5 days of lag are significantly associated with higher risk of mortality (i.e. heat on a particular day also increases the risk of mortality several days later).[[22]](#endnote-22) There is mixed evidence that lags of 7+ days can reveal a decreased risk of mortality. Studies which find this, attribute it to the “harvesting effect” where individuals who would have died a week or two after a heat event instead die sooner as a result of the heat event. Therefore, slightly lower mortality can be observed 1-2 weeks after a heat event.[[23]](#endnote-23)



**Table 1:** Lists papers examining the association between heat and health outcomes that were reviewed and the variables each paper used to quantify exposure to heat. Daily Mean Temperature is the most common. For clarity, all variables which were used in only 1 paper were excluded. For a full, longer version of the table and associated notes, see Appendix A.

There is disagreement about whether or not it is necessary to control for air pollution separately in studies assessing the relationship between health outcomes and heat. Reid et al. (2012) shows that air pollutants are more likely to be mediators between temperature and health rather than confounders.[[24]](#endnote-24) Therefore, since heat can be associated with ozone and particulate matter levels, not controlling for air pollution explicitly results in a calculation of the total impact of heat, including the indirect impact through air pollution. However, Li et al. (2013), which assessed the relationship between heat and all-cause mortality attempted to control for air pollution found that it did not make a substantial difference in the results.[[25]](#endnote-25)

***I.2. Past associations between heat with mortality and morbidity***

Past research has taken various approaches to limiting their data to specific causes of mortality and examining the cause specific effects of heat. Some studies examine the association between heat and all-cause mortality or all-cause emergency department visits.[[26]](#endnote-26) Most commonly, however, studies linking temperature and hospital admissions consider only all “non-external” or “non-accidental” causes, both of which refer to International Classification of Diseases (ICD) diagnosis codes ICD-9 0-799 and ICD-10 A00-R99.[[27]](#endnote-27) One notable weakness of this classification is that it excludes two highly relevant diagnosis codes because they are considered external causes: ICD-9 E900 which is an accident caused by excessive heat due to weather conditions and ICD-9 992 which refers to the effects of heat and light when they require medical treatment. Studies which examine the association of heat with all-cause hospital admissions or all admissions with non-accidental causes does not reveal the

Studies examining cause-specific associations with heat are less common, since there is rarely a sufficient sample size within individual categories of disease. However, there are some exceptions. Kingsley et al. (2016) analyzes the relationship between heat and emergency department (ED) admissions including non-external as defined above plus ICD-9 codes E900 and 992 specifically. Additionally, Kingsley et al. (2016) flags dehydration (276.51) as a “heat-related cause,” and they examine the relationship of several other cause-specific categories of diagnoses to heat including heat stroke, dehydration, cardiovascular disease, renal disease, acute renal failure, respiratory disease, and asthma. They analyze both heat-related ED admissions, ED admissions for these other specific causes, all-cause ED admissions, and all cause death, stating that the sample size was too small to consider the association between heat and specific causes of death. They observed significant increases in all-cause ED admissions and all-cause mortality with increases in heat, and even more substantial increases in heat-related ED visits and mortality. Additionally, they found a positive association of heat with cardiovascular and respiratory diseases, but it was not statistically significant. [[28]](#endnote-28)

Weinberger et al. (2021) examines the association of extreme heat with seven specific causes of admission for Medicare patients: septicemia, diabetes mellitus with complications, fluid and electrolyte disorders, peripheral vascular disease, renal failure, urinary tract infections, and heat stroke and other external causes. Weinberger et al. (2021) chose these conditions specifically because Bobb et al. (2014), which also analyzed the association between heat and causes of admission had shown these to be particularly affected by extreme heat (but had not attempted to quantify the resulting excess hospital admissions). They found statistically significant increased risk for five out of the seven specific causes they examined: all except diabetes and peripheral vascular disease.[[29]](#endnote-29)

The decision to examine the relationship of heat to all-cause mortality or ED admissions rather than only those with diagnosis codes related to heat is often justified because these researchers are confident that their statistical methods sufficiently control for other factors to extract the relationship with heat specifically. Additionally, Ebi et al. (2004) hypothesizes that medical examiners may not always be able to accurately determine whether extreme temperatures are an underlying cause (e.g., they may be unable to determine if a case of cardiovascular disease was caused/triggered by heat).[[30]](#endnote-30)

***I.3. The Urban Heat Island Effect***

The urban heat island (UHI) effect refers to higher temperatures that exist within cities as a result of the built environment and human activity. This effect can amount to up to an 8.0° C (14.4° F) temperature difference in New York City.[[31]](#endnote-31) The main causes of the UHI are increases in heat-absorbing surfaces and structures, increasing population density and human activity, and the removal of natural vegetation. Increasing urban green space with parks, street trees, and green roofs has consistently been shown to mitigate the UHI effect.[[32]](#endnote-32)

Heat risks to human health are expected to be greater in urban areas due to UHI effects.[[33]](#endnote-33) However, there has been little research on whether interactions between heatwaves and the UHI effect lead to significantly higher mortality rates within cities. One paper examining Helsinki and its surrounding area found that there was significantly higher mortality within the city.[[34]](#endnote-34) However, there is still a need to explore this question in a wider array of climates and geographic areas and for specific disease and socio-economic profiles.

***I.4. Background on liver and kidney transplant patients***

Transplant recipients of solid organs are at greater risk for infection because of immunosuppressive medications to prevent the rejection of a transplanted organ.[[35]](#endnote-35) A study at the Third Xiangya Hospital in China found that incidences of tuberculosis were 7 times higher in kidney and liver transplant recipients than the general population.[[36]](#endnote-36) Research also suggests that immunosuppressive medications also increase the risk of cancer in solid organ transplant recipients. Liver transplant recipients have been found to be at higher risk for lung, liver, and kidney cancers; kidney transplant recipients have been found to be at higher risk for lung cancer and kidney cancer.[[37]](#endnote-37)

One measure of the severity of a patient’s liver disease pre-transplant is their Model of End Stage Liver Disease Sodium (MELD-Na) score. The MELD-Na score is based on a patient’s serum bilirubin, serum creatinine, institutional normalized ratio, and serum sodium; before 2016 a MELD score using the same variables except serum sodium was used. There is limited evidence suggesting a higher MELD score increases the risk of post-transplant mortality. However, it has been shown that a higher proportion of this score attributable to creatinine is associated with higher all-cause mortality post-transplant. High serum creatine and consequently a high proportion of MELD-Na attributable to creatinine indicates poor kidney health.[[38]](#endnote-38) Additionally, it has been shown that patients in the ICU prior to their transplant have significantly higher short-term post-transplant mortality.[[39]](#endnote-39)

There has been limited research on the impact of specific environmental factors on the health of transplant recipients. It is not clear how their risk from exposure to environmental hazards differs from the general population or whether these may magnify the other health risks they face, largely as a result of immunosuppressive medications.

1. **Methods**

***II.1. Data Sources***

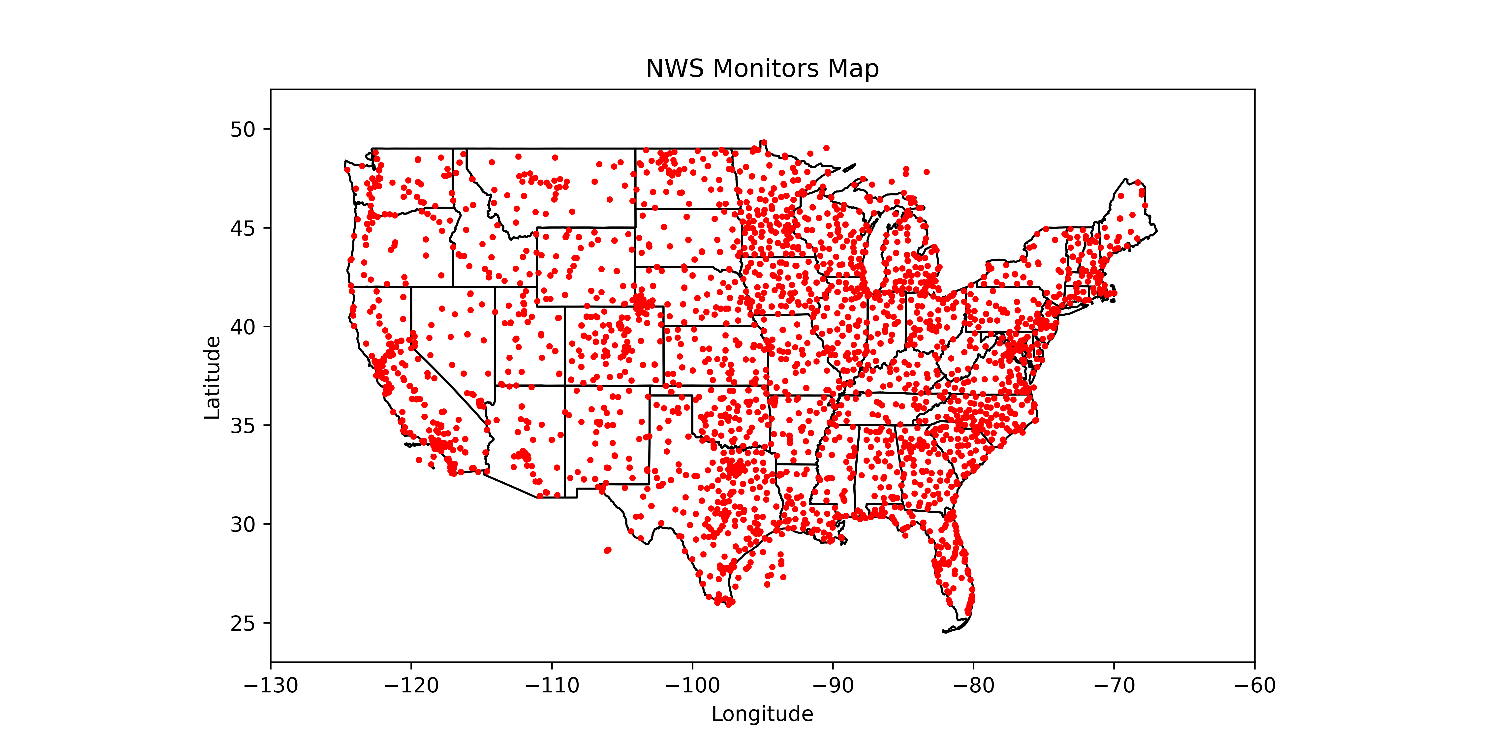
**Health Data**: Daily mortality for kidney and liver transplant recipients was obtained from the United Network for Organ Sharing (UNOS). The available dataset contained the gender, race, age, zip code, and death date for kidney and liver transplant recipients for the entire continental US between 2005-01-01 and 2023-01-31. In total, about 80,000 instances of mortality were documented in this dataset. This dataset was chosen because it was publicly available and had high spatial (zip code level) resolution and daily temporal resolution.

**Climatological Data**: Hourly measurements of surface temperature and dew point from the National Weather Service (NWS) stations across the US were downloaded from the National Ocean Atmospheric Administration (NOAA) integrated surface database.[[40]](#endnote-40) All measurements from 2005-01-01 to 2023-01-31 were downloaded, corresponding to the date range of our UNOS health data. Relative humidity was calculated from temperature and dew point. Then, heat index was calculated using temperature and relative humidity. The formulas used for these conversions are shown in appendix B.The daily mean temperature and heat index at each station on each day were calculated. Each station had latitude and longitude coordinates. Details about the processing of the data from NWS stations and associated python code are available in Appendices C and E.

**Geographical Data**: Data describing the population density and land cover in zip codes were used as indicators of the urban heat island effect. Population density by zip code as calculated initially by Bitner (2014) based on the 2010 US census was downloaded.[[41]](#endnote-41) Population density ranges from 0 in unpopulated zip codes up to 150,000 in the densest zip codes. Zip codes at this end of the distribution are often incredibly small such as zip code 20052 which is only 0.003 square miles but has 470 residents. The 2016 Land Cover dataset was downloaded from the National Land Cover Database, as compiled by the Multi-Resolution Land Characteristics Consortium.[[42]](#endnote-42) This raster specified the type of landcover (e.g. developed or green space) for the entirety of the continental US on a 20m by 20m grid. Separately, a shapefile of zip codes corresponding to the 2010 Census was downloaded from the US Census Bureau.[[43]](#endnote-43) The percent of developed space and green space in each zip code was then calculated. Associated python code, and details on the definitions used for developed and green space are in Appendix D.

***II.2. Analysis and Regression***

A table was constructed containing one row for each zip code for each day between 2005-01-01 and 2023-01-31. This table contained all zip codes where mortality was observed in the UNOS data. For each zip code, we tabulate the date, the number of deaths from the UNOS dataset in that zip code on that day, the fraction of green space and fraction of developed space in that zip code (constant in time), the population density in that zip code (constant in time). Additionally, for each zip code in the table we record the daily average temperature and heat index recorded at the nearest NWS station, as well as the aggregated population density, green space, and developed space in the zip codes nearest to the NWS station taking the measurements. Note that not all the stations are operational for the entire time period, hence the closest station to a given zip code changes in some instances over the period considered. The python code used to construct this table and further notes can be found in Appendix F.

****

**Figure 1:** Map of NWS stations in the continental U.S. from which data was downloaded for this project from the NOAA Integrated Surface Database. This consists of all NWS stations in the continental U.S. which measure both temperature and dew point and were operational at any point from 2005-01-01 to 2023-01-31.

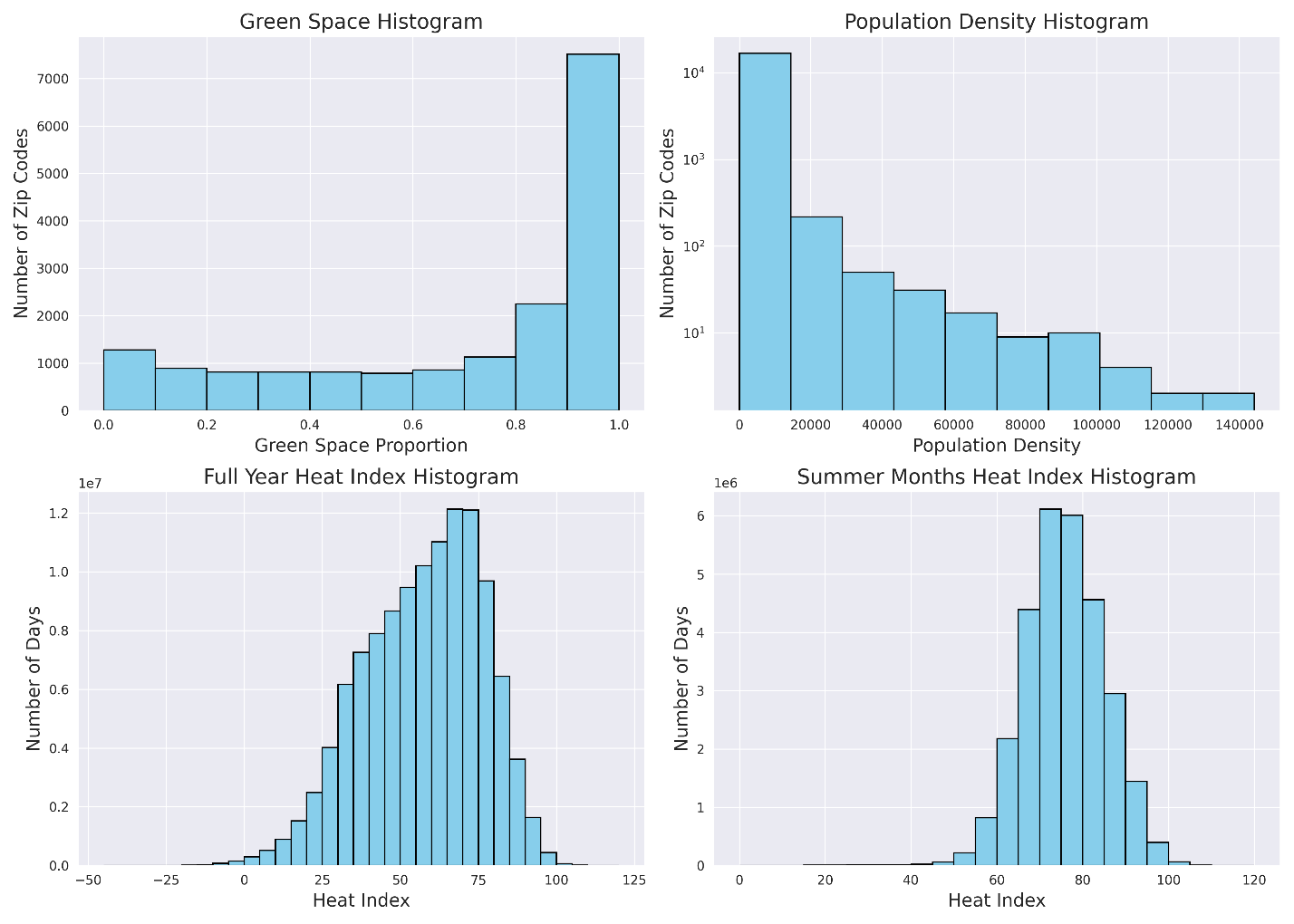
This table was used to create three Poisson Generalized Additive Models (GAM). A Poisson model is appropriate since it allows only positive values for the dependent variable, in our case, the number of observed deaths. Additionally, Poisson models are suitable when the number of observations is high, and the event of interest has low probability. This reflects the high number of days for which we have temperature observations and the relatively low number of observed deaths. A Poisson model represents the natural logarithm of the response variable as a linear combination of the predictors (see Equation 1). This is useful for modeling the small amounts of variation in the number of deaths. A GAM is a model in which the response variable depends linearly on unknown smooth functions called splines of predictor variables, which it infers from the data. Although this study focuses on extreme heat, extreme cold is also hazardous to human health, and so a nonlinear U-shaped relationship between temperature and mortality is expected.

All models used population density and green space in the zip code where deaths were recorded in addition to heat index as measured at the nearest NWS station. They also included terms to model the interaction between heat index and population density and the interaction between heat index and green space. Therefore, they model expected mortality according to Equation 1.

ln(Expected Mortality) = f₁(Heat Index) + f₂(Population Density) + f₃(Green Space) + f₄(Heat Index, Population Density) + f₅(Heat Index, Green Space) + c

**Equation 1:** The equation used to model temperature in model 1. The model learns the functions f1 through f5 from the data and also adds a constant intercept term, c. As it is a Poisson model, we predict the natural logarithm of Expected Mortality.

Model 1 encompasses data from the entire year, excluding outliers with a heat index falling outside the -10 to 120 range. Such outliers account for only 0.03% of the total data. Model 2, in contrast, focuses exclusively on the summer months from June through August, a period where we anticipate the most pronounced impact of extreme heat. In this model, outliers exhibiting a heat index outside the 40 to 120 range have been excluded.



**Figure 2:** This figure presents a series of histograms illustrating the distribution of various key parameters within our dataset. In the top-left, we visualize the distribution of green space across zip codes in 10% increments. The top-right histogram demonstrates the distribution of population densities across zip codes, displayed on a logarithmic scale. This scale was necessitated by the significant skew in the data, with only 0.6% of zip codes exhibiting a population density greater than 30,000 people per square mile. The bottom-left histogram offers insight into the frequency of different average heat index values across all recorded days. Lastly, the bottom-right histogram narrows this perspective, focusing exclusively on the distribution of average heat index values during the summer months of June, July, and August. The code used to create these plots is contained in Appendix G.

Average temperature and developed space were not included in these models because of their high correlation with average heat index and green space, respectively. Average temperature and heat index had a correlation of 0.998, and green space and developed space had a correlation of -0.97, and so very little information is lost by excluding temperature and developed space from the models. Table 2 displays a correlation matrix for the columns of interest for these models.



**Table 2:** Correlation Matrix of variables in table with temperature and heat index measurements (daily data for all zip codes combined, using only zip codes where deaths were observed). This table displays the Pearson correlation coefficient between all pairs of variables. “Difference” variables refer to the difference between the value in the death location and the value at the location of the NWS station that measured temperature and dew point. The difference variables are not used in any of the models which are discussed in the body of this paper. Instead, they were used in an attempt to adjust for the potential impact of the UHI on the NWS stations recording temperature measurements, which is discussed in appendix H.

Poisson GAM models have several parameters that can be selected/altered including the number of splines used to model the relationship between the predictor variables and the response variable and the penalty coefficient (lambda) used for smoothing the splines. In order to select values for these parameters we used a grid search over multiple possibilities with k-fold cross validation.

K-fold cross validation refers to a method where the data is randomly divided into a number ‘k’ samples (folds) of equal size. The model is then trained and evaluated k times. Each time it uses k-1 folds to create the model and then uses the last fold to evaluate the performance of the model. The average performance across the k iterations is used as the overall performance estimate of a model. For our modelling, we used k = 5.[[44]](#endnote-44)

A grid search is a method to systematically explore a set of possible model parameter values by creating and evaluating models with all possible combinations of these values and selecting the model with the best performance. We searched over models using between 4 and 7 splines for each predictor variable and using penalty coefficients between -3 and 3. Four is the minimum number of splines possible when using cubic splines. We then defined the ‘best’ model to be the one with the best overall performance in 5-fold cross validation. Selecting the model that performs best on this metric limits our ability to bias the results and ensures that we are not simply able to select whichever model best supports our hypothesis.[[45]](#endnote-45)

This process was conducted for both models. For each model, the best parameters were determined to be 4 splines for all variables with a penalty coefficient of 0.001. These models are discussed in the following section. Details about the regression process, grid search and cross validation, creation of plots, and all associated Python code can all be found in Appendix G.

1. **Results and Discussion**

***III.1. Interpreting Poisson GAM Regression Output***

Models were fit to the data using the parameters found to be optimal by the grid search. To analyze the output of these models it is most helpful to visually examine the resulting “partial dependence” plots. These are plots of the individual functions f1 through f5 which are fit by the model, mentioned in Equation 1.

