Capstone

Jane Doe

1/25/23

Table of contents

Data Summary

This section is to satisfy L04 Assignment: Summary of Data Use

Summary of Data Sources
as of 2-5-2023

Variable	Unit	Source	Year(s)
Age-adjusted mortality rates: all-cause	County	CDC WONDER	2014 - 2016 (3-year rates)
Age-adjusted mortality rates: ICD J00 - ICD J99	County	CDC WONDER	2014 - 2016 (3-year rates)
Age-adjusted mortality rates: ICD J00 - ICD J99 (65 and over)	County	CDC WONDER	2014 - 2016 (3-year rates)
Tobacco use (SMOKER3 in 1,2,3)	County	CDC Behavioral Risk Factor Surveillance System (BRFSS)	2016
Percent Black	County	American Community Survey, 5-year	2016
Percent Hispanic	County	American Community Survey, 5-year	2016
Percent White	County	American Community Survey, 5-year	2016
Percent Poverty	County	American Community Survey, 5-year	2016
Highest Degree	County	American Community Survey, 5-year	2016
Land Cover	Raster	NLCD	2019
Annual Air Quality: PM 2.5	Monitor	EPA	2014 - 2016
Annual Air Quality: Ozone?	Monitor	EPA	2014 - 2016

Figure 1: Data Summary

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. 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. Survey data from the CDC's Behavioral Risk Factor Surveillance System (BRFSS) is used to understand smoking use by county. One limitation to the BRFSS is that data is collected fro surveys then extrapolated to characterize county-level summararies. 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). The ACS extrapolates based on TALK ABOUT HOW ACS RESULTS ARE GENERATED. 5-year rates for the survey ended in 2016 are used in this model. CURRENTLY TARGETING THESE VARIABLES: Percent poverty, percent Black, percent Hispanic… what else? 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 downloaded from the Multi-Resolution Land Characteristics Consortium website 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 two exposure assumptions: exposure spaces and exposure time. 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. MORE ABOUT THIS. Another assumption, which may be of particular interest to this study, deals with exposure over time. Ambient air pollution may take time to … Indoor versus outdoor.

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, mitigation, and proliferation. 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 respiratory diseases (defined as ICD-10 codes J00 - J99) as the respiratory 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 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 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, five regions were selected for this research, which represent different climate areas and a wide variety of land cover types.

Figure 2 plots the 2015 annual PM 2.5 measures on different scales by region to better evaluate the measurement trends. The west region has the highest annual PM 2.5 readings in two monitors in particular. KEEP EXPLAINING.

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.

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 (NOTATE data integrity as a limitation later). Finally, a third data set limited the respiratory disease age-adjusted mortality rates to the 65 and over population. [ADD GRAPHS for age-adjusted mortality rates and explain.]

