

Capstone

Jane Doe

1/25/23

Table of contents

Index	3
Methods	6
1 Introduction	7
1.1 Relationship Between Health and Place	9
1.2 Data and Methods	10
2 Summary	21
References	22

Index

UPDATE 2/27: THIS SECTION IS NOT UP TO DATE. DO NOT READ.

A	B	C	D	E	F
1	Summary of Data Sources				
2	as of 2-5-2023				
3					
4	Variable	Unit	Source	Year(s)	
5	Age-adjusted mortality rates: all-cause	County	CDC WONDER	2014 - 2016 (3-year rates)	
6	Age-adjusted mortality rates: ICD J00 - ICD J99	County	CDC WONDER	2014 - 2016 (3-year rates)	
7	Age-adjusted mortality rates: ICD J00 - ICD J99 (65 and over)	County	CDC WONDER	2014 - 2016 (3-year rates)	
8	Tobacco use (SMOKER3 in 1,2,3)	County	CDC Behavioral Risk Factor Surveillance System (BRFSS)	2016	
9	Percent Black	County	American Community Survey, 5-year	2016	
10	Percent Hispanic	County	American Community Survey, 5-year	2016	
11	Percent White	County	American Community Survey, 5-year	2016	
12	Percent Poverty	County	American Community Survey, 5-year	2016	
13	Highest Degree	County	American Community Survey, 5-year	2016	
14	Land Cover	Raster	NLCD	2019	
15	Annual Air Quality: PM 2.5	Monitor	EPA	2014 - 2016	
16	Annual Air Quality: Ozone?	Monitor	EPA	2014 - 2016	
17					

UPDATE 2/27: THIS SECTION IS NOT UP TO DATE. DO NOT READ.

Data fall into four major categories: health, demographic, air pollution, and land use. However, since this is a spatial analysis, the common thread in all data sources is the spatial linkage. Spatial evaluation of air pollution's impact on health is considered on a regional level. Following the approach that Simon and colleagues take to evaluate ozone trends in the U.S., this analysis also uses NOAA climate regions in the analysis of air pollution (2015). Evaluation of data on a regional scale lowers the model's sensitivity to factors such as wind, climate, and pollutants' chemical reactions with other molecules in the air that can cause the concentration of air pollution to change (Simon et al. 2015, pp 186, pp. 188). To this end, five regions were selected for this research, which represent different climate areas and a wide variety of land cover types.

The Environmental Protection Agency (EPA) publishes extensive amounts of air quality data through their Air Data site. This research focuses on PM 2.5 and Ozone data. PM 2.5 is a measure of the ambient air pollution... EXPLAIN MORE. EPA's annual summaries of air quality metrics by monitoring site for the years 2014 through 2016 are used in this model. Location data on monitoring sites are included in the data set. Spatial interpolation techniques are used to infer air pollution estimates for unobserved locations (see: [Introduction to spatial data analysis | Data Science \(stanford.edu\)](#)). MORE ABOUT THIS AND WRAP. One limitation to this technique is that annual summaries, while useful, may not be sensitive enough to capture acute changes in air quality, which pose a health risk (CITE?). This relationship may not be

captured in the model if poor air quality days are infrequent enough to impact the mean. Additionally, some monitors capture and report air quality more frequently than others, so there could be some variation due to the frequency of measurement capture. For simplicity, this research does not take seasonality into consideration and instead uses annual averages, which does not accurately reflect variation in air pollution ranges. Finally, the biggest limitation as it relates to our spatial model is that monitors are not normally distributed throughout space. Spatial bias is introduced due to the distribution of data across the measurement areas.

Respiratory health is measured by age-adjusted mortality rates by county, accessed through the Center for Disease Control and Prevention's (CDC) WONDER database. The first data set extracted is all-age, all-cause age-adjusted mortality rate by county for the three-year period from 2014 through 2016. Data were further filtered based on primary reported cause of death based on the ICD-10 code range J00-J99, which encompasses respiratory disease. Finally, a third data set limited the respiratory disease age-adjusted mortality rates to the 65 and over population. One limitation of data on the primary cause of death is that health can be influenced by a variety of issues. When there are multiple causes of death, those multiple causes are noted but for this purpose, only the primary cause is used to characterize the respiratory disease-related death population. Therefore, the data would be missing those who had respiratory illnesses as a contributing factor but not the primary cause. It would be interesting to better analyze this population to see if there are trends that would support the inclusion of respiratory disease as a cause of death (and not just the primary cause), but for this research, honing in on the primary cause of death should be sufficient.

To further characterize county populations, demographic and behavioral data are consulted. Data on age-adjusted adult smoking rates by county are extracted from the University of Wisconsin Public Health Institute's County Health Rankings. This data is consulted to normalize out any distortions caused by smoking habits by county. However, prevalence rates do not capture the severity of tobacco use by individual. Severity of smoking habits are not considered in this model. County Health Rankings source for its smoking data is the CDC's Behavioral Risk Factor Surveillance System (BRFSS) ([CITE Adult Smoking | County Health Rankings & Roadmaps](#)). One limitation to the BRFSS is that data is collected from surveys then extrapolated to characterize county-level summaries. The conduction of surveys and extrapolation process leave room for error so is noted for transparency. Demographic data are obtained from the U.S. Census Bureau's American Community Survey (ACS). 5-year rates for the survey ended in 2016 are used in this model. CURRENTLY TARGETING THESE VARIABLES: Percent poverty, percent Black, percent Hispanic, urbanization are used in this model. ACS data are obtained through the Census API, accessed through R Studio.

Land cover data used in this model are from the National Land Cover Database (NLCD). Data were obtained from the FedData R package and further manipulated in R to differentiate primary land cover characteristics by county and by National Ocean and Atmospheric Administration (NOAA)-defined climate regions. HOW ARE DATA CAPTURED?

Using a spatial approach to find an association to a health outcome introduces assumptions about populations' exposure to pollutants. Since we are using county level data to characterize

the populations, one assumption is that individuals' activity spaces are primarily in their county of residence. This assumption is fallible, particularly for those living on the fringes of boundaries. Exposure to air pollution differs for those who spend more time outdoors than those who spend significant time indoors. HOW DO I ACCOUNT FOR THIS IN THE MODEL?

Methods

This research uses regression modeling on three NOAA climate regions to test if creating models for each region produces significant results. The region-based modeling using NOAA's climatically consistent region definitions draws from research on by Simon and colleagues. Simon and colleagues justify their use of NOAA climate regions by acknowledging that ozone patterns are partially determined by temperature and precipitation (2015, 188). Additionally, they note that climate regions generally group areas with similar emissions types and similar trends in temporal variation due to daylight and season (Simon et al. 2015, 188). By modeling for each region, I attempt to normalize out some of the variation due to climate, seasonality, and emissions type which should remove some noise from the models.

