Walkability Index Visualized

Adithya Sunil — IMT2021068
Raj Bunsha — IMT2021010
Tanmay Jain — IMT2021015

Instructor: Prof. Jaya Sreevalsan Nair *Date:* Sep 18, 2023

Contents

1 What is Walkability Index?

We shall begin our report by giving a brief definition for walkability index. Walkability index is a quantitative measure used to assess how pedestrian-friendly a neighborhood, city, or area is for walking. It takes into account various factors and features that can influence the ease and safety of walking, and it provides a numerical score or ranking to help policymakers, urban planners, and researchers evaluate and compare different locations in terms of their walkability. Walkability indexes typically consider factors such as:

- Proximity to amenities
- Street connectivity
- Pedestrian infrastructure
- Traffic and road conditions
- Safety of location
- Environmental Factors

Walkability indexes may use different methodologies and weightings for these factors, resulting in a numeric score or ranking that indicates the overall walkability of an area. Higher scores indicate greater walkability, meaning that it is easier and more convenient for people to walk as a primary mode of transportation in that location. These indexes are valuable tools for urban planning and development to create more pedestrian-friendly and livable communities.

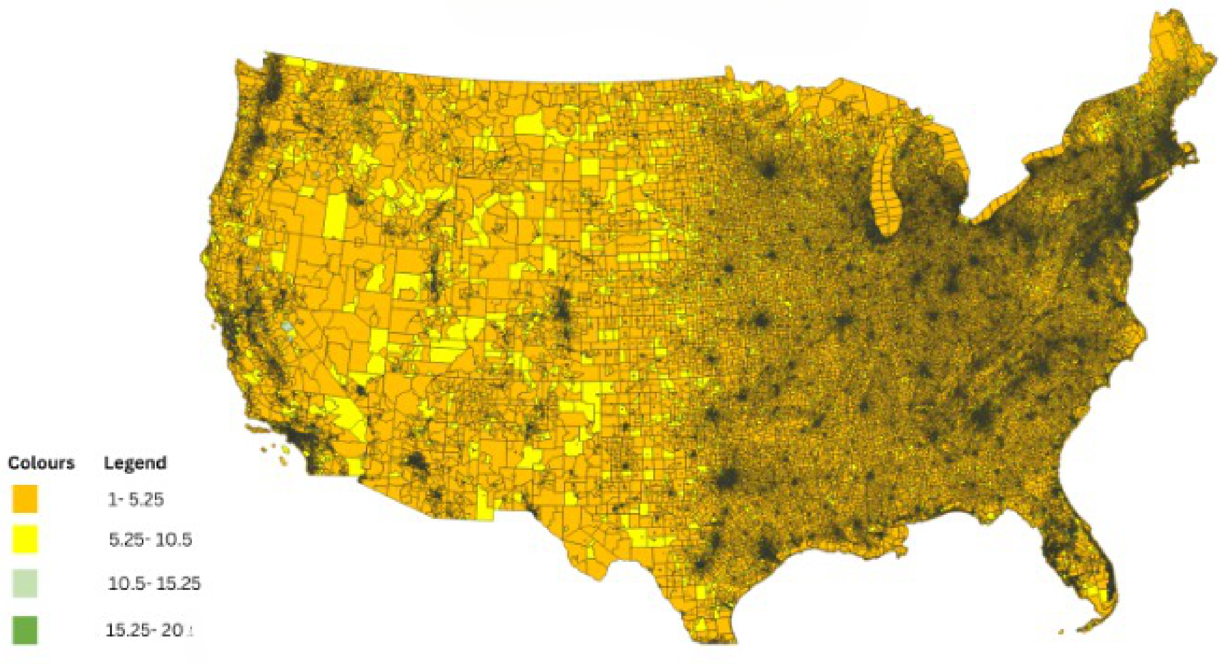

Figure 1: Block-wise Walkability Index on the USA map

2 The Dataset

Our dataset was taken from the official website of the US government, and consists of 183 columns and roughly 200000 rows. The data contains several fields related to population, employment, location, and their correlation to walkability index taken across many cities in the United States. Clearly it was impractical to include all the fields, so we first categorized the columns and selected only the essential features.

	A	B	C	D	E
1	Field Name	Type	Alias	Details	
2	GEOID10	esriFieldTypeString	Census block group 12-digit FIPS code (2010)	length: 12	
3	GEOID20	esriFieldTypeString	Census block group 12-digit FIPS code (2018)	length: 50	
4	STATEFP	esriFieldTypeString	State FIPS code	length: 2	
5	COUNTYFP	esriFieldTypeString	County FIPS code	length: 3	
6	TRACTCE	esriFieldTypeString	Census tract FIPS code in which CBG resides	length: 6	
7	BLKGRPCE	esriFieldTypeString	Census block group FIPS code in which CBG resides	length: 1	
8	CSA	esriFieldTypeString	Combined Statistical Area (CSA) Code	length: 3	
9	CSA_Name	esriFieldTypeString	Name of CSA in which CBG resides	length: 100	
10	CBSA	esriFieldTypeString	FIPS for Core-Based Statistical Area (CBSA) in which CBG resides	length: 5	
11	CBSA_Name	esriFieldTypeString	Name of CBSA in which CBG resides	length: 100	
12	Ac_Total	esriFieldTypeDouble	Total geometric area (acres) of the CBG	N/A	
13	Ac_Water	esriFieldTypeDouble	Total water area (acres)	N/A	
14	Ac_Land	esriFieldTypeDouble	Total land area (acres)	N/A	
15	Ac_Unpr	esriFieldTypeDouble	Total land area (acres) that is not protected from development (i.e.	not a park	
16	TotPop	esriFieldTypeInteger	Population	2018	
17	CountHU	esriFieldTypeInteger	Housing units	2018	
18	HH	esriFieldTypeInteger	Households (occupied housing units)	2018	
19	Workers	esriFieldTypeInteger	Count of workers in CBG (home location)	2017	
20	D2B_E8MIXA	esriFieldTypeDouble	8-tier employment entropy (denominator set to the static 8 employment types in the CBG)	N/A	
21	D2A_EPHHM	esriFieldTypeDouble	Employment and household entropy	N/A	
22	D3B	esriFieldTypeDouble	Street intersection density (weighted	auto-oriented intersections eliminated)	
23	D4A	esriFieldTypeDouble	Distance from the population-weighted centroid to nearest transit stop (meters)	N/A	
24	D2A_Ranked	esriFieldTypeDouble	Quantile ranked order (1-20) of [D2a_EpHHm] from lowest to highest	N/A	
25	D2B_Ranked	esriFieldTypeDouble	Quantile ranked order (1-20) of [D2b_E8MixA] from lowest to highest	N/A	
26	D3B_Ranked	esriFieldTypeDouble	Quantile ranked order (1-20) of [D3b] from lowest to highest	N/A	
27	D4A_Ranked	esriFieldTypeDouble	Quantile ranked order (1-20) of [D4a] from lowest to highest	N/A	
28	NatWalkInd	esriFieldTypeDouble	Walkability Index	N/A	
29	Shape_Length	esriFieldTypeDouble	Shape_Length	N/A	
30	Shape_Area	esriFieldTypeDouble	Shape_Area	N/A	
31	OBJECTID	esriFieldTypeOID	OBJECTID	N/A	

Figure 2: The colour coded excel file with the selected fields

2.1 Categorizing the Fields

In the end, we chose 30 fields to include in our visualization. These were put in an excel file, and colour coded as shown in Figure 2 above. Each field is described by its name in the dataset, its datatype, and an alias which provides a brief description. Additional details are also present in some of the fields, such as the year in which the data was collected, or the length of the data. The different colours are described below:

- The yellow section represents fields which describe location. Each state, county, and census block has unique numerical tags, known as FIPS codes. These codes can be used to pinpoint its location on a map.
- The green section represents fields which describe area based statistics, such as the total land area and water area.
- The blue section contains population based statistics collected in the year of 2018.
- The pink section contains employment related statistics, along with street intersection density, and proximity to transit stops.
- Finally, the red section can be considered the most important, as it takes the above statistics and ranks them into quantiles. These are the values which are finally used to calculate walkability index represented as "NatWalkInd" in the dataset.

2.2 Calculating Walkability Index

We will also mention in brief, how the dataset uses the above fields to calculate the walkability index. The four variables taken into consideration are employment mix, employment and household mix, street intersection density, and proximity to transit stops. These fields can be seen in the pink section in the above picture, and have field names D2B, D2A, D3B, and D4A respectively. For each data block, these four variables are ranked with respect to other data blocks and assigned a score from 1 to 20. These fields are shown in the red section. Finally the data block is assigned a walkability index given by the below formula:

$$Index = (D2A_{Ranked} + D2B_{Ranked})/6 + (D3B_{Ranked} + D4A_{Ranked})/3$$

```python
import pandas as pd
import numpy as np

# Load data from the first CSV file into df1
df1 = pd.read_csv('walk.csv', usecols=['NatWalkInd', 'STATEFP', 'COUNTYFP'])

# Load data from the second CSV file into df2
df2 = pd.read_csv('fp.csv')

# Add an additional column to df1 by concatenating the first two columns
df1['MERGEFP'] = df1['STATEFP'].astype(str) + df1['COUNTYFP'].astype(str)
df2['fips_code'] = df2['fips_code'].astype(str)

# Merge both datasets on the basis of the FIPS code
result_df = pd.merge(df1, df2, left_on='MERGEFP', right_on='fips_code', how='inner')
# Save the result_df dataframe to a CSV file
result_df[['NatWalkInd', 'lng', 'lat', 'name']].to_csv('result.csv', index=False)

```

Figure 3: Python script to convert FIPS codes to latitude and longitude

3 Data Preprocessing

After choosing the essential features, the plan was to visualize the different fields on the United States map, along with various graphs to capture relations. However, there were a few issues with the data which needed to be fixed with preprocessing.

3.1 Removing Null Values

Upon initial examination of the data, we noticed some rows which contained null values. This would pose a problem incase we needed to perform calculations on the data. One of the options was to drop all rows with null values, but this was unnecessary as these null values were restricted to only few columns (The CBSA_Name column for example). Dropping the entire row would mean that the data in the other columns would be dropped too. Therefore we just decided to replace all null value entries with 0.

3.2 Generating Latitude and Longitude

As the location data was given in terms of FIPS codes instead of latitude and longitude, it was difficult to project it onto a United States map. To solve this problem, we found another dataset online which mapped FIPS codes to latitude and longitude coordinates. So we created a python script (refer to Figure 3 above), which uses the pandas library to merge both the datasets on the basis of the STATEFP and COUNTYFP columns.

Unfortunately, some of the FIPS codes in the walkability dataset did not match with those in the latitude and longitude dataset. Upon performing the inner join, any row without a matching field in both datasets would be ignored. This effectively reduced the size of our dataset from 200000 rows to 60000. Therefore we needed to choose an alternative approach to project the data.

3.3 Using QGIS and Tableau

Finally, we decided to use an online block census API which would convert the provided FIPS codes into precise spatial coordinates. Subsequently these coordinates were inputted into QGIS, a software used to manipulate and visualize geospatial data. This approach yielded much better results, allowing us to generate the maps shown in below sections. In addition to QGIS, tableau was used to generate the graphs and scatter plots. Together, these tools provided us with a data story which is described in the following section.

4 Walkability and Urbanization: The Data Story

4.1 The Question

The main question that we want to answer is *"How Does Walkability Influence Urban Life? An Exploration of its Impact on Population, Employment, Street Infrastructure, and Housing Patterns Across Different States."* To solve this question, we shall compare each factor individually, and try to observe the common results between them. The three factors are:

- Walkability and Population
- Walkability and Street Intersection Density
- Walkability and Housing & Employment

4.2 Walkability and Population

4.2.1 Introduction

Urban areas tend to be more populated as compared to rural areas. In order to find a relation between walkability and urbanization, it is logical to first compare walkability index and population. For this visualization, we will mainly consider the percentage of population in a state having a certain value of walkability index.

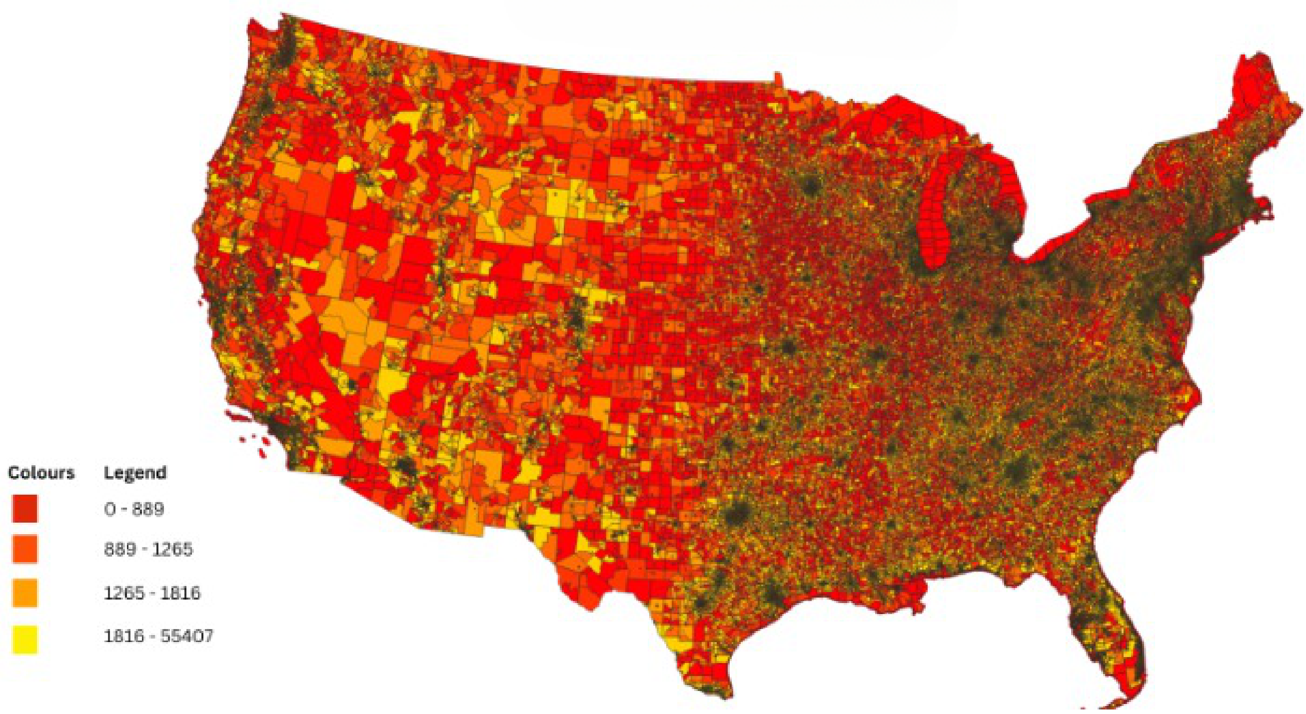

Figure 4: Block-wise Population on the USA map

4.2.2 Line Graph

We began by plotting the population percent of each state against its respective walkability index as a line graph, where each state is a different coloured line (as shown in Figure 5). From the graph we can say that the percentage of population having extreme values of walkability index (1 or 20) is very low. However, due to the inclusion of all 50 states in this plot, there were not many other inferences to be made. The resulting graph, while technically correct, was unable to show any meaningful result. This was primarily due to the densely packed lines and data points on the graph, which made it practically impossible to discern clear patterns, trends, or correlations.

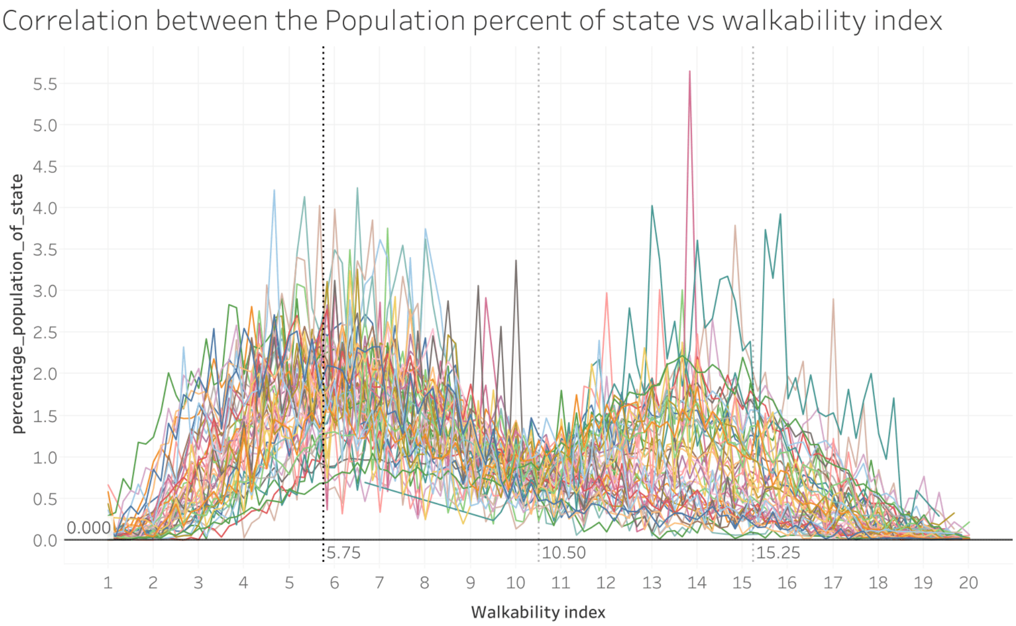

Figure 5: Line Graph

4.2.3 Dot Plot

To enhance the readability of our visualization, we switched to a dot plot (as shown in Figure 6), and constrained the domain into four distinct categories: Least Walkable," "Below Average," "Above Average," and "Highly Walkable". The advantage of having such a dot plot, is that we can easily compare states with each other. In order to demonstrate this, we will look at the case of five most walkable, and least walkable states in the United States.

4.2.4 The Most Walkable States

In the context of the dot plot we've examined, which focuses on five specific states (California, Washington DC, Florida, Massachusetts, and Nevada), a notable pattern emerges. This pattern centers around the "Above Average" walkable category, indicating that more than 50% of the population living in these states lives in area exhibiting an above-average walkability index. Conversely, fewer than 15% of the population in these states fall into the least walkable category. These statistics suggest a compelling association between the walkability index and urbanization trends in these states.

However, before we draw any firm conclusions regarding the correlation between walkability and urbanization, it's crucial to also investigate the characteristics of the least urbanized states. By doing so, we can gain a more comprehensive understanding of walkability index.

4.2.5 The Least Walkable States

In the dot plot, we observe a distinctive pattern in the states of Arkansas, Maine, Mississippi, Montana, and West Virginia. Figure 7 (below) illustrates a notable concentration of data points within the "Least Walkable" and "Below Average" categories for these states. Interestingly, the data also reveals that fewer than

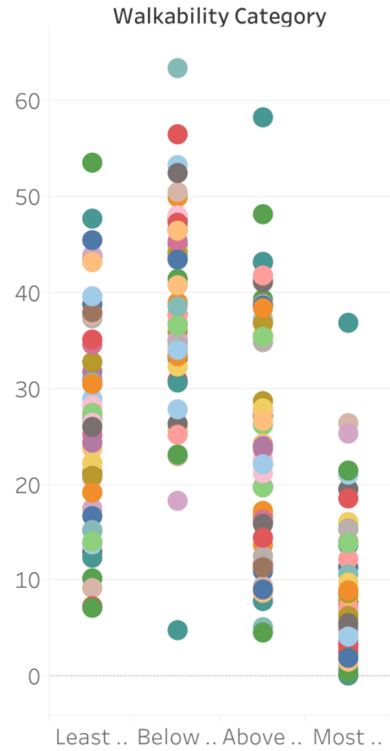

Figure 6: Dot Plot (y-axis represents the percent of population living in the category)

10% of the population in these areas falls within the most walkable category. This observation points to a significant variation in walkability within these states.

Figure 7: Most Urbanized States (Left) and Least Urbanized States(Right)

4.2.6 Comparing all the States

While the dot plot makes it easier to compare few states at a time, it is still quite difficult to analyze all the states together. To improve the visualization further, we experimented on a bar graph, where the x-axis is the state FIPS code, and the y-axis shows population percentage. In addition to this, four different colours are used to represent the different categories of walkability. This graph is shown below in Figure 8 (Note that there are no state codes beyond 56).

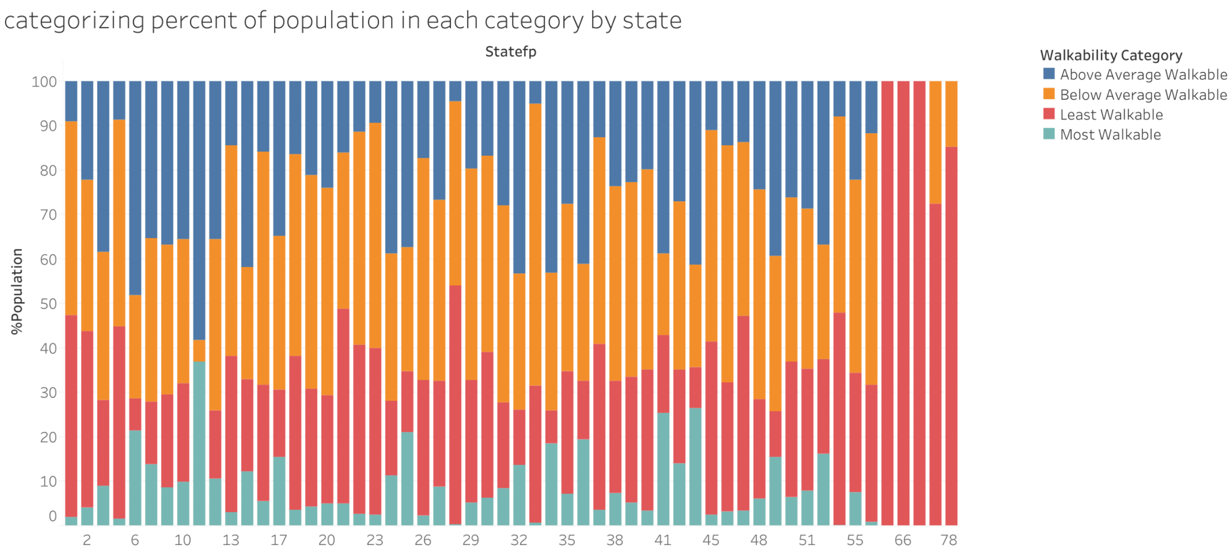

Figure 8: State-wise Walkability index distribution into 4 Categories

We can then compare these values with the urbanization percentage graph of the United States, taken from Wikipedia.

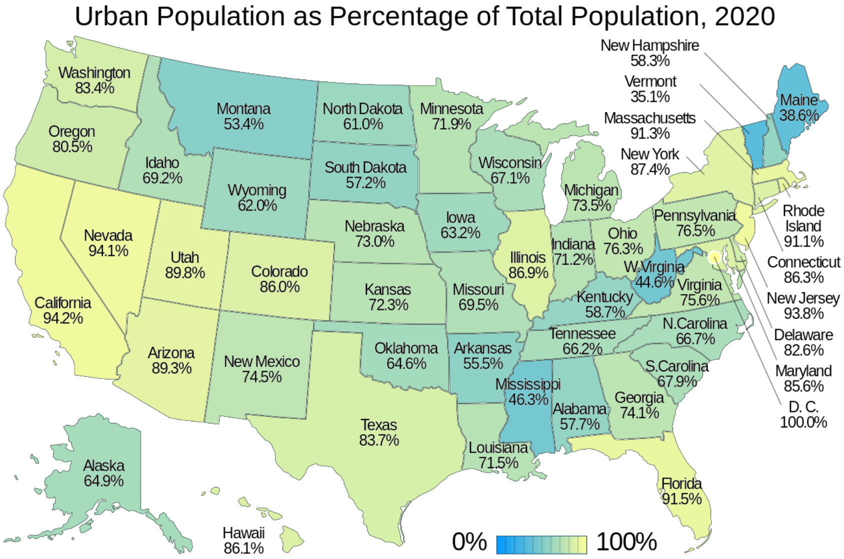

Figure 9: State-wise Percentage Urban Population (Source: Wikipedia)

4.2.7 Key Findings:

From the above graphs, some conclusions can be drawn about the relationship between walkability and the population which are listed below:

- Walkability Index Values: For all the states, the percentage of population having extreme values of walkability (1 or 20) is very low. Most of the values are distributed in between with no explicit pattern or trend. Some states have a mix of both high and low walkability areas.
- Urbanization Impact: States with **higher** urban population tend to show **higher** values of walkablitiy index, whereas states with **lower** urban population tend to show **lower** values of walkability index. This strengthens that the fact that as urbanization of a place increases the walkability index increases as well.

4.3 Walkability and Street Intersection Density

4.3.1 Introduction

Urban cities are often characterized by their closely interconnected streets. A high street intersection density indicates that roads intersect frequently, resulting in a more grid-like street pattern. Conversely, a low street intersection density suggests that roads are more spaced out, resulting in a less interconnected and potentially more hierarchical road network. The structure of roads also plays a role in the walkability index, so in this section we will compare the walkability index and the street intersection density of the states

4.3.2 Line Graph

To begin, we plotted the average value of street intersection density against the walkability index for all the data. The y-axis contains the average value of field "D3B" from the dataset. This field contains the weighted street intersection density. The resulting visualization is a line graph which seems to increase with the rise in walkability index. Infact, it resembles a linear equation with constant positive slope. (Refer to Figure 10 below)

4.3.3 Scatter Plot

To confirm this trend, we next considered each state individually as we did for the population, using a scatter plot. Here each point represents a state with x coordinate equal to the average walkability index, and the y coordinate its average street intersection density. Once again the plot showed clear correlation between the two variables, a point with higher street intersection density showed higher walkability index. To show this correlation we have drawn an approximate line of best fit for the scatter plot, in Figure 11.

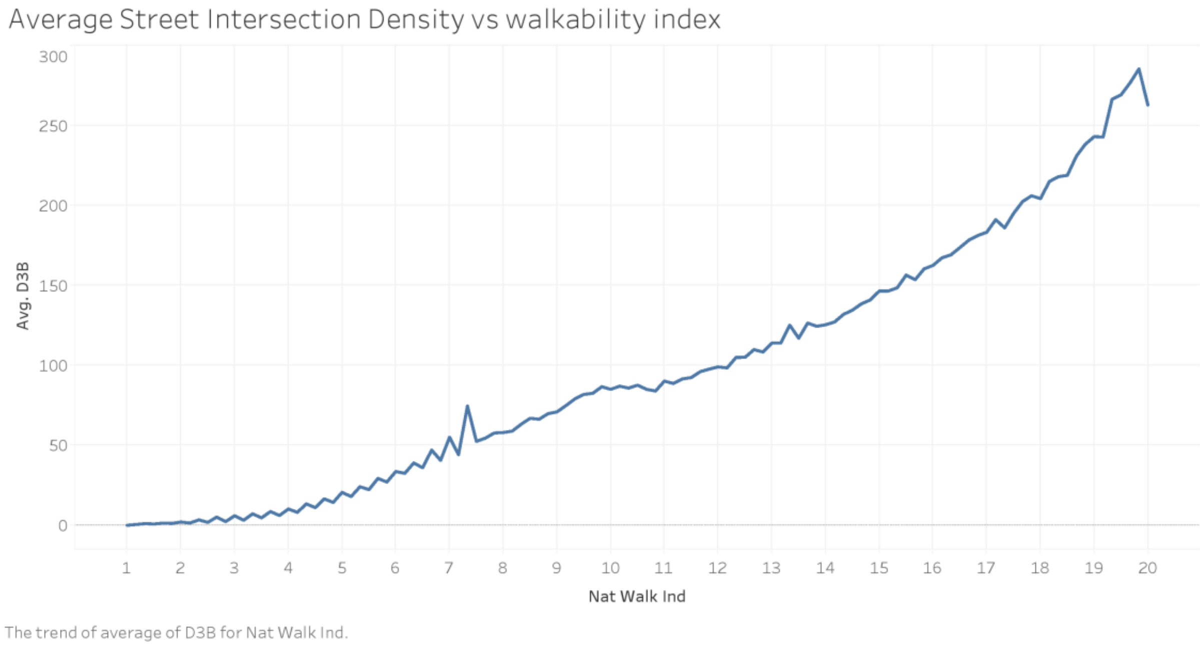

Figure 10: Line Graph

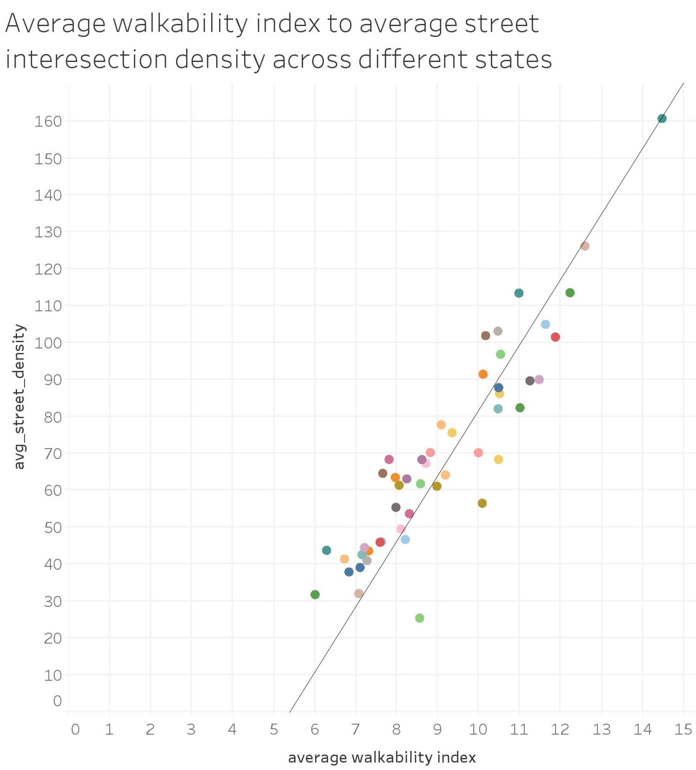

Figure 11: Scatter Plot with approximate line of best fit

4.3.4 Street Intersection Density and Urbanization

As we mentioned above, urban cities tend to have higher street intersection density. To confirm this, we fed the FIPS codes and the street intersection density (D3B_Ranked) columns from the dataset into QGIS. We assigned a colour code, where orange meant low street intersection density and green meant high street intersection density. The software generated a map of the United States from the given data which is shown in Figure 12.

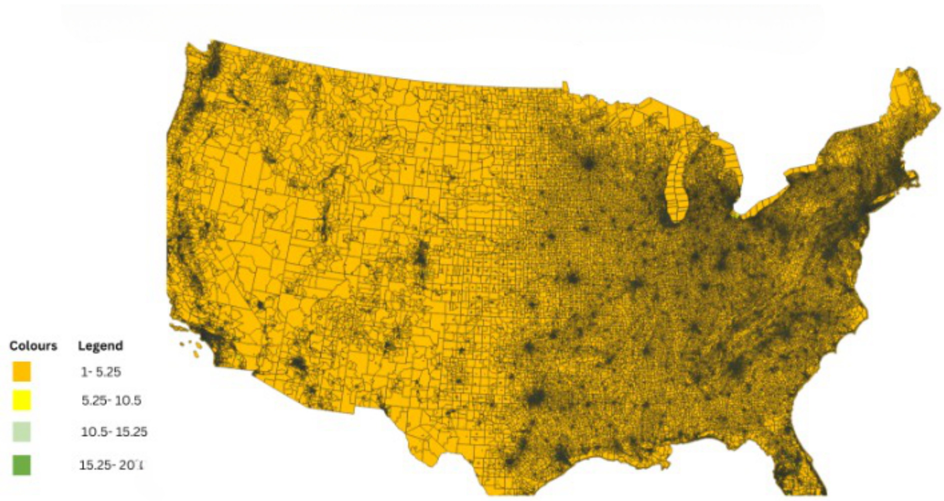

Figure 12: Street Intersection Density Map of United States

Unfortunately, taking the entire United states map doesn't show much, as most of the cities are densely packed together and just appear black in colour. For improved visualization we must zoom in, and analyze a particular city. For this example we have taken New York City, which is definitely ranked high in terms of urbanization. According to our claim, we would expect high street intersection density in this area. This is shown in Figure 13 below, and as expected we can see the colours start to turn from orange to yellow to green as we converge in, on New York City. Therefore we can safely assume that urban cities have higher street intersection density.

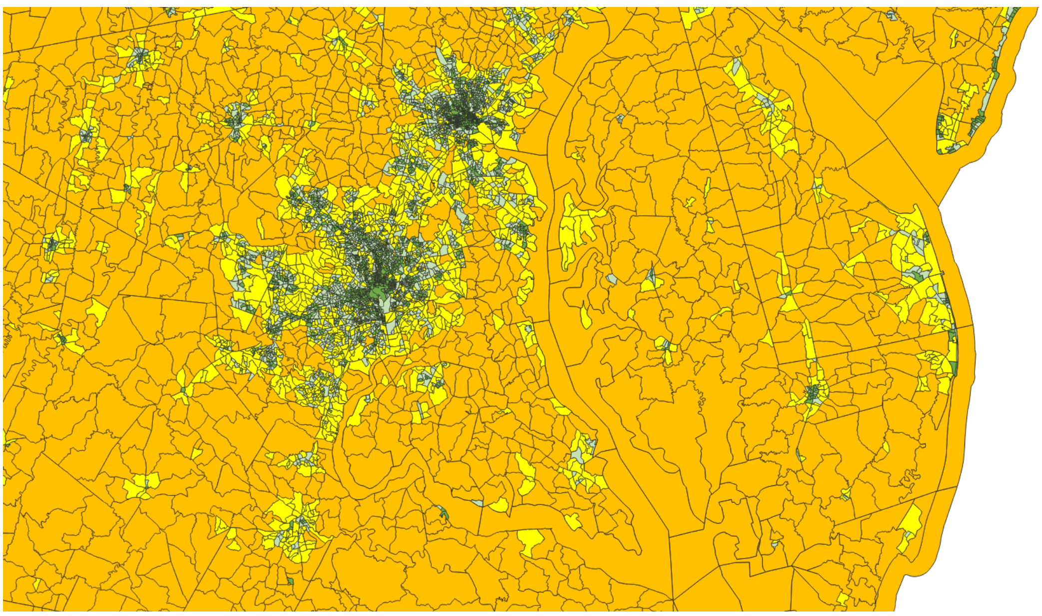

Figure 13: Street Intersection Density Map of New York City

4.3.5 Key Findings

- Urbanization and Street Intersection Density: From the colour map of the USA which was generated, we noticed that urban cities such as New York City had a large concentration of green colour, which indicated high street intersection density. This relation was expected as urban cities tend to have more number of intersecting roads.
- Walkability Index and Street Intersection Density: In our analysis of street intersection density and its correlation with the average walkability index across the states, we observed a noteworthy pattern. Generally, we found that as street intersection density increases, the average walkability index tends to increase as well, indicating a positive correlation, as shown by the line of best fit. This positive correlation is consistent with the expectation that areas with a higher density of intersections often offer more pedestrian-friendly environments, contributing to higher walkability.
- Walkability Index and Urbanization: Using both of the above findings, we can assume that **more** urbanization leads to **higher** walkability index which is the same inference we made in the walkability and population section.

4.4 Walkability and Housing & Employment

4.4.1 Introduction

In the exploration of urban dynamics, the interplay between housing and employment mix stands as a pivotal factor contributing to urbanization. To elucidate this relationship, we embark on a comprehensive analysis of the walkability index in conjunction with housing and employment patterns. Much like the prior examination of walkability and population, we will primarily focus on discerning the intricate percentage breakdowns of states in relation to their walkability index values. This endeavor seeks to shed light on the symbiotic connection between housing, employment, and the evolution of urban landscapes.

4.4.2 Line Graph

We further embarked on the visual exploration of the relationship between housing units and walkability index across the United States. Our initial approach involved plotting the average housing units for each state against its corresponding walkability index. The resulting line graph, encompassing all states, reveals a striking pattern—a convergence of lines, tightly clustered within a defined spectrum. This spectrum serves as a compelling representation, demonstrating that the housing units remains remarkably consistent across the majority of states. Notably, only a select few states deviate from this uniformity, possibly signifying outliers in this analysis.

For a more detailed visualization of this phenomenon, please refer to Figure 14 in this report.

4.4.3 Validation through Median Housing Units vs. Walkability Analysis

To fortify our earlier assertion regarding the consistency of housing units in relation to walkability, we conducted an additional analