Modeling Access to City Walkability

Kyler Shu, James Roberts

Abstract

Walkable areas benefit citizens' physical health, boost the economy, and create a cleaner environment. Thus, walkability has become a critical factor in city planning. However, measuring the walkability of a block is a complex process that requires highly granular observations like the proximity to public transportation. Because of this, we sought to find a way to approximate the walkability of blocks based on standard metrics that are easier to collect. We also explored the relationship between specific metrics, such as income, and how they may relate to walkability. We found that walkability can, in fact, be effectively modeled with simpler metrics.

Research Question

To what extent can city walkability be understood from income and other non-spatial dependent data?

According to a 2023 survey from the National Association of Realtors, being an easy walk away from amenities is an important factor in deciding where to live for 79% of Americans. This number is skewed even higher for the younger generations.

Thus, given the high demand for walkability and the complex nature of its quantification, we aim to explore accessibility to walkable spaces and the extent to which walkability can be modeled from simple, non-spatial data.
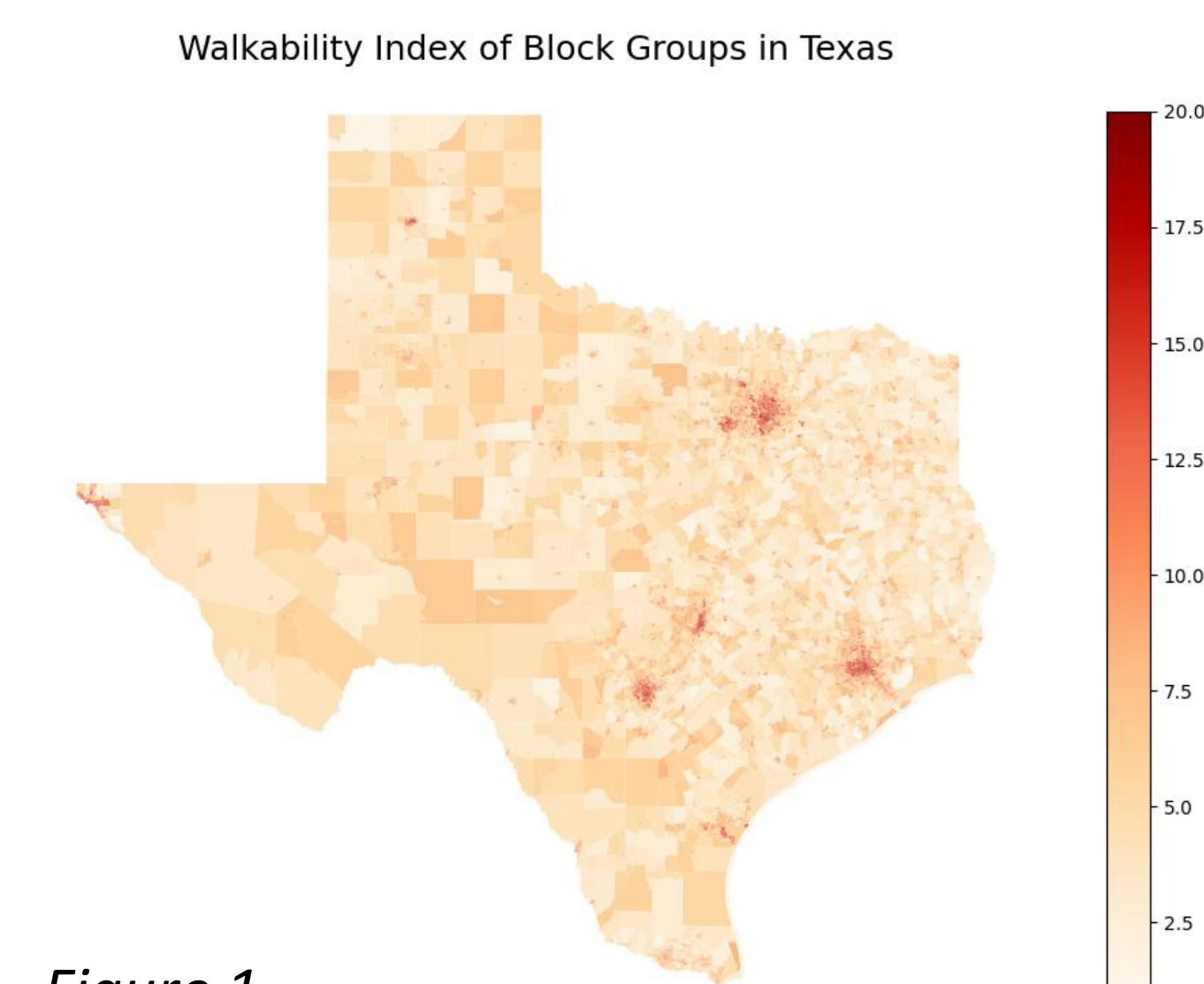
Data Collection

Our data collection process involved merging separate datasets and working with geospatial data.