Data are first approximated at a county level. Air quality data is captured by monitors, whose locations are indicated in the EPA's data files. In order to spread air quality metrics over space, spatial interpolation is used (add specifications). Land cover data is also approximated at the county level by using R to quantify the area of major land cover types that are measured in each county.

The dependent variable that is being used as a proxy for health is age-adjusted mortality rate for respiratory diseases. After a review of the dependent and independent variables, OLS regression models are consulted. The OLS regression model will estimate the effect of the independent variables on the dependent variable. Different combinations of variables are tested to evaluate what creates the best model, and results are evaluated for statistical significance.

A linear regression model is then evaluated for spatial autocorrelation. The regression model depends on independence and heteroskedasticity of residuals. If spatial autocorrelation is present, then both rules are broken. Spatial autocorrelation is tested using the Global Moran's I test, followed by mapping the Local Moran's I. Spatial clustering is revealed by the Local Indicator of Spatial Association (LISA) map. The Lagrange Multiplier test is consulted to determine if spatial error or spatial lag is driving spatial autocorrelation in the model. Based on the Lagrange Multiplier results, the spatial [lag/error] model is used to try to produce a better model.

1 Introduction

Air pollution's adverse impact on respiratory health is a widely known issue. The World Health Organization acknowledged the significance of this risk factor in its 68th conference, where they note that “exposure to air pollutants, including fine particulate matter, is a leading risk factor for noncommunicable diseases in adults, including ischaemic heart disease, stroke, chronic obstructive pulmonary disease, asthma and cancer, and poses a considerable health threat to current and future generations,” (2015). Further, numerous studies have substantiated the risk that air pollution poses to population health, many linking air pollution to industrial areas and urbanization. On the other hand, other research recognize the variable impacts of interactions between urban built environments, air pollution, and health. This group of research considers factors ranging from the built environment's impact on promoting walking and other forms of exercise, to the impact of exercising in areas with high levels of air pollution, and to the role that urban greenspace plays in reducing air pollution (Hankey, Marshall, and Brauer 2012), (Lee et al. 2022),(Alcock et al. 2017). The study of air pollution's impact of respiratory health is complicated by the interactions between the natural environment, built environment, and human behavior. Therefore, approaching the problem using spatial methods may reveal prevailing patterns that result from the exchanges between all factors. This paper attempts to create a spatial model that predicts respiratory health patterns using air pollution and land cover data in the United States. By incorporating land cover data, I attempt to consider human interactions with the natural and built environment, and its impact on respiratory health, through direct or indirect means.

Associations between land use and health impacts have also been studied in the past, with varying findings. Wang and colleagues found a negative correlation between Chronic Obstructive Pulmonary Disease (COPD) mortality and land use mix in most neighborhoods in China's Jing'an district except for northwest areas. In northwest neighborhoods, COPD mortality was positively correlated with land use mix (Wang et al. 2019, 7). Other studies, such as Alcock and colleagues' analysis of asthma hospitalizations in the UK, find that increases in urban green space and tree density are associated with reductions in asthma hospitalizations (Alcock et al. 2017, 39). However, models that included air pollutant exposure variables complicated this correlation (Alcock et al. 2017, 39). Another study that analyzes variation in spatial distribution of lung cancer incidence rates in Shanghai, China found that industrial parks and urban-rural mixed areas show higher risk of lung cancer (Wang et al. 2022, 13). Wang and colleagues also postulate that high lung cancer rates in high-density urban areas away from industrial pollution exposures may be due to traffic and other pollutants (2022, 13).

The complex results of past studies underscore the symbiotic nature of human interactions with the built environment, and the downstream implications on individuals' health status. Wang and colleagues explained the changes in relationship between land use mix and COPD mortality by linking the existence of arterial roadways, which may increase opportunities for generation of air pollutants and negatively impact health, to their use to improve access to care (2019). The shift in balance between the two factors, they say, is why results varied in highly urban areas as compared to highly rural regions (Wang et al. 2019, 8). Kim and colleagues also address the dichotomy of the urban environment's impact on traffic, air pollution and behavior by considering walkability and land use diversity in their model of the association between the built environment and asthma in Los Angeles, California (2023). Their model builds upon urban planning ideas that mixed-use land reduces motorized travel through promotion of ride-sharing, walking, and shortened distances to destinations, while also promoting better physical through dedicated open spaces and recreational areas (Kim et al. 2023, 58). One limitation of Kim and colleagues' model is the limited spatial scope. In this paper, I build upon elements of Kim and colleagues' model but incorporate other scale considerations.

Scale considerations are a challenge in all spatial analyses, therefore it is not surprising that scale is listed as limitation listed in most spatial analyses of land use, pollution, and respiratory health. Huang and colleagues address the modifiable areal unit problem (MAUP) by analyzing the impacts of land use data at different buffer sizes to determine trends across different spatial units (2021). They also select data from different regions that are representative of different environments and climates to determine if regional factors influence their air pollution models (Huang et al. 2021). Their results suggest that land cover types have varying significance across spatial scales, concluding that air pollutant levels are primarily affected by regional land cover types (Huang et al. 2021, 7). This paper builds upon their approach by acknowledging the role that climate plays in air pollution distribution, proliferation, and mitigation. Separating U.S. land areas based on climate region will acknowledge the spillover effects of air pollution that may not be appropriately captured in analyses performed at local scales (Simon et al. 2015).

This research uses U.S. age-adjusted mortality rates from cardiovascular and respiratory diseases (defined as ICD-10 codes I00-I99 and J00 - J98) as the health indicator. In Mueller and colleagues' literature analysis on greenspace and respiratory health research, they found that respiratory mortality had the most consistent positive evidence as compared to research measuring other health indicators, including asthma, lung function, hospital admissions, among other measures (Mueller et al. 2022, 28). Using mortality rates from cardiovascular and respiratory diseases will capture acute and chronic conditions that can be impacted by instances of severe drops in air quality and more moderate yet sustained air quality issues.

1.1 Relationship Between Health and Place

One reason why it is difficult to prove a causal relationship between environmental risk factors such as air pollution and a health outcome is because health outcomes can be cumulative results of a combination of factors. When considering air pollution's impact on health, one such factor are air pollution's exposure pathways. Exposure can vary depending on distance from sources of pollution and the amount of time spent breathing in pollutants, both of which can fluctuate when considering climate, ventilation, time spent outdoors, among other factors. Arguably the simplest way to measure exposure is by measuring proximity to pollutant sources. However, spatial analysis of epidemiological phenomena should consider more than just geographic location. This is because an individual's activity spaces are not simply a feature of that person, but a reflection of a myriad of influences that result in an individuals' choice to frequent those spaces. Entwistle highlights this topic in *Putting People into Place*, arguing that there is a critical need to "incorporate agency in theories of neighborhood and health," (Entwistle 2007, 699). Entwistle identifies the symbiotic relationship between individuals' opportunities, social interactions, BLAH that inform decisions to move to or remain in a space, and how that informs health decisions. She also acknowledges that individuals can influence their communities. Bernard and colleagues build on this notion by formalizing a conceptualization of neighborhoods. Their framework identifies four primary influences on access to health-related resources: "proximity, rights, prices, and informal reciprocity," (Bernard et al. 2007, 1849). WRAP ABOUT HOW THIS RELATES.