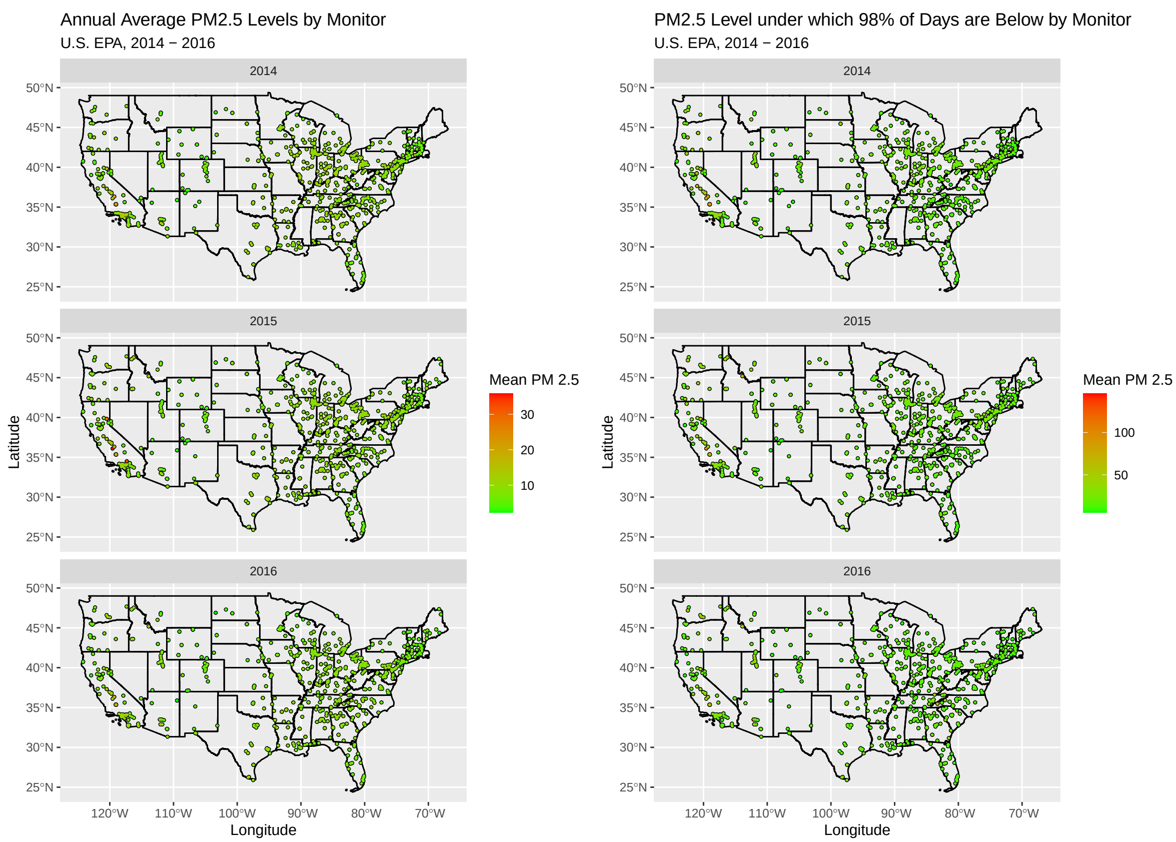


Figure 1.1: Figure 1; Annual summary of air quality metrics by monitoring site from 2014 - 2016

