Understanding Walkability: Calculating the Index and Investigating Its Connection to Diabetes Prevalence

A Preprint

Arkaprabho Bose
Undergraduate Program in Department of Computer Science
Texas A & M University
College Station, TX, 77843
abose0267@tamu.edu

Sebastian Oberg
Undergraduate Program in Department of Computer Science
Texas A & M University
College Station, TX, 77843

Abhinav Cheruvu
Undergraduate Program in Department of Mathematics
Texas A & M University
College Station, TX, 77843

May 6, 2024

Abstract

The diabetes epidemic in the United States presents a nuanced public health challenge, shaped by factors such as socioeconomic status and climate. While the influence of these factors on diabetes is well-established, the role of walkability in managing diabetes prevalence remains contested. This study revisits the relationship between walkability and diabetes in the U.S., using walkability indexes calculated from CDC data. Contrary to some studies suggesting that increased walkability reduces diabetes prevalence, our findings, analyzed through Geographically Weighted Regression (GWR), reveal that walkability is not a significant predictor of diabetes prevalence and exhibits notable regional anomalies. Further analysis using Monte Carlo simulations, Global I Moran's Test, and Variance Inflation Ratio (VIR) supports these results. Our study also critiques the current methods of calculating the walkability index, proposing a revised model that incorporates additional relevant variables from the CDC. This nuanced understanding underscores the need for region-specific urban planning and public health strategies that recognize the complex interplay between walkability, environmental, and socioeconomic factors.

1 Introduction

Diabetes is a common chronic illness that is caused due to consistently high blood sugar levels, and can be prevented through sugar intake management, exercise and dieting. In a study done on 2016 and 2017 National Center for Health Statistics data, it was shown that among adults in the United States, there was a prevalence of 9.7% (Xu, et. al). This high prevalence can impact humans on a daily basis by directly impacting the quality of life both physically and mentally. Diabetes can affect organs all around the body such as the eyes (Lutty,2013). In addition to having direct impact on people, high prevalence of diabetes puts stress on the existing healthcare systems by forcing hospitals and doctors to put resources into solving issues that are preventable.

In recent years, there have been speculations that lifestyle changes, specifically walkability of a region can impact the prevalence of diabetes in that given region. The Environmental Protection Agency has developed a standardized scale on which regions can be ranked based on how walkable it is. The scale ranges from 1-20 with 1 being the least walkable and 20 being the most walkable. It takes into account various things such as intersection density, and proximity to transit (Glazier et al., 2014). According to a temporal analysis study done in 2016, areas with highest walkability score, which is a value calculated had lower rates of diabetes prevalence (Creatore et. al). An area being walkable results in less reliance on cars, and forces the population to walk which is a form of exercise that is often overlooked and can have a meaningful impact on ones health.

As of right now, there is no global standard for calculating something like a walkability index score. This leads to different organizations and institutions coming up with their own, and causing discrepancies in their results. As we will discuss later in the article, the United States standard for walkability index is created by the Environmental Protection Agency. If this really is being used at a government level as the standard, it is important for the methodology they use to calculate their scores to be thoroughly investigated to see if it is truly representative of a places walkability, and whether there are ways to improve their model.

It is crucial to understand the availability index, so that the correct actions can be taken to decrease the prevalence of diabetes in the necessary regions. If regions are showing increase in disease prevalence, and their walkability index is showing it to be incorrect, then people like urban planners will not take the necessary steps to solve the issues.

2 Related Works

2.1 Significance of Walkability on Diabetes Prevalence

The study by Gebel et al. (2019) titled "Walkability and its association with prevalent and incident diabetes in a pooled sample from five German cohorts" highlights the crucial impact of walkability on diabetes prevalence. By integrating Geographically Weighted Regression (GWR) to analyze the relationship between neighborhood walkability and diabetes across Germany, the study finds that higher walkability scores are generally associated with lower rates of diabetes. This supports our hypothesis that improving walkability in U.S. counties could serve as an effective preventive measure against diabetes. These findings are directly applicable to our research, which suggests that enhancements in urban design that prioritize walkability could significantly contribute to public health improvements.

2.2 Comparative Analysis with Similar Research

The study by Booth et al. (2020) "Neighborhood walkability and pre-diabetes incidence in a multiethnic population" provides a significant comparative analysis to our work. It examines the potential of walkable neighborhoods to mitigate early stages of diabetes, using a comprehensive dataset from Southern Ontario, Canada. The use of GWR to map spatial variances in pre-diabetes incidence reveals that walkability can significantly influence diabetes management across diverse populations. This parallels our findings and methodological approach, which emphasizes the importance of considering ethnic and demographic factors when assessing the health impacts of urban walkability.

2.3 Understanding the EPA's National Walkability Index

The "National Walkability Index" developed by the Environmental Protection Agency (EPA) is a comprehensive tool for assessing walkability across the United States. These variables include:

1. **Intersection Density**: Measures the potential connectivity of pedestrian paths by calculating the number of intersections per square kilometer. High intersection density generally indicates more walkable areas as it suggests greater connectivity and options for pedestrian movement.
2. **Proximity to Transit Stops**: Evaluates the accessibility of public transit by measuring the distance from the population center of a block group to the nearest transit stop. Closer proximity implies better walkability as it eases access to public transportation.
3. **Employment Mix**: Assesses the diversity of job types within a walking distance. A diverse employment mix within close proximity can enhance walkability by reducing the need for long commutes and encouraging walking to work.
4. **Employment and Household Mix**: Analyzes the balance between residential locations and job locations. An optimal mix promotes walking by enabling residents to live closer to where they work, which promotes a walkable environment.

By using these variables, the EPA's index not only measures the physical infrastructure that supports walking but also considers how these elements interact with socio-economic factors to affect public health outcomes positively.