Our first dataset, from the U.S. Government's open data website, contained data on walkability, income, and demographics but lacked geospatial information. Essentially, we had the walkability of the blocks within the United States but lacked any way to visualize the data.

To fix this, we gathered another data set from the U.S. Census Bureau containing a shapefile of the block groups within the United States. Then, we had to assemble GEOIDs from FIPS codes within our first dataset to join the two together in order to visualize our data in mapping.
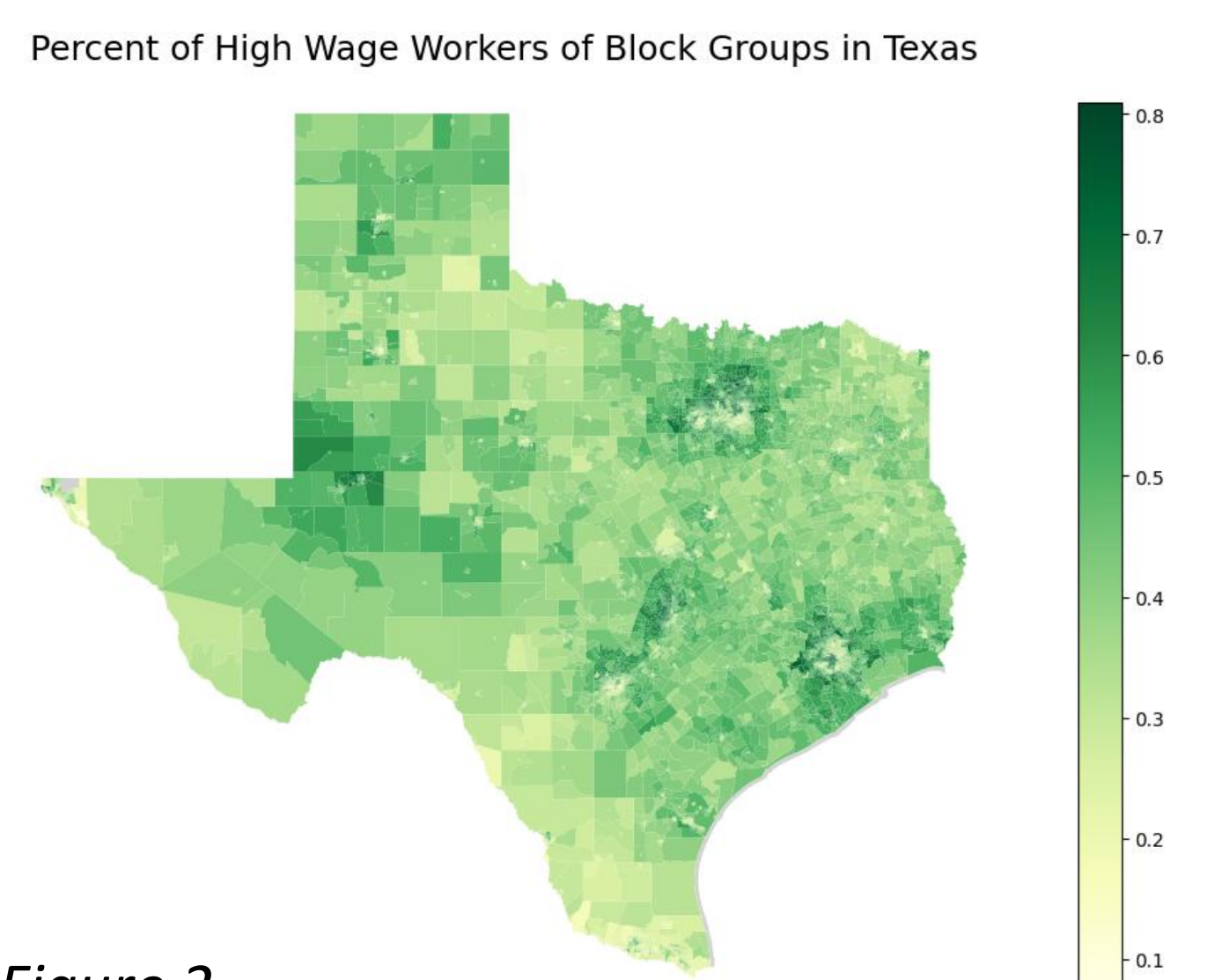
Data Exploration

Figure 1 maps the walkability of block groups in Texas, revealing that the most walkable areas are typically in the centers of large cities like Dallas, Houston, and El Paso.