This research acknowledges the role that individual decisions play in informing their activity spaces by including demographic and quality of life variables. The inclusion of these variables help to contextualize individuals within each county. These variables range from race to income, education, access to care, and health behaviors. By including demographic and quality of life variables, the model attempts to characterize those occupying the spaces, account for factors that influence interactions with health related resources, and acknowledge the role of positive and negative health behaviors on mortality risk.

The fact that this research fails to acknowledge the role that individuals' play in influencing their localities is a shortcoming that can be addressed in future iterations of this model. Local communities are constantly being reshaped in large and small ways. Agency over change in space (use the vocabulary that Browning uses) can be evaluated by reviewing changes in localities over time. In this research, individual influence over their communities may manifest in improvements to regulations to reduce industrial air pollution, improvement to ventilation systems, and increases in indoor recreation alternatives, to name a few examples. Such interventions could lower community exposure to air pollution and potentially reduce adverse health impacts. A future enhancement to this model would incorporate variables over a longer time horizon to measure sensitivity of mortality rates to changes in localities over time.

1.2 Data and Methods

Data fall into four major categories: health, demographic, air pollution, and land use. However, since this is a spatial analysis, the common thread in all data sources is the spatial linkage. Spatial evaluation of air pollution's impact on health is considered on a regional level. Following the approach that Simon and colleagues take to evaluate ozone trends in the U.S., this analysis also uses NOAA climate regions in the analysis of air pollution (2015). Evaluation of data on a regional scale lowers the model's sensitivity to factors such as wind, climate, and pollutants' chemical reactions with other molecules in the air that can cause the concentration of air pollution to change (Simon et al. 2015, pp 186, pp. 188). To this end, four regions were selected for this research, which represent different climate areas and a wide variety of land cover types.

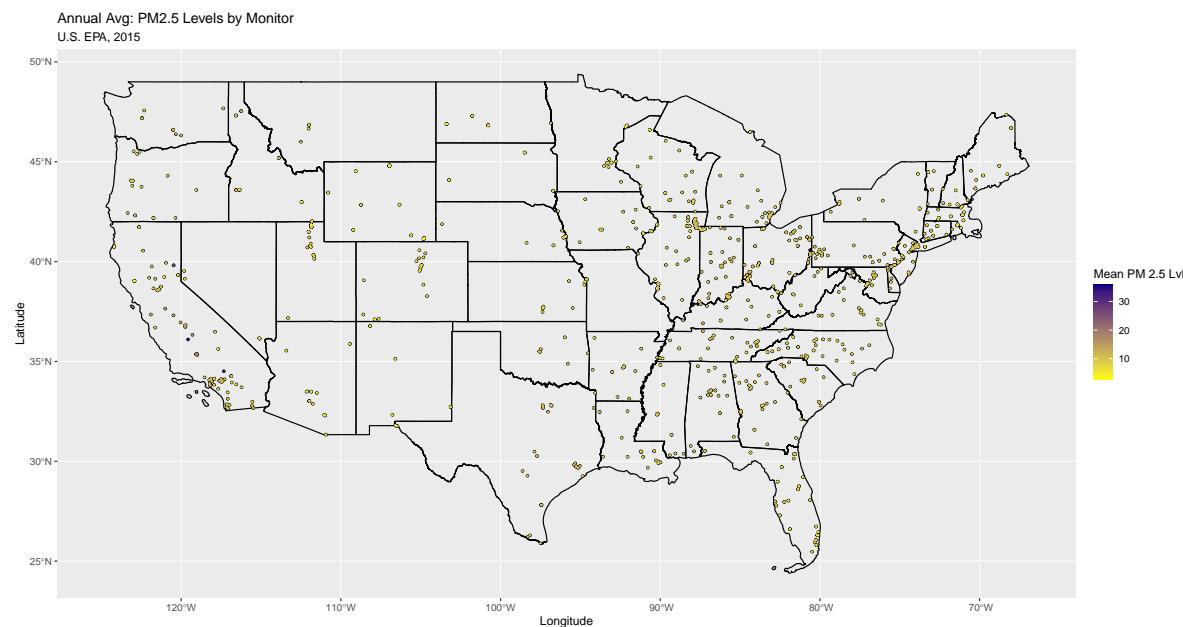


Figure 1.1: Figure 1; Annual summary of air quality metrics by monitoring site from 2014 - 2016

This research focuses on particulate matter (PM) 2.5 measures from 2015, as captured by the Environmental Protection Agency (EPA) and further manipulated by the Centers for Disease Control and Prevention (CDC). PM 2.5 are “fine, inhalable particles with diameters that are generally 2.5 micrometers or smaller,” that have been linked to various complications with lung

and heart functions (US EPA 2016). The EPA publishes extensive amounts of air quality data, which are accessible through their Air Data website. Available data include annual summaries by monitor.

Figure 1 plots the 2015 annual PM 2.5 measures on region-dependent scales by NOAA Climate region to provide a geographic reference for monitor locations and start to visually represent the variation in PM 2.5 measurements by region. The entire contiguous U.S. is represented in Figure 1, but it is worth noting that only the Northwest, West, Ohio Valley, and Southeast NOAA Climate Regions were selected for this study. These regions were selected due to their climatic variation, regional air quality variation, and either high presence or decent distribution of air quality monitors. A visual inspection of Figure 1 reveals how monitors are distributed throughout each region and their individual readings. It is noted that though the West region has a relative abundance of monitors in California, monitors are sparse in Nevada. This observation leads to the acknowledgement of potential spatial bias: monitors are not normally distributed throughout space, therefore spatial bias may be introduced due to the distribution of data across the measurement areas. However, the level of spatial bias difficult to measure due to the use of the CDC's modeled data.

The CDC provides measured and modeled PM 2.5 data by county for the U.S., as depicted in Figure 2. They derive their data from daily PM 2.5 measurements produced by the EPA, then further model to infer measurements for areas where there are gaps in the data. Derivations are primarily based on the Downscaler (DS) model, which is “a statistical fusion of monitoring data and Community Multiscale Air Quality (CMAQ) outputs,” [@geograph]. While use of this data is convenient, a future iteration of this research will produce and employ a PM 2.5 inference model, which would be produced specifically for this research. Doing so would control for any unknowns that are introduced due to the limitations of using modeled data that is not easily reproducible.