2.4 Variations in Walkability Index Methodologies

In the study by Poelman (2022) "Development of an objectively measured walkability index for the Netherlands," a methodology is used that differs from the EPA's approach. This method uses circular buffers, which are zones around a given point, usually the geographic center of a postal code area, to measure how many important amenities are within walking distance. These buffers help in accurately assessing the accessibility of essential services like groceries, parks, and schools, which are significant for day-to-day living. A tailored approach like this highlights the importance of considering localized settings and cultural norms in walkability assessments to ensure that the indices are reflective of the actual conditions experienced by residents.

2.5 Comparative Approaches to Walkability Assessments

In "Planning and Design Support Tools for Walkability," Glazier et al. (2020) piece walkability into categories such as transport and land use, which examines the infrastructural and spatial arrangements that facilitate walking; urban health, focusing on how walkable environments affect physical and mental well-being; and livable cities, which looks at the overall quality of life in urban settings. Each category uses different analytical tools, from spatial analysis by means of Geographic Information Systems (GIS) for detailed mapping, to qualitative methods like surveys to gather resident experiences. This diverse methodology provides urban analysts with robust tools for specific applications and supports comprehensive urban planning that fosters sustainable and healthy communities.

2.6 Innovations in Walkability Index Calculations

The study "Spatial Pattern of the Walkability Index, Walk Score and Walk Score Modification for Elderly" by Schuurman et al. (2020) highlights innovative adaptations of common walkability indices to better cater to the elderly. This adaptation considers factors like gait speed and proximity to age-appropriate facilities, altering traditional indices to reflect the mobility and access needs of older adults. These considerations are important for creating inclusive urban environments that support the independence and health of all citizens, especially as demographic profiles shift towards older populations.

2.7 Context-Specific Walkability Indices

The "Development of a Child-Focused Walkability Index" by Giles-Corti et al. (2013) introduces a specialized approach by highlighting features like traffic volume and the availability of sidewalks. These factors are significant in ensuring that children have safe routes for school and recreational activities, which promotes physical activity from a young age. By focusing on the specific needs of children, this index highlights the necessity for planning urban spaces that cater to the youngest residents, ensuring their safety and well-being in city environments.

3 Methods

3.1 Data

3.1.1 Walkability Index

Walkability Index is a measurement of relative walkability that is developed and calculated by the Environmental Protection Agency. The goal of this measurement is to analyze different parts of the United States on a common scale, and see trends related to walkability. The Environmental Protection Agency calculates this walkability on a scale 1-20 with 1 being the least walkable, and 20 being the most. The actual value on the walkability index scale is calculated using a few different metrics that the EPA collects. The formula used is

$$\text{Walkability Index} = \frac{w}{3} + \frac{x}{3} + \frac{y}{6} + \frac{z}{6}$$

- w is a block group's intersection density, which is calculated by analyzing the number of different types of intersections. Intersections were defined by the types of roads the they connect. The types of road were defined as
  - Auto Oriented: This type of road includes all roads that are geared for automotive use. Examples of this type of road includes highways and tollways where you would not expect pedestrians or bicyclists.
  - Multi-Modal: Roads that are designed so that both automotive and pedestrians can simultaneously use it. This would include roads with separate bike lanes, and roads that fall under a certain speed limit depending on whether the road is one-way or two-way

  - Pedestrian Oriented: Roads that are made primarily for pedestrian use, this can include neighborhood roads, and even paths and trails that auto-motives are not permitted on.

  Based on these road types, intersection densities were calculated based on what types of road any given intersection is connecting. Once these intersection densities are calculated, an overall intersection density for the block group is calculated using a weighted sum of the components. This weighted sum penalizes intersections that are barriers for pedestrians. For example, intersections that connect two auto roads were given a weight of zero since it is expected to have little to no pedestrians. Similarly intersections connecting pedestrian oriented roads were given a high weight since it is expected to encourage and be easy for pedestrians to use

- x is a measurement of the distance to the closest transit stop in meters. Specifically, geoprocessing models were used to calculate the distance from the block group centroid to the nearest transit stop. This value can directly impact walkability since people who live in places with higher walkability tend to rely less on cars and more on public transport whether it's by bus or train.
- y is a block group's employment density mix. This is a variable that calculates the entropy of a block group based on the different jobs that are available in a certain area. This is one of the chosen factors for calculating walkability index since it takes into account the diversity of the area. In this case, a higher entropy would indicate a more diverse distribution of employment groups in the block group. The idea is that it will be a highly walkable place if there is a lot of different types of stores and businesses.
- z is a block group's employment and household mix. This is a measurement of the diversity in a given region, in terms of businesses and occupied housing. Similar to the employment density mix, this measurement uses the entropy of businesses and housing to calculate an overall entropy value. This can be used to calculate a walkability score since it implies that with businesses and homes being close together, it is not necessary for employees to take a car to get to work, but rather take a form of public transit.

All formulas and variables were derived from the Environmental Protection Agency Smart Location Database(Chapman et. al, 2021)

3.1.2 Other Covariates

In addition to walkability, we felt it was important to account for a few different variables when conducting our study. We decided that based on our research, it would be important to take into account several different health factors that can impact diabetes prevalence. According to an article by Mokdad et. al, there is a strong association between diabetes and obesity, and obesity is tied into several other health risks such as high blood pressure, and low physical activity. As a response to this research, we decided to take common health issues into account as covariates for our model. The health complications that we focused on was the obesity, high blood pressure, smoking prevalence, and low-physical activity. All of the chosen factors have some sort of researched impact on diabetes prevalence and will allow for more thorough results. In addition to this, we took into account the median household income for all of the counties that we were working with, as well as the average temperature. The health factors data was provided by the CDC as part of their annual report (Center for Disease Control and Prevention, 2023), and the median household income data was provided by the United States Census Bureau, as part of their Small Area Income and Poverty Estimates Program (United States Census Bureau, 2022). The average temperature data was found from the GIS for Racial Equity Website (Gilbert, 2023). Together these values will help lead to our model fitting to be more complete and allow for more of the data to be explained by the model.