Recall that in Poisson regression the natural logarithm of expected mortality is modeled as the sum of f1 through f5. The relative change of these functions in response to changes in the predictors, heat index, population density, and green space is far more important than the exact values they output. This is because any of these functions should be shifted upwards or downwards by a constant value without effecting the final output as long as the intercept term is adjusted by the same value in the opposite direction. That is to say, for instance, we can create model 1′ by replacing f1 with f1′ = f1 + b where b is a constant and replacing c with c′ = c – b while preserving the same output ‘expected mortality’ for each of our inputs:

ln(Expected Mortality) =

f₁(Heat Index) + f₂(Population Density) + f₃(Green Space) + f₄(Heat Index, Population Density) + f₅(Heat Index, Green Space) + c =

f₁′(Heat Index) + f₂(Population Density) + f₃(Green Space) + f₄(Heat Index, Population Density) + f₅(Heat Index, Green Space) + c′

Model 1′ and model 1 can be thought of as equivalent since they produce the same output for all predictors. f1 is shifted vertically by a value of ‘b’ between the two models, however the response of f1 to heat index is preserved. That is to say, for any two particular heat index values HI1 and HI2:

f1(HI2) – f1(HI1) = f₁′(HI2) - f₁′(HI1)

This shows that a particular value of f1 is less important for understanding the output of the model, since it can be adjusted without changing the model by inversely adjusting the intercept term. The shape of f1 cannot be changed without changing the model. Therefore, the shape of the function contains the most important information. Since the sum of functions f1 through f5 and the intercept term models the natural logarithm of expected mortality, an increase in any of these functions by 1 (while holding the others constant) represents an increase in expected mortality by a factor of *e* (Euler’s number, *e* = 2.718…).

***III.2. Model 1 Results: Full Year***

The regression output of model 1 is displayed in Figure 3. One feature to note includes the high significance of all variables. Unfortunately, the python library used to conduct this regression has a known bug which causes p-values to appear lower than they should. In some cases this causes the p-value to be lower than the minimum value that can be stored and as a result it is reported as exactly 0. However, confidence intervals for the partial dependence functions are calculated correctly and can give an indication of the degree of certainty about whether the modelled relationship between the predictor and target variables reflects a true relationship or is due to random chance.

Additionally, note the pseudo R-squared of 0.0198. R-squared is a measure of the amount of variation in the dependent variable that is explained by the independent variables. In this case the dependent variable is mortality, which has many causes not related to heat index, population density or greenspace. Therefore, we expect to explain only a small amount of variation in the data is expected. Past research has found that about 7-8% of deaths to be “temperature-attributable.”[[46]](#endnote-46) Therefore, under ideal conditions we would expect to achieve an R-squared around or slightly above 0.08. However, given the sparse nature of the UNOS data and limited number of deaths used for creating the models, 0.0198 is a satisfactory result.

A screenshot of a computer

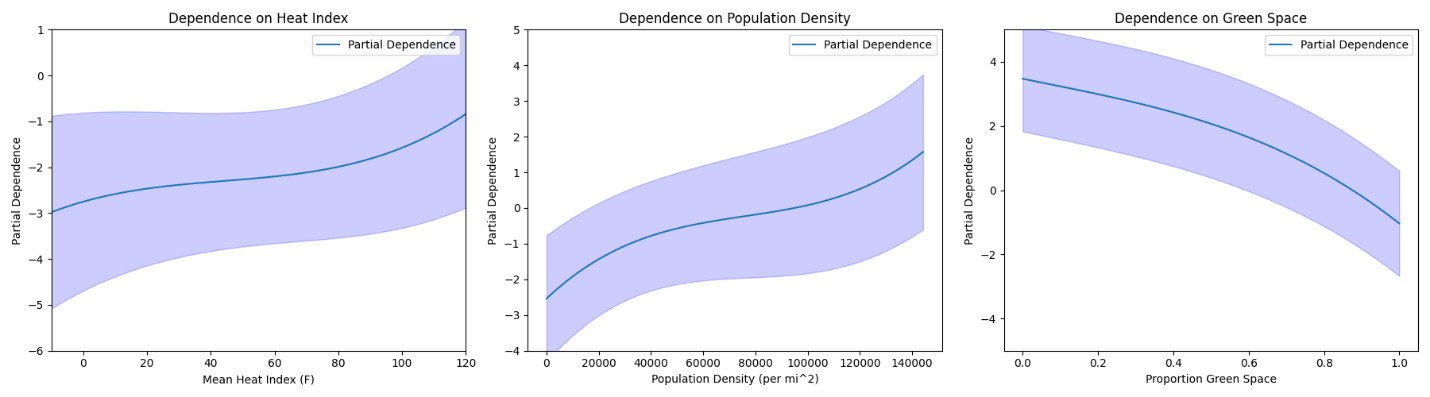
Description automatically generated

**Figure 3:** Regression output of model 1. This is the result of fitting the following equation:

ln(Expected Mortality) = f₁(Heat Index) + f₂(Population Density) + f₃(Green Space) + f₄(Heat Index, Population Density) + f₅(Heat Index, Green Space) + c

A known bug in the PyGAM library causes p-values to appear lower than they should. The significance of variables can better be assessed by examining the partial dependence plots and their confidence intervals. Note a pseudo R-squared of 0.0198 which is reasonably in line with what is expected based on past research which has sought to determine the effect of temperature on all-cause mortality.

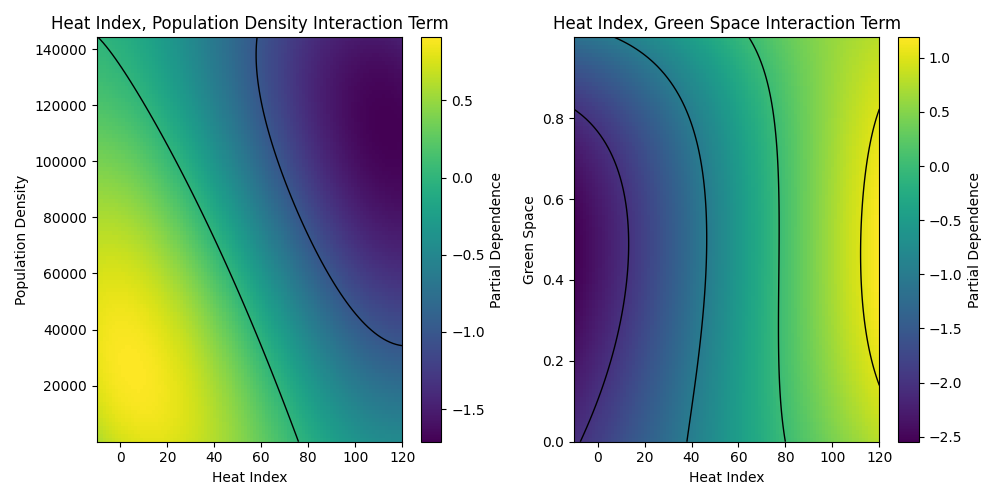
Figure 4 shows the partial dependence functions for the first three terms in the model. Contrary to the expectation of a U-shaped curve indicating an increase in expected mortality with both extremely low and high temperatures, we see expected mortality increasing steadily as heat index increases. The second plot displays increasing expected mortality with increasing population density. Finally, we notice strongly decreasing expected mortality with increasing green space.



**Figure 4:** Plots of the partial dependence of expected mortality on the three individual terms in model 1. From left to right these show f₁(Heat Index), f₂(Population Density), f₃(Green Space). Note steadily increasing expected mortality with increasing heat index, increasing expected mortality with increasing population density, and decreasing expected mortality with increasing green space. Recall that the Poisson model fits the following relationship:

ln(Expected Mortality) = f₁(Heat Index) + f₂(Population Density) + f₃(Green Space) + f₄(Heat Index, Population Density) + f₅(Heat Index, Green Space).

Figure 5 displays plots of the interaction terms in model 1. These are included for the sake of completion (to show plots of each function f1 through f5), but it is necessary to be cautious interpreting these plots since they do not fully include the effects of the individual predictors heat index, population density and green space. Instead, these can be thought of as adjustments to the individual terms plotted in Figure 4. For instance, the positive values in the bottom left corner of the first plot in Figure 5 indicates that mortality in zip codes with low population density on days with a low average heat index is higher than would be expected based on the plots in Figure 4.

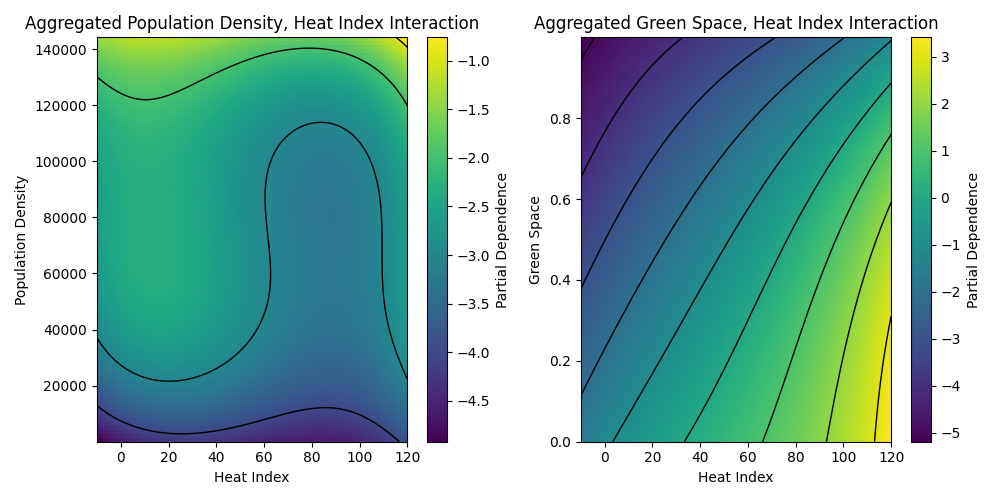


**Figure 5:** Plots of the partial dependence of expected mortality on the two interaction terms in model 1. These show f4(Heat Index, Population Density) and f5(Heat Index, Green Space) respectively. Contours are plotted for integer values, indicating that movement from one contour to another represents a change in expected mortality by a factor of *e*. Recall that the Poisson model fits the following relationship:

ln(Expected Mortality) = f₁(Heat Index) + f₂(Population Density) + f₃(Green Space) + f₄(Heat Index, Population Density) + f₅(Heat Index, Green Space).

To fully inspect the effect of interactions of heat index and population density on expected mortality we examine the sum of all functions in the model using these variables f1(Heat Index) + f2(Population Density) + f4(Heat Index, Population Density). Likewise, to fully inspect the effect of interactions of heat index and greenspace on expected mortality, we examine f1(Heat Index) + f2(Green Space) + f5(Heat Index, Green Space). These aggregated functions are plotted in Figure 6.

Figure 6 shows that expected mortality is dramatically higher in zip codes with low amounts of green space on days with hot temperatures. Decreases in green space and increases in heat index both contribute to increasing expected mortality. Additionally, the gradient is slightly higher with lower values of green space. This suggests that zip codes with the least green space are slightly more sensitive to temperature increases. The dependence of expected mortality on population density and heat index together is less dramatic and less straightforward. It is difficult to make definitive inferences about how these variables are related based on these results.



**Figure 6:** Left – A plot of the partial dependence of expected mortality on the interaction heat index and population density using all relevant terms: f₁(Heat Index) + f₂(Population Density) + f₄(Heat Index, Population Density).

Right – A plot of the partial dependence of expected mortality on the interaction heat index and green space using all relevant terms: f₁(Heat Index) + f3(Green Space) + f5(Heat Index, Green Space).

Contours are plotted for integer values, indicating that movement from one contour to another represents a change in expected mortality by a factor of *e*. Recall that the Poisson model fits the following relationship:

ln(Expected Mortality) = f₁(Heat Index) + f₂(Population Density) + f₃(Green Space) + f₄(Heat Index, Population Density) + f₅(Heat Index, Green Space).

***III.3 Model 2 Results, Summer Months***

The regression summary of model 2 is displayed in figure 7. This model is created using observations only from the summer months, June through August. A slightly higher pseudo R-squared than model 1 indicates that a higher portion of mortality in summer months is attributable to the predictors of average heat index, population density, and green space.

A screenshot of a computer

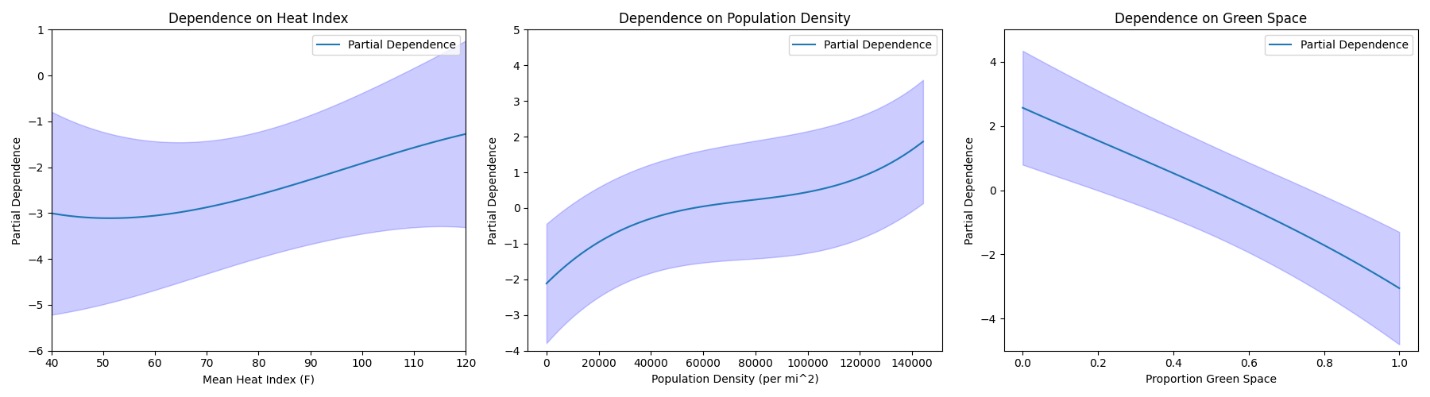
Description automatically generated

**Figure 7:**  Summary of the regression results for model 2. Model 2 uses only data in months June through August with outliers with heat index values outside of the range 40-120 removed. This model is the result of fitting the following equation:

ln(Expected Mortality) = f₁(Heat Index) + f₂(Population Density) + f₃(Green Space) + f₄(Heat Index, Population Density) + f₅(Heat Index, Green Space) + c

A known bug in the PyGAM library causes p-values to appear lower than they should. The significance of variables can better be assessed by examining the partial dependence plots and their confidence intervals. Note a pseudo R-squared of 0.0206, slightly higher than in model 1.

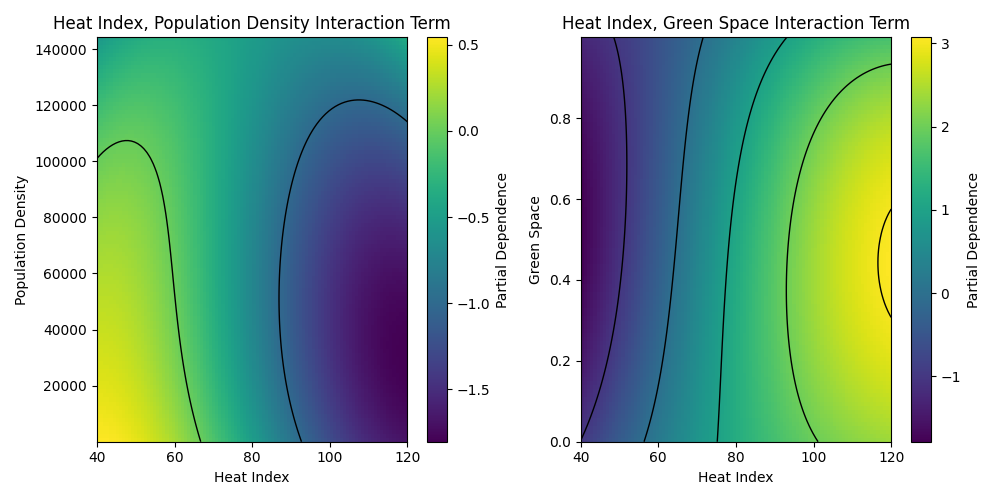
Figure 8 shows the partial dependence functions for the first three terms in the model. In contrast with model 1, the first plot shows a flat or slightly decreasing expected mortality with rising heat index up to about 60 degrees. This is more in line with what has been shown by past research, which has observed minimal mortality occurs in many cities between 60-70°F.[[47]](#endnote-47) The effect of population density on expected mortality is very similar to model 1. However, when we look exclusively at summer months, we see a much more significant decline in mortality with increasing green space.



**Figure 8:** Plots of the partial dependence of expected mortality on the three individual terms in model 2, created using only data from June-August. From left to right these show f₁(Heat Index), f₂(Population Density), f₃(Green Space). Note increasing expected mortality with increasing average heat index, increasing population density, and decreasing green space. Expected mortality decreases much more rapidly with increasing green space in this model compared to model 1. Recall that the Poisson model fits the following relationship:

ln(Expected Mortality) = f₁(Heat Index) + f₂(Population Density) + f₃(Green Space) + f₄(Heat Index, Population Density) + f₅(Heat Index, Green Space).

Figure 9 displays plots of the interaction terms in model 2. Interestingly, in low population density zip codes we observe decreasing expected mortality with increasing average heat index, while in high population density zip codes there appears to be no effect. Additionally, the interaction term between green space and heat index is more influential in this model compared to model 1. Here values of the interaction term range from below -1.5 to 3, whereas the interaction term varied from -2.5 to 1 in model 1. This is further evidence that green space does more to decrease mortality in summer months.



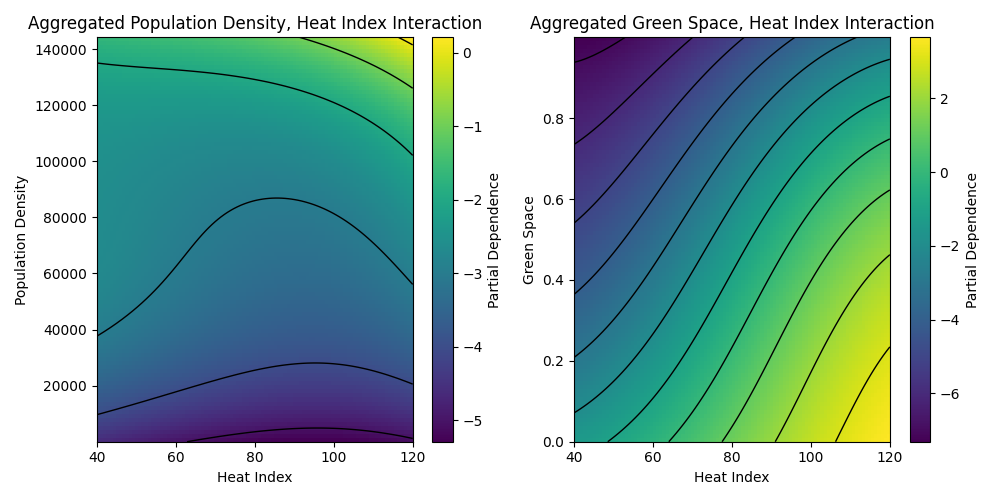
**Figure 6:** Left – A plot of the partial dependence of expected mortality on the interaction heat index and population density using all relevant terms: f₁(Heat Index) + f₂(Population Density) + f₄(Heat Index, Population Density).

Right – A plot of the partial dependence of expected mortality on the interaction heat index and green space using all relevant terms: f₁(Heat Index) + f3(Green Space) + f5(Heat Index, Green Space).

Contours are plotted for integer values, indicating that movement from one contour to another represents a change in expected mortality by a factor of *e*. Recall that the Poisson model fits the following relationship:

ln(Expected Mortality) = f₁(Heat Index) + f₂(Population Density) + f₃(Green Space) + f₄(Heat Index, Population Density) + f₅(Heat Index, Green Space).

Figure 10 shows the aggregated relationships of heat index with population density and green space respectively. When all functions are examined together, expected mortality appears to increase slightly with increasing heat index in the zip codes with the highest population densities in summer months, but there appears to be close to no dependence on heat index in the vast majority of zip codes. Additionally, we observe a very strong relationship between expected mortality and the interaction of green space and heat index. This relationship is stronger than in model 1. Furthermore, in model 2, expected mortality increases most rapidly with heat index in zip codes with a moderate amount of green space (35-70%), whereas in model 1 it increased most rapidly in zip codes with low green space (< 25%). This may suggest that the risk of heat increases most significantly in areas with moderate amounts of green space in the summer, relative to the rest of the year.



**Figure 10:** Left – A plot of the partial dependence of expected mortality on the interaction heat index and population density using all relevant terms: f₁(Heat Index) + f₂(Population Density) + f₄(Heat Index, Population Density).

Right – A plot of the partial dependence of expected mortality on the interaction heat index and green space using all relevant terms: f₁(Heat Index) + f3(Green Space) + f5(Heat Index, Green Space).

Contours are plotted for integer values, indicating that movement from one contour to another represents a change in expected mortality by a factor of *e*. Recall that the Poisson model fits the following relationship:

ln(Expected Mortality) = f₁(Heat Index) + f₂(Population Density) + f₃(Green Space) + f₄(Heat Index, Population Density) + f₅(Heat Index, Green Space).

**III.4 Discussion**

The results of our models indicate that lack of green space and exposure to high average heat index are both associated independently with higher expected mortality of liver and kidney transplant recipients. Together, high heat index and lack of green space create dramatically higher mortality risk. In the summer months, increasing heat index causes mortality risk to rise most rapidly in areas with a moderate amount of green space. However, over the course of the entire year, high heat index causes the greatest increase in mortality risk in areas with low green space. One possible explanation is that areas with less green space are more likely to exceed dangerous thresholds for heat even in non-summer months; although, more research is necessary to definitively explain this observation.

Both models show an increased expectation of mortality in zip codes with higher population density holding all else equal (Figure 4,8). However, the interaction of this relationship with heat index is more complicated as lower density zip codes have higher expected mortality at lower temperatures in both models (Figures 6, 10). Negative health effects from extreme cold cannot be used to explain this relationship, since it holds even on days with an average heat index between 40-80 in summer months (Figure 10).

Future research should consider both controlling for and examining associations with other social determinants of health. Access to green space often correlates with other socioeconomic factors. For instance, a study of urban green space accessibility in Atlanta, Georgia found that African American and Asian populations had significantly poorer access to green spaces.[[48]](#endnote-48) Future research could investigate how this may contribute to disparate health outcomes. Conversely, high population density tends to correlate with closer access to healthcare, and this may make it more difficult to determine population density’s impact on health as a result of its impact on the UHI effect.

Additionally, future research could both control for and examine associations with other risk factors. For instance, the risk of extreme heat is much higher for the elderly. Likewise, as previously discussed liver and kidney transplant recipients have a higher risk for infections and cancer due to the immunosuppressive medication they take. It could also be informative to examine the effect of heat on short-term vs. long-term mortality in transplant recipients. In the short-term, transplant recipients may face unique risks related to their operation and the disease which caused them to need the transplant. It is possible that heat could interact with and elevate these risks. In the long-term, it would be expected that their risk of mortality as a result of heat would be more similar to the general population, but more research is necessary to confirm this.