One limitation of the air quality data used in this model is the use of annually summarized data. While useful, annual summaries may not be sensitive enough to capture acute changes in air quality, which pose a health risk (CITE?). This relationship may not be captured in the model if poor air quality days are infrequent enough to impact the mean. For simplicity, this research does not take seasonality into consideration and instead uses annual averages, which does not accurately reflect variation in air pollution ranges.

DESCRIBE FIGURE 2

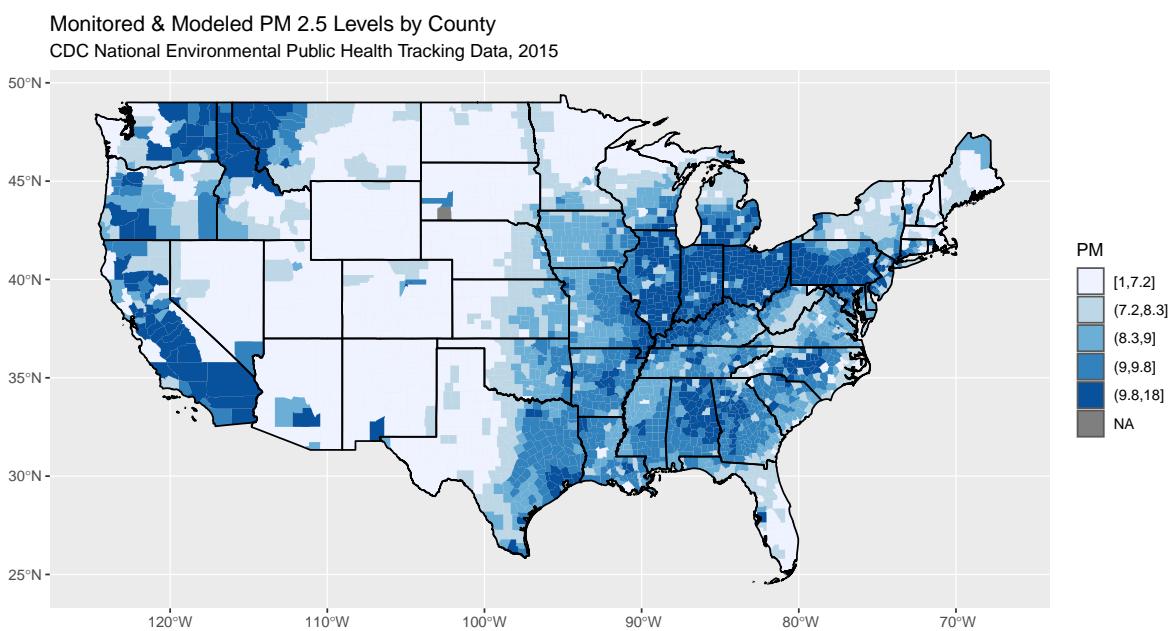
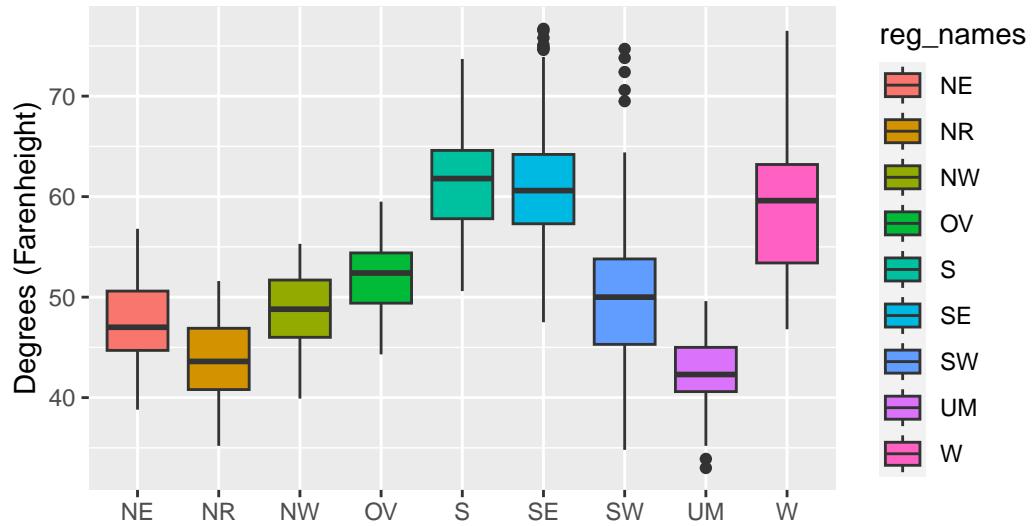
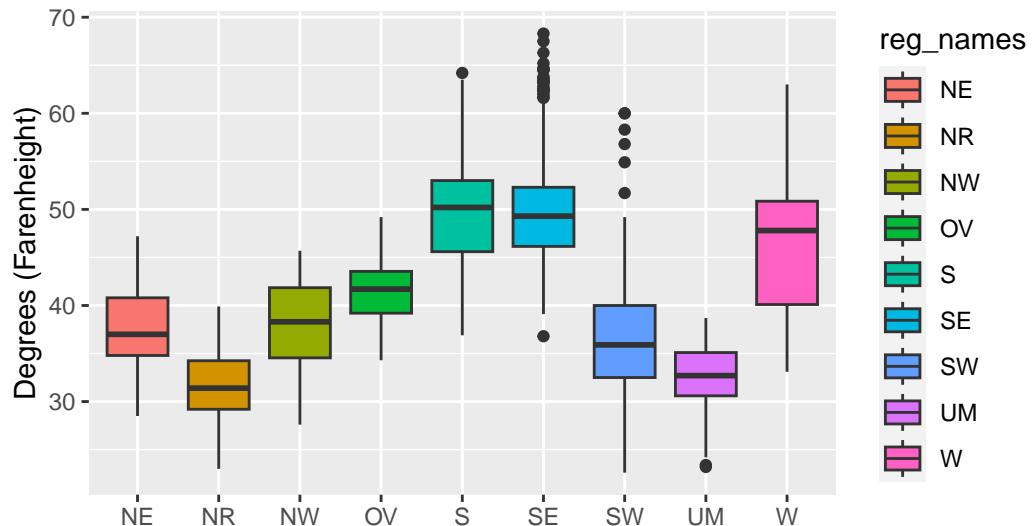


Figure 1.2: Figure 3; County-level air pollution statistics. Region selection limited by availability of data