3.1.3 Data Cleaning

Originally, we had started working with 4 different datasets, which we then had to aggregate into one large dataset. The datasets had a lot of data since most came from annual government reports, so it was important to select only the relevant ones which were the spatial information and the chosen covariates. Each of the datasets that we had decided to use had originally been at the block group level. This was an issue as it led to at least one value of relevant data being missing in some block groups. Our solution for this issue was to average out the relevant columns from each dataset at the county level. This took care of the missing values as if one block group in a county was missing data, it was dropped from the calculation and there were other block groups which could be used as part of the average. In order to better fit median household income as a covariate, we log transformed the values to make the values on a more reasonable scale along with the rest of our predictors.

3.2 Analysis

3.2.1 Correlation Matrix

In order to get a general understanding of the data at hand, we generated a correlation matrix to see the relationships between each of the variables. This was also used as an exploratory tool to see if there was any multicollinearity within our data. In the case of our project, we used a threshold of 0.8 to find variables with concerning similarities, and depending on the values in the correlation matrix, conducted further analysis to ensure that the values that were highly correlated were not impacting the results of the model heavily.

3.2.2 Spatial Analysis

Based on the type of the data that was available, we chose to do spatial analysis because we believed it would yield the best result. The model we used in this study was Geographically Weighted Regression. Specifically, we used the model provided by the GWModel Package (Gollini et. al, 2015). This model, originally created by Brunsdon, is based on the formula

$$y_i = \alpha_0 + \sum_{k=1}^{m} \alpha_k x_{ik} + \varepsilon_i$$

- In the case of this formula:
  - y_i is the ith observation of the dependent variable
  - α_0 is the intercept
  - x_{ik} is the ith observation of the kth independent variable
  - ε_i is the normally distributed independent error terms
  - α_k is estimated from n observations

This model was chosen for this specific problem because it is able to improve on regular regression models like global regression, and account for spatial heterogeneity. In the case of our data, county level data exhibits spatial heterogeneity due to how small the counties are relative to the size of the United States. Initially, we had wondered whether it would make sense to use this model in a predictive context, but eventually decided that in order to investigate the relationship of walkability on the diabetes prevalence, it would make more sense to use the model in an inferential context. The GWModel package allowed for us to set a formula where we set the diabetes prevalence as the response variable and the covariates as the independent variables. The package provides a function that, based on the inputted data identifies an ideal bandwidth to use for the model. This bandwidth can then be used to fit the model. Once the model was fit, it was important to adjust the P-values to deal with insignificant coefficients to make the visualization of the results more intuitive. Using the GWModel package, the p-values were adjusted using the Fotheringham-Byrne procedures (Gollini et. al,2015). Once this process was completed, coefficients that had a significance level greater than 0.05 were zeroed out for plotting purposes to show only the statistically significant coefficients in the impact plot.

3.2.3 Impact Plots

Once the insignificant coefficients were set to 0, we wanted to understand how each covariate impacted the fitted diabetes prevalence value in the GWR model. In order to do this, we used the original observed diabetes data for each county, and multiplied it by the predicted coefficients. It is important to note that since we adjusted the P-values, there were a lot of cases where the impact was calculated as 0 due to the adjusted coefficient. Once this calculation was done, it was plotted spatially. Once these plots were made, we were able to make inferences about the walkability index and it's relationship to diabetes.

3.3 Validation

3.3.1 Simulation Study

We conducted a simulation study to ensure that the model we had chosen was truly a good model to fit on our type of data, and was not a random fluke due to specific values. We conducted the simulation study by using the original dataset along with spatially varying functions. The goal here was to create realistic spatially varying coefficients for each county, and that would be considered the "true" coefficient. The coefficients were generated using a method that we wrote, which accepts the latitude and longitude of a location, and generates realistic coefficients for each covariate with both the longitude and latitude being part of the calculation. This means that as each county's spatial area changes, the impact of that covariate will change proportionally to the distance. Once these "true" coefficients were calculated, we calculated our "true" diabetes prevalence which was done by summing the products of the "true" coefficient and the observed value for that county. This "true" diabetes prevalence was then passed into a GWR model as the response with the original observed values as the independent variables. Success of this model was measured using the adjusted Rˆ2 values, as well as a mean absolute error plot of the "true" coefficients and the predicted coefficients

3.3.2 Diagnostics

In addition to the simulation study, we also ran a few different tests to not only verify the validity of our model, but also as a way to investigate why some of our results looked the way they did. The first test that we ran was a Monte

Carlo simulation on the residuals of our model. The Monte Carlo simulation is conducted to identify which variables are statistically significant. We analyzed the p-values to see which of the chosen covariates are significantly spatially varying to see if any of the covariates used could have been left out. In addition to this, we used a Moran's I Test as a way to test for spatial autocorrelation.

4 Results

4.1 Validation

4.1.1 Monte Carlo Significance Test

We initially conducted a Monte Carlo Significance Test to assess spatial variation in each of our covariates. Table 1 below presents the results of this test. It reveals that the National Walkability Index has a p-value of 0.35, indicating that it is not significantly spatially varying. Consequently, if we were to model the National Walkability Index on a global scale, it would not substantially alter our model's results. This finding reinforces our argument that the Walkability Index provided by the CDC may overlook other crucial factors that could accurately explain the relationship between walkability and diabetes prevalence.

Table 1: Results of Monte Carlo Simulation

Covariates	P_Value
Intercept	0.39
National Walkability Index	0.35
Obesity Prevalence	0.01
High Blood Pressure Prevalence	0.00
Low Physical Activity Prevalence	0.00
Current Smoking Prevalence	0.00
Median Household Income	0.94
Average Temperature	0.00

4.1.2 Global Moran's I Test on Residuals

Next we conducted a Global Moran's I Test on the Residuals of our GWR model to check for spatial autocorrelation. From the Global Moran's I Test in Table 2 we obtained a Moran's I Statistic of 0.045. This indicates that there is slight positive autocorrelation in our model. We expect this to occur, as some of our covariates are similar in nature(eg. obesity, smoking, high blood pressure,etc.). Ultimately, there is not significant autocorrelation or an indication of high spatial dependence further showing that our model is indeed accurate.

Table 2: Moran's I Test Results

	Value
Moran I statistic	0.0448723
Expectation	-0.0003249
Variance	0.0001135

4.1.3 Multicollinearity

In addition, we examined if there was any significant multicollinearity in our GWR model. From the Correlation Matrix from Figure 2, we can see that the only highly correlated variables included the health risk factors such as diabetes, obesity rates, etc. However, these health risk factors were incorporated into our model, providing valuable insights into the multifaceted nature of diabetes prevalence and its associated risks. These covariates can account for spatial variation and capture an accurate representation of diabetes prevalence. In addition, the importance of including these risk factors is further explained by the paper M. I. Creatore et al., "Association of Neighborhood Walkability With Change in Overweight, Obesity, and Diabetes". In addition, we tested the multicollinearity of our GWR model by checking the Variance Inflation Factor(VIF) for each of our covariates. For each of our covariates, none of our values exceeded a VIF score of 13. Thus, from the Correlation Matrix and when checking the Variance Inflation Factor of each of the covariates, there is not significant multicollinearity that poses an issue in our GWR model , further indicating the validity of our model.Lastly, Figure 5 depicted in the Appendix illustrates evenly dispersed residuals across the United States, confirming the validity of our GWR model.

4.1.4 Residuals

The residual plot in the Appendix shows a fairly evenly scattered distribution of predicted values around zero, indicating a well-fitted model. This further emphasizes the validity of our model.

4.2 Simulation Study using Artificial Data

From the simulation study using artificial data, we obtained results that demonstrate the strength of our GWR model. Our results depicted below provide valid estimates and coefficients when fitting a GWR on our simulated data. In our simulation study, the Mean Absolute Error ranged from 0.13 to 1.41 for all covariates aside from Median Household Income. The Mean Absolute Error of the predicted coefficient for Median Household Income from our simulated data when fitting a GWR model was 14.22. This relatively high Mean Absolute Error we assume is from the discrepancy in coastal areas where there are extremely high cost of living areas relative to the average cost of living. This also supports our previous diagnostics where the model struggled to fit the Walkability Index Score and the Median Household Income as they were both not significantly spatially varying. However, the other covariates had a relatively low Mean Absolute Error close to 0 suggesting that they are significantly spatially varying, suggesting the validity of our GWR model on real world. Additionally, we obtained a relatively high R-squared value of 0.91 when fitting a GWR model on our artificial data, demonstrating strong explanatory power.

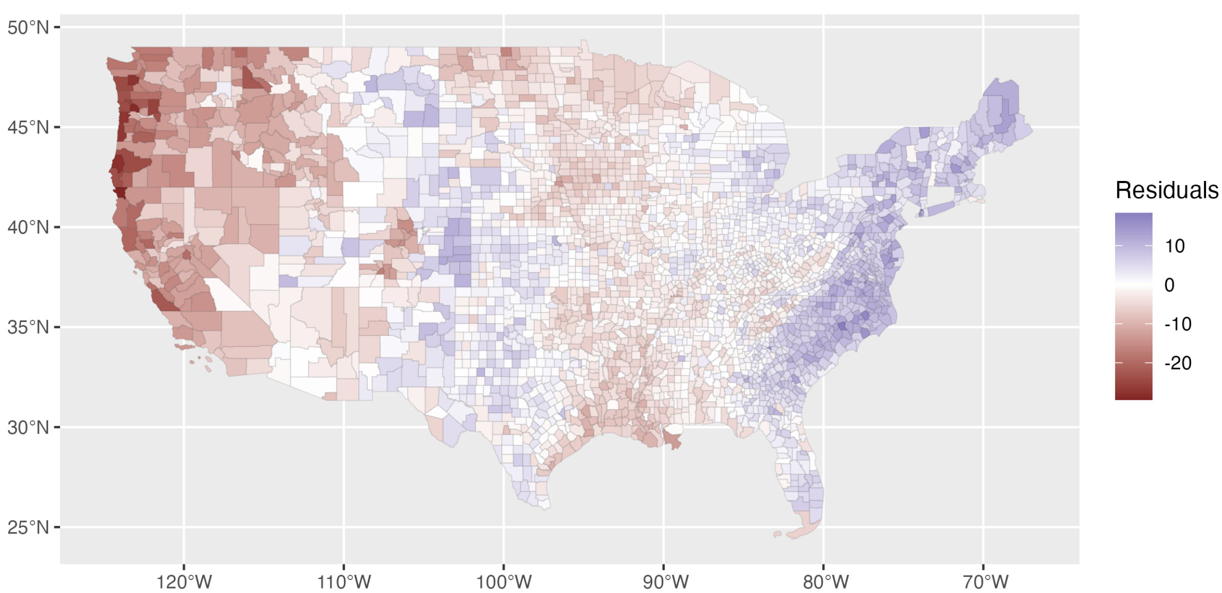

Figure 1: Residuals of GWR Model Results on Simulated Data

Table 3: Mean Absolute Error of Coefficient Predictions for Simulated Data

Covariates	Mean Absolute Error
National Walkability Index	1.4125221
Obesity Crude Prevalence	0.1359228

Covariates	Mean Absolute Error
High Blood Pressure Crude Prevalence	0.1878284
Low Physical Activity Crude Prevalence	0.4585690
Smoking Crude Prevalence	0.6344249
Average Temperature	0.9496559
Median Household Income	14.2233776