Finally, in order to evaluate the relationship between heat index and factors influencing the urban heat island effect in the general population rather than only kidney and liver transplant recipients, the ‘restricted-use linked mortality files’ associated with the National Center for Health Statistics. This is a linkage between National Center for Health Statistics Survey data and National Death Index data containing date of death, cause of death, and location of death. Therefore, this research could also investigate which specific causes of death are associated with heat.

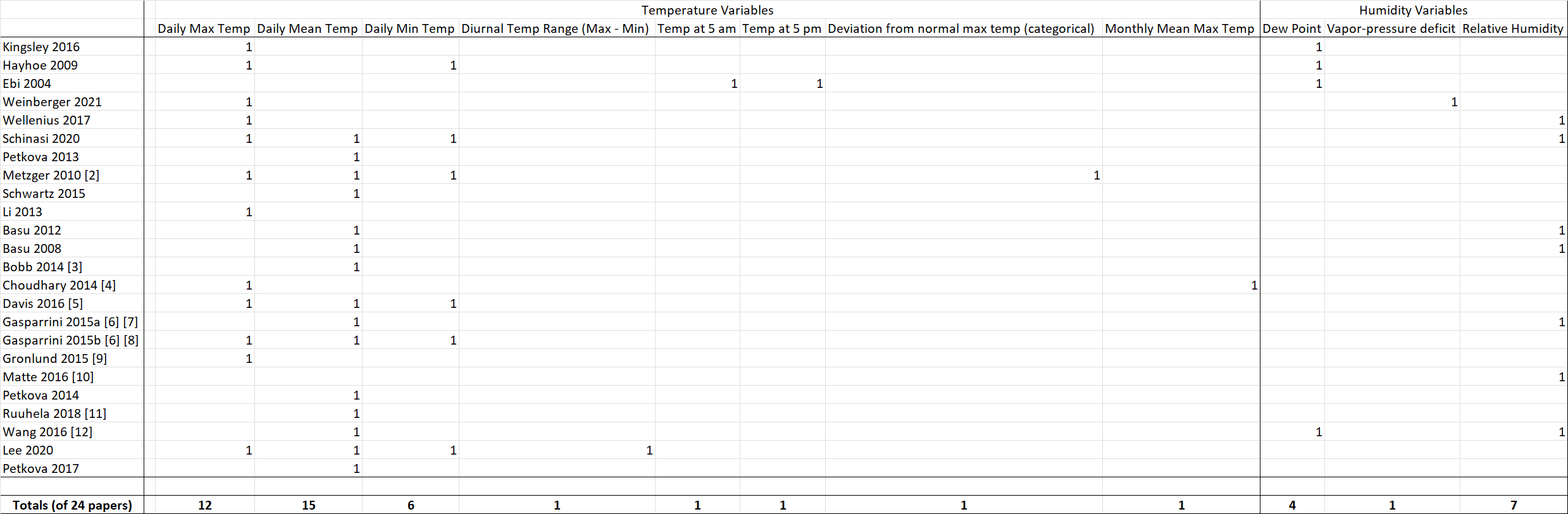
1. **Conclusions**

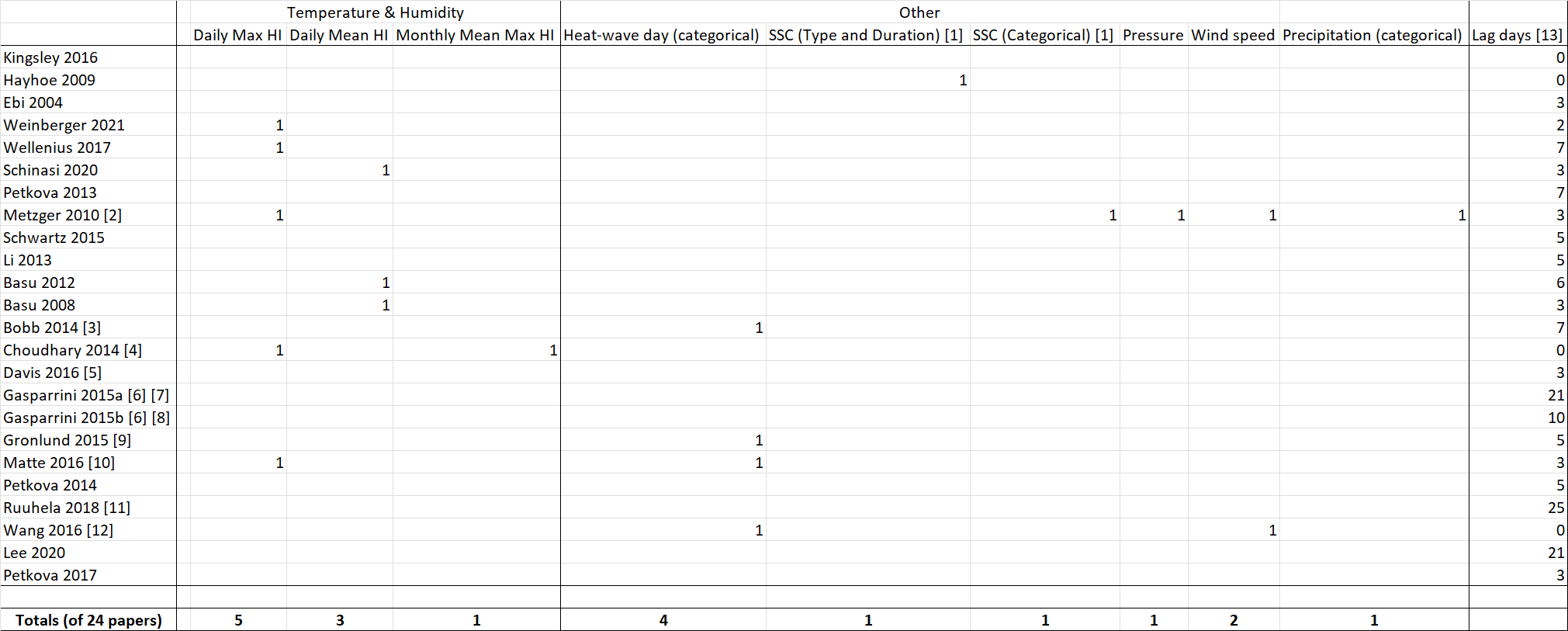
We find both a strong association between increasing average daily heat index and decreasing green space with higher expected mortality in kidney and liver transplant recipients from 2005 to January 2023. This relationship is most pronounced in summer months. Overall, this suggests that the UHI effect, which is often driven by a lack of green space, significantly increases mortality due to high temperatures within cities. This is consistent with past research which has shown increasing risk of mortality on days with high heat index, and research which has shown green space to be associated with improved health. Additionally, over the course of the whole year, zip codes with low amounts of green space (< 25%) appear to be most sensitive to increasing average heat index. During the summer months specifically, zip codes with moderate amounts of green space (35-70%) appear to be most sensitive.

As global temperatures rise due to climate change, it will be important to take action to mitigate the worst negative health outcomes of extreme heat. It is necessary inform people of the risks of extreme heat; past research has shown that early warning systems reduce mortality during periods of extreme heat.[[49]](#endnote-49) Additionally, it will be necessary to ensure access to air conditioning, construct and retrofit buildings to be more efficient and require less cooling. This will allow vulnerable populations to limit their exposure to heat. Finally, this research suggests that expanding green spaces may mitigate the UHI effect and limit the worst negative health outcomes.

**Appendix A: Partial Review of Past Research**

Full version of Table 1 (on two lines):

****

****

Notes:

1. SSC = spatial synoptic classification (air mass type)
2. Metzger et al. 2010 defines max HI as the greater of max temp and max heat index. Pressure, wind speed, and precipitation were used only in sensitivity analysis, and not in the main model.
3. Bobb et al. 2014 explores multiple definitions of heatwaves, each based on a number of consecutive days (2-4 days) with temperatures at the high end of distribution of daily temps for a given county (97th, 98th, 99th percentile)
4. Choudhary et al. 2014 uses public health data that is monthly rather than daily, so they use convert from Daily max temp and HI to monthly mean max temp and HI
5. Davis et al. 2016 calculates daily mean temp in three different ways: the average of max and min at a 15 minute resolution, the average of max and min at a 1-hr resolution, and the average of 24 hourly measurements
6. Gasparrini et al. 2015a = Mortality risk attributable to high and low ambient temperature. Gasparrini et al. 2015b = Temporal Variation in Heat-Mortality Associations
7. Humidity was only used in sensitivity analysis. Mean temp was the only variable in the primary model.
8. Max and min temp were used in sensitivity analysis. Mean temp was used in the primary model.
9. Heatwave day defined as having max temp > 97th percentile
10. Matte et al. 2016 defines a heatwave day as a any day with HI > 100 degrees F or consecutive days with HI > 95 degrees
11. The authors say they use spatially averaged daily temperature. However, It is not clear if it is mean or max temp or other.
12. Heat wave day is defined as consecutive days with daily mean temp greater than 97th percentile mean temperature of all days. Compared to stricter definitions: 98th or 99th percentile and requiring 4 consecutive days. (Similar to Bobb et al. (2014))
13. Listed here is the max number of lag days considered by a particular study. Many studies test up to this many days of lag, but in their primary model, only use as many days as are statistically significant.

**Appendix B: Relative Humidity and Heat Index Formulas**

Relative Humidity as calculated in Alduchov et al. (1996) [[50]](#endnote-50)

Heat Index as defined by the National Centers for Environmental Prediction (including adjustments):[[51]](#endnote-51)

If the **RH** is less than 13% and the temperature is between 80 and 112 degrees F, then subtract:

if the **RH**is greater than 85% and the temperature is between 80 and 87 degrees F, then add:

If the above results in an HI value less than 80, it is replaced by the following:

**Appendix C: Preprocessing National Weather Service (NWS) Data**

Summary:

* Take .csv and .csv.gz files downloaded from NOAA
* Extract the relevant columns,
* Calculate corresponding heat index for each measurement of heat index and dew point
* Calculate daily averages of heat index and temperature

Combine all data into one single file called Aggregated\_NWS.csv

Final Product: Aggregated\_NWS.csv file contains daily averages of temperature and heat indices based on measurements at each NWS station with data available. It also includes each station’s longitude and latitude.

Raw files as downloaded from the NOAA Integrated Surface Database Available here:

<https://upenn.box.com/s/s7ktb46iokiumzutone0eajc5sb3yiip>

Requests for these files were made from March 28-31, 2023. From the following location:

<https://www.ncei.noaa.gov/access/search/data-search/global-hourly>

Aggregated\_NWS.csv and Intermediate files (which are created by the following code) available here:

<https://upenn.box.com/s/ezuim96kf0de2um01hzda5jbvn1z02pc>

Python Code

**import** numpy **as** np

**import** pandas **as** pd

**import** os

**import** matplotlib**.**pyplot **as** plt

**import** pickle

**import** math

# T is temperature and TD is dew point in Celsius

# Formula from University of Miami

# https://bmcnoldy.rsmas.miami.edu/Humidity.html

**def** relative\_humidity**(**T**,** TD**):**

**return** 100 **\*** np**.**exp**((**17.625 **\*** TD**)** **/** **(**243.04 **+** TD**))** **/** np**.**exp**((**17.625 **\*** T**)** **/** **(**243.04 **+** T**))**

# Based on HI equation from NOAA

# T should be in celsius, converted to F within the function

# https://www.wpc.ncep.noaa.gov/html/heatindex\_equation.shtml

**def** heat\_index**(**T**,** RH**):**

T **=** 32 **+** **(**T**\***1.8**)** # Convert to F

hi **=** 0.5 **\*** **(**T **+** 61.0 **+** **((**T **-** 68.0**)** **\*** 1.2**)** **+** **(**RH **\*** 0.094**))**

**if** **(**hi**+**T**)/**2 **<** 80**:**

**return** T

**else:**

hi **=** **-**42.379 **+** **(**2.04901523**\***T**)** **+** **(**10.14333127**\***RH**)** **-** **(**0.22475541**\***T**\***RH**)** **-** **(**0.00683783**\***T**\*\***2**)** **-** **(**0.05481717**\*(**RH**\*\***2**))** **+** **(**0.00122874**\*(**T**\*\***2**)\***RH**)** **+** **(**0.00085282**\***T**\*(**RH**\*\***2**))** **-** 0.00000199**\*(**T**\*\***2**)\*(**RH**\*\***2**)**

**if** RH **<** 13 **and** T **>=** 80 **and** T **<=** 112**:**

adjustment **=** **((**13 **-** RH**)** **/** 4**)** **\*** np**.**sqrt**((**17 **-** **abs(**T **-** 95**))/**17**)**

hi **-=** adjustment

**elif** RH **>** 85 **and** T **>=** 80 **and** T **<=** 87**:**

adjustment **=** **((**RH **-** 85**)** **/** 10**)** **\*** **((**87 **-** T**)** **/** 5**)**

hi **+=** adjustment

**return** hi

# input df with temp and dewPt columns

# optional arguments if those columns have different names

**def** heat\_index\_df**(**df**,** temp**=**'temp'**,** dewPt**=**'dewPt'**):**

df**[**'RH'**]** **=** relative\_humidity**(**df**[**temp**],**df**[**dewPt**])**

df**[**'HI'**]** **=** **[**heat\_index**(**T**,**RH**)** **for** T**,** RH **in** **zip(**df**[**temp**],**df**[**'RH'**])]**

**return** df

# input pandas series following the NWS format (e.g. TMP, DEW columns)

# These cols have a specific notation

# chars 0-4 are the value multiplied by 10

# char 6 gives info about quality and source of measurement

# Note that TMP and DEW are in celsius

# page 10-11: https://www.ncei.noaa.gov/data/global-hourly/doc/isd-format-document.pdf

**def** extract\_vals**(**ser**):**

validation\_char **=** ser**.str[**6**]**

ser **=** ser**.str[:**5**]**

**if** validation\_char **in** **[**'0'**,**'1'**,**'4'**,**'5'**,**'9'**,**'A'**,**'I'**,**'M'**,**'P'**,**'R'**,**'U'**]:**

ser **=** pd**.**to\_numeric**(**ser**,** errors**=**'coerce'**)**

ser **/=** 10.0 #NWS notation is scaled by a factor of 10

ser**.**replace**(**999.9**,** pd**.**NA**,** inplace**=True)** # 999.9 indicates missing value

**elif** validation\_char **==** 'C'**:** # in this case reported in whole degrees C instead of scaled by 10

ser **=** pd**.**to\_numeric**(**ser**,** errors**=**'coerce'**)**

ser**.**replace**(**999.9**,** pd**.**NA**,** inplace**=True)**

**return** ser

# Process from raw files downloaded from NOAA Integrated surface database into 9 CSV files with just the necessary variables

# Processed 15 or fewer at a time to avoid running out of memory

# Processed files were manually moved out of C:/NWS after each run of this program.

# Similarly X in the output file 'NWS\_dataX' was manually incremented

os**.**chdir**(**'C:/NWS'**)**

NWS\_data **=** pd**.**DataFrame**()**

i**=**0

**for** file **in** os**.**listdir**():**

**if** i **<** 15**:**

**if** file**.**endswith**(**'.csv'**):**

df **=** pd**.**read\_csv**(**file**,** header**=**0**)[[**'LATITUDE'**,** 'LONGITUDE'**,** 'DATE'**,** 'TMP'**,** 'DEW'**]]**

**elif** file**.**endswith**(**'.csv.gz'**):**

df **=** pd**.**read\_csv**(**file**,** compression**=**'gzip'**,** header**=**0**)[[**'LATITUDE'**,**'LONGITUDE'**,**'DATE'**,**'TMP'**,**'DEW'**]]**

**else:**

**print(**'unexpected file'**)**

**continue**

df**[**'LATITUDE'**]** **=** pd**.**to\_numeric**(**df**[**'LATITUDE'**],** downcast**=**'float'**)**

df**[**'LONGITUDE'**]** **=** pd**.**to\_numeric**(**df**[**'LONGITUDE'**],** downcast**=**'float'**)**

df**[**'DATE'**]** **=** pd**.**to\_datetime**(**df**[**'DATE'**]).**dt**.**date

df**[**'TMP'**]** **=** extract\_vals**(**df**[**'TMP'**])**

df**[**'DEW'**]** **=** extract\_vals**(**df**[**'DEW'**])**

df**.**dropna**(**subset**=[**'TMP'**,**'DEW'**],** axis**=**0**,** inplace**=True)**

NWS\_data **=** pd**.**concat**([**NWS\_data**,** df**])**

i**+=**1

NWS\_data**.**to\_csv**(**'NWS\_data9.csv'**,** index**=False)**

# Read the created CSV files

# Calculate heat index

# Average the measurements from each monitor for each day

# Combine into a single dataframe

NWS\_groupby **=** pd**.**DataFrame**()**

#os.chdir('NWS\_data')

**for** file **in** os**.**listdir**():**

**if** **(**file**[:**3**]** **==** 'NWS'**):**

df **=** pd**.**read\_csv**(**file**,** header**=**0**)[[**'LATITUDE'**,**'LONGITUDE'**,**'DATE'**,**'TMP'**,**'DEW'**]]**

df **=** heat\_index\_df**(**df**,**temp**=**'TMP'**,**dewPt**=**'DEW'**)**

df\_groupby **=** df**.**groupby**(**by**=[**'LATITUDE'**,**'LONGITUDE'**,**'DATE'**]).**agg**({**'HI'**:** 'mean'**,** 'TMP'**:** 'mean'**}).**reset\_index**()**

NWS\_groupby **=** pd**.**concat**([**NWS\_groupby**,** df\_groupby**])**

NWS\_groupby **=** NWS\_groupby**.**rename**(**columns**={**'TMP'**:** 'MEAN\_TMP'**,** 'HI'**:**'MEAN\_HI'**})**

# Data from some monitors was downloaded twice due to nature of NOAA ISDB

# Count duplicate rows, drop them

num\_duplicates **=** NWS\_groupby**.**duplicated**().sum()**

**print(**f'Duplicate rows: {num\_duplicates}'**)**

NWS\_groupby**.**drop\_duplicates**(**inplace**=True)**

# Output into a single CSV

NWS\_groupby**.**to\_csv**(**'Aggregated\_NWS.csv'**,** index**=False)**

Notes

* As indicated by the comments in the code, the intermediate files NWS\_data1 through NWS\_data10 were created by manually running a portion of this code multiple times to process up to 15 files at a time. This was due to memory constraints.
* Air temperature and dew point are both reported as strings with the format “+0218,1” where the first 5 characters denote the temperature multiplied by 10 and the seventh character is a quality code. Therefore, the example “+0218,1” indicates that the temperature is positive 21.8°C and that the measurement passed all quality control checks. The extract\_vals function extracts the numeric values of temperature from these strings.
* The extract\_vals function excludes measurements with quality codes of 2 and 6 which are ‘suspect’ and codes of 3 and 7 which are ‘erroneous,’ and includes all others. A code of C indicates that measurements are reported in whole degrees Celsius instead of scaled by 10, and so if this code is present, values are not divided by 10.[[52]](#endnote-52)

**Appendix D: Process Landcover Data**

Summary:

* Read raster of landcover data.
* Read and reproject shapefile of zip code data.
* Count number of developed, ‘green,’ and total pixels in each zip code

Final Product: zip\_landcover\_stats.csv which records each zip code and the number of all pixels, developed pixels, and ‘green’ pixels within it.

Data Available Here: <https://upenn.box.com/s/b0zku9s3z0zi1h3y92q2zo8zqvxg5wob>

Raw landcover data downloaded from: <https://www.mrlc.gov/data/nlcd-2016-land-cover-conus>

Zip code shapefile downloaded from: <https://catalog.data.gov/dataset/tiger-line-shapefile-2019-2010-nation-u-s-2010-census-5-digit-zip-code-tabulation-area-zcta5-na>

**import** numpy **as** np

**import** rasterio

**import** geopandas **as** gpd

**from** rasterstats **import** zonal\_stats

**with** rasterio**.open(**"Land Cover/nlcd\_2016\_land\_cover\_l48\_20210604.img"**)** **as** src**:**

# Read the metadata

metadata **=** src**.**meta

# Read the raster data as a NumPy array

data **=** src**.**read**(**1**)**

shapefile\_path **=** "ZIP Code Shapefile/tl\_2019\_us\_zcta510.shp"

gdf **=** gpd**.**read\_file**(**shapefile\_path**)**

# Reproject the shapefile to the raster's CRS

gdf **=** gdf**.**to\_crs**(**metadata**[**'crs'**])**

# Legend here: https://www.mrlc.gov/data/legends/national-land-cover-database-class-legend-and-description

# 22, 23, and 24 are low, medium, and high-intensity development. Each low is defined by >20% impervious surfaces

**def** count\_developed\_pixels**(**x**):**

**return** **(**x **==** 22**).sum()** **+** **(**x **==** 23**).sum()** **+** **(**x **==** 24**).sum()**

# Includes essentially everything except development, water, snow, desert

**def** count\_green\_pixels**(**x**):**

**return** **((**x**==**41**).sum()+(**x**==**42**).sum()+(**x**==**43**).sum()+(**x**==**21**).sum()+(**x**==**51**).sum()+(**x**==**52**).sum()+**

**(**x**==**71**).sum()+(**x**==**72**).sum()+(**x**==**73**).sum()+(**x**==**74**).sum()+(**x**==**81**).sum()+(**x**==**82**).sum()+**

**(**x**==**90**).sum()+(**x**==**95**).sum())**

# Calculate zonal statistics for each polygon

stats **=** zonal\_stats**(**gdf**,** data**,** affine**=**src**.**transform**,** stats**=[**'count'**],** add\_stats**={**'developed\_count'**:** count\_developed\_pixels**,** 'green\_count'**:** count\_green\_pixels**})**

# Add the zonal statistics to the GeoDataFrame

gdf\_stats **=** gdf**.**copy**()**

gdf\_stats**[**'count'**]** **=** **[**s**[**'count'**]** **for** s **in** stats**]**

gdf\_stats**[**'developed\_count'**]** **=** **[**s**[**'developed\_count'**]** **for** s **in** stats**]**

gdf\_stats**[**'green\_count'**]** **=** **[**s**[**'green\_count'**]** **for** s **in** stats**]**

gdf\_stats**.**to\_csv**(**'zip\_landcover\_stats.csv'**,** index**=False)**

Notes

* In defining green space I excluded low, medium, and high intensity development. I also excluded open water, perennial ice/snow, and barren land (rock/sand/clay). I was hesitant about excluding these natural, non-developed categories because I suspect that when most people conceptualize “green space,” really what they mean is non-developed space, especially in an urban context. This was a major part of the motivation for creating a separate category for “developed space.”
* In defining “developed space” I made the decision to exclude “developed, open space” because I believed it would include much of what we consider “urban green space.” From the classification description, this category includes: “areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses… These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control or aesthetic purposes.”