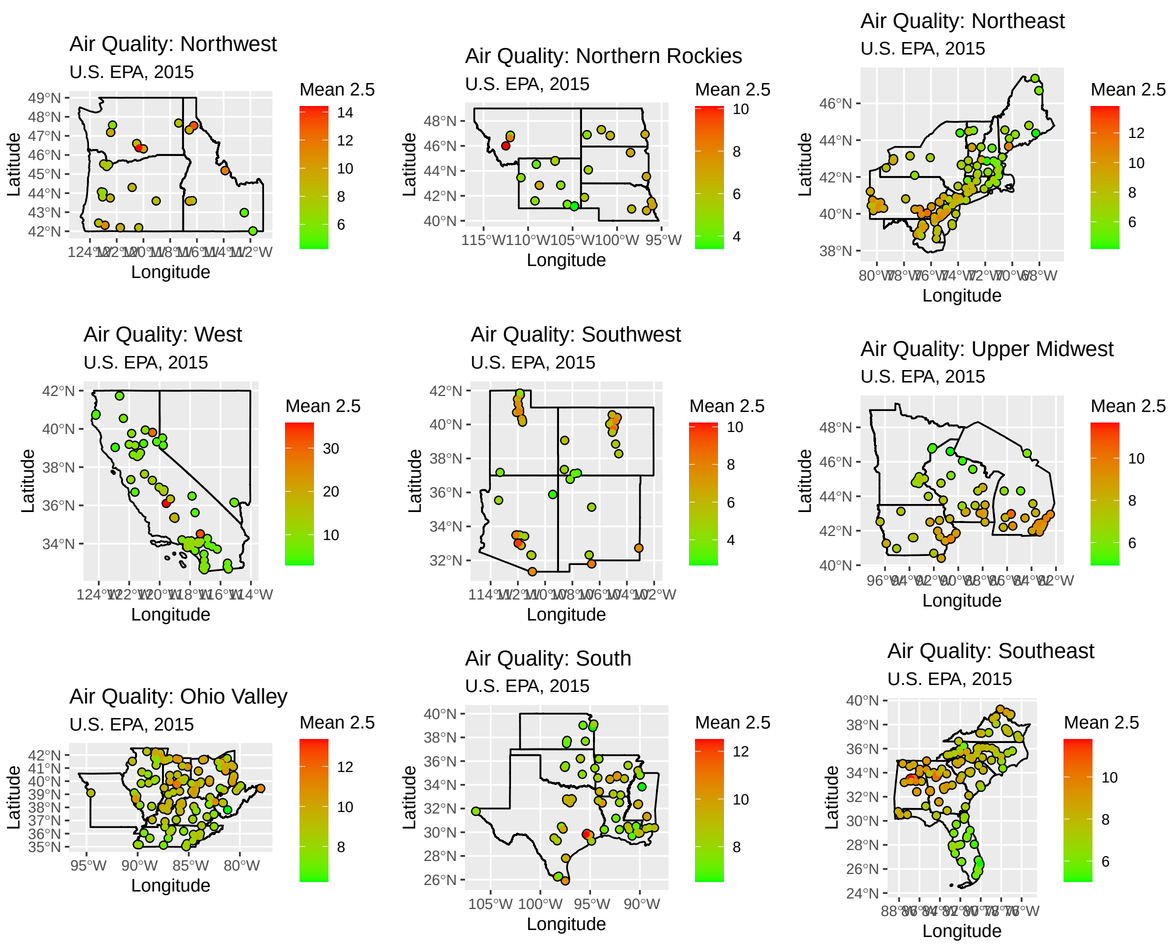
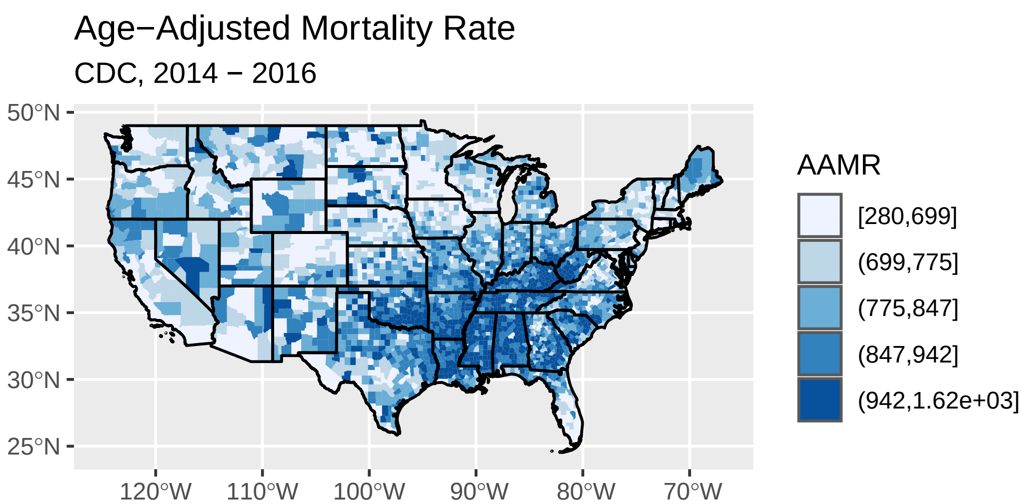
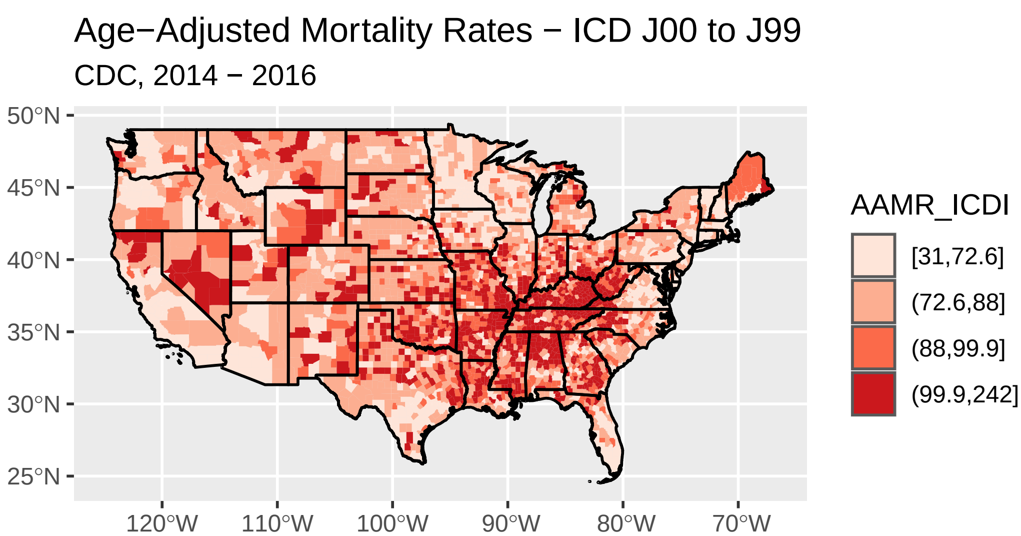
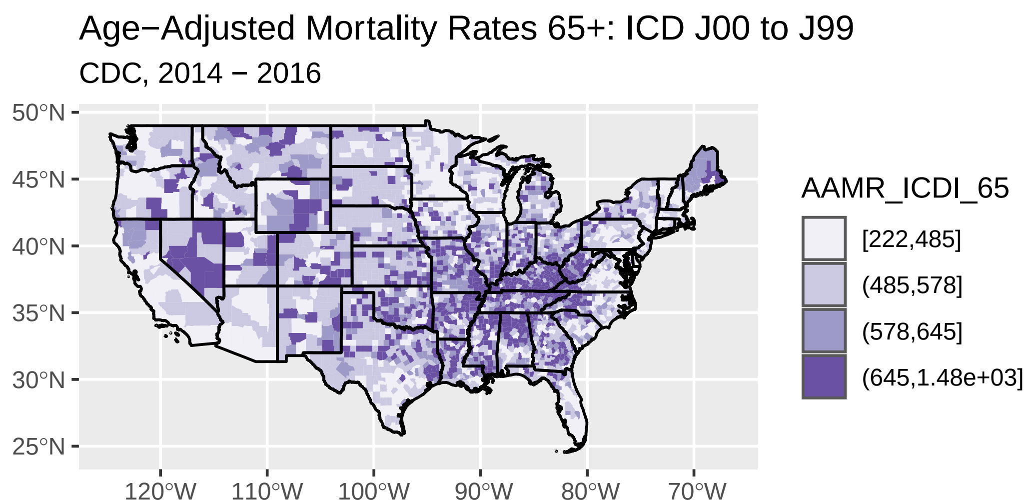