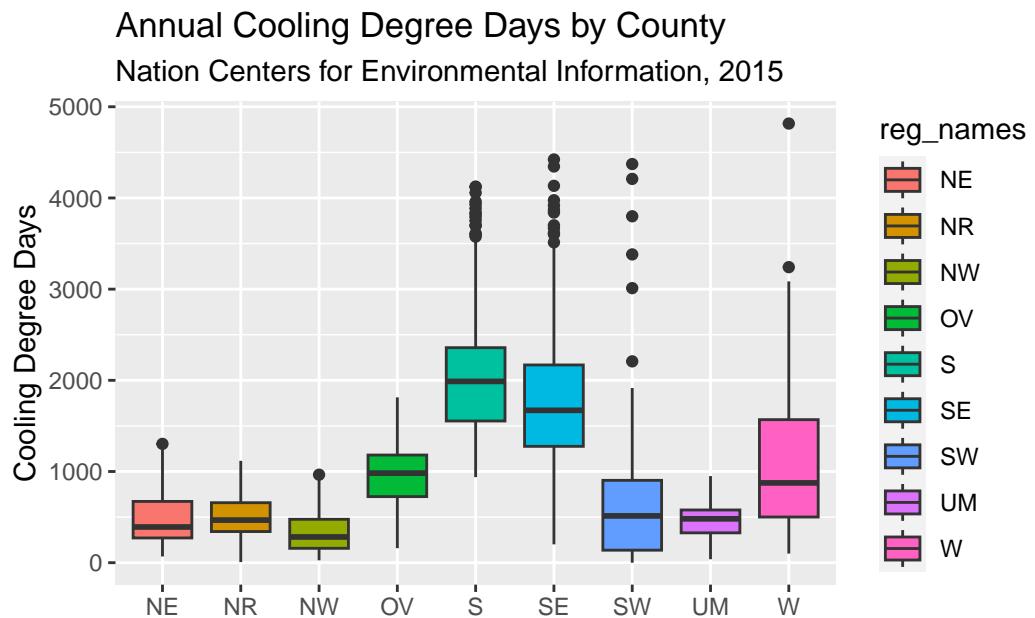
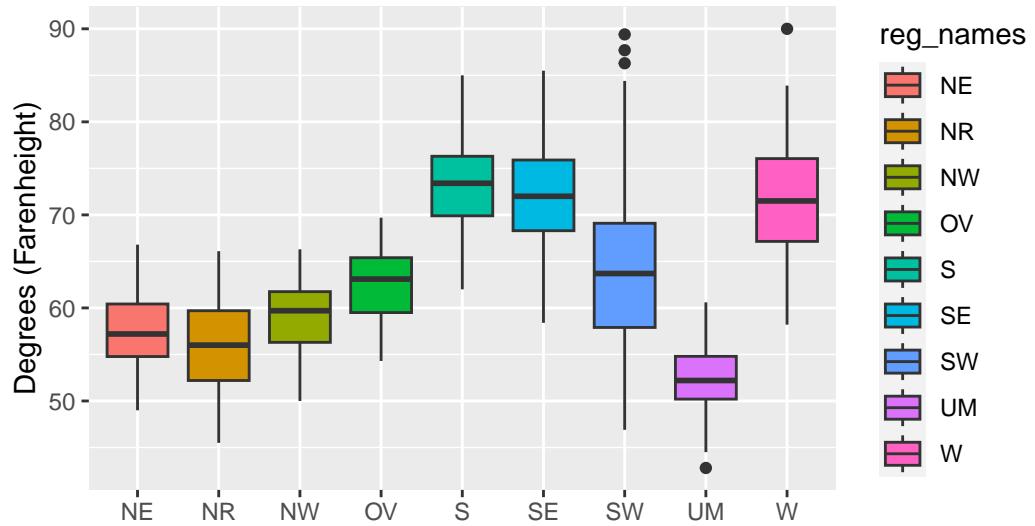
Average Annual Temperature by County
Nation Centers for Environmental Information, 2015