**Appendix E: Create Dictionary of NWS Measurements, Population Density, and Green Space percentage around NWS stations**

Summary:

* This transforms the NWS data into a dictionary data structure to make it quicker to retrieve data in later portions of the code. Keys are latitude and longitude coordinates of each station, and values are a list of all average temperatures and heat indices for the days where that station has measurements.
* Additionally, the estimated landcover and population density surrounding the NWS station is calculated and included with its measurements, so that these values can be controlled for in model 2..

Final Product: dict\_nws.pkl which contains the dictionary created here.

For each station this records the variables: average temp and average heat index for all days when that monitor took measurements, aggregated population density and green space in a 5 mile radius around the stations.

Data Available Here: <https://upenn.box.com/s/by9j836evwuig5ki1y1hybyi2522srfh>

**import** numpy **as** np

**import** pandas **as** pd

**import** os

**import** matplotlib**.**pyplot **as** plt

**import** pickle

**import** math

NWS\_groupby **=** pd**.**read\_csv**(**'Aggregated\_NWS.csv'**,** header**=**0**)**

# Read ZIP code centroids

zips **=** pd**.**read\_csv**(**'US\_ZIP\_codes\_to\_longitude\_and\_latitude.csv'**)**

zips**[**'Latitude'**]** **=** pd**.**to\_numeric**(**zips**[**'Latitude'**])**

zips**[**'Longitude'**]** **=** pd**.**to\_numeric**(**zips**[**'Longitude'**])**

zips**.**reset\_index**(**drop**=True,**inplace**=True)**

# Read population density data (mercifully clean)

dtypes **=** **{**'Zip/ZCTA'**:int,** '2010 Population'**:int,** 'Land-Sq-Mi'**:float,**'Density Per Sq Mile'**:float}**

names **=** **{**'Zip/ZCTA'**:**'zip'**,** '2010 Population'**:**'pop'**,** 'Land-Sq-Mi'**:**'area'**,**'Density Per Sq Mile'**:**'density'**}**

pop\_dens **=** pd**.**read\_csv**(**'Zipcode-ZCTA-Population-Density-And-Area-Unsorted.csv'**,** header**=**0**,** dtype**=**dtypes**)**

pop\_dens**.**rename**(**columns**=**names**,** inplace**=True)**

pop\_dens\_nonzero **=** pop\_dens**[**pop\_dens**[**'density'**]** **!=** 0**]**

# Calculate population density z-scores

pop\_dens\_nonzero**[**'density\_norm'**]** **=** **(**pop\_dens\_nonzero**[**'density'**]** **-** pop\_dens\_nonzero**[**'density'**].**mean**())/(**pop\_dens\_nonzero**[**'density'**].**std**())**

zips\_pop\_dens **=** zips**.**merge**(**pop\_dens**,** how **=** 'inner'**,** left\_on **=** 'Zip'**,** right\_on **=** 'zip'**)**

zips\_dens\_landcover **=** zips\_pop\_dens**.**merge**(**landcover**,**how**=**'inner'**,**left\_on**=**'zip'**,**right\_on**=**'ZCTA5CE10'**)**

zips\_dens\_landcover **=** zips\_dens\_landcover**.**merge**(**pop\_dens\_nonzero**[[**'zip'**,**'density\_norm'**]],**how**=**'left'**,**on**=**'zip'**)**

ips\_landcover **=** zips**.**merge**(**landcover**,**how **=** 'inner'**,**left\_on**=**'Zip'**,**right\_on**=**'ZCTA5CE10'**)**

# Get list of lat/lon coords from NWS data

NWS\_LAT\_LON**=**NWS\_groupby**.**drop\_duplicates**([**'LATITUDE'**,**'LONGITUDE'**],** keep**=**'first'**)[[**'LATITUDE'**,**'LONGITUDE'**,**'DATE'**]]**

NWS\_LAT\_LON**.**reset\_index**(**drop**=True,**inplace**=True)**

# nws\_lat\_lon is a df with one line of coordinates for each monitor

# zips\_pop\_dens is df with zip code centroid coordinates, population density, and landcover data

# This function returns nws\_lat\_lon with several columns added

# nearest\_zip is the zip code centroid that the monitor is closest to

# nearest\_zips\_pop\_dens is the population density of up to the nearest num\_to\_check zip codes within

# max\_dist miles of the monitor. If there are no zips within max distance, it is the population density

# of the closest zip.

# furthest\_distance is the distance to the centroid of the furthest zip code used to calculate nearest\_zips\_pop\_dens

**def** find\_closest\_zips**(**nws\_lat\_lon**,** zips\_pop\_dens**,** num\_to\_check**=**5**,** max\_dist**=**5**):**

nws\_coords **=** nws\_lat\_lon**[[**'LATITUDE'**,** 'LONGITUDE'**]].**to\_numpy**()**

zip\_coords **=** zips\_pop\_dens**[[**'Latitude'**,** 'Longitude'**]].**to\_numpy**()**

# This calculates L2 distance which should be a good approximation here since our points are close enough that

# they are not affected by the curvature of the Earth.

# Originally used Haversine distance, but it is much slower and not practical for the size of our data

dists **=** np**.**linalg**.**norm**(**nws\_coords**[:,** np**.**newaxis**]** **-** zip\_coords**,** axis**=**2**)** **\*** 69.172 # Convert from degrees to miles

closest\_zip\_indices **=** np**.**argpartition**(**dists**,** **range(**num\_to\_check**),** axis**=**1**)[:,** **:**num\_to\_check**]**

nearest\_zips\_data **=** **[]**

# indices are the indices of num\_to\_check closest monitors

**for** i**,** indices **in** **enumerate(**closest\_zip\_indices**):**

**if** max\_dist **is** **not** **None:**

indices **=** **[**idx **for** idx **in** indices **if** dists**[**i**,** idx**]** **<=** max\_dist **or** idx **==** indices**[**0**]]**

nearest\_pop\_dens **=** zips\_pop\_dens**.**iloc**[**indices**]**

nearest\_pop\_dens\_sum **=** **(**nearest\_pop\_dens**[**'pop'**].sum()** **/** nearest\_pop\_dens**[**'area'**].sum())**

landcover\_count **=** nearest\_pop\_dens**[**'count'**].sum()**

nearest\_green\_prop **=** **(**nearest\_pop\_dens**[**'green\_count'**].sum()/**landcover\_count**)**

nearest\_dev\_prop **=** **(**nearest\_pop\_dens**[**'developed\_count'**].sum()/**landcover\_count**)**

furthest\_zip\_index **=** indices**[-**1**]**

furthest\_distance **=** dists**[**i**,** furthest\_zip\_index**]**

nearest\_zip **=** zips\_pop\_dens**[**'Zip'**].**iloc**[**indices**[**0**]]**

nearest\_zips\_data**.**append**((**nearest\_zip**,** nearest\_pop\_dens\_sum**,** nearest\_green\_prop**,** nearest\_dev\_prop **,**furthest\_distance**))**

nearest\_zips\_df **=** pd**.**DataFrame**(**nearest\_zips\_data**,** columns**=[**'nearest\_zip'**,** 'nearest\_zips\_pop\_dens'**,** 'nearest\_zips\_green\_prop'**,**'nearest\_zips\_dev\_prop'**,** 'furthest\_distance'**])**

nws\_lat\_lon **=** pd**.**concat**([**nws\_lat\_lon**,** nearest\_zips\_df**],** axis**=**1**)**

**return** nws\_lat\_lon

NWS\_LAT\_LON **=** find\_closest\_zips**(**NWS\_LAT\_LON**,** zips\_dens\_landcover**)**

# Merge to NWS Monitors

# First add to the list of monitors, then add to all the measurements we have.

# NWS\_dens = NWS\_LAT\_LON.merge(pop\_dens\_nonzero[['zip', 'density\_norm']],how='left',on='zip')

mu **=** pop\_dens\_nonzero**[**'density'**].**mean**()**

sigma **=** pop\_dens\_nonzero**[**'density'**].**std**()**

NWS\_LAT\_LON**[**'density\_norm'**]** **=** **[(**d**-**mu**)/**sigma **for** d **in** NWS\_LAT\_LON**[**'nearest\_zips\_pop\_dens'**]]**

NWS\_groupby\_zips\_dens **=** NWS\_groupby**.**merge**(**NWS\_LAT\_LON**[[**'LATITUDE'**,**'LONGITUDE'**,**'nearest\_zip'**,**'density\_norm'**,**'nearest\_zips\_green\_prop'**,**'nearest\_zips\_dev\_prop'**]],**how**=**'left'**,**on**=[**'LATITUDE'**,**'LONGITUDE'**])**

# Verify that these are 0

**print(**NWS\_LAT\_LON**[**'density\_norm'**].**isna**().sum())**

**print(**NWS\_groupby\_zips\_dens**[**'density\_norm'**].**isna**().sum())**

# Create dict out of nws data. Super fast to retrieve data from.

# Takes 10+ min to create

nws\_dict **=** **{}**

**for** \_**,** row **in** NWS\_groupby\_zips\_dens**.**iterrows**():**

lat\_lon **=** **(**row**[**'LATITUDE'**],** row**[**'LONGITUDE'**])**

**if** lat\_lon **not** **in** nws\_dict**:**

nws\_dict**[**lat\_lon**]** **=** **[]**

nws\_dict**[**lat\_lon**].**append**((**row**[**'DATE'**],** row**[**'MEAN\_TMP'**],** row**[**'MEAN\_HI'**],** row**[**'density\_norm'**],** row**[**'nearest\_zips\_green\_prop'**],** row**[**'nearest\_zips\_dev\_prop'**]))**

# Save dict in file so it will be quicker to read in the future

**with** **open(**'dict\_nws.pkl'**,** 'wb'**)** **as** f**:**

pickle**.**dump**(**nws\_dict**,** f**)**

Notes

* In this section I estimated landcover and population density around NWS stations. The reason for this is that I wanted to compare the population density surrounding stations where measurements were taken with the population density of zip codes where deaths were observed. If a measurement was taken in a high-density area and a death occurs in a low-density area, it is likely that the measurement was affected by the UHI effect more than the individual who died. Therefore, I wanted to see if we would observe stronger relationships by using the population density in the zip code of a death minus the population density in the zip code where temperature was recorded as an independent variable rather than just looking at the population density in the zip code of the death alone.
* In this section, I calculate z-scores of population density for each zip code. This refers to the number of standard deviations away from the mean of the population density in each zip code. I chose to use the z-scores before subtracting because based on what is known about the UHI effect, I expected the relationship to be nonlinear. In retrospect, however, z-scoring does not solve this issue since it is a linear transformation. The non-linearity is instead accounted for by using splines in the Poisson GAM regression. Using raw, non-z-scored population density would not have changed the results.
* In the find\_closest\_zips function I aggregate landcover and population density data from up to ‘num\_to\_check’ zip codes within ‘max\_dist’ miles of the NWS station. The reason for this is that NWS stations can often be located in very small zip codes that contain only businesses and therefore technically have a population density of 0 despite the surrounding area in reality having a large, dense population. I believe that aggregating the values from the nearest few zip codes gives a more accurate picture of the conditions surrounding the station. For this analysis the default values of up to 5 zip codes within 5 miles were used.
* A .json file, which is often used to store dictionaries could not be used since the keys are tuples (latitude, longitude), and so it is stored in a .pkl file instead.

**Appendix F: Create ZIP code table with Temperature and Heat Indices for all days**

Summary:

* This code records in a single table the temperature and heat index on each day for each zip code where a death was observed in the UNOS data.
* It also records the landcover and population density statistics for the NWS station which took each measurement.

Final Product: Aggregated\_ZIP\_TEMPS.csv which includes the average temperature and heat index for each zip code on each day from 2005-01-01 to 2023-01-31, population density and green space in those zip codes, and the population density and landcover statistics at the station which took each temperature measurement. Importantly this table also includes the number of deaths observed in the UNOS data during this range, which is used as the dependent variable in the following regression analysis.

Data Available Here:

* Zip code table and intermediate files: <https://upenn.box.com/s/yt0nbfzgo0xj2qgiviprgdjyecsi6r5i>
* UNOS: <https://upenn.box.com/s/e3oze49hj97auzp2fke2u38rdu0ul2ww>

Sources:

* UNOS data provided directly by Dr. Therese Bittermann, which is also available here: <https://unos.org/data/>
* NWS Data downloaded here: <https://www.ncei.noaa.gov/access/search/data-search/global-hourly>
* Landcover data: <https://www.mrlc.gov/data/nlcd-2016-land-cover-conus>
* Zip code shapefile: <https://catalog.data.gov/dataset/tiger-line-shapefile-2019-2010-nation-u-s-2010-census-5-digit-zip-code-tabulation-area-zcta5-na>

Python Code

**import** numpy **as** np

**import** pandas **as** pd

**import** os

**import** matplotlib**.**pyplot **as** plt

**import** pickle

**import** math

NWS\_groupby **=** pd**.**read\_csv**(**'Aggregated\_NWS.csv'**,** header**=**0**)**

# Read ZIP code centroids

zips **=** pd**.**read\_csv**(**'US\_ZIP\_codes\_to\_longitude\_and\_latitude.csv'**)**

zips**[**'Latitude'**]** **=** pd**.**to\_numeric**(**zips**[**'Latitude'**])**

zips**[**'Longitude'**]** **=** pd**.**to\_numeric**(**zips**[**'Longitude'**])**

zips**.**reset\_index**(**drop**=True,**inplace**=True)**

# Read Kidney Transplant data

kidn\_trans **=** pd**.**read\_csv**(**'Kidney transplant dataset CSV.csv'**)**

kidn\_trans**[**'death\_date'**]** **=** pd**.**to\_datetime**(**kidn\_trans**[**'composite\_death\_date'**])** # Cast death to datetime

kidn\_trans**[**'perm\_zip\_trr'**]** **=** pd**.**to\_numeric**(**kidn\_trans**[**'perm\_zip\_trr'**],** errors**=**'coerce'**,** downcast**=**'integer'**)**

# Record and print number of rows with missing or non-numeric zip (should not be many), which are dropped

num\_dropped **=** kidn\_trans**[**'perm\_zip\_trr'**].**isna**().sum()**

**print(**f"Dropped {num\_dropped} out of {**len(**kidn\_trans**)**} rows from kidney data because of non-numeric zip."**)**

kidn\_trans **=** kidn\_trans**.**dropna**(**subset **=** **[**'perm\_zip\_trr'**,**'composite\_death\_date'**],** axis **=** 0**)**

# Read liver transplant data

liver\_trans **=** pd**.**read\_csv**(**'Liver transplant dataset CSV.csv'**)**

liver\_trans**[**'death\_date'**]** **=** pd**.**to\_datetime**(**liver\_trans**[**'composite\_death\_date'**])** #cast death to datetime

liver\_trans**[**'perm\_zip\_trr'**]** **=** pd**.**to\_numeric**(**liver\_trans**[**'perm\_zip\_trr'**],** errors**=**'coerce'**,** downcast**=**'integer'**)**

# Record and print number of rows with missing or non-numeric zip (should not be many), which are dropped

num\_dropped **=** liver\_trans**[**'perm\_zip\_trr'**].**isna**().sum()**

**print(**f"Dropped {num\_dropped} out of {**len(**liver\_trans**)**} rows from liver data because of non-numeric zip."**)**

liver\_trans **=** liver\_trans**.**dropna**(**subset **=** **[**'perm\_zip\_trr'**,**'composite\_death\_date'**],** axis **=** 0**)**

# Combine transplant data and count deaths occuring on each date in ecah zip code

combined\_df **=** pd**.**concat**([**kidn\_trans**,** liver\_trans**],** axis**=**0**)**

combined\_df\_group **=** combined\_df**.**groupby**([**'perm\_zip\_trr'**,**'death\_date'**])**

ct\_combined\_df **=** combined\_df\_group**.**count**()**

ct\_combined\_df **=** ct\_combined\_df**.**reset\_index**()**

# Join transplant data to zip code centroids

counts\_lat\_lon **=** ct\_combined\_df**.**merge**(**zips**,** how**=**'left'**,** left\_on**=**'perm\_zip\_trr' **,** right\_on**=**'Zip'**)**

num\_to\_drop **=** counts\_lat\_lon**[**'Latitude'**].**isna**().sum()**

# Drop necessary rows

**print(**f"Dropped {num\_to\_drop} out of {**len(**counts\_lat\_lon**)**} rows because zip did not exist in centroid data."**)**

counts\_lat\_lon **=** counts\_lat\_lon**.**dropna**(**subset **=** **[**'Latitude'**,** 'Longitude'**],** axis **=** 0**)**

# Get list of lat/lon from transplant data zips

ZIP\_LAT\_LON**=**counts\_lat\_lon**.**drop\_duplicates**([**'Latitude'**,**'Longitude'**])[[**'Latitude'**,**'Longitude'**,**'Zip'**]]**

ZIP\_LAT\_LON**.**reset\_index**(**drop**=True,**inplace**=True)**

date\_range **=** pd**.**date\_range**(**start**=**'2005-01-01'**,** end**=**'2023-01-31'**)**

# lat\_lon is tuple location of NWS monitor

# zip\_row is a row of the zip\_lat\_lon df

# ZIP\_TEMPS is the new numpy array we're building

# i is the index we're keeping track of

# Assigns to each zip code data from the nearest monitor for each available date range and linearly interpolates

# missing values

**def** resample\_and\_add**(**lat\_lon**,** zip\_row**,** ZIP\_TEMPS**,** i**,** dates\_after **=** pd**.**to\_datetime**(**'2004-12-31'**),** dates\_before **=** pd**.**to\_datetime**(**'2023-02-01'**)):**

i0**=**i

temp\_data **=** pd**.**DataFrame**(**nws\_dict**[**lat\_lon**],** columns**=[**'DATE'**,** 'MEAN\_TMP'**,** 'MEAN\_HI'**,** 'density\_norm'**,**'nearest\_zips\_green\_prop'**,**'nearest\_zips\_dev\_prop'**])**

temp\_data**[**'DATE'**]** **=** pd**.**to\_datetime**(**temp\_data**[**'DATE'**])**

temp\_data**.**drop\_duplicates**(**subset**=[**'DATE'**],** inplace**=True)**

temp\_data**.**set\_index**(**'DATE'**,** inplace**=True)**

# Interpolate missing values at daily frequency

temp\_data **=** temp\_data**.**resample**(**'D'**).**interpolate**(**method**=**'linear'**)**

temp\_data**.**reset\_index**(**inplace**=True)**

temp\_data**[**'zip'**]** **=** zip\_row**[**'Zip'**]**

# Columns here are zip, date, mean temp, mean HI, density norm

**for** \_**,**t\_row **in** temp\_data**.**iterrows**():**

**if** **(**t\_row**[**'DATE'**]** **>** dates\_after **and** t\_row**[**'DATE'**]** **<** dates\_before**):**

ZIP\_TEMPS**[**i**]** **=** **(**zip\_row**[**'Zip'**],** t\_row**[**'DATE'**],** t\_row**[**'MEAN\_TMP'**],** t\_row**[**'MEAN\_HI'**],** t\_row**[**'density\_norm'**],**t\_row**[**'nearest\_zips\_green\_prop'**],**t\_row**[**'nearest\_zips\_dev\_prop'**])**

i **+=** 1

**if** i **!=** i0 **+** **(**dates\_before**-**dates\_after**).**days **-** 1**:print(**f'dates before: {dates\_before}. dates after: {dates\_after}. i: {i}. i0: {i0}. days in range: {**(**dates\_before**-**dates\_after**).**days **+** 1}. Days added: {i**-**i0}. Temp data begin: {temp\_data**[**"DATE"**].**iloc**[**0**]**}. Temp data end: {temp\_data**[**"DATE"**].**iloc**[-**1**]**}.'**)**

**return** ZIP\_TEMPS**,** i

date\_range **=** pd**.**date\_range**(**'2005-01-01'**,** '2023-01-31'**,** freq**=**'D'**)**

nws\_date\_range **=** NWS\_groupby**.**groupby**([**'LATITUDE'**,** 'LONGITUDE'**]).**agg**({**'DATE'**:** **[**'min'**,** 'max'**]}).**reset\_index**()**

nws\_date\_range**.**columns **=** **[**'\_'**.**join**(**col**).**strip**()** **for** col **in** nws\_date\_range**.**columns**.**values**]**

nws\_date\_range**.**rename**(**columns**={**'DATE\_min'**:** 'nws\_fst\_date'**,** 'DATE\_max'**:** 'nws\_lst\_date'**},** inplace**=True)**

nws\_date\_range**.**sort\_values**(**by**=[**'LATITUDE\_'**],** ascending**=True,** inplace**=True)**

nws\_date\_range**.**reset\_index**(**drop**=True,** inplace**=True)**

nws\_date\_range\_coords **=** nws\_date\_range**[[**'LATITUDE\_'**,** 'LONGITUDE\_'**]].**to\_numpy**()**

# Batch ZIPS to do 1000 at a time so we don't run out of memory

**for** j **in** **range(**math**.**ceil**(len(**ZIP\_LAT\_LON**)/**1000**)):**

ZIP\_TEMPS **=** np**.**empty**((**1000 **\*** **(len(**date\_range**)),** 7**),** dtype**=object)**

i**=**0 # tracks index within new zip\_temps array

**for** k **in** **range(**1000**):**

zip\_idx **=** **(**j**\***1000**)** **+** k

**if(**zip\_idx **==** **len(**ZIP\_LAT\_LON**)):** **break** # stops when we have finished all zips

zip\_coords **=** ZIP\_LAT\_LON**[**'Latitude'**][**zip\_idx**],** ZIP\_LAT\_LON**[**'Longitude'**][**zip\_idx**]**

distances **=** np**.**linalg**.**norm**(**nws\_date\_range\_coords **-** zip\_coords**,** axis**=**1**)**

sorted\_indices **=** np**.**argsort**(**distances**)**

# Record first and last dates from the closest monitor

first **=** nws\_date\_range**[**'nws\_fst\_date'**].**iloc**[**sorted\_indices**[**0**]]**

last **=** nws\_date\_range**[**'nws\_lst\_date'**].**iloc**[**sorted\_indices**[**0**]]**

needed\_ranges **=** **[]**

**if** first **!=** '2005-01-01'**:** needed\_ranges**.**append**((**pd**.**to\_datetime**(**'2005-01-01'**),**pd**.**to\_datetime**(**first**)-**pd**.**Timedelta**(**days**=**1**)))**