Figure 1.2: Figure 2; Average Annual PM 2.5 by region. Regional P.M 2.5 levels plotted on different scales.


Age−Adjusted Mortality Rate
CDC, 2014 − 2016
50°N
45°N
40°N
35°N
30°N
25°N
120°W
110°W
100°W
90°W
80°W
70°W
AAMR
[280,699]
(699,775]
(775,847]
(847,942]
(942,1.62e+03]


Age−Adjusted Mortality Rates − ICD J00 to J99
CDC, 2014 − 2016
50°N
45°N
40°N
35°N
30°N
25°N
120°W
110°W
100°W
90°W
80°W
70°W
AAMR_ICDI
[31,72.6]
(72.6,88]
(88,99.9]
(99.9,242]

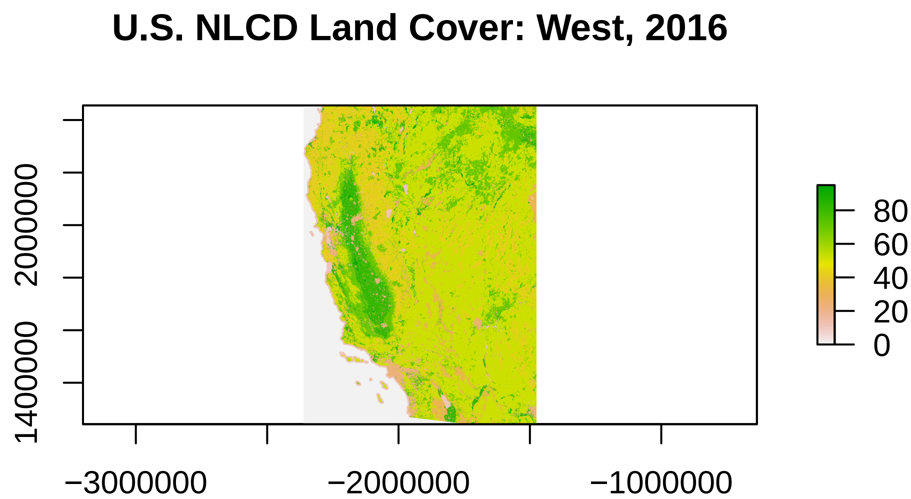


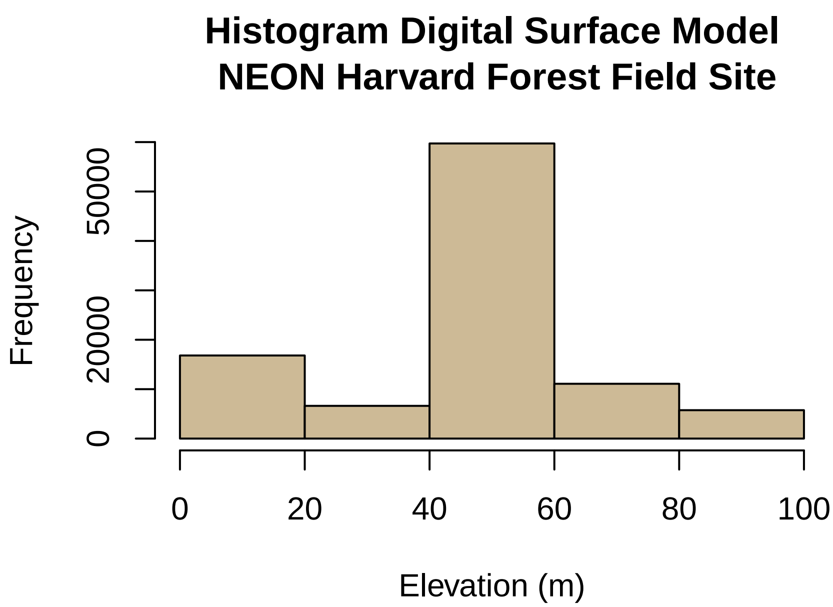
ALSO WHERE CAN I GET TOBACCO USE DATA?

Demographic data are obtained from the U.S. Census Bureau's American Community Survey (ACS). The ACS extrapolates based on TALK ABOUT HOW ACS RESULTS ARE GENERATED. 5-year rates for the survey ended in 2016 are used in this model. CURRENTLY TARGETING THESE VARIABLES: Percent poverty, percent Black, percent Hispanic… what else? 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 downloaded from the Multi-Resolution Land Characteristics Consortium website and further manipulated in R to differentiate primary land cover characteristics by county and by National Ocean and Atmospheric Administration (NOAA)-defined climate regions.

NEED TO SUMMARIZE BELOW… less explanatory?





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. The first model evaluated includes the percent of urban households within a county to distinguish differences in county makeup.

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 model is used to try to produce a better model. GWR model?

2 Summary

In summary, this book has no content whatsoever.

```
1 + 1
```

```
[1] 2
```

References

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.

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/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.

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.

Wang, Lan, Rui Chen, Wenyao Sun, Xiaoming Yang, and Xinhu 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.

World Health Assembly, 68. 2015. “Health and the Environment: Addressing the Health Impact of Air Pollution.”