Minimum Temperature by County
Nation Centers for Environmental Information, 2015

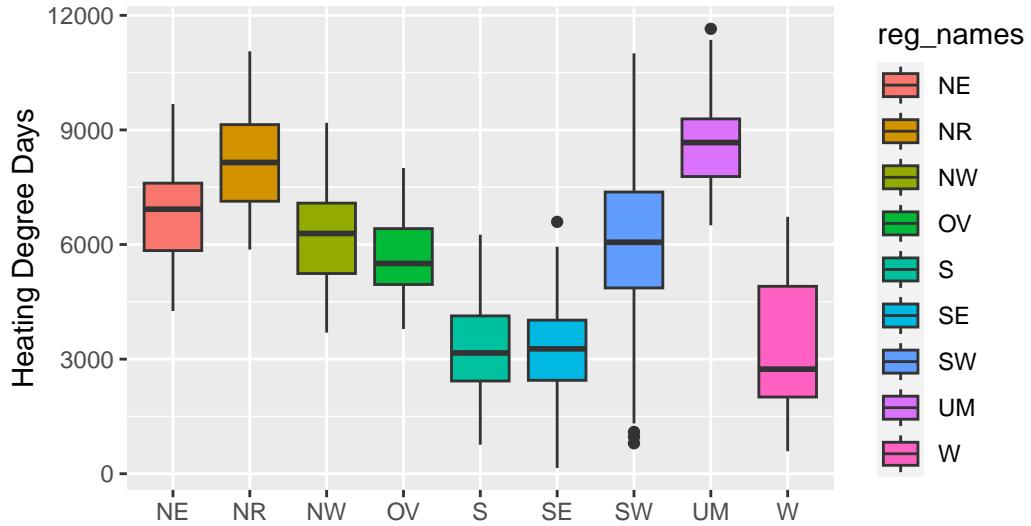


Maximum Temperature by County
Nation Centers for Environmental Information, 2015



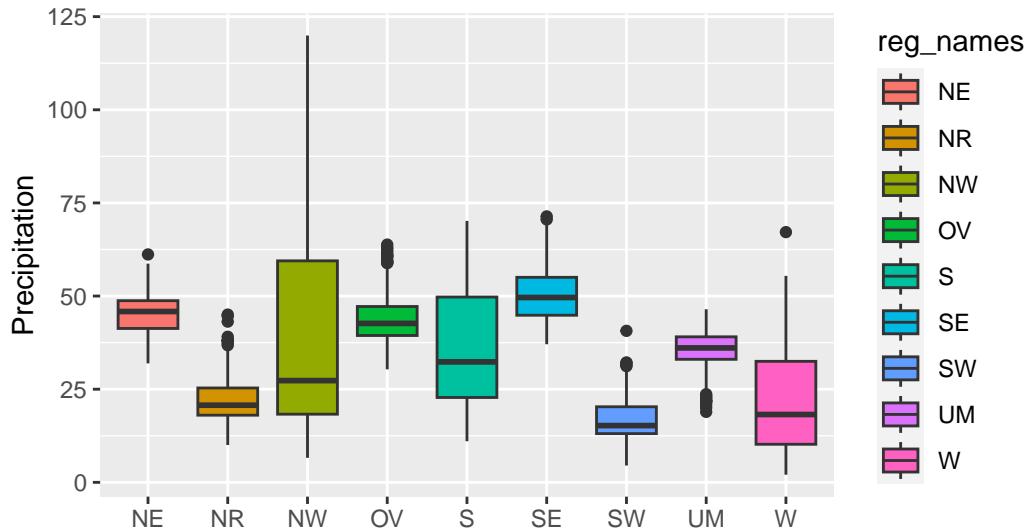
Annual Heating Degree Days by County

Nation Centers for Environmental Information, 2015



Annual Precipitation by County

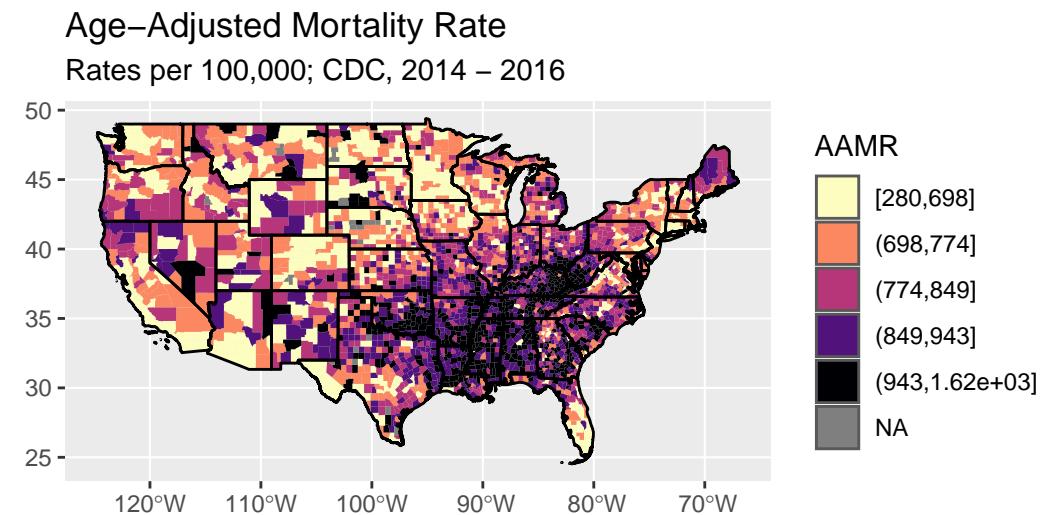
Nation Centers for Environmental Information, 2015



Respiratory health is measured by age-adjusted mortality rates by county, accessed through the Center for Disease Control and Prevention's (CDC) WONDER database. The first data set extracted is all-age, all-cause age-adjusted mortality rate by county for the three-year period from 2014 through 2016. Data were further filtered based on primary reported cause of death

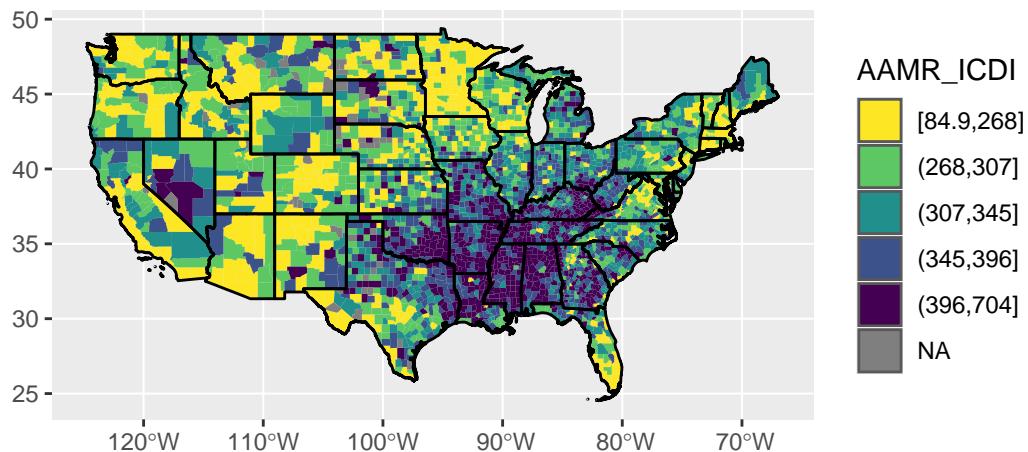
based on the ICD-10 code range I00-I99, which encompasses cardiovascular diseases and J00-J98, respiratory diseases (“CDC WONDER,” n.d.). Previous research exposing correlations between air pollution and mortality have focused on the same set of diagnostic codes (O’Neill et al. 2008), (Wong et al. 2008). One limitation of data on the primary cause of death is that health can be influenced by a variety of issues. When there are multiple causes of death, those multiple causes are noted but for this purpose, only the primary cause is used to characterize the cardiovascular or respiratory disease-related death population. Therefore, the data would be missing those who had respiratory illnesses as a contributing factor but not the primary cause. It would be interesting to better analyze this population to see if there are trends that would support the inclusion of respiratory disease as a cause of death (and not just the primary cause), but for this research, honing in on the primary cause of death should be sufficient. Another limitation is that the diagnostic classes used in this model are broad. It is an intentional choice to cast a wide net on mortality causes because air pollution’s impact on individual health can manifest in a variety of ways. However, the shortcoming of this tactic is that it introduces more data that may potentially be unrelated to air pollution. This is something to be mindful of when reviewing the results of data modeling.

DESCRIBE DATA



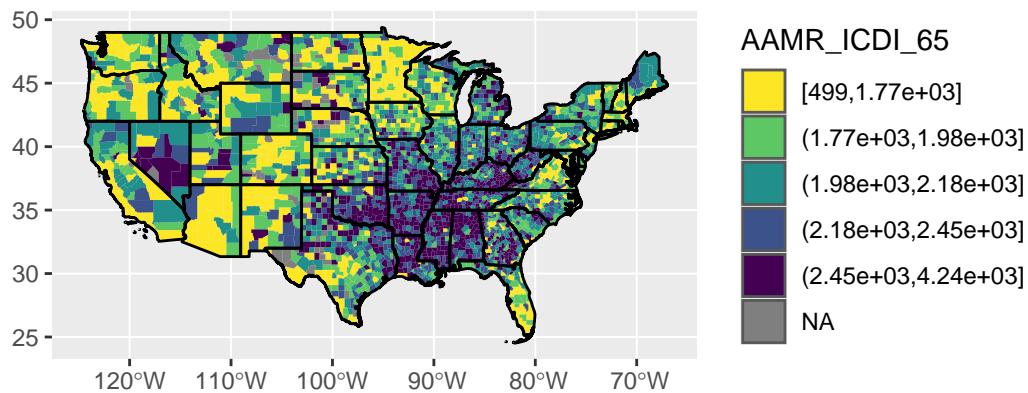
Age-Adjusted Mortality Rates – Cardiovascular & Respiratory Disease

Rates per 100,000; CDC, 2014 – 2016



Age-Adjusted Mortality Rates 65+: Cardiovascular & Respiratory Disease

Rates per 100,000; CDC, 2014 – 2016



To further characterize county populations, demographic and behavioral data are consulted. 2015 data on age-adjusted adult smoking rates by county are extracted from the University of Wisconsin Public Health Institute's County Health Rankings. This data are consulted to normalize out any distortions caused by smoking habits by county. However, prevalence

rates do not capture the severity of tobacco use by individual. Severity of smoking habits are not considered in this model. County Health Rankings' source for its smoking data is the CDC's Behavioral Risk Factor Surveillance System (BRFSS) ("Adult Smoking," n.d.). County Health Rankings data aggregates other useful variables that are used in their model. Adult obesity rates by county, sourced from BRFSS data, are also used. One limitation to the BRFSS is that data is collected from surveys, then extrapolated to characterize county-level summaries. The conduction of surveys and extrapolation process leave room for error so is noted for transparency. Another data point extracted from County Health Rankings is the percent of the population that is uninsured (Source: U.S. Census Bureau's Small Area Health Insurance Estimates (SAHIE)) ("2022 Measures," n.d.). SAHIE's percent uninsured estimates the percent of the population that is 65 and under who do not have health insurance ("Uninsured," n.d.). According to the U.S. Census Bureau, "SAHIE uses statistical models that combine survey data from the American Community Survey (ACS) with administrate records data and Census 2010 data," (Bureau, n.d.). The benefit of SAHIE's modeling approach is that it attempts to control for estimates for small areas where data is more sparse (Bureau, n.d.). As with any modeled data, a notable risk is the reliance on the robustness of the U.S. Census Bureau's model, though the Census Bureau does provide margins of error for their model results. Other demographic data are obtained from the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates. 2016 data for following variables are used in this model: Percent poverty, percent Black, percent Hispanic, urbanization, and education statistics. ACS data are obtained through the Census API, accessed through R Studio.

Land cover data used in this model are originally sourced from the National Land Cover Database's (NLCD). The International Public Use Microdata Series (IPUMS) released county-level summaries of NLCD land cover data, which are displayed in the ensuing plots. Per IPUMS' annotations on methodology, NLCD data were layered with census geographic polygons to calculate the area covered by each land class ("Environmental Summaries | IPUMS NHGIS," n.d.). Land cover classifications are determined by NLCD using a modified Anderson Land Cover Classification System ("National Land Cover Database Class Legend and Description | Multi-Resolution Land Characteristics (MRLC) Consortium," n.d.). A description of NLCD's land cover classifications are denoted in the data table below. Data used in this model reflect 2011 land cover data summarized by county geographic boundaries in the U.S. Census Bureau's 2015 data set.

This research uses regression modeling on three NOAA climate regions to test if creating models for each region produces significant results. The region-based modeling using NOAA's climatically consistent region definitions draws from research on by Simon and colleagues. Simon and colleagues justify their use of NOAA climate regions by acknowledging that ozone patterns are partially determined by temperature and precipitation (2015, 188). Additionally, they note that climate regions generally group areas with similar emissions types and similar trends in temporal variation due to daylight and season (Simon et al. 2015, 188). By modeling for each region, I attempt to normalize out some of the variation due to climate, seasonality, and emissions type which should remove some noise from the models.

Class value	Class name	Description
11	Open water	Areas of open water, generally with less than 25% cover of vegetation or soil.
12	Perennial ice/snow	Areas characterized by a perennial cover of ice and/or snow, generally greater than 25% of total cover.
21	Developed, open space	Areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20% of total cover. 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.
22	Developed, low intensity	Areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20% to 49% of total cover. These areas most commonly include single-family housing units.
23	Developed, medium intensity	Areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50% to 79% of the total cover. These areas most commonly include single-family housing units.
24	Developed, high intensity	Highly developed areas where people reside or work in high numbers. Examples include apartment complexes, row houses and commercial/industrial. Impervious surfaces account for 80% to 100% of the
31	Barren land (rock/sand/clay)	Areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits and other accumulations of earthen material. Generally, vegetation accounts for less than 15% of total cover.
41	Deciduous forest	Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75% of the tree species shed foliage simultaneously in response to seasonal change.
42	Evergreen forest	Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75% of the tree species maintain their leaves all year. Canopy is never without green.
43	Mixed forest	Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. Neither deciduous nor evergreen species are greater than 75% of total tree cover.
52	Shrub/scrub	Areas dominated by shrubs; less than 5 meters tall with shrub canopy typically greater than 20% of total vegetation. This class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions.
71	Grassland/herbaceous	Areas dominated by gramanoid or herbaceous vegetation, generally greater than 80% of total vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing.
81	Pasture/hay	Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20% of total vegetation.
82	Cultivated crops	Areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops such as orchards and vineyards. Crop vegetation accounts for greater than 20% of total vegetation. This class also includes all land being actively tilled.
90	Woody wetlands	Areas where forest or shrubland vegetation accounts for greater than 20% of vegetative cover and the soil or substrate is periodically saturated with or covered with water.
95	Emergent herbaceous wetlands	Areas where perennial herbaceous vegetation accounts for greater than 80% of vegetative cover and the soil or substrate is periodically saturated with or covered with water.

Figure 1.3: NLCD Land Cover Types

Data are first approximated at a county level. Air quality data is captured by monitors, whose locations are indicated in the EPA's data files. In order to spread air quality metrics over space, spatial interpolation is used (add specifications). Land cover data is also approximated at the county level by using R to quantify the area of major land cover types that are measured in each county.REMOVE?

The dependent variable that is being used as a proxy for health is age-adjusted mortality rate for respiratory diseases. After a review of the dependent and independent variables, OLS regression models are consulted. The OLS regression model will estimate the effect of the independent variables on the dependent variable. Different combinations of variables are tested to evaluate what creates the best model, and results are evaluated for statistical significance.

A linear regression model is then evaluated for spatial autocorrelation. The regression model depends on independence and heteroskedasticity of residuals. If spatial autocorrelation is present, then both rules are broken. Spatial autocorrelation is tested using the Global Moran's I test, followed by mapping the Local Moran's I. Spatial clustering is revealed by the Local Indicator of Spatial Association (LISA) map. The Lagrange Multiplier test is consulted to determine if spatial error or spatial lag is driving spatial autocorrelation in the model. Based on the Lagrange Multiplier results, the spatial [lag/error] model is used to try to produce a better model.

```
{# {r} # #prep data # health_c <- read_excel('data files/2016 county health  
v3.xlsx') # # c_prep5<-left_join(c_prep4,health_c, by="cofips") # c_prep5<-subset(c_prep5,s  
# # #START TO MAP VARIABLES IN LISA #
```

2 Summary

In summary, this book has no content whatsoever.

1 + 1

[1] 2

References

- “2022 Measures.” n.d. <https://www.countyhealthrankings.org/explore-health-rankings/county-health-rankings-measures>.
- “Adult Smoking.” n.d. <https://www.countyhealthrankings.org/explore-health-rankings/county-health-rankings-model/health-factors/health-behaviors/tobacco-use/adult-smoking>.
- Alcock, Ian, Mathew White, Mark Cherrie, Benedict Wheeler, Jonathon Taylor, Rachel McInnes, Eveline Otte im Kampe, et al. 2017. “Land Cover and Air Pollution Are Associated with Asthma Hospitalisations: A Cross-Sectional Study.” *Environment International* 109 (December): 29–41. <https://doi.org/10.1016/j.envint.2017.08.009>.
- Bernard, Paul, Rana Charafeddine, Katherine L. Frohlich, Mark Daniel, Yan Kestens, and Louise Potvin. 2007. “Health Inequalities and Place: A Theoretical Conception of Neighbourhood.” *Social Science & Medicine* 65 (9): 1839–52. <https://doi.org/10.1016/j.socscimed.2007.05.037>.
- Bureau, US Census. n.d. “SAHIE 2008 - 2020 Demographic and Income Model Methodology: Summary for Counties and for States.” <https://www.census.gov/programs-surveys/sahie/technical-documentation/methodology/demographic-income-model.html>.
- “CDC WONDER.” n.d. <https://wonder.cdc.gov/>.
- Entwistle, Barbara. 2007. “Putting People into Place.” *Demography* 44 (4): 687–703. <https://doi.org/10.1353/dem.2007.0045>.
- “Environmental Summaries | IPUMS NHGIS.” n.d. <https://www.nhgis.org/user-resources/environmental-summaries>.
- Hankey, Steve, Julian D. Marshall, and Michael Brauer. 2012. “Health Impacts of the Built Environment: Within-Urban Variability in Physical Inactivity, Air Pollution, and Ischemic Heart Disease Mortality.” *Environmental Health Perspectives* 120 (2): 247–53. <https://doi.org/10.1289/ehp.1103806>.
- Huang, Dian, Bing He, Lai Wei, Liqun Sun, Yangzhong Li, Zengxiang Yan, Xiaoxue Wang, Yuanlei Chen, Qinglan Li, and Shengzhong Feng. 2021. “Impact of Land Cover on Air Pollution at Different Spatial Scales in the Vicinity of Metropolitan Areas.” *Ecological Indicators* 132 (December): 108313. <https://doi.org/10.1016/j.ecolind.2021.108313>.
- Kim, Yonsu, John Cho, Frank Wen, and Simon Choi. 2023. “The Built Environment and Asthma: Los Angeles Case Study.” *Journal of Public Health* 31 (1): 57–64. <https://doi.org/10.1007/s10389-020-01417-6>.
- Lee, Kang-Yun, Sheng-Ming Wu, Hsiao-Yun Kou, Kuan-Yuan Chen, Hsiao-Chi Chuang, Po-Hao Feng, Kian Fan Chung, et al. 2022. “Association of Air Pollution Exposure with Exercise-Induced Oxygen Desaturation in COPD.” *Respiratory Research* 23 (1): 77. <https://doi.org/10.1186/s13601-022-02080-w>.

- //doi.org/10.1186/s12931-022-02000-1.
- Mueller, William, James Milner, Miranda Loh, Sotiris Vardoulakis, and Paul Wilkinson. 2022. “Exposure to Urban Greenspace and Pathways to Respiratory Health: An Exploratory Systematic Review.” *Science of The Total Environment* 829 (July): 154447. <https://doi.org/10.1016/j.scitotenv.2022.154447>.
- “National Land Cover Database Class Legend and Description | Multi-Resolution Land Characteristics (MRLC) Consortium.” n.d. <https://www.mrlc.gov/data/legends/national-land-cover-database-class-legend-and-description>.
- O’Neill, Marie S., Michelle L. Bell, Nalini Ranjit, Luis A. Cifuentes, Dana Loomis, Nelson Gouveia, and Victor H. Borja-Aburto. 2008. “Air Pollution and Mortality in Latin America: The Role of Education.” *Epidemiology* 19 (6): 810–19. <https://doi.org/10.1097/EDE.0b013e3181816528>.
- Simon, Heather, Adam Reff, Benjamin Wells, Jia Xing, and Neil Frank. 2015. “Ozone Trends Across the United States over a Period of Decreasing NOx and VOC Emissions.” *Environmental Science & Technology* 49 (1): 186–95. <https://doi.org/10.1021/es504514z>.
- “Uninsured.” n.d. <https://www.countyhealthrankings.org/explore-health-rankings/county-health-rankings-model/health-factors/clinical-care/access-to-care/uninsured>.
- US EPA, OAR. 2016. “Particulate Matter (PM) Basics.” <https://www.epa.gov/pm-pollution/particulate-matter-pm-basics>.
- Wang, Lan, Rui Chen, Wenyao Sun, Xiaoming Yang, and Xinhua Li. 2019. “Impact of High-Density Urban Built Environment on Chronic Obstructive Pulmonary Disease: A Case Study of Jing’an District, Shanghai.” *International Journal of Environmental Research and Public Health* 17 (1): 252. <https://doi.org/10.3390/ijerph17010252>.
- Wang, Lan, Wenyao Sun, Anne Vernez Moudon, Yong-Guan Zhu, Jinfeng Wang, Pingping Bao, Xiaojing Zhao, et al. 2022. “Deciphering the Impact of Urban Built Environment Density on Respiratory Health Using a Quasi-Cohort Analysis of 5495 Non-Smoking Lung Cancer Cases.” *Science of The Total Environment* 850 (December): 158014. <https://doi.org/10.1016/j.scitotenv.2022.158014>.
- Wong, Chit-Ming, Nuntavarn Vichit-Vadakan, Haidong Kan, and Zhengmin Qian. 2008. “Public Health and Air Pollution in Asia (PAPA): A Multicity Study of Short-Term Effects of Air Pollution on Mortality.” *Environmental Health Perspectives* 116 (9): 1195–1202. <https://doi.org/10.1289/ehp.11257>.
- World Health Assembly, 68. 2015. “Health and the Environment: Addressing the Health Impact of Air Pollution.”