**if** last **!=** '2023-01-31'**:** needed\_ranges**.**append**((**pd**.**to\_datetime**(**last**)+**pd**.**Timedelta**(**days**=**1**),** pd**.**to\_datetime**(**'2023-01-31'**)))**

# Add measurements from all days from the closest monitor

closest\_lat\_lon **=** **(**nws\_date\_range\_coords**[**sorted\_indices**[**0**],**0**],** nws\_date\_range\_coords**[**sorted\_indices**[**0**],**1**])**

i0 **=** i

ZIP\_TEMPS**,** i **=** resample\_and\_add**(**closest\_lat\_lon**,** ZIP\_LAT\_LON**.**iloc**[**zip\_idx**],** ZIP\_TEMPS**,** i**,** pd**.**to\_datetime**(**first**)-**pd**.**Timedelta**(**days**=**1**),** pd**.**to\_datetime**(**last**)+**pd**.**Timedelta**(**days**=**1**))**

**assert** i **==** **(**i0 **+** **(**pd**.**to\_datetime**(**last**)-**pd**.**to\_datetime**(**first**)).**days **+** 1**),** f'not adding the right number of vals. Off by: {**(**i0 **+** **(**pd**.**to\_datetime**(**last**)-**pd**.**to\_datetime**(**first**)).**days **+** 1**)-**i}'

**for** s **in** sorted\_indices**:**

**if** needed\_ranges**:**

monitor\_begin**,** monitor\_end **=** pd**.**to\_datetime**(**nws\_date\_range**[**'nws\_fst\_date'**].**iloc**[**s**]),** pd**.**to\_datetime**(**nws\_date\_range**[**'nws\_lst\_date'**].**iloc**[**s**])**

new\_needed\_ranges **=** **[]**

**for** begin\_range**,** end\_range **in** needed\_ranges**:**

#begin\_range, end\_range = pd.to\_datetime(begin\_range), pd.to\_datetime(end\_range)

**if** monitor\_begin **<=** begin\_range **and** monitor\_end **>=** end\_range**:** # Monitor data fully covers the needed range

lat\_lon **=** **(**nws\_date\_range\_coords**[**s**,**0**],** nws\_date\_range\_coords**[**s**,**1**])**

ZIP\_TEMPS**,** i **=** resample\_and\_add**(**lat\_lon**,** ZIP\_LAT\_LON**.**iloc**[**zip\_idx**],** ZIP\_TEMPS**,** i**,** begin\_range**-**pd**.**Timedelta**(**days**=**1**),** end\_range**+**pd**.**Timedelta**(**days**=**1**))**

**elif** monitor\_begin **<=** begin\_range **and** monitor\_end **>=** begin\_range**:** # Monitor data partially covers the needed range from the beginning

lat\_lon **=** **(**nws\_date\_range\_coords**[**s**,**0**],** nws\_date\_range\_coords**[**s**,**1**])**

ZIP\_TEMPS**,** i **=** resample\_and\_add**(**lat\_lon**,** ZIP\_LAT\_LON**.**iloc**[**zip\_idx**],** ZIP\_TEMPS**,** i**,** begin\_range**-**pd**.**Timedelta**(**days**=**1**),** monitor\_end**+**pd**.**Timedelta**(**days**=**1**))**

new\_needed\_ranges**.**append**((**monitor\_end**+**pd**.**Timedelta**(**days**=**1**),** end\_range**))**

**elif** monitor\_begin **<=** end\_range **and** monitor\_end **>=** end\_range**:** # Monitor data partially covers the needed range from the end

lat\_lon **=** **(**nws\_date\_range\_coords**[**s**,**0**],** nws\_date\_range\_coords**[**s**,**1**])**

ZIP\_TEMPS**,** i **=** resample\_and\_add**(**lat\_lon**,** ZIP\_LAT\_LON**.**iloc**[**zip\_idx**],** ZIP\_TEMPS**,** i**,** monitor\_begin**-**pd**.**Timedelta**(**days**=**1**),** end\_range**+**pd**.**Timedelta**(**days**=**1**))**

new\_needed\_ranges**.**append**((**begin\_range**,** monitor\_begin**-**pd**.**Timedelta**(**days**=**1**)))**

**else:** # No overlap between monitor data and needed range

new\_needed\_ranges**.**append**((**begin\_range**,** end\_range**))**

needed\_ranges **=** new\_needed\_ranges

**else:**

**break** # exit loop when there are no more ranges to fill

**if** i **!=** **(**i0 **+** **len(**date\_range**)):** **print(**f'value not added for each day in range. Zip: {zip\_idx}'**)**

**if** **(**zip\_idx**+**1**)** **%** 100 **==** 0**:** **print(**f'{zip\_idx**+**1} out of {**len(**ZIP\_LAT\_LON**)**} zips completed'**)**

ZIP\_TEMPS **=** pd**.**DataFrame**(**ZIP\_TEMPS**)**

ZIP\_TEMPS**.**to\_csv**(**f'ZIP\_TEMPS\_{j}.csv'**,** index**=False)**

# Read in ZIP temp files instead of creating from scratch every time

ZIP\_TEMPS **=** pd**.**DataFrame**()**

#os.chdir('NWS\_data')

**for** file **in** os**.**listdir**():**

**if** **(**file**[:**9**]** **==** 'ZIP\_TEMPS'**):**

df **=** pd**.**read\_csv**(**file**,** header**=**0**,** names **=** **[**'Zip'**,**'DATE'**,**'MEAN\_TEMP'**,**'MEAN\_HI'**,**'nws\_density\_norm'**,**'nws\_green\_prop'**,**'nws\_dev\_prop'**])**

ZIP\_TEMPS **=** pd**.**concat**([**ZIP\_TEMPS**,** df**])**

# Validation. Missing should be the same number as expected.

num\_missing **=** ZIP\_TEMPS**.**isna**().sum(**axis**=**0**)**

**print(**f'{num\_missing}, expected {**(**18000**-**17884**)\*len(**date\_range**)**}'**)**

ZIP\_TEMPS**.**dropna**(**inplace**=True)**

ZIP\_TEMPS **=** ZIP\_TEMPS**.**astype**({**'Zip'**:int,** 'DATE'**:str,** 'MEAN\_TEMP'**:float,**'MEAN\_HI'**:float,**'nws\_density\_norm'**:float,**'nws\_green\_prop'**:float,**'nws\_dev\_prop'**:float})**

ZIP\_TEMPS**[**'DATE'**]** **=** pd**.**to\_datetime**(**ZIP\_TEMPS**[**'DATE'**])**

ZIP\_TEMPS**.**reset\_index**(**inplace**=True)**

ZIP\_TEMPS **=** ZIP\_TEMPS**.**rename**(**columns**={**'density\_norm'**:** 'nws\_density\_norm'**})**

# Save this since it takes forever to create. 8-9 min to run

ZIP\_TEMPS**.**to\_csv**(**'Aggregated\_ZIP\_TEMPS.csv'**,** index**=False)**

Notes

* This code creates intermediate files: e.g. ZIP\_TEMPS\_0 for each 1000 zip codes to avoid running out of memory.
* Many NWS stations were not operational for the entirety of the range of dates 2005-01-01 to 2023-01-31. Therefore, the closest operational station to any given zip code in our data changes throughout this range. This is the reason for the complicated logic in this function and the necessity of keeping track of “needed\_ranges.” It is also why density and landcover characteristics of the stations are included when creating the ZIP\_TEMPS dataframe instead of being added using a merge based on the closest station to each zip code.
* If NWS stations have no measurements for a day within the time they were operational (which happens occasionally) average heat index and temperature values were linearly interpolated. This may not be ideal, but it should offer a strong approximation in most cases.

**Appendix G: Regression and Analysis**

Summary:

* Join UNOS data to landcover data, and weather data based on zip code.
* Conduct a grid search to determine optimal parameters for the models.
* Create the models
* Create plots

Final Product: Model summaries and plots

Samples mentioned in the code available here: <https://upenn.box.com/s/6v5sxl6c2w3gyw01ou5ckfjf9use2415>

Data Sources:

* UNOS data provided directly by Dr. Therese Bittermann, which is also available here: <https://unos.org/data/>
* NWS Data downloaded here: <https://www.ncei.noaa.gov/access/search/data-search/global-hourly>
* Landcover data: <https://www.mrlc.gov/data/nlcd-2016-land-cover-conus>

Python Code:

**import** numpy **as** np

**import** pandas **as** pd

**import** os

**import** matplotlib**.**pyplot **as** plt

**import** pickle

**import** math

**from** pygam **import** LinearGAM**,** PoissonGAM**,** s**,** l**,** te

**import** folium

**import** geopandas **as** gpd

**import** seaborn **as** sns

**from** sklearn**.**model\_selection **import** GridSearchCV

**def** create\_plots**(**feature\_names**,** titles**,** model**,** ylim**=None,** xlim**=None,** desc**=**''**):**

fig**,** axes **=** plt**.**subplots**(**nrows**=**1**,** ncols**=**3**,** figsize**=(**18**,** 5**))**

**for** i**,** term **in** **enumerate(**model**.**terms**):**

**if** i **<=** 2**:**

# Compute the partial dependence and the confidence intervals

XX **=** model**.**generate\_X\_grid**(**term**=**i**)**

pdep**,** confi **=** model**.**partial\_dependence**(**term**=**i**,** X**=**XX**,** width**=**0.95**)**

# Create a DataFrame for the partial dependence data

data **=** **{**'x'**:** XX**[:,** term**.**feature**],** 'y'**:** pdep**[:],** 'lower'**:** confi**[:,** 0**],** 'upper'**:** confi**[:,** 1**]}**

df **=** pd**.**DataFrame**(**data**)**

# Create a Seaborn line plot for the partial dependence

sns**.**lineplot**(**x**=**'x'**,** y**=**'y'**,** data**=**df**,** errorbar**=None,** label**=**'Partial Dependence'**,** ax**=**axes**[**i**])**

# Add the confidence intervals as a shaded area

axes**[**i**].**fill\_between**(**df**[**'x'**],** df**[**'lower'**],** df**[**'upper'**],** color**=**'blue'**,** alpha**=**0.2**)**

# Set the title and labels for the plot

axes**[**i**].**set\_title**(**f'Dependence on {titles**[**i**]**}'**)**

axes**[**i**].**set\_xlabel**(**feature\_names**[**term**.**feature**])**

axes**[**i**].**set\_ylabel**(**'Partial Dependence'**)**

**if** ylim **is** **not** **None:**

ylim1**,** ylim2 **=** ylim**[**i**]**

axes**[**i**].**set\_ylim**(**ylim1**,** ylim2**)**

**if** xlim **is** **not** **None:**

**if** xlim**[**i**]** **is** **not** **None:**

xlim1**,** xlim2 **=** xlim**[**i**]**

axes**[**i**].**set\_xlim**(**xlim1**,** xlim2**)**

plt**.**tight\_layout**()**

plt**.**savefig**(**f'plots\_{desc}.png'**)**

plt**.**show**()**

# Terms is a list of the indices in model.terms which we want to plot

**def** create\_heatmap\_plot**(**terms**,** ylabels**,** model**,** titles**,** xlim**=None,** desc**=**''**):**

fig**,** axes **=** plt**.**subplots**(**nrows**=**1**,** ncols**=len(**terms**),** figsize**=(**5**\*len(**terms**),** 5**))**

**for** i**,** t **in** **enumerate(**terms**):**

XX **=** model**.**generate\_X\_grid**(**term**=**t**,** meshgrid**=True)**

Z **=** model**.**partial\_dependence**(**term**=**t**,** X**=**XX**,** meshgrid**=True)**

x\_vals **=** XX**[**0**][:,**0**]**

y\_vals **=** XX**[**1**][**0**]**

**if** xlim **is** **not** **None:**

xlim1**,** xlim2 **=** xlim**[**i**]**

im **=** axes**[**i**].**imshow**(**Z**,** cmap**=**'viridis'**,** aspect**=**'auto'**,** origin**=**'lower'**,**

extent**=[**xlim1**,** xlim2**,** y\_vals**.min(),** y\_vals**.max()])**

levels **=** np**.**arange**(**np**.**floor**(**np**.min(**Z**)),** np**.**ceil**(**np**.max(**Z**)+**1**),** 1**)**

contour **=** axes**[**i**].**contour**(**Z**,** colors**=**'k'**,** levels**=**levels**,** extent**=[**xlim1**,** xlim2**,** y\_vals**.min(),** y\_vals**.max()],** linestyles**=**'solid'**,**linewidths**=**1**)**

**else:**

im **=** axes**[**i**].**imshow**(**Z**,** cmap**=**'viridis'**,** aspect**=**'auto'**,** origin**=**'lower'**,**

extent**=[**x\_vals**.min(),** x\_vals**.max(),** y\_vals**.min(),** y\_vals**.max()])**

levels **=** np**.**arange**(**np**.**floor**(**np**.min(**Z**)),** np**.**ceil**(**np**.max(**Z**)+**1**),** 1**)**

contour **=** axes**[**i**].**contour**(**Z**,** colors**=**'k'**,** levels**=**levels**,** extent**=[**x\_vals**.min(),** x\_vals**.max(),** y\_vals**.min(),** y\_vals**.max()],** linestyles**=**'solid'**,**linewidths**=**1**)**

fig**.**colorbar**(**im**,** ax**=**axes**[**i**],** label**=**'Partial Dependence'**)**

# Set the appropriate axis labels

axes**[**i**].**set\_xlabel**(**'Heat Index'**)**

axes**[**i**].**set\_ylabel**(**ylabels**[**i**])**

axes**[**i**].**set\_title**(**titles**[**i**])**

plt**.**tight\_layout**()**

plt**.**savefig**(**f'plot\_heatmap\_{desc}.png'**)**

plt**.**show**()**

# xterm, yterm, tensorterm are lists indices of the terms to be ploted

**def** create\_additive\_heatmap\_subplots**(**xterm**,** yterm**,** tensorterm**,** ylabels**,** titles**,** model**,**xlim**=None,** desc**=**''**):**

fig**,** ax **=** plt**.**subplots**(**nrows**=**1**,** ncols**=len(**xterm**),** figsize**=(**5**\*len(**xterm**),** 5**))**

**for** i **in** **range(len(**xterm**)):**

XX0 **=** model**.**generate\_X\_grid**(**term**=**xterm**[**i**])**

XX1 **=** model**.**generate\_X\_grid**(**term**=**yterm**[**i**])**

XX2 **=** model**.**generate\_X\_grid**(**term**=**tensorterm**[**i**],** meshgrid**=True)**

Z0 **=** model**.**partial\_dependence**(**term**=**xterm**[**i**],** X**=**XX0**)**

Z1 **=** model**.**partial\_dependence**(**term**=**yterm**[**i**],** X**=**XX1**)**

Z2 **=** model**.**partial\_dependence**(**term**=**tensorterm**[**i**],** X**=**XX2**,** meshgrid**=True)**

x\_vals **=** XX2**[**0**][:,**0**]**

y\_vals **=** XX2**[**1**][**0**]**

# Z0 and Z1 are vectors length 100, Z2 is a 100x100 grid

# Z2[i] gives the ith row from the bottom (left to right)

# Z2[:,i] gives the ith column from the left (bottom to top)

# We want to add Z0 to all the rows and Z1 to all the columns

Z2 **=** Z2 **+** Z0 # adds to all the rows

Z2 **=** Z2 **+** Z1**[:,**np**.**newaxis**]** # pivots to a column vector, adds to all the columns

**if** xlim **is** **not** **None:**

xlim1**,**xlim2 **=** xlim**[**i**]**

im **=** ax**[**i**].**imshow**(**Z2**,** cmap**=**'viridis'**,** aspect**=**'auto'**,** origin**=**'lower'**,**

extent**=[**xlim1**,** xlim2**,** y\_vals**.min(),** y\_vals**.max()])**

levels **=** np**.**arange**(**np**.**floor**(**np**.min(**Z2**)),** np**.**ceil**(**np**.max(**Z2**)+**1**),** 1**)**

contour **=** ax**[**i**].**contour**(**Z2**,** colors**=**'k'**,** levels**=**levels**,** extent**=[**xlim1**,** xlim2**,** y\_vals**.min(),** y\_vals**.max()],** linestyles**=**'solid'**,**linewidths**=**1**)**

**else:**

im **=** ax**[**i**].**imshow**(**Z2**,** cmap**=**'viridis'**,** aspect**=**'auto'**,** origin**=**'lower'**,**

extent**=[**x\_vals**.min(),** x\_vals**.max(),** y\_vals**.min(),** y\_vals**.max()])**

levels **=** np**.**arange**(**np**.**floor**(**np**.min(**Z2**)),** np**.**ceil**(**np**.max(**Z2**)+**1**),** 1**)**

contour **=** ax**[**i**].**contour**(**Z2**,** colors**=**'k'**,** levels**=**levels**,** extent**=[**x\_vals**.min(),** x\_vals**.max(),** y\_vals**.min(),** y\_vals**.max()],** linestyles**=**'solid'**,**linewidths**=**1**)**

fig**.**colorbar**(**im**,** ax**=**ax**[**i**],** label**=**'Partial Dependence'**)**

ax**[**i**].**set\_xlabel**(**'Heat Index'**)**

ax**[**i**].**set\_ylabel**(**ylabels**[**i**])**

ax**[**i**].**set\_title**(**titles**[**i**])**

plt**.**tight\_layout**()**

plt**.**savefig**(**f'plot\_additive\_heatmaps\_{desc}.png'**)**

plt**.**show**()**

**def** create\_poisson\_gam\_model**(**n1**=**4**,** n2**=**4**,** n3**=**4**,** l**=**0.6**):**

model **=** PoissonGAM**(**s**(**0**,** n\_splines**=**n1**,** lam**=**l**)** **+**

s**(**1**,** n\_splines**=**n2**,** lam**=**l**)** **+**

s**(**2**,** n\_splines**=**n3**,** lam**=**l**)** **+**

te**(**0**,** 1**,** n\_splines**=[**n1**,** n2**],** lam**=[**l**,** l**])** **+**

te**(**0**,** 2**,** n\_splines**=[**n1**,** n3**],** lam**=[**l**,** l**]))**

**return** model

**def** grid\_search**(**df**,** subsample\_size**):**

df\_sub **=** df**.**sample**(**frac**=**subsample\_size**)** # downsample further necessary for gridsearch

# Define the search space for the number of splines and lambda

num\_splines\_range **=** np**.**arange**(**4**,** 8**,** 1**)**

lambda\_range **=** np**.**logspace**(-**3**,** 3**,** 7**)**

# Create a parameter grid

param\_grid **=** **{**'n1'**:** num\_splines\_range**,**

'n2'**:** num\_splines\_range**,**

'n3'**:** num\_splines\_range**,**

'l'**:** lambda\_range

**}**

# Model 1 grid search

# Set up the GridSearchCV with K-fold cross-validation

cv **=** GridSearchCV**(**create\_poisson\_gam\_model**(),** param\_grid**,** scoring**=**"neg\_mean\_squared\_error"**,** cv**=**5**,** verbose**=**1**,** n\_jobs**=-**1**)**

X **=** df\_sub**[[**'MEAN\_HI'**,**'density'**,**'green\_prop'**]]**

y **=** df\_sub**[**'death\_ct'**]**

# Fit the model and find the best parameters

cv**.**fit**(**X**,** y**)**

# Get the best model

**print(**"Best params:"**,** cv**.**best\_params\_**)**

**def** regression\_summary\_with\_var\_names**(**summary**):**

custom\_names **=** **{**'s(0) '**:** 'Heat Index'**,** 's(1) '**:** 'Population Density'**,** 's(2) '**:** 'Green Space'**,** 'te(0, 1) '**:** 'Heat Index\*Population Density'**,** 'te(0, 2) '**:** 'Heat Index\*Green Space'**}**

**for** old\_name**,** new\_name **in** custom\_names**.**items**():**

summary **=** summary**.**replace**(**old\_name**,** new\_name**)**

**print(**summary**)**

**return**

# Read ZIP code centroids

zips **=** pd**.**read\_csv**(**'US\_ZIP\_codes\_to\_longitude\_and\_latitude.csv'**)**

zips**[**'Latitude'**]** **=** pd**.**to\_numeric**(**zips**[**'Latitude'**])**

zips**[**'Longitude'**]** **=** pd**.**to\_numeric**(**zips**[**'Longitude'**])**

zips**.**reset\_index**(**drop**=True,**inplace**=True)**

# Read population density data (mercifully clean)

dtypes **=** **{**'Zip/ZCTA'**:int,** '2010 Population'**:int,** 'Land-Sq-Mi'**:float,**'Density Per Sq Mile'**:float}**

names **=** **{**'Zip/ZCTA'**:**'zip'**,** '2010 Population'**:**'pop'**,** 'Land-Sq-Mi'**:**'area'**,**'Density Per Sq Mile'**:**'density'**}**

pop\_dens **=** pd**.**read\_csv**(**'Zipcode-ZCTA-Population-Density-And-Area-Unsorted.csv'**,** header**=**0**,** dtype**=**dtypes**)**

pop\_dens**.**rename**(**columns**=**names**,** inplace**=True)**

pop\_dens\_nonzero **=** pop\_dens**[**pop\_dens**[**'density'**]** **!=** 0**]**

# Calculate population density z-scores based on all zip codes that are populated

pop\_dens\_nonzero**[**'density\_norm'**]** **=** **(**pop\_dens\_nonzero**[**'density'**]** **-** pop\_dens\_nonzero**[**'density'**].**mean**())/(**pop\_dens\_nonzero**[**'density'**].**std**())**

# Read landcover data

dtypes **=** **{**'ZCTA5CE10'**:int,** 'GEOID10'**:int,** 'CLASSFP10'**:str,**'MTFCC10'**:str,**'FUNCSTAT10'**:str,**'ALAND10'**:int,**

'AWATER10'**:int,**'INTPTLAT10'**:float,**'INTPTLON10'**:float,**'geometry'**:str,**'count'**:int,**'developed\_count'**:int,**

'green\_count'**:int}**

landcover **=** pd**.**read\_csv**(**'zip\_landcover\_stats.csv'**,**header**=**0**,**dtype**=**dtypes**)**

landcover**[**'green\_prop'**]** **=** landcover**[**'green\_count'**]/**landcover**[**'count'**]**

landcover**[**'developed\_prop'**]** **=** landcover**[**'developed\_count'**]/**landcover**[**'count'**]**

# These proportions already range 0-1 so z-scoring is not necessary

zips\_pop\_dens **=** zips**.**merge**(**pop\_dens**,** how **=** 'inner'**,** left\_on **=** 'Zip'**,** right\_on **=** 'zip'**)**

zips\_dens\_landcover **=** zips\_pop\_dens**.**merge**(**landcover**,**how**=**'inner'**,**left\_on**=**'zip'**,**right\_on**=**'ZCTA5CE10'**)**