4.3 Correlation Matrix

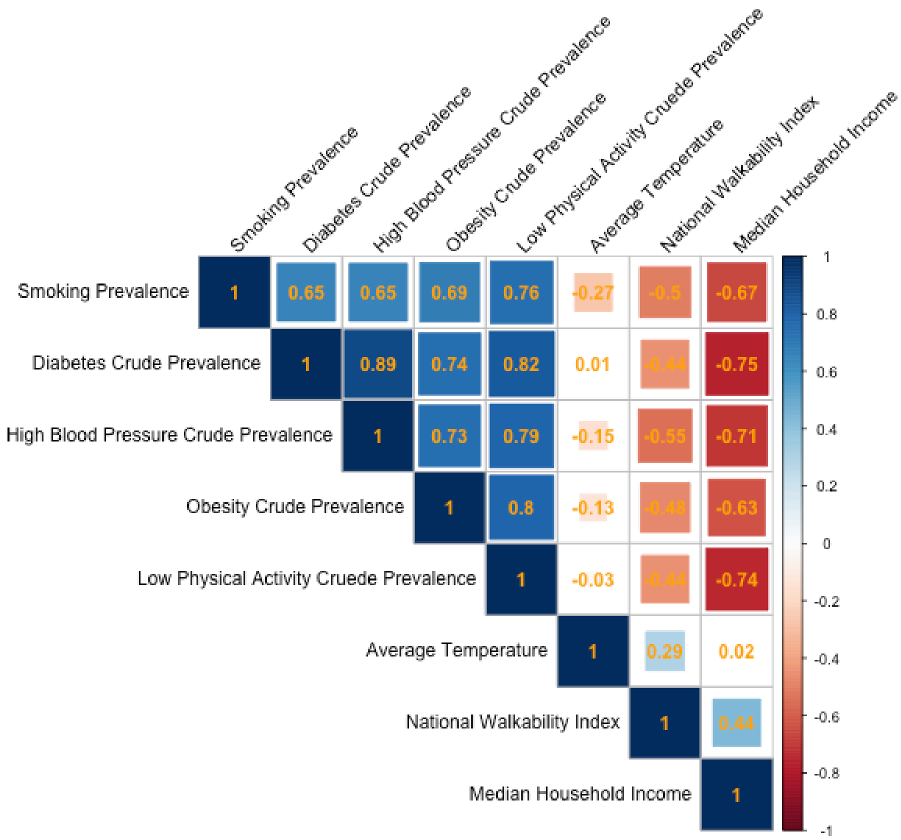

Figure 2: Correlation Matrix of Chosen Variables

From the Analysis of our Correlation Matrix plot above in Figure 2, we observed several factors highly correlated with Diabetes Prevalence, notably High Blood Pressure (correlation of 0.89) and Obesity (correlation of 0.74). These health risk factors were incorporated into our model, providing valuable insights into the multifaceted nature of diabetes prevalence and its associated risks. While these factors demonstrate strong positive associations, they are still important to include. As explained previously in our validation section and in the paper M. I. Creatore et al., "Association of Neighborhood Walkability With Change in Overweight, Obesity, and Diabetes", these factors are indeed important to include in our GWR model. Other covariates such as Average Temperature (correlation of 0.01) and National Walkability Index (correlation of -0.44) were also included in our analysis. Despite their lower correlations with Diabetes Prevalence, these variables offer diverse perspectives and contribute to a comprehensive understanding of the relationships between various covariates and their impacts on diabetes prevalence in the U.S.. Including these factors, allows us to explore the intricate interplay between different factors, providing a more nuanced understanding in shaping diabetes prevalence trends across the entirety of the U.S.

4.4 The Impact of Walkability on Diabetes Prevalence

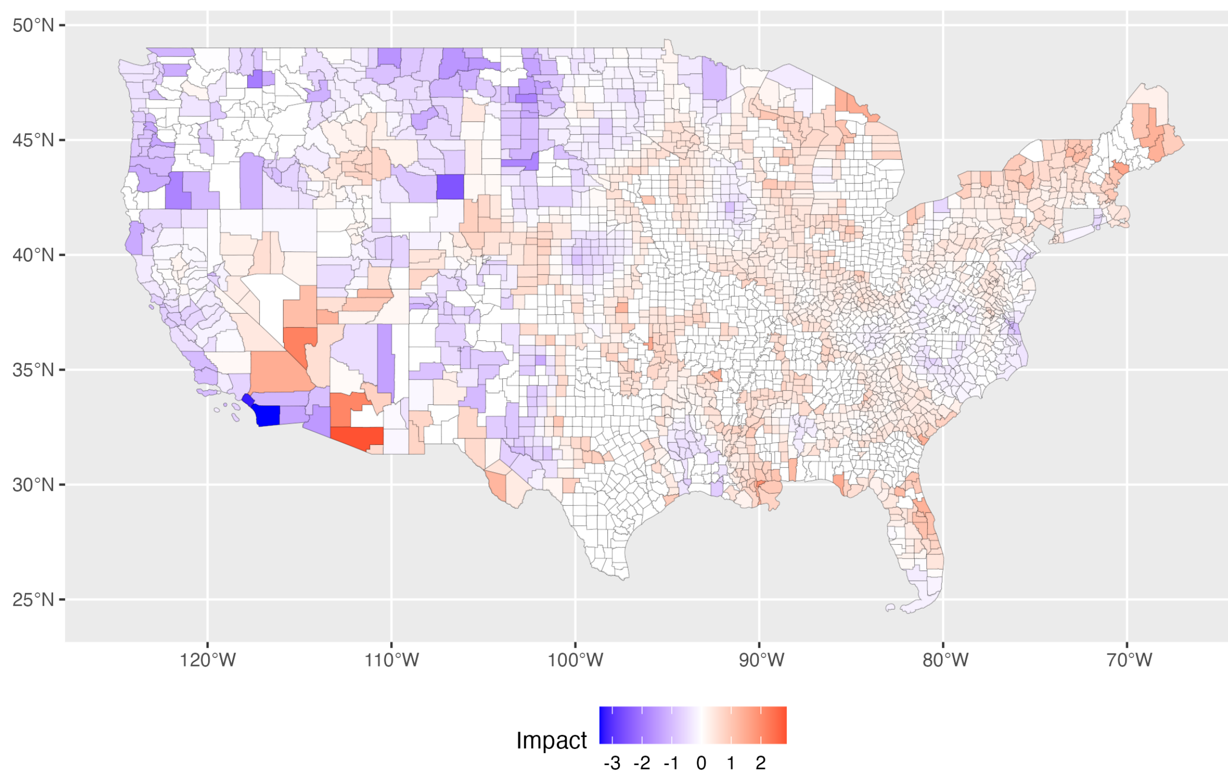

Figure 3: Impact of Walkability on Diabetes Prevalence: Observed * Predicted Coefficients

When fitting the GWR model on our real-world data, we expected there to be a consistent trend between walkability and diabetes prevalence across the U.S. Accounting for walkability, multiple health risk factors, temperature, income,etc, we expected there to be a consistent trend between walkability and diabetes prevalence. However, from Figure 1 above, we can see that this is not the case. As seen in the West Coast, Pacific North West, and the Mountainous Regions in the Northern part of the U.S, there is a negative correlation between walkability and diabetes prevalence. However, in the East Coast and Southern Regions of the U.S there is a positive correlation between walkability and diabetes prevalence. Ultimately, since there is a lack of consistency across the U.S we believe that the walkability index score calculated by the CDC is inaccurate and fails to account for certain additional factors to explain the true impact of walkability on diabetes prevalence.

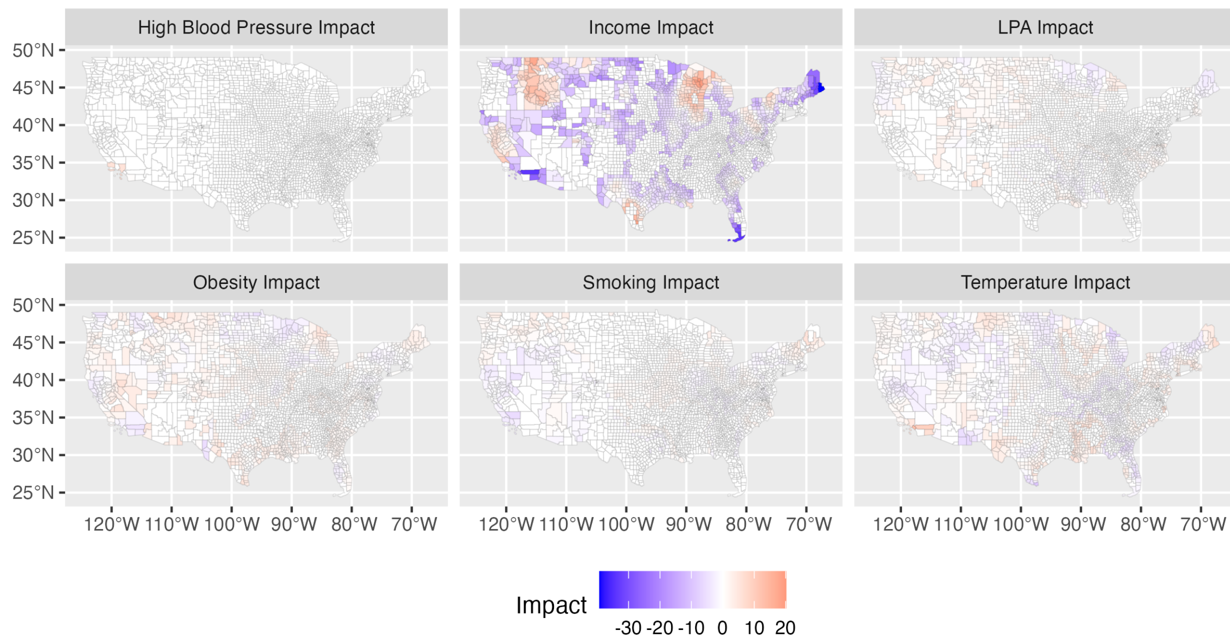

Figure 4: Impact of Covariates on Diabetes Prevalence: Observed * Predicted Coefficients

In summary, the insights from our model, both from simulated and real-world data, shed light on the complex interplay between walkability, various risk factors, and diabetes prevalence across different regions. The consistency between metrics obtained from our simulation study and real data, coupled with the absence of multicollinearity issues and spatial dependence , underscores the reliability and validity of our findings, supporting the robustness of our approach. Based on the results of the model along with the plots that we made, we believe that the current formula used by the EPA and other government agencies fail to consider many extraneous factors, and the formula should be revised.

5 Discussion

5.1 Possible explanations for The Variation of Diabetes Prevalence in Different Regions

The correlation between walkability and diabetes prevalence in the South can be attributed to various factors, with higher temperatures emerging as a key consideration. In warmer climates, such as those prevalent in the South, the positive relationship between walkability and diabetes may be influenced by people spending more time indoors to avoid the heat. This reduction in outdoor activity diminishes walkability and could potentially contribute to higher diabetes rates.Conversely, in colder regions like the west coast and the Pacific Northwest, the impact of walkability on diabetes prevalence appears to be negative, as depicted in the plot. This suggests that regional differences, including climate variations, play a significant role in shaping the relationship between walkability and diabetes. Moreover, the analysis identified various additional risk factors, notably health-related ones, contributing to elevated diabetes prevalence nationwide. From the facet plot depicted below, factors like smoking and obesity showed clear associations with higher rates of diabetes, as expected given their impact on overall health and predisposition to chronic conditions like diabetes.

5.2 Revisiting the EPA's Walkability Index in Light of Regional Variations

Our analysis shows that the walkability index, as formulated by the EPA, does not consistently predict diabetes outcomes across different U.S. regions. This finding reflects the challenges mentioned in the "National Walkability Index User Guide and Methodology," from the Environmental Protection Agency where variables like intersection density and land use diversity may not sufficiently capture regional factors that influence health outcomes.The related work "Development of an objectively measured walkability index for the Netherlands" suggests that walkability assessments must use local urban layouts and societal norms to enhance their relevance and accuracy. Taking this into account, our study proposes that additional variables specific to regional characteristics in the U.S. should be considered to improve the predictive power of the walkability index concerning diabetes prevalence.

5.3 Aligning Methodological Approaches with Urban Health Outcomes

The inconsistencies we observed in walkability's impact on diabetes prevalence across different U.S. regions highlight the need for a more nuanced approach to walkability assessments, as suggested by the diverse methodologies mention in "Planning and Design Support Tools for Walkability." (Glazier et. al,2020). This source suggests using tools that adapt to different city environments, emphasizing the need to customize walkability indices to both physical and social aspects of cities. In addition, the study "Spatial Pattern of the Walkability Index, Walk Score, and Walk Score Modification for Elderly" shows the importance of adapting indices to specific demographic needs, such as adjusting for elderly populations. Similarly, our findings advocate for the adaptation of walkability indices to better address the health implications of walkability in regions with different urban dynamics and demographic profiles.

5.4 Enhancing Walkability Indices through Tailored Assessments

Our research further confirms the belief from "Development of a Child-Focused Walkability Index" that walkability assessments should consider demographic-specific needs (Giles-Corti et al.,2013). Just as traffic volume and sidewalk presence are crucial for assessing children mobility in urban areas, our results suggest that factors like community health services access, food landscapes, and recreational spaces might be significant predictors of diabetes prevalence related to walkability. This perspective aligns with our observation that the existing walkability indices may overlook critical elements that influence health outcomes in varied urban settings.

5.5 Proposed Modifications to Current Walkability Calculations

Given the regional anomalies noted in our study, we propose several modifications to the current methods of calculating the walkability index. These include incorporating health-related facilities access and community socio-economic status to provide a more comprehensive method of how walkability influences diabetes prevalence. This approach is inspired by the modifications discussed in the "Spatial Pattern of the Walkability Index, Walk Score and Walk Score Modification for Elderly," which adapted traditional metrics to better suit elderly residents by considering their specific mobility and access needs (Schuurman et al.,2020).

5.5.1 Conclusion

In conclusion, our study shows that we need to improve the walkability index by including more detailed factors that consider both the physical layout and the social and economic conditions of neighborhoods. By adding these details to walkability assessments, city planners and health officials can better understand how city design affects public health, and work more effectively to reduce health risks related to areas that are hard to walk in. Our research points out the complex ways in which city planning and public health are connected, suggesting that city planning should take a more detailed and specific approach that meets the diverse needs of different communities. Looking ahead, more studies are needed that follow how changes in city planning affect public health over time. This will provide solid evidence to guide decisions in making cities healthier and more walkable.

