STAT4000 Final Project: An analysis on walkability in the US

Sean Shi

December 15, 2023

1 Motivation

Walking, biking, and other forms of micro-transportation have many health benefits and produces less pollution compared opposed to driving. With all these benefits, these other forms of transportation(i.e. not driving) should see a lot of use. However, especially in the United States, walkability often varies greatly from location to location, and this can have a detrimental effect on the number of people who benefit from walking or other forms of micro-transportation.

In this paper, we seek to analyze the state of walkability across the United States using the National Walkability Index(see below) and its correlation to other variables in the Smart Location Database. [1] By additionally grouping the data into a group with good walkability and one with bad walkability, we will seek to analyze the differences between these blocks

The dataset used in this analysis is the Smart Location Database [2] first released by the EPA in 2011 and most recently updated in 2021. This database aggregates data from multiple sources including previous censuses, geological surveys, and HERE maps. Each entry in the database represents a Census block group. In this paper, we will seek to analyze the distributions of various factors and their correlations with walkability.

2 Methods

2.1 Analysis of the National Walkabilty Index

The National Walkability Index(NatWlkInd) referenced in this report is calculated with the following formula described in the *National Walkability Index:Methodology and User Guide* [3]

w = block group's ranked score for intersection density
x = block group's ranked score for proximity to transit stops
y = block group's ranked score for employment mix
z = block group's ranked score for employment and household mix

$$w/3 + x/3 + y/6 + z/6$$

A histogram of the walkability index(see below) shows that the distribution for the index is bimodal rather than normal. Additionally, note that more than half of the observations are Considered Below average walkable or lower.

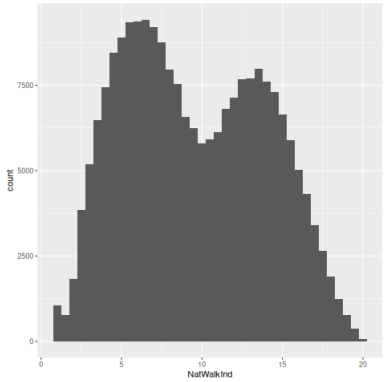
[1] An explanation of all variables in the database can be found here: https://www.epa.gov/system/files/documents/2023-10/epa_sld_3.0_technicaldocumentationuserguide_may2021_0.pdf

[2] "Walkability Index." Walkability Index, Responsible Party U.S. Environmental Protection Agency, Office of Sustainable Communities (Point of Contact), 30 Aug. 2023, catalog.data.gov/dataset/walkability-index1.

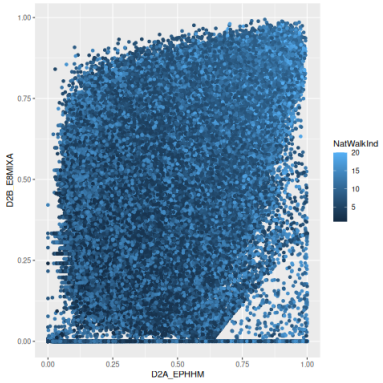
[3] Thomas, John, et al. "National Walkability Index Methodology and User Guide." *National Walkability Index: Methodology and User Guide*, June 2021, www.epa.gov/sites/default/files/2021-06/documents/national_walkability_index_methodology_and_user_guide_june2021.pdf.

Table 1: Interpretation of Walkability *taken from National Walkability Index User Guide and Methodology* [2]

1 – 5.75	Least walkable
5.76 – 10.5	Below average walkable
10.51 – 15.25	Above average walkable
15.26 – 20	Most walkable

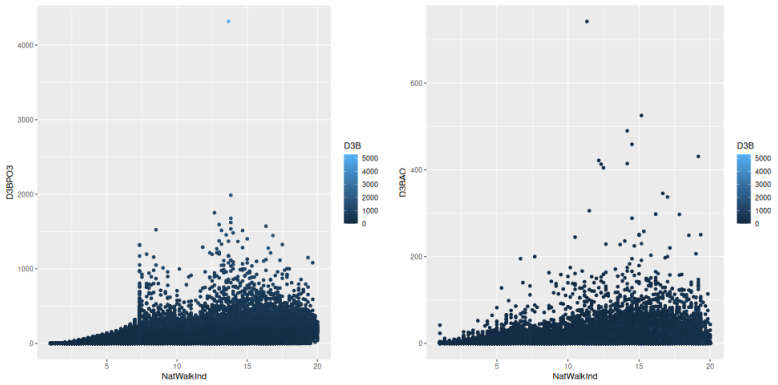


Additionally, note that the variables used to create the index are not independent. In particular, employment mix and employment and household mix have a correlation of 0.504 which would would suggest that a significant amount of the variance in one can be explained by the other.

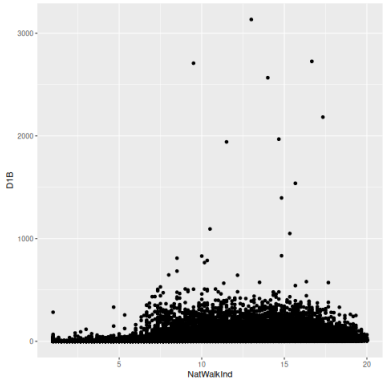


2.2 Relation with other variables

When analyzing the correlation between the National Walkability index and other variables, we can see some interesting trends. For example, The national walkability index has a moderate correlation(0.501) with the density of pedestrian intersections per square mile(D3bpo3), but a very weak correlation(0.147) with the density of automotive intersections per square mile(D3bao) as shown in the graphs below



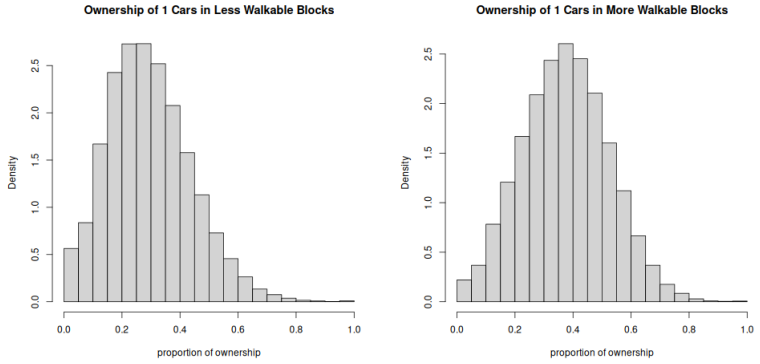
Interestingly the correlation between population density and walkability is weak(0.268) which indicates that these 2 variables are fairly independent. The plot between these is shown below, and it suggests that if walkability is low, density is low, but not the other way around.



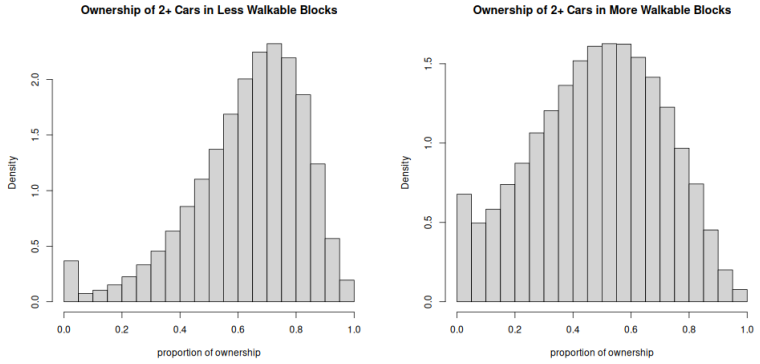
2.3 Differences Between Low and High Walkability

In this section, we seek to examine the difference between more and less walkable blocks which are split based on if the Walkability index of a block is above or below 10.5.

Examining the proportion of households with 1 automobile, we can see that this is approximately normally distributed for both groups.



The proportion of households owning 2 or more automobiles is also approximately normally distributed for both.



We can see from the graphs that the proportion of 1 car households seems greater for the second group and the proportion of 2 or more cars households seems greater for the first. Since these are approximately normal and assuming the observations are independent, we can use a t-test for difference in means to test these hypotheses. With these tests, we get a 95% confidence interval for the difference between the less walkable group and more walkable group for proportion of 1 car households to be $(-\inf, -0.07890386)$ which does not include 0. thus, we can say we are 95% confident the true proportion of households owning 2 or more cars is greater for households in less walkable blocks. Similarly, we get a 95% confidence interval of $(0.1476029, \inf)$ and so can say we are 95% confident that the average proportion of households owning 1 car is greater for less walkable blocks.

Note that a difference in means between the groups for the proportion of households owning 0 cars would have to use different methods, because the data is not normally distributed, but rather heavily skewed right with the mode of each distribution at 0.

3 Conclusions

as shown in the methodology, there is a positive correlation between density of pedestrian oriented intersections and walkability. This may be due to its correlation with street intersection density. There is also very little correlation between density of auto-oriented intersections and walkability which indicates that having better walkability requires special infrastructure and not just that required for automotives.

When analyzing the difference between data with a low walkability and that with a higher walkability, we found several variables which are distinct between the two. notably, car ownership tended to be on average less in the group with higher walkability possibly because alternative means of transport made a car less necessary. A lower number of cars could also incentivize the development of more infrastructure for pedestrians.

3.1 Further Research

Walkability is a very complex issue, and factors for the walkability index from the dataset used in this analysis were chosen to be small in number and consistent/available for all census blocks in the dataset. The index itself was also calculated by weighting each variable similarly (Diversity of land use was calculated by 2 different variables) and this is likely not reflective of the actual influence of each of these factors. However, this meant the index and this analysis do not consider factors such as the presence of pedestrian paths, ease of driving, or quality and availability of sidewalks and bike lanes, (which are often disconnected or entirely missing).

Another further step in this research would be to analyze how the score given by the national walkability index correlates to trips made on foot. The variables in the analysis were chosen based on an meta-analysis done in 2010. Thus, it might not necessarily correlate with the actual number of trips made on foot, so analyzing how the score and other factors analyzed in this analysis would correlate to data of foot traffic gathered by a survey would be interesting.