# density norm column hasn't been added yet

zips\_dens\_landcover **=** zips\_dens\_landcover**.**merge**(**pop\_dens\_nonzero**[[**'zip'**,**'density\_norm'**]],**how**=**'left'**,**on**=**'zip'**)**

# Read Kidney Transplant data

kidn\_trans **=** pd**.**read\_csv**(**'Kidney transplant dataset CSV.csv'**)**

kidn\_trans**[**'death\_date'**]** **=** pd**.**to\_datetime**(**kidn\_trans**[**'composite\_death\_date'**])** # Cast death to datetime

kidn\_trans**[**'perm\_zip\_trr'**]** **=** pd**.**to\_numeric**(**kidn\_trans**[**'perm\_zip\_trr'**],** errors**=**'coerce'**,** downcast**=**'integer'**)**

# Record and print number of rows with missing or non-numeric zip (should not be many), which are dropped

num\_dropped **=** kidn\_trans**[**'perm\_zip\_trr'**].**isna**().sum()**

**print(**f"Dropped {num\_dropped} out of {**len(**kidn\_trans**)**} rows from kidney data because of non-numeric zip."**)**

kidn\_trans **=** kidn\_trans**.**dropna**(**subset **=** **[**'perm\_zip\_trr'**,**'composite\_death\_date'**],** axis **=** 0**)**

# Read liver transplant data

liver\_trans **=** pd**.**read\_csv**(**'Liver transplant dataset CSV.csv'**)**

liver\_trans**[**'death\_date'**]** **=** pd**.**to\_datetime**(**liver\_trans**[**'composite\_death\_date'**])** #cast death to datetime

liver\_trans**[**'perm\_zip\_trr'**]** **=** pd**.**to\_numeric**(**liver\_trans**[**'perm\_zip\_trr'**],** errors**=**'coerce'**,** downcast**=**'integer'**)**

# Record and print number of rows with missing or non-numeric zip (should not be many), which are dropped

num\_dropped **=** liver\_trans**[**'perm\_zip\_trr'**].**isna**().sum()**

**print(**f"Dropped {num\_dropped} out of {**len(**liver\_trans**)**} rows from liver data because of non-numeric zip."**)**

liver\_trans **=** liver\_trans**.**dropna**(**subset **=** **[**'perm\_zip\_trr'**,**'composite\_death\_date'**],** axis **=** 0**)**

# Combine transplant data and count deaths occuring on each date in ecah zip code

combined\_df **=** pd**.**concat**([**kidn\_trans**,** liver\_trans**],** axis**=**0**)**

combined\_df\_group **=** combined\_df**.**groupby**([**'perm\_zip\_trr'**,**'death\_date'**])**

ct\_combined\_df **=** combined\_df\_group**.**count**()**

ct\_combined\_df **=** ct\_combined\_df**.**reset\_index**()**

ct\_combined\_df**.**rename**(**columns**={**'composite\_death\_date'**:** 'death\_ct'**},** inplace**=True)**

# Join transplant data to zip code centroids

counts\_lat\_lon **=** ct\_combined\_df**.**merge**(**zips**,** how**=**'left'**,** left\_on**=**'perm\_zip\_trr' **,** right\_on**=**'Zip'**)**

num\_to\_drop **=** counts\_lat\_lon**[**'Latitude'**].**isna**().sum()**

# Validation checks about what we're going to drop

zips\_missing **=** counts\_lat\_lon**[**'perm\_zip\_trr'**][**counts\_lat\_lon**[**'Latitude'**].**isna**()]**

zip\_counts **=** zips\_missing**.**value\_counts**().**sort\_values**(**ascending**=False)**

zips\_missing**.**drop\_duplicates**(**inplace**=True)**

#print(f'Missing zips: {zip\_counts.to\_string()}')

# Drop necessary rows

**print(**f"Dropped {num\_to\_drop} out of {**len(**counts\_lat\_lon**)**} rows because zip did not exist in centroid data."**)**

counts\_lat\_lon **=** counts\_lat\_lon**.**dropna**(**subset **=** **[**'Latitude'**,** 'Longitude'**],** axis **=** 0**)**

ZIP\_TEMPS **=** pd**.**read\_csv**(**'Aggregated\_ZIP\_TEMPS.csv'**,** header**=**0**)**

ZIP\_TEMPS **=** ZIP\_TEMPS**.**astype**({**'Zip'**:int,** 'DATE'**:str,** 'MEAN\_TEMP'**:float,**'MEAN\_HI'**:float,**'nws\_density\_norm'**:float,**'nws\_green\_prop'**:float,**'nws\_dev\_prop'**:float})**

ZIP\_TEMPS**[**'DATE'**]** **=** pd**.**to\_datetime**(**ZIP\_TEMPS**[**'DATE'**])**

ZIP\_TEMPS\_UNOS **=** ZIP\_TEMPS**.**merge**(**counts\_lat\_lon**[[**'Zip'**,**'death\_date'**,**'death\_ct'**]],**how**=**'left'**,**left\_on**=[**'Zip'**,**'DATE'**],**right\_on**=[**'Zip'**,**'death\_date'**])**

**del** ZIP\_TEMPS # free up memory

ZIP\_TEMPS\_UNOS**[**'death\_ct'**].**fillna**(**0**,**inplace**=True)**

# ZIP\_TEMPS was national, but raster with landcover is only the lower 48, so here I'm assuming 0 for both means it wasn't in the raster range and dropping it.

mask **=** **(**ZIP\_TEMPS\_UNOS**[**'nws\_green\_prop'**]** **==** 0**)** **&** **(**ZIP\_TEMPS\_UNOS**[**'nws\_dev\_prop'**]** **==** 0**)**

ZIP\_TEMPS\_UNOS**.**drop**(**ZIP\_TEMPS\_UNOS**[**mask**].**index**,** inplace**=True)**

ZIP\_TEMPS\_UNOS**.**reset\_index**(**inplace**=True,**drop**=True)**

ZIP\_TEMPS\_UNOS**=**ZIP\_TEMPS\_UNOS**.**merge**(**zips\_dens\_landcover**[[**'Zip'**,**'green\_prop'**,**'developed\_prop'**,**'density'**,**'density\_norm'**]],**how**=**'left'**,**on**=**'Zip'**)**

ZIP\_TEMPS\_UNOS**[**'green\_prop\_dif'**]=**ZIP\_TEMPS\_UNOS**[**'green\_prop'**]-**ZIP\_TEMPS\_UNOS**[**'nws\_green\_prop'**]**

ZIP\_TEMPS\_UNOS**[**'dev\_prop\_dif'**]=**ZIP\_TEMPS\_UNOS**[**'developed\_prop'**]-**ZIP\_TEMPS\_UNOS**[**'nws\_dev\_prop'**]**

ZIP\_TEMPS\_UNOS**[**'density\_norm\_dif'**]=**ZIP\_TEMPS\_UNOS**[**'density\_norm'**]-**ZIP\_TEMPS\_UNOS**[**'nws\_density\_norm'**]**

# Create correlation matrix and export for table

corr\_matrix **=** ZIP\_TEMPS\_UNOS**[[**'MEAN\_TEMP'**,**'MEAN\_HI'**,**'density'**,**'green\_prop'**,**'developed\_prop'**,** 'green\_prop\_dif'**,**'dev\_prop\_dif'**,**'density\_norm\_dif'**]].**corr**()** # Creates correlation matrix

writer **=** pd**.**ExcelWriter**(**'table2.xlsx'**,** engine**=**'xlsxwriter'**)**

corr\_matrix**.**to\_excel**(**writer**,** sheet\_name**=**'corr\_matrix'**)**

writer**.**close**()**

# Create Histograms

distinct\_zips **=** ZIP\_TEMPS\_UNOS**[**'Zip'**].**drop\_duplicates**().**copy**()**

data\_zips\_dens\_landcover **=** pd**.**merge**(**zips\_dens\_landcover**,** distinct\_zips**,** on**=**'Zip'**,** how**=**'inner'**)**

sns**.**set\_style**(**"darkgrid"**)**

fig**,** axs **=** plt**.**subplots**(**2**,** 2**,** figsize**=(**14**,** 10**))**

# Green Space Histogram

axs**[**0**,** 0**].**hist**(**data\_zips\_dens\_landcover**[**'green\_prop'**],** color**=**'skyblue'**,** edgecolor**=**'black'**)**

axs**[**0**,** 0**].**set\_title**(**"Green Space Histogram"**,** fontsize**=**16**)**

axs**[**0**,** 0**].**set\_xlabel**(**'Green Space Proportion'**,** fontsize**=**14**)**

axs**[**0**,** 0**].**set\_ylabel**(**'Number of Zip Codes'**,** fontsize**=**14**)**

# Population Density Histogram

axs**[**0**,** 1**].**hist**(**data\_zips\_dens\_landcover**[**'density'**],** color**=**'skyblue'**,** edgecolor**=**'black'**)**

axs**[**0**,** 1**].**set\_title**(**'Population Density Histogram'**,** fontsize**=**16**)**

axs**[**0**,** 1**].**set\_xlabel**(**'Population Density'**,** fontsize**=**14**)**

axs**[**0**,** 1**].**set\_ylabel**(**'Number of Zip Codes'**,** fontsize**=**14**)**

axs**[**0**,** 1**].**set\_yscale**(**'log'**)**

# Full Year Heat Index Histogram

axs**[**1**,** 0**].**hist**(**ZIP\_TEMPS\_UNOS**[**'MEAN\_HI'**],** bins**=**np**.**arange**(-**45**,**125**,**5**),** color**=**'skyblue'**,** edgecolor**=**'black'**)**

axs**[**1**,** 0**].**set\_title**(**'Full Year Heat Index Histogram'**,** fontsize**=**16**)**

axs**[**1**,** 0**].**set\_xlabel**(**'Heat Index'**,** fontsize**=**14**)**

axs**[**1**,** 0**].**set\_ylabel**(**'Number of Days'**,** fontsize**=**14**)**

# Summer Months Heat Index Histogram

ZIP\_TEMPS\_UNOS\_SUMMER **=** ZIP\_TEMPS\_UNOS**[**ZIP\_TEMPS\_UNOS**[**'DATE'**].**dt**.**month**.**isin**([**6**,** 7**,** 8**])]**

axs**[**1**,** 1**].**hist**(**ZIP\_TEMPS\_UNOS\_SUMMER**[**'MEAN\_HI'**],** bins**=**np**.**arange**(**0**,**125**,**5**),** color**=**'skyblue'**,** edgecolor**=**'black'**)**

axs**[**1**,** 1**].**set\_title**(**'Summer Months Heat Index Histogram'**,** fontsize**=**16**)**

axs**[**1**,** 1**].**set\_xlabel**(**'Heat Index'**,** fontsize**=**14**)**

axs**[**1**,** 1**].**set\_ylabel**(**'Number of Days'**,** fontsize**=**14**)**

plt**.**tight\_layout**()**

# Save the figure

plt**.**savefig**(**'histogram\_plots.png'**,** dpi**=**300**)**

plt**.**show**()**

downsampled\_data2 **=** ZIP\_TEMPS\_UNOS**.**sample**(**frac**=**0.06**)**

downsampled\_data2**.**to\_csv**(**'0505\_6pct\_downsample'**)**

downsampled\_data **=** pd**.**read\_csv**(**'0505\_6pct\_downsample'**)**

# Full Year, Remove Outliers

downsampled\_data\_limited\_domain **=** downsampled\_data**[(**downsampled\_data**[**'MEAN\_HI'**]** **>=** **-**10**)** **&** **(**downsampled\_data**[**'MEAN\_HI'**]** **<=** 120**)]**

**print(len(**downsampled\_data\_limited\_domain**)/len(**downsampled\_data**))** # Only 0.03% of the data removed

# Model 1 Grid Search

grid\_search**(**downsampled\_data\_limited\_domain**,** 0.2**)**

# Returns: Best params: {'l': 0.001, 'n1': 4, 'n2': 4, 'n3': 4}

# Using the full 6pct downsample for model creation

model1 **=** PoissonGAM**(**s**(**0**,** n\_splines**=**4**,** lam**=**0.001**)** **+**

s**(**1**,** n\_splines**=**4**,** lam**=**0.001**)** **+**

s**(**2**,** n\_splines**=**4**,** lam**=**0.001**)** **+**

te**(**0**,** 1**,** n\_splines**=[**4**,** 4**],** lam**=[**0.001**,** 0.001**])** **+**

te**(**0**,** 2**,** n\_splines**=[**4**,** 4**],** lam**=[**0.001**,** 0.001**]))**

model1**.**fit**(**downsampled\_data\_limited\_domain**[[**'MEAN\_HI'**,** 'density'**,** 'green\_prop'**]],** downsampled\_data\_limited\_domain**[**'death\_ct'**])**

model1**.**summary**()**

regression\_summary\_with\_var\_names**(**"""PoissonGAM

=============================================== ==========================================================

Distribution: PoissonDist Effective DoF: 15.2562

Link Function: LogLink Log Likelihood: -37631.8391

Number of Samples: 6774920 AIC: 75294.1906

AICc: 75294.1907

UBRE: 2.0098

Scale: 1.0

Pseudo R-Squared: 0.0198

==========================================================================================================

Feature Function Lambda Rank EDoF P > x Sig. Code

================================= ==================== ============ ============ ============ ============

s(0) [0.001] 4 3.9 2.87e-03 \*\*

s(1) [0.001] 4 2.9 4.11e-13 \*\*\*

s(2) [0.001] 4 3.0 5.25e-14 \*\*\*

te(0, 1) [0.001 0.001] 16 2.8 0.00e+00 \*\*\*

te(0, 2) [0.001 0.001] 16 2.7 0.00e+00 \*\*\*

intercept 1 0.0 5.58e-06 \*\*\*

==========================================================================================================

Significance codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1"""**)**

f1r**=(-**6**,**1**)**

f2r**=(-**4**,**5**)**

f3r**=(-**5**,**5**)**

feature\_names1 **=** **[**'Mean Heat Index (F)'**,** 'Population Density (per mi^2)'**,** 'Proportion Green Space'**]**

titles1 **=** **[**'Heat Index'**,**'Population Density'**,**'Green Space'**]**

create\_plots**(**feature\_names1**,** titles1**,** model1**,**ylim**=[**f1r**,**f2r**,**f3r**],**xlim**=[(-**10**,**120**),None,None],** desc**=**'1\_0512'**)**

create\_heatmap\_plot**([**3**,**4**],** **[**'Population Density'**,**'Green Space'**]** **,**model1**,**

**[**'Heat Index, Population Density Interaction Term'**,**'Heat Index, Green Space Interaction Term'**],**

xlim**=[(-**10**,**120**),(-**10**,**120**)],**desc**=**'1\_0512'**)**

create\_additive\_heatmap\_subplots**([**0**,**0**],[**1**,**2**],[**3**,**4**],[**'Population Density'**,**'Green Space'**],**

**[**'Aggregated Population Density, Heat Index Interaction'**,**'Aggregated Green Space, Heat Index Interaction'**],**

model1**,** xlim**=[(-**10**,**120**),(-**10**,**120**)],** desc**=**'1\_0512'**)**

downsampled\_data4 **=** ZIP\_TEMPS\_UNOS**.**sample**(**frac**=**0.24**)**

downsampled\_data4 **=** downsampled\_data4**[**downsampled\_data4**[**'DATE'**].**dt**.**month**.**isin**([**6**,** 7**,** 8**])]**

downsampled\_data4**.**to\_csv**(**'downsampled\_data4.csv'**)**

downsampled\_data2 **=** pd**.**read\_csv**(**'downsampled\_data4.csv'**)**

# This includes only the summer months of a 24pct sample, so it is about 6pct of the total data, the same size as the other downsample

# Summer Months, Remove Outliers

downsampled\_data2\_limited\_domain **=** downsampled\_data2**[(**downsampled\_data2**[**'MEAN\_HI'**]** **>=** 40**)** **&** **(**downsampled\_data2**[**'MEAN\_HI'**]** **<=** 120**)]**

**print(len(**downsampled\_data2\_limited\_domain**)/len(**downsampled\_data2**))** # Only 0.19% of the data removed

grid\_search**(**downsampled\_data2\_limited\_domain**,** 0.2**)**

# Returns: Best params: {'l': 0.001, 'n1': 4, 'n2': 4, 'n3': 4}

model2 **=** PoissonGAM**(**s**(**0**,** n\_splines**=**4**,** lam**=**0.001**)** **+**

s**(**1**,** n\_splines**=**4**,** lam**=**0.001**)** **+**

s**(**2**,** n\_splines**=**4**,** lam**=**0.001**)** **+**

te**(**0**,** 1**,** n\_splines**=[**4**,** 4**],** lam**=[**0.001**,** 0.001**])** **+**

te**(**0**,** 2**,** n\_splines**=[**4**,** 4**],** lam**=[**0.001**,** 0.001**]))**

model2**.**fit**(**downsampled\_data2\_limited\_domain**[[**'MEAN\_HI'**,** 'density'**,** 'green\_prop'**]],** downsampled\_data2\_limited\_domain**[**'death\_ct'**])**

model2**.**summary**()**

regression\_summary\_with\_var\_names**(**"""PoissonGAM

=============================================== ==========================================================

Distribution: PoissonDist Effective DoF: 14.4921

Link Function: LogLink Log Likelihood: -35679.8376

Number of Samples: 6781069 AIC: 71388.6592

AICc: 71388.6593

UBRE: 2.0092

Scale: 1.0

Pseudo R-Squared: 0.0206

==========================================================================================================

Feature Function Lambda Rank EDoF P > x Sig. Code

================================= ==================== ============ ============ ============ ============

s(0) [0.001] 4 3.8 5.39e-03 \*\*

s(1) [0.001] 4 2.9 6.66e-16 \*\*\*

s(2) [0.001] 4 3.0 2.96e-07 \*\*\*

te(0, 1) [0.001 0.001] 16 2.5 0.00e+00 \*\*\*

te(0, 2) [0.001 0.001] 16 2.3 0.00e+00 \*\*\*

intercept 1 0.0 6.23e-05 \*\*\*

==========================================================================================================

Significance codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1"""**)**

create\_plots**(**feature\_names1**,** titles1**,** model2**,**ylim**=[**f1r**,**f2r**,**f3r**],**xlim**=[(**40**,**120**),None,None],** desc**=**'2\_0512'**)**

create\_heatmap\_plot**([**3**,**4**],** **[**'Population Density'**,**'Green Space'**]** **,**model2**,[**'Heat Index, Population Density Interaction Term'**,**'Heat Index, Green Space Interaction Term'**],**

xlim**=[(**40**,**120**),(**40**,**120**)],**desc**=**'2\_0512'**)**

create\_additive\_heatmap\_subplots**([**0**,**0**],[**1**,**2**],[**3**,**4**],[**'Population Density'**,**'Green Space'**],**

**[**'Aggregated Population Density, Heat Index Interaction'**,**'Aggregated Green Space, Heat Index Interaction'**],**model2**,**

xlim**=[(**40**,**120**),(**40**,**120**)],**desc**=**'2\_0512'**)**

Notes:

* Some code for loading data here is repeated from other sections, so that they can be run as discrete sections.
* It was necessary to use a random subset of the data for modeling because creating GAM models is extremely memory intensive and the virtual machine with 52 GB of RAM used for running this program would crash when trying to use the full data. A smaller portion of the data still was used for the grid search as at necessitates fitting a model to the data 2240 times.
* The regression summaries appear in the code because model.summary() prints the output without variable names, and also returns ‘None,’ so the variable names are added manually after the fact.

**Appendix H: Additional model attempting to control for the impact of the UHI on NWS measurements.**

**Note:** This model was less successful than the other models in the paper. However, I felt the need to include it in the appendices because without this model, much of the code used for processing the data (particularly in Appendix E and F) would seem superfluous because it would not be necessary to keep track of the population density and green space in the zip codes near NWS monitors. Additionally, several columns of the correlation matrix (Table 2) are used only in this model, and not in the body of the paper.

In many cases in our data, the nearest NWS station to a zip code is relatively distant. Therefore we were concerned that using the heat index as measured at the nearest NWS station, and population density and green space in the location where a death occurs could obfuscate the relationship between these variables and mortality (the response variable). Ideally, we would like to adjust measurements of heat index to reflect the true value in the zip code where a death occurs. However, there is no method to do this reliably. Therefore, we instead adjusted for the amount of population density and the green space around the NWS station based on the (rather bold) assumption that these factors indicate the extent to which the NWS station’s measurement was impacted by the UHI affect, and therefore indicate the difference in temperature between the NWS station and the zip code where the deaths occurred. To be more concrete, if a measurement was taken in a high density area and a death occurs in a low density area, it is likely that the measurement was affected by the UHI effect more than the individual who died. Therefore, using the difference between these values at the death location and station location may allow the model to be more accurate. Based on this logic, another model (referred to here as model 3) was constructed to fit the following equation:

ln(Expected Mortality) = f₁(Heat Index) +

f₂(Population Density at location of Mortality – Population Density at NWS station) +

f₃(Green Space at location of Mortality – Green Space at location of NWS station) +

f₄(Heat Index, Population Density at location of Mortality – Population Density at NWS station) +

f₅(Heat Index, Green Space at location of Mortality – Green Space at location of NWS station) + c

**Equation 2:** The equation used to model temperature in model 3. The model learns the functions f1 through f5 from the data and also adds an intercept term c. As it is a Poisson model, we predict the natural logarithm of Expected Mortality. In this model, the predictor variables of Population Density and Green Space at the location where mortality is observed are adjusted for their values at the NWS station which took the Heat Index measurement.

The regression output for model 3 is shown in Figure 11. Note the significantly lower pseudo R-squared. The model is able to fit a much smaller portion of the variation present in the data. Essentially, the attempt to control for conditions around NWS stations can be seen as adding unnecessary noise to the population density and green space predictor variables. As a result, the relationships between the predictors and expected mortality are consistently more ambiguous than in models included in the main body of the paper. These results show that subtracting the population density and green space in the area around NWS stations from the values at another location of interest is not an effective way to control for the impact of the UHI effect on NWS measurements.

A screenshot of a computer screen

Description automatically generated with low confidence

**Figure 11:** Regression output of model 3. This is the result of fitting the following equation:

ln(Expected Mortality) = f₁(Heat Index) + f₂(Population Density at location of Mortality – Population Density at NWS station) + f₃(Green Space at location of Mortality – Green Space at location of NWS station) + f₄(Heat Index, Population Density at location of Mortality – Population Density at NWS station) + f₅(Heat Index, Green Space at location of Mortality – Green Space at location of NWS station) + c

A known bug in the PyGAM library causes p-values to appear lower than they should. The significance of variables can better be assessed by examining the partial dependence plots and their confidence intervals. Note a pseudo R-squared of 0.0045 which is substantially lower than in the models in the main body of the paper and is likely because the attempt to control for conditions at the NWS stations taking measurements introduced unnecessary noise into the data.

For the sake of consistency, all the same types of plots are reported for model 3 as for the models in the main body of the paper. The partial dependence plots for individual predictors in model 3 are included in Figure 12. The partial dependence plots for the interaction terms are included in Figure 13. Finally, the aggregated plots showing the partial dependence of expected mortality on heat index in conjunction with each population density and green space are shown in Figure 14. Despite the worse fit overall, we still observe increasing expected mortality with increasing temperature, increasing expected mortality with increasing population density in the zip code where a death occurred (relative to the area of the nearest NWS station), and decreasing expected mortality with increasing greenspace in the zip code where a death occurred (relative to the area of the nearest NWS station) in Figure 12. Additionally, the aggregated plot for population density in Figure 14 appear to show that zip codes with the highest population density relative to the NWS stations which took their heat index measurements have greater expected mortality and that this is most acute at high heat indices. Finally, the aggregated plot for green space in Figure 14 shows a similar relationship to the version of the plot generated by model 1 (in Figure 6). Altogether, the models in the body of the paper and their associated plots are still more suitable for making inferences given their substantially higher pseudo R-squared.

A picture containing plot, text, line, diagram

Description automatically generated

**Figure 12:** Plots of the partial dependence of expected mortality on the three individual terms in model 3, in which we attempt to control for the conditions around the NWS stations which take the measurements of Heat Index. From left to right these show f₁(Heat Index), f₂( Population Density at location of Mortality – Population Density at NWS station), f₃( Green Space at location of Mortality – Green Space at location of NWS station). Recall that this Poisson model fits the following relationship:

ln(Expected Mortality) = f₁(Heat Index) + f₂(Population Density at location of Mortality – Population Density at NWS station) + f₃(Green Space at location of Mortality – Green Space at location of NWS station) + f₄(Heat Index, Population Density at location of Mortality – Population Density at NWS station) + f₅(Heat Index, Green Space at location of Mortality – Green Space at location of NWS station) + c

A picture containing text, screenshot, colorfulness, diagram

Description automatically generated

**Figure 13:** Plots of the partial dependence of expected mortality on the two interaction terms in model 3, in which we attempt to control for the conditions around the NWS stations which take the measurements of Heat Index. From left to right these show f₄(Heat Index, Population Density at location of Mortality – Population Density at NWS station) and f₅(Heat Index, Green Space at location of Mortality – Green Space at location of NWS station). Contours are plotted for integer values, indicating that movement from one contour to another represents a change in expected mortality by a factor of *e*. Recall that this Poisson model fits the following relationship:

ln(Expected Mortality) = f₁(Heat Index) + f₂(Population Density at location of Mortality – Population Density at NWS station) + f₃(Green Space at location of Mortality – Green Space at location of NWS station) + f₄(Heat Index, Population Density at location of Mortality – Population Density at NWS station) + f₅(Heat Index, Green Space at location of Mortality – Green Space at location of NWS station) + c

A picture containing text, screenshot, colorfulness, plot

Description automatically generated

**Figure 14:** Left – A plot of the partial dependence of expected mortality on the interaction heat index and population density using all relevant terms: f₁(Heat Index) + f₂(Population Density at location of Mortality – Population Density at NWS station) + f₄(Heat Index, Population Density at location of Mortality – Population Density at NWS station).

Right – A plot of the partial dependence of expected mortality on the interaction heat index and green space using all relevant terms: f₁(Heat Index) + f₃(Green Space at location of Mortality – Green Space at location of NWS station) + f₅(Heat Index, Green Space at location of Mortality – Green Space at location of NWS station).

Contours are plotted for multiples of two, so movement from one contour to another represents a change in expected mortality by a factor of *e*2. Recall that the Poisson model fits the following relationship:

ln(Expected Mortality) = f₁(Heat Index) + f₂(Population Density at location of Mortality – Population Density at NWS station) + f₃(Green Space at location of Mortality – Green Space at location of NWS station) + f₄(Heat Index, Population Density at location of Mortality – Population Density at NWS station) + f₅(Heat Index, Green Space at location of Mortality – Green Space at location of NWS station) + c

**Notes**

1. World Health Organization. “Climate Change and Health.” <https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health> [↑](#endnote-ref-1)
2. World Health Organization. Climate change and health. <https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health> [↑](#endnote-ref-2)
3. Jennifer F. Bobb et al., “Cause-Specific Risk of Hospital Admission Related to Extreme Heat in Older Adults,” *JAMA* 312, no. 24 (2014): 2659, <https://doi.org/10.1001/jama.2014.15715>.

   Samantha L. Kingsley et al., “Current and Projected Heat-Related Morbidity and Mortality in Rhode Island,” *Environmental Health Perspectives* 124, no. 4 (2016): 460–67, <https://doi.org/10.1289/ehp.1408826>. [↑](#endnote-ref-3)
4. World Health Organization. 2014.*Quantitative Risk Assessment of the Effects of Climate Change on Selected Causes of Death, 2030s and 2050s* World Health Organization. <http://www.who.int/iris/handle/10665/134014.> [↑](#endnote-ref-4)
5. Cissé, G., R. McLeman, H. Adams, P. Aldunce, K. Bowen, D. Campbell-Lendrum, S. Clayton, K.L. Ebi, J. Hess, C. Huang, Q. Liu, G. McGregor, J. Semenza, and M.C. Tirado, 2022: Health, Wellbeing, and the Changing Structure of Communities. In:*Climate Change 2022: Impacts, Adaptation, and Vulnerability.*Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 1041-1170, doi:10.1017/9781009325844.009. [↑](#endnote-ref-5)
6. Cissé, G., R. McLeman, H. Adams, P. Aldunce, K. Bowen, D. Campbell-Lendrum, S. Clayton, K.L. Ebi, J. Hess, C. Huang, Q. Liu, G. McGregor, J. Semenza, and M.C. Tirado, 2022: Health, Wellbeing, and the Changing Structure of Communities. In:*Climate Change 2022: Impacts, Adaptation, and Vulnerability.* pp. 1041-1170, [↑](#endnote-ref-6)
7. Antonio Gasparrini et al., “Mortality Risk Attributable to High and Low Ambient Temperature: A Multicountry Observational Study,” *The Lancet* 386, no. 9991 (2015): 369–75, https://doi.org/10.1016/s0140-6736(14)62114-0. [↑](#endnote-ref-7)
8. Krishnan Bhaskaran et al., “Time Series Regression Studies in Environmental Epidemiology,” *International Journal of Epidemiology* 42, no. 4 (2013): 1187–95, <https://doi.org/10.1093/ije/dyt092>. [↑](#endnote-ref-8)
9. Ekta Choudhary and Ambarish Vaidyanathan. 2014. "Heat Stress Illness Hospitalizations — Environmental Public Health Tracking Program, 20 States, 2001–2010." *MMWR. Surveillance Summaries* 63 (13): 1-10. <https://www.jstor.org/stable/24806265.>

   Robert E. Davis, David M. Hondula, and Anjali P. Patel, “Temperature Observation Time and Type Influence Estimates of Heat-Related Mortality in Seven U.S. Cities,” *Environmental Health Perspectives* 124, no. 6 (2016): 795–804, <https://doi.org/10.1289/ehp.1509946>.

   Antonio Gasparrini et al., “Temporal Variation in Heat–Mortality Associations: A Multicountry Study,” *Environmental Health Perspectives* 123, no. 11 (2015): 1200–1207, <https://doi.org/10.1289/ehp.1409070.doi:10.1289/ehp.1409070>.

   Carina J. Gronlund et al., “Vulnerability to Renal, Heat and Respiratory Hospitalizations during Extreme Heat among U.S. Elderly,” *Climatic Change* 136, no. 3–4 (2016): 631–45, <https://doi.org/10.1007/s10584-016-1638-9>.

   Katharine Hayhoe et al., “Climate Change, Heat Waves, and Mortality Projections for Chicago,” *Journal of Great Lakes Research* 36 (2010): 65–73, <https://doi.org/10.1016/j.jglr.2009.12.009>.

   Samantha L. Kingsley et al., *Environmental Health Perspectives* (2016)

   Whanhee Lee et al., “Projections of Diurnal Temperature Range and Diurnal Temperature Range-Related Excess Mortality under Climate Change,” *Environmental Epidemiology* 3 (2019): 226–27, <https://doi.org/10.1097/01.ee9.0000608348.49159.75>.

   Tiantian Li, Radley M. Horton, and Patrick L. Kinney, “Projections of Seasonal Patterns in Temperature- Related Deaths for Manhattan, New York,” *Nature Climate Change* 3, no. 8 (2013): 717–21, <https://doi.org/10.1038/nclimate1902>.

   Kristina B. Metzger, Kazuhiko Ito, and Thomas D. Matte, “Summer Heat and Mortality in New York City: How Hot Is Too Hot?,” *Environmental Health Perspectives* 118, no. 1 (2010): 80–86, <https://doi.org/10.1289/ehp.0900906>.

   Leah H. Schinasi et al., “High Ambient Temperature and Infant Mortality in Philadelphia, Pennsylvania: A Case–Crossover Study,” *American Journal of Public Health* 110, no. 2 (2020): 189–95, <https://doi.org/10.2105/ajph.2019.305442>.

   Kate R. Weinberger et al., “Heat Warnings, Mortality, and Hospital Admissions among Older Adults in the United States,” *Environment International* 157 (2021): 106834, <https://doi.org/10.1016/j.envint.2021.106834>.

   Gregory A. Wellenius et al., “Heat-Related Morbidity and Mortality in New England: Evidence for Local Policy,” *Environmental Research* 156 (2017): 845–53, <https://doi.org/10.1016/j.envres.2017.02.005>.  [↑](#endnote-ref-9)
10. Rupa Basu, Wen-Ying Feng, and Bart D. Ostro, “Characterizing Temperature and Mortality in Nine California Counties,” *Epidemiology* 19, no. 1 (2008): 138–45, <https://doi.org/10.1097/ede.0b013e31815c1da7>.

    Rupa Basu et al., “The Effect of High Ambient Temperature on Emergency Room Visits,” *Epidemiology* 23, no. 6 (2012): 813–20, <https://doi.org/10.1097/ede.0b013e31826b7f97>.

    Jennifer F. Bobb et al., *JAMA* (2014): 2659-2667

    Robert E. Davis, David M. Hondula, and Anjali P. Patel, *Environmental Health Perspectives* 124, (2016): 795–804.

    Antonio Gasparrini et al., *The Lancet* (2015): 369–75

    Antonio Gasparrini et al., *Environmental Health Perspectives* 123, no. 11 (2015): 1200–1207.

    Whanhee Lee et al., *Environmental Epidemiology* 3 (2019): 226–27.

    Kristina B. Metzger, Kazuhiko Ito, and Thomas D. Matte, *Environmental Health Perspectives* 118, (2010): 80–86.

    Elisaveta Petkova et al., “Projected Heat-Related Mortality in the U.S. Urban Northeast,” *International Journal of Environmental Research and Public Health* 10, no. 12 (2013): 6734–47, <https://doi.org/10.3390/ijerph10126734>. 1. Elisaveta P. Petkova, Antonio Gasparrini, and Patrick L. Kinney, “Heat and Mortality in New York City since the Beginning of the 20th Century,” *Epidemiology* 25, no. 4 (2014): 554–60, <https://doi.org/10.1097/ede.0000000000000123>.

    Elisaveta P. Petkova et al., “Towards More Comprehensive Projections of Urban Heat-Related Mortality: Estimates for New York City under Multiple Population, Adaptation, and Climate Scenarios,” *Environmental Health Perspectives* 125, no. 1 (2017): 47–55, <https://doi.org/10.1289/ehp166>.

    Reija Ruuhela, Otto Hyvärinen, and Kirsti Jylhä, “Regional Assessment of Temperature-Related Mortality in Finland,” *International Journal of Environmental Research and Public Health* 15, no. 3 (2018): 406, <https://doi.org/10.3390/ijerph15030406>.

    Leah H. Schinasi et al., *American Journal of Public Health* 110, no. 2 (2020): 189–95.

    Joel D. Schwartz et al., “Projections of Temperature-Attributable Premature Deaths in 209 U.S. Cities Using a Cluster-Based Poisson Approach,” *Environmental Health* 14, no. 1 (2015), <https://doi.org/10.1186/s12940-015-0071-2>.

    Yan Wang et al., “Heat Stroke Admissions during Heat Waves in 1,916 US Counties for the Period from 1999 to 2010 and Their Effect Modifiers,” *Environmental Health* 15, no. 1 (2016), <https://doi.org/10.1186/s12940-016-0167-3>. [↑](#endnote-ref-10)
11. Robert E. Davis, David M. Hondula, and Anjali P. Patel, *Environmental Health Perspectives* 124, no. 6 (2016): 795–804.

    Antonio Gasparrini et al., *Environmental Health Perspectives* 123, no. 11 (2015): 1200–1207.

    Katharine Hayhoe et al., *Journal of Great Lakes Research* 36 (2010): 65–73.

    Whanhee Lee et al., *Environmental Epidemiology* 3 (2019): 226–27.

    Kristina B. Metzger, Kazuhiko Ito, and Thomas D. Matte, *Environmental Health Perspectives* 118, (2010): 80–86.

    Leah H. Schinasi et al., *American Journal of Public Health* 110, no. 2 (2020): 189–95. [↑](#endnote-ref-11)
12. Kristie L. Ebi et al., “Heat Watch/Warning Systems Save Lives: Estimated Costs and Benefits for Philadelphia 1995–98,” *Bulletin of the American Meteorological Society* 85, no. 8 (2004): 1067–74, <https://doi.org/10.1175/bams-85-8-1067>.

    Katharine Hayhoe et al., *Journal of Great Lakes Research* 36 (2010): 65–73.

    Samantha L. Kingsley et al., *Environmental Health Perspectives* (2016)

    Yan Wang et al., *Environmental Health* 15, no. 1 (2016). [↑](#endnote-ref-12)
13. Kate R. Weinberger et al., *Environment International* 157 (2021): 106834. [↑](#endnote-ref-13)
14. Rupa Basu, Wen-Ying Feng, and Bart D. Ostro, *Epidemiology* 19, no. 1 (2008): 138–45.

    Rupa Basu et al., *Epidemiology* 23, no. 6 (2012): 813–20.

    Antonio Gasparrini et al., *The Lancet* (2015): 369–75

    Thomas D. Matte, Kathryn Lane, and Kazuhiko Ito, “Excess Mortality Attributable to Extreme Heat in New York City, 1997-2013,” *Health Security* 14, no. 2 (2016): 64–70, <https://doi.org/10.1089/hs.2015.0059>

    Leah H. Schinasi et al., *American Journal of Public Health* 110, no. 2 (2020): 189–95.

    Yan Wang et al., *Environmental Health* 15, no. 1 (2016).

    Gregory A. Wellenius et al., *Environmental Research* 156 (2017): 845–53. [↑](#endnote-ref-14)
15. Equations and sources Listed in Appendix B [↑](#endnote-ref-15)
16. Rupa Basu, Wen-Ying Feng, and Bart D. Ostro, *Epidemiology* 19, no. 1 (2008): 138–45.

    Rupa Basu et al., *Epidemiology* 23, no. 6 (2012): 813–20.

    Ekta Choudhary and Ambarish Vaidyanathan. 2014. *MMWR. Surveillance Summaries* 63 (13): 1-10.

    Kristina B. Metzger, Kazuhiko Ito, and Thomas D. Matte, *Environmental Health Perspectives* 118, (2010): 80–86.

    Leah H. Schinasi et al., *American Journal of Public Health* 110, no. 2 (2020): 189–95.

    Kate R. Weinberger et al., *Environment International* 157 (2021): 106834.

    Gregory A. Wellenius et al., *Environmental Research* 156 (2017): 845–53. [↑](#endnote-ref-16)
17. Jennifer F. Bobb et al., *JAMA* (2014): 2659-2667

    Yan Wang et al., *Environmental Health* 15, no. 1 (2016). [↑](#endnote-ref-17)
18. Carina J. Gronlund et al., *Climatic Change* 136, no. 3–4 (2016): 631–45. [↑](#endnote-ref-18)
19. Thomas D. Matte, Kathryn Lane, and Kazuhiko Ito, *Health Security* 14, no. 2 (2016): 64–70. [↑](#endnote-ref-19)
20. Katharine Hayhoe et al., *Journal of Great Lakes Research* 36 (2010): 65–73 [↑](#endnote-ref-20)
21. Rupa Basu, Wen-Ying Feng, and Bart D. Ostro, *Epidemiology* 19, no. 1 (2008): 138–45.

    Rupa Basu et al., *Epidemiology* 23, no. 6 (2012): 813–20.

    Jennifer F. Bobb et al., *JAMA* (2014): 2659-2667

    Robert E. Davis, David M. Hondula, and Anjali P. Patel, *Environmental Health Perspectives* 124, no. 6 (2016): 795–804.

    Kristie L. Ebi et al., *Bulletin of the American Meteorological Society* 85, no. 8 (2004): 1067–74.

    Antonio Gasparrini et al., *The Lancet* (2015): 369–75.

    Antonio Gasparrini et al., *Environmental Health Perspectives* 123, no. 11 (2015): 1200–1207.

    Carina J. Gronlund et al., *Climatic Change* 136, no. 3–4 (2016): 631–45.

    Whanhee Lee et al., *Environmental Epidemiology* 3 (2019): 226–27.

    Tiantian Li, Radley M. Horton, and Patrick L. Kinney, *Nature Climate Change* 3, no. 8 (2013): 717–21.

    Thomas D. Matte, Kathryn Lane, and Kazuhiko Ito, *Health Security* 14, no. 2 (2016): 64–70.

    Kristina B. Metzger, Kazuhiko Ito, and Thomas D. Matte, *Environmental Health Perspectives* 118, (2010): 80–86.

    Elisaveta Petkova et al., *International Journal of Environmental Research and Public Health* 10, no. 12 (2013): 6734–47.

    Elisaveta P. Petkova, Antonio Gasparrini, and Patrick L. Kinney, *Epidemiology* 25, no. 4 (2014): 554–60.

    Elisaveta P. Petkova et al., “Towards More Comprehensive Projections of Urban Heat-Related Mortality: Estimates for New York City under Multiple Population, Adaptation, and Climate Scenarios,” *Environmental Health Perspectives* 125, no. 1 (2017): 47–55, <https://doi.org/10.1289/ehp166>.

    Reija Ruuhela, Otto Hyvärinen, and Kirsti Jylhä, *International Journal of Environmental Research and Public Health* 15, no. 3 (2018): 406.

    Leah H. Schinasi et al., *American Journal of Public Health* 110, no. 2 (2020): 189–95.

    Leah H. Schinasi et al., *American Journal of Public Health* 110, no. 2 (2020): 189–95.

    Joel D. Schwartz et al., *Environmental Health* 14, no. 1 (2015).

    Kate R. Weinberger et al., *Environment International* 157 (2021): 106834.

    Gregory A. Wellenius et al., *Environmental Research* 156 (2017): 845–53. [↑](#endnote-ref-21)
22. Kristina B. Metzger, Kazuhiko Ito, and Thomas D. Matte, *Environmental Health Perspectives* 118, (2010): 80–86.

    Gregory A. Wellenius et al., *Environmental Research* 156 (2017): 845–53. [↑](#endnote-ref-22)
23. Antonio Gasparrini et al., *The Lancet* (2015): 369–75. [↑](#endnote-ref-23)
24. Colleen E. Reid et al., “The Role of Ambient Ozone in Epidemiologic Studies of Heat-Related Mortality,” *Environmental Health Perspectives* 120, no. 12 (2012): 1627–30, <https://doi.org/10.1289/ehp.1205251>. [↑](#endnote-ref-24)
25. Tiantian Li, Radley M. Horton, and Patrick L. Kinney, *Nature Climate Change* 3, no. 8 (2013): 717–21. [↑](#endnote-ref-25)
26. Katharine Hayhoe et al., *Journal of Great Lakes Research* 36 (2010): 65–73

    Elisaveta Petkova et al., *International Journal of Environmental Research and Public Health* 10, no. 12 (2013): 6734–47. [↑](#endnote-ref-26)
27. Antonio Gasparrini et al., *The Lancet* (2015): 369–75

    Kristina B. Metzger, Kazuhiko Ito, and Thomas D. Matte, *Environmental Health Perspectives* 118, (2010): 80–86.

    Joel D. Schwartz et al., *Environmental Health* 14, no. 1 (2015). [↑](#endnote-ref-27)
28. Samantha L. Kingsley et al., *Environmental Health Perspectives* (2016) [↑](#endnote-ref-28)
29. Jennifer F. Bobb et al., *JAMA* (2014): 2659-2667

    Kate R. Weinberger et al., *Environment International* 157 (2021): 106834. [↑](#endnote-ref-29)
30. Kristie L. Ebi et al., *Bulletin of the American Meteorological Society* 85, no. 8 (2004): 1067–74. [↑](#endnote-ref-30)
31. Li Yang et al., “Research on Urban Heat-Island Effect,” *Procedia Engineering* 169 (2016): pp. 11-18, <https://doi.org/10.1016/j.proeng.2016.10.002>. [↑](#endnote-ref-31)
32. Abbas Mohajerani, Jason Bakaric, and Tristan Jeffrey-Bailey, “The Urban Heat Island Effect, Its Causes, and Mitigation, with Reference to the Thermal Properties of Asphalt Concrete,” *Journal of Environmental Management* 197 (April 14, 2017): pp. 522-538, <https://doi.org/10.1016/j.jenvman.2017.03.095>. [↑](#endnote-ref-32)
33. Cissé, G., R. McLeman, H. Adams, P. Aldunce, K. Bowen, D. Campbell-Lendrum, S. Clayton, K.L. Ebi, J. Hess, C. Huang, Q. Liu, G. McGregor, J. Semenza, and M.C. Tirado, 2022: Health, Wellbeing, and the Changing Structure of Communities. In:*Climate Change 2022: Impacts, Adaptation, and Vulnerability.* pp. 1041-1170

    Tran Ngoc Dang et al., “Green Space and Deaths Attributable to the Urban Heat Island Effect in Ho Chi Minh City,” *American Journal of Public Health* 108, no. S2 (2018): pp. 137-143, https://doi.org/10.2105/ajph.2017.304123.,

    Wouters, Hendrik, Koen De ridder, Lien Poelmans, Patrick Willems, Johan Brouwers, Parisa Hosseinzadehtalaei, Hossein Tabari, Sam Vanden broucke, Nicole P. M. Van lipzig, and Matthias Demuzere. 2017.*Heat Stress Increase Under Climate Change Twice as Large in Cities as in Rural Areas: A Study for a Densely Populated Midlatitude Maritime Region*. Vol. 44 American Geophysical Union (AGU). doi:10.1002/2017gl074889,

    Lei Zhao et al., “Global Multi-Model Projections of Local Urban Climates,” *Nature Climate Change* 11, no. 2 (2021): 152–57, <https://doi.org/10.1038/s41558-020-00958-8>. [↑](#endnote-ref-33)
34. Reija Ruuhela et al., “Temperature-Related Mortality in Helsinki Compared to Its Surrounding Region over Two Decades, with Special Emphasis on Intensive Heatwaves,” *Atmosphere* 12, no. 1 (2020): 46, <https://doi.org/10.3390/atmos12010046>. [↑](#endnote-ref-34)
35. Jay A. Fishman, “Infection in Solid-Organ Transplant Recipients,” *New England Journal of Medicine* 357, no. 25 (2007): pp. 2601-2614, <https://doi.org/10.1056/nejmra064928>. [↑](#endnote-ref-35)
36. Jia Liu et al., “The Risk Factors for Tuberculosis in Liver or Kidney Transplant Recipients,” *BMC Infectious Diseases* 14, no. 1 (November 2014), <https://doi.org/10.1186/1471-2334-14-387>. [↑](#endnote-ref-36)
37. Eric A Engles et al., “Spectrum of Cancer Risk among US Solid Organ Transplant Recipients,” *JAMA: Journal of the American Medical Association* 2012 (2012): pp. 105-106, <https://doi.org/10.1016/j.ysur.2012.05.001>. [↑](#endnote-ref-37)
38. Therese Bittermann et al., “The Proportion of Model for End‐stage Liver Disease Sodium Score Attributable to Creatinine Independently Predicts Post‐transplant Survival and Renal Complications,” *Clinical Transplantation* 34, no. 3 (2020), <https://doi.org/10.1111/ctr.13817>. [↑](#endnote-ref-38)
39. Therese Bittermann, George Makar, and David S. Goldberg, “Early Post-Transplant Survival: Interaction of MELD Score and Hospitalization Status,” *Journal of Hepatology* 63, no. 3 (2015): 601–8, <https://doi.org/10.1016/j.jhep.2015.03.034>. [↑](#endnote-ref-39)
40. “Data Access Integrated Surface Dataset (Global),” National Centers for Environmental Information (NCEI) (National Oceanic and Atmospheric Administration (NOAA), accessed May 6, 2023, https://www.ncei.noaa.gov/access/search/data-search/global-hourly. [↑](#endnote-ref-40)
41. Jon Bittner, “Free US Population Density and Unemployment Rate by ZIP Code,” The Splitwise Blog (Splitwise, January 6, 2014), <https://blog.splitwise.com/2014/01/06/free-us-population-density-and-unemployment-rate-by-zip-code/>. [↑](#endnote-ref-41)
42. “NLCD 2016 Land Cover (CONUS),” NLCD 2016 Land Cover (CONUS) | Multi-Resolution Land Characteristics (MRLC) Consortium (Multi-Resolution Land Characteristics (MRLC) Consortium), accessed May 3, 2023, <https://www.mrlc.gov/data/nlcd-2016-land-cover-conus>. [↑](#endnote-ref-42)
43. “Tiger/Line Shapefile, 2019, 2010 Nation, U.S., 2010 Census 5-Digit ZIP Code Tabulation Area (ZCTA5) National,” DATA.GOV (U.S. Census Bureau, Department of Commerce, November 1, 2022), <https://catalog.data.gov/dataset/tiger-line-shapefile-2019-2010-nation-u-s-2010-census-5-digit-zip-code-tabulation-area-zcta5-na>. [↑](#endnote-ref-43)
44. Gareth James et al., *An Introduction to Statistical Learning: With Applications in R* (Boston, MA: Springer, 2022). pp. 203-205 [↑](#endnote-ref-44)
45. “Sklearn.model\_selection.GridSearchCV,” scikit-learn (scikit-learn), accessed May 3, 2023, <https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html>. [↑](#endnote-ref-45)
46. Antonio Gasparrini et al., *The Lancet* (2015): 369–75 [↑](#endnote-ref-46)
47. Antonio Gasparrini et al., *The Lancet* (2015): 369–75 [↑](#endnote-ref-47)
48. Dajun Dai, “Racial/Ethnic and Socioeconomic Disparities in Urban Green Space Accessibility: Where to Intervene?,” *Landscape and Urban Planning* 102, no. 4 (2011): pp. 234-244, <https://doi.org/10.1016/j.landurbplan.2011.05.002>. [↑](#endnote-ref-48)
49. Kristie L. Ebi et al., *Bulletin of the American Meteorological Society* 85, no. 8 (2004): 1067–74.

    Kate R. Weinberger et al., “Effectiveness of National Weather Service Heat Alerts in Preventing Mortality in 20 US Cities,” *Environment International* 116 (2018): 30–38, <https://doi.org/10.1016/j.envint.2018.03.028>. [↑](#endnote-ref-49)
50. “Temperature, Dewpoint, and Relative Humidity Calculator,” Calculate Temperature, Dewpoint, or Relative Humidity (University of Miami), accessed May 3, 2023, <https://bmcnoldy.rsmas.miami.edu/Humidity.html>. [↑](#endnote-ref-50)
51. “The Heat Index Equation,” Heat index equation (NOAA/National Weather Service, National Centers for Environmental Prediction, Weather Prediction Center, May 12, 2022), <https://www.wpc.ncep.noaa.gov/html/heatindex_equation.shtml>. [↑](#endnote-ref-51)
52. “Federal Climate Complex Data Documentation for Integrated Surface Data (ISD),” National Centers for Environmental Information (NCEI) (NOAA - National Centers for Environmental Information, January 12, 2018), 11 <https://www.ncei.noaa.gov/data/global-hourly/doc/isd-format-document.pdf>.

    **Bibliography**

    Basu, Rupa, Wen-Ying Feng, and Bart Ostro. 2008. "Characterizing Temperature and Mortality in Nine California Counties." *Epidemiology (Cambridge, Mass.)* 19 (1): 138-145. doi:10.1097/EDE.0b013e31815c1da7. <https://www.jstor.org/stable/20486505.>

    Basu, Rupa, Dharshani Pearson, Brian Malig, Rachel Broadwin, and Rochelle Green. 2012. "The Effect of High Ambient Temperature on Emergency Room Visits." *Epidemiology (Cambridge, Mass.)* 23 (6): 813-820. doi:10.1097/EDE.0b013e31826b7f97. <https://www.jstor.org/stable/41739678.>

    Bhaskaran, Krishnan, Antonio Gasparrini, Shakoor Hajat, Liam Smeeth, and Ben Armstrong. 2013. "Time Series Regression Studies in Environmental Epidemiology." *International Journal of Epidemiology* 42 (4): 1187-1195. doi:10.1093/ije/dyt092. <https://www.ncbi.nlm.nih.gov/pubmed/23760528.>

    Bittermann, Therese, George Makar, and David S. Goldberg. “Early Post-Transplant Survival: Interaction of MELD Score and Hospitalization Status.” *Journal of Hepatology* 63, no. 3 (2015): 601–8. <https://doi.org/10.1016/j.jhep.2015.03.034>.

    Bittermann, Therese, Rebecca A. Hubbard, James D. Lewis, and David S. Goldberg. “The Proportion of Model for End‐stage Liver Disease Sodium Score Attributable to Creatinine Independently Predicts Post‐transplant Survival and Renal Complications.” *Clinical Transplantation* 34, no. 3 (2020). <https://doi.org/10.1111/ctr.13817>.

    Bittner, Jon. “Free US Population Density and Unemployment Rate by ZIP Code.” The Splitwise Blog. Splitwise, January 6, 2014. <https://blog.splitwise.com/2014/01/06/free-us-population-density-and-unemployment-rate-by-zip-code/>.

    Bobb, Jennifer F., Ziad Obermeyer, Yun Wang, and Francesca Dominici. 2014. "Cause-Specific Risk of Hospital Admission Related to Extreme Heat in Older Adults." *JAMA : The Journal of the American Medical Association* 312 (24): 2659-2667. doi:10.1001/jama.2014.15715. <http://dx.doi.org/10.1001/jama.2014.15715.>

    Choudhary, Ekta and Ambarish Vaidyanathan. 2014. "Heat Stress Illness Hospitalizations — Environmental Public Health Tracking Program, 20 States, 2001–2010." *MMWR. Surveillance Summaries* 63 (13): 1-10. <https://www.jstor.org/stable/24806265.>

    “Data Access Integrated Surface Dataset (Global).” National Centers for Environmental Information (NCEI). National Oceanic and Atmospheric Administration (NOAA). Accessed May 6, 2023. https://www.ncei.noaa.gov/access/search/data-search/global-hourly.

    Dai, Dajun. “Racial/Ethnic and Socioeconomic Disparities in Urban Green Space Accessibility: Where to Intervene?” *Landscape and Urban Planning* 102, no. 4 (2011): 234–44. https://doi.org/10.1016/j.landurbplan.2011.05.002.

    Davis, Robert E., David M. Hondula, and Anjali P. Patel. 2016. "Temperature Observation Time and Type Influence Estimates of Heat-Related Mortality in Seven U.S. Cities." *Environmental Health Perspectives* 124 (6): 795-804. doi:10.1289/ehp.1509946. <https://www.ncbi.nlm.nih.gov/pubmed/26636734.>

    Ebi, Kristie L., Thomas J. Teisberg, Laurence S. Kalkstein, Lawrence Robinson, and Rodney F. Weiher. 2004. "Heat Watch/Warning Systems Save Lives." *Bulletin of the American Meteorological Society* 85 (8): 1067-1073. doi:10.1175/BAMS-85-8-1067. <https://www.jstor.org/stable/26221144.>

    Engles, Eric A, Ruth M Pfeiffer, Joseph F Fraumeni, Bertram L Kasiske, Ajay K Israni, Jon J Snyder, Robert A Wolfe, et al. “Spectrum of Cancer Risk among US Solid Organ Transplant Recipients.” *JAMA: Journal of the American Medical Association* 2012 (2012): 105–6. https://doi.org/10.1016/j.ysur.2012.05.001.

    “Federal Climate Complex Data Documentation for Integrated Surface Data (ISD).” National Centers for Environmental Information (NCEI). NOAA - National Centers for Environmental Information, January 12, 2018. <https://www.ncei.noaa.gov/data/global-hourly/doc/isd-format-document.pdf>.

    Fishman, Jay A. “Infection in Solid-Organ Transplant Recipients.” *New England Journal of Medicine* 357, no. 25 (2007): 2601–14. https://doi.org/10.1056/nejmra064928.

    Gasparrini, Antonio, Dr, Yuming Guo PhD, Masahiro Hashizume Prof, Eric Lavigne PhD, Antonella Zanobetti PhD, Joel Schwartz Prof, Aurelio Tobias PhD, et al. 2015. "Mortality Risk Attributable to High and Low Ambient Temperature: A Multicountry Observational Study." *The Lancet* 386 (9991): 369-375. doi:10.1016/S0140-6736(14)62114-0. <https://www.clinicalkey.es/playcontent/1-s2.0-S0140673614621140.>

    Gasparrini, Antonio, Yuming Guo, Masahiro Hashizume, Patrick L. Kinney, Elisaveta P. Petkova, Eric Lavigne, Antonella Zanobetti, et al. 2015. "Temporal Variation in Heat–Mortality Associations: A Multicountry Study." *Environmental Health Perspectives* 123 (11): 1200-1207. doi:10.1289/ehp.1409070. <https://www.ncbi.nlm.nih.gov/pubmed/25933359.>

    Gronlund, Carina J., Antonella Zanobetti, Gregory A. Wellenius, Joel D. Schwartz, and Marie S. O’Neill. 2016. "Vulnerability to Renal, Heat and Respiratory Hospitalizations during Extreme Heat among U.S. Elderly." *Climatic Change* 136 (3-4): 631-645. doi:10.1007/s10584-016-1638-9. <https://link.springer.com/article/10.1007/s10584-016-1638-9.>

    Hayhoe, Katharine, Scott Sheridan, Laurence Kalkstein, and Scott Greene. “Climate Change, Heat Waves, and Mortality Projections for Chicago.” *Journal of Great Lakes Research* 36 (2010): 65–73. https://doi.org/10.1016/j.jglr.2009.12.009.

    James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Boston, MA: Springer, 2022.

    Kingsley, Samantha L., Melissa N. Eliot, Julia Gold, Robert R. Vanderslice, and Gregory A. Wellenius. 2016. "Current and Projected Heat-Related Morbidity and Mortality in Rhode Island." *Environmental Health Perspectives* 124 (4): 460-467. doi:10.1289/ehp.1408826. <https://www.ncbi.nlm.nih.gov/pubmed/26251954.>

    Lee, Whanhee, Yoonhee Kim, Francesco Sera, Antonio Gasparrini, Rokjin Park, Hayon Michelle Choi, Kristi Prifti, et al. “Projections of Diurnal Temperature Range and Diurnal Temperature Range-Related Excess Mortality under Climate Change.” *Environmental Epidemiology* 3 (2019): 226–27. <https://doi.org/10.1097/01.ee9.0000608348.49159.75>.

    Li, Tiantian, Radley M. Horton, and Patrick Kinney. 2013. "Future Projections of Seasonal Patterns in Temperature-Related Deaths for Manhattan." *Nature Climate Change* 3: 717-721. doi:10.1038/nclimate1902. <https://www.ncbi.nlm.nih.gov/pubmed/24910717.>

    Liu, Jia, Jin Yan, Qiquan Wan, Qifa Ye, and Yisheng Huang. “The Risk Factors for Tuberculosis in Liver or Kidney Transplant Recipients.” *BMC Infectious Diseases* 14, no. 1 (2014). https://doi.org/10.1186/1471-2334-14-387.

    Matte, Thomas D., Kathryn Lane, and Kazuhiko Ito. 2016. "Excess Mortality Attributable to Extreme Heat in New York City, 1997-2013." *Health Security* 14 (2): 64-70. doi:10.1089/hs.2015.0059. <https://www.liebertpub.com/doi/abs/10.1089/hs.2015.0059.>

    Metzger, Kristina B., Kazuhiko Ito, and Thomas D. Matte. 2010. "Summer Heat and Mortality in New York City: How Hot is Too Hot?" *Environmental Health Perspectives* 118 (1): 80-86. doi:10.1289/ehp.0900906. <https://www.jstor.org/stable/30249909.>

    Mohajerani, Abbas, Jason Bakaric, and Tristan Jeffrey-Bailey. “The Urban Heat Island Effect, Its Causes, and Mitigation, with Reference to the Thermal Properties of Asphalt Concrete.” *Journal of Environmental Management* 197 (April 14, 2017): 522–38. https://doi.org/10.1016/j.jenvman.2017.03.095.

    “NLCD 2016 Land Cover (CONUS).” NLCD 2016 Land Cover (CONUS) | Multi-Resolution Land Characteristics (MRLC) Consortium. Multi-Resolution Land Characteristics (MRLC) Consortium. Accessed May 3, 2023. https://www.mrlc.gov/data/nlcd-2016-land-cover-conus.

    Petkova, Elisaveta P., Jan K. Vink, Radley M. Horton, Antonio Gasparrini, Daniel A. Bader, Joe D. Francis, and Patrick L. Kinney. 2016. "Towards More Comprehensive Projections of Urban Heat-Related Mortality: Estimates for New York City Under Multiple Population, Adaptation, and Climate Scenarios." *Environmental Health Perspectives* 125 (1): 47-55. doi:10.1289/EHP166. <https://ntrs.nasa.gov/citations/20170006034.>

    Petkova, Elisaveta P., Radley M. Horton, Daniel A. Bader, and Patrick L. Kinney. 2013. "Projected Heat-Related Mortality in the U.S. Urban Northeast." *International Journal of Environmental Research and Public Health* 10 (12): 6734-6747. doi:10.3390/ijerph10126734. <https://www.ncbi.nlm.nih.gov/pubmed/24300074.>

    Petkova, Elisaveta, Antonio Gasparrini, and Patrick Kinney. 2014. "Heat and Mortality in New York City since the Beginning of the 20th Century." *Epidemiology (Cambridge, Mass.)* 25 (4): 554-560. doi:10.1097/EDE.0000000000000123. <https://www.jstor.org/stable/24759160.>

    Reid, Colleen E., Jonathan M. Snowden, Caitlin Kontgis, and Ira B. Tager. 2012. "The Role of Ambient Ozone in Epidemiologic Studies of Heat-Related Mortality." *Environmental Health Perspectives* 120 (12): 1627-1630. doi:10.1289/ehp.1205251. <https://www.jstor.org/stable/23323088.>

    Ruuhela, Reija, Otto Hyvärinen, and Kirsti Jylhä. 2018. "Regional Assessment of Temperature-Related Mortality in Finland." *International Journal of Environmental Research and Public Health* 15 (3): 406. doi:10.3390/ijerph15030406. <https://www.ncbi.nlm.nih.gov/pubmed/29495454.>

    Ruuhela, Reija, Athanasios Votsis, Jaakko Kukkonen, Kirsti Jylhä, Susanna Kankaanpää, and Adriaan Perrels. 2020. "Temperature-Related Mortality in Helsinki Compared to its Surrounding Region Over Two Decades, with Special Emphasis on Intensive Heatwaves." *Atmosphere* 12 (1): 46. doi:10.3390/atmos12010046. <https://search.proquest.com/docview/2524458714.>

    Schinasi, Leah H., Joan Rosen Bloch, Steven Melly, Yuzhe Zhao, Kari Moore, and Anneclaire J. De Roos. 2020. "High Ambient Temperature and Infant Mortality in Philadelphia, Pennsylvania: A Case–Crossover Study." *American Journal of Public Health (1971)* 110 (2): 189-195. doi:10.2105/AJPH.2019.305442. <https://www.ncbi.nlm.nih.gov/pubmed/31855483.>

    Schwartz, Joel D., Mihye Lee, Patrick L. Kinney, Suijia Yang, David Mills, Marcus C. Sarofim, Russell Jones, et al. 2015. "Projections of Temperature-Attributable Premature Deaths in 209 U.S. Cities using a Cluster-Based Poisson Approach." *Environmental Health* 14 (1): 85. doi:10.1186/s12940-015-0071-2. <https://ntrs.nasa.gov/citations/20170005277.>

    “Sklearn.model\_selection.GridSearchCV.” scikit-learn. scikit-learn. Accessed May 3, 2023. https://scikit-learn.org/stable/modules/generated/sklearn.model\_selection.GridSearchCV.html.

    “Temperature, Dewpoint, and Relative Humidity Calculator.” Calculate Temperature, Dewpoint, or Relative Humidity. University of Miami. Accessed May 3, 2023. <https://bmcnoldy.rsmas.miami.edu/Humidity.html>.

    “The Heat Index Equation.” Heat index equation. NOAA/National Weather Service, National Centers for Environmental Prediction, Weather Prediction Center, May 12, 2022. https://www.wpc.ncep.noaa.gov/html/heatindex\_equation.shtml.

    “Tiger/Line Shapefile, 2019, 2010 Nation, U.S., 2010 Census 5-Digit ZIP Code Tabulation Area (ZCTA5) National.” DATA.GOV. U.S. Census Bureau, Department of Commerce, November 1, 2022. https://catalog.data.gov/dataset/tiger-line-shapefile-2019-2010-nation-u-s-2010-census-5-digit-zip-code-tabulation-area-zcta5-na.

    Wang, Yan, Jennifer F. Bobb, Bianca Papi, Yun Wang, Anna Kosheleva, Qian Di, Joel D. Schwartz, and Francesca Dominici. 2016. "Heat Stroke Admissions during Heat Waves in 1,916 US Counties for the Period from 1999 to 2010 and their Effect Modifiers." *Environmental Health* 15 (1): 83. doi:10.1186/s12940-016-0167-3. <https://www.ncbi.nlm.nih.gov/pubmed/27503399.>

    Weinberger, Kate R., Xiao Wu, Shengzhi Sun, Keith R. Spangler, Amruta Nori-Sarma, Joel Schwartz, Weeberb Requia, et al. 2021. "Heat Warnings, Mortality, and Hospital Admissions among Older Adults in the United States." *Environment International* 157: 106834. doi:10.1016/j.envint.2021.106834. <https://dx.doi.org/10.1016/j.envint.2021.106834.>

    Weinberger, Kate R., Antonella Zanobetti, Joel Schwartz, and Gregory A. Wellenius. 2018. "Effectiveness of National Weather Service Heat Alerts in Preventing Mortality in 20 US Cities." *Environment International* 116: 30-38. doi:10.1016/j.envint.2018.03.028. <https://dx.doi.org/10.1016/j.envint.2018.03.028.>

    Wellenius, Gregory A., Melissa N. Eliot, Kathleen F. Bush, Dennis Holt, Rebecca A. Lincoln, Andy E. Smith, and Julia Gold. 2017. "Heat-Related Morbidity and Mortality in New England: Evidence for Local Policy." *Environmental Research* 156: 845-853. doi:10.1016/j.envres.2017.02.005. <https://dx.doi.org/10.1016/j.envres.2017.02.005.>

    World Health Organization. 2014.*Quantitative Risk Assessment of the Effects of Climate Change on Selected Causes of Death, 2030s and 2050s* World Health Organization. <http://www.who.int/iris/handle/10665/134014.>

    Wouters, Hendrik, Koen De ridder, Lien Poelmans, Patrick Willems, Johan Brouwers, Parisa Hosseinzadehtalaei, Hossein Tabari, Sam Vanden broucke, Nicole P. M. Van lipzig, and Matthias Demuzere. 2017.*Heat Stress Increase Under Climate Change Twice as Large in Cities as in Rural Areas: A Study for a Densely Populated Midlatitude Maritime Region*. Vol. 44 American Geophysical Union (AGU). doi:10.1002/2017gl074889.

    Zhao, Lei, Keith Oleson, Elie Bou-Zeid, E. Scott Krayenhoff, Andrew Bray, Qing Zhu, Zhonghua Zheng, Chen Chen, and Michael Oppenheimer. 2021. "Global Multi-Model Projections of Local Urban Climates." *Nature Climate Change* 11 (2): 152-157. doi:10.1038/s41558-020-00958-8. <https://search.proquest.com/docview/2488188362.> [↑](#endnote-ref-52)