6 Appendix

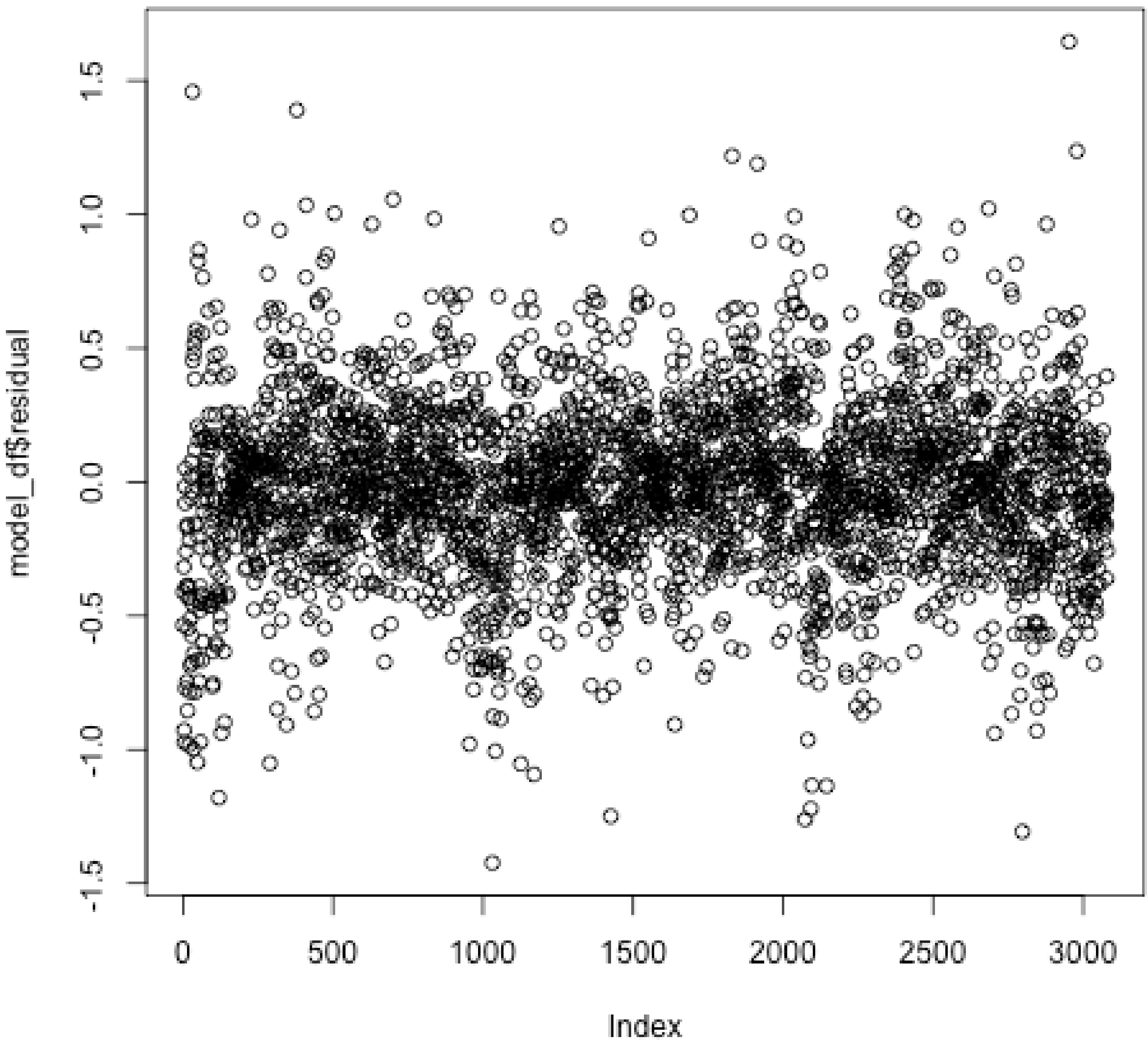

Figure 5: Resdiual Plot of GWR Model on Real Data

7 References

- G. Xu et al., "Prevalence of diagnosed type 1 and type 2 diabetes among US adults in 2016 and 2017: population based study,"
- M. I. Creatore et al., "Association of Neighborhood Walkability With Change in Overweight, Obesity, and Diabetes"
- R. H. Glazier et al., "Density, Destinations or Both? A Comparison of Measures of Walkability in Relation to Transportation Behaviors, Obesity and Diabetes in Toronto, Canada," PLoS ONE, vol. 9, no. 1, p. e85295, Jan. 2014, doi: 10.1371/journal.pone.0085295.
- M. A. Lazar, "How Obesity Causes Diabetes: Not a Tall Tale," Science, vol. 307, no. 5708, pp. 373–375, Jan. 2005, doi: 10.1126/science.1104342.
- C. Brunsdon, A. S. Fotheringham, and M. E. Charlton, "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity," Geographical Analysis, vol. 28, no. 4, pp. 281–298, Oct. 1996, doi: 10.1111/j.1538-4632.1996.tb00936.x.
- Byrne, Graeme & Charlton, Martin & Fotheringham, Alexander. (2009). Multiple Dependent Hypothesis Tests in Geographically Weighted Regression.
- Bureau, US Census. "Small Area Income and Poverty Estimates (SAIPE) Program." Census.Gov, 1 July 2022, www.census.gov/programs-surveys/saipe.html.
- U.S. Environmental Protection Agency, "National Walkability Index User Guide and Methodology." Available: EPA Walkability Index User Guide and Methodology
- M. Poelman, "Development of an objectively measured walkability index for the Netherlands," International Journal of Behavioral Nutrition and Physical Activity, 2022. Available: Development of an objectively measured walkability index for the Netherlands
- R. H. Glazier et al., "Planning and Design Support Tools for Walkability," Sustainability, vol. 12, no. 11, May 2020. Available: Planning and Design Support Tools for Walkability
- V. Schuurman et al., "Spatial Pattern of the Walkability Index, Walk Score and Walk Score Modification for Elderly," International Journal of Health Geographics, vol. 11, no. 5, 2020. Available: Spatial Pattern of the Walkability Index, Walk Score and Walk Score Modification for Elderly
- L. Giles-Corti et al., "Development of a Child-Focused Walkability Index," International Journal of Health Geographics, vol. 12, 2013. Available: Development of a Child-Focused Walkability Index
- A. H. Mokdad et al., "Prevalence of Obesity, Diabetes, and Obesity-Related Health Risk Factors, 2001," JAMA, vol. 289, no. 1, p. 76, Jan. 2003, doi: 10.1001/jama.289.1.76.
- J. Chapman, E. Fox, W. Bachman, L. Frank, J. Thomas, and A. Rourk Reyes, "Smart Location Database Technical Documentation and User Guide Version 3.0," Jun. 2021, [Online]. Available: EPA Smart Location Database
- I. Gollini, B. Lu, M. Charlton, C. Brunsdon, and P. Harris, "GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models," J. Stat. Soft., vol. 63, no. 17, 2015, doi: 10.18637/jss.v063.i17.
- "Places: Census Tract Data (GIS Friendly Format), 2021 Release." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 2023, data.cdc.gov/500-Cities-Places/PLACES-Census-Tract-Data-GIS-Friendly-Format-2021-/mb5y-ytti/about_data.
- Gilbert, Mark. "Heat Health Census Tracts." GIS For Racial Equity, gis-for-racialequity.hub.arcgis.com/datasets/climatesolutions::heat-health-census-tracts/about. Accessed 3 May 2024.
- S. Gebel, M. Rapp, A. Voigtländer, A. Schneider, R. Döring, and S. Karrasch, "Walkability and its association with prevalent and incident diabetes in a pooled sample from five German cohorts," BMC Endocrine Disorders, vol. 19, no. 101, 2019, doi: 10.1186/s12902-019-0485-x
- J. Booth, J. R. Parks, and T. L. Frost, "Neighborhood walkability and pre-diabetes incidence in a multiethnic population," BMJ Open Diabetes Research & Care, vol. 8, no. 1, 2020, doi: 10.1136/bmjdrc-2020-001533 -G. A. Lutty, "Effects of Diabetes on the Eye," Invest. Ophthalmol. Vis. Sci., vol. 54, no. 14, p. ORSF81, Dec. 2013, doi: 10.1167/iovs.13-12979.