Figure 2 maps the proportion of high-wage workers who live in a block group. This was calculated by dividing the number of high-wage workers by all of the workers in a block group. This figure reveals that the areas with the highest percentage high-wage workers tend to be suburbs, surrounding large cities. The cities tend to have a lower proportion of high-wage workers, likely because they also have a significant amount of lower-wage workers.
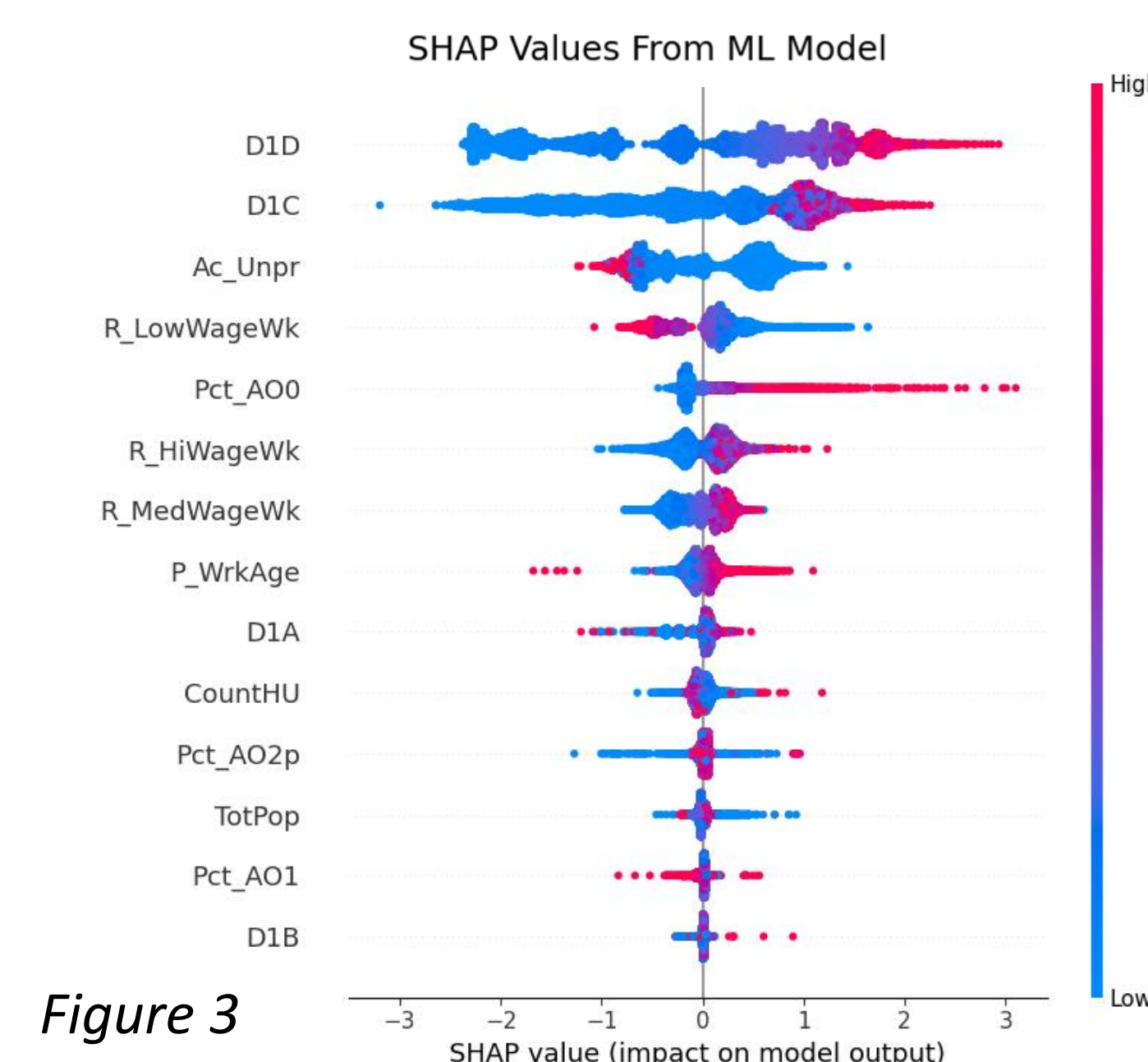
Data Analysis

Training a gradient-boosted regressor model on only non-spatial dependent data, we achieved a cross-validated RMSE of 2.30, with walkability scores ranging from 1 to 20.

By further applying SHapley Additive exPlanations (SHAP values), we were able to quantify which variables were most important in our model's predictions and in what direction they influenced output. Among the most impactful variables were employment density (D1C), residential density (D1D), car ownership (Pct_AO0), and income, as seen in *Figure 3*.

Conclusion

We found that walkability can be effectively modeled without data on the underlying geography of a location, greatly simplifying its quantification. Additionally, our investigation reveals a correlation between walkability and high incomes, suggesting the difficulty of accessing walkable communities. Further research could explore housing prices, and analysis could be extended to the broader United States.

Walkability Index of Block Groups in Texas

Figure 1

Percent of High Wage Workers of Block Groups in Texas

Figure 2

SHAP Values From ML Model

Figure 3

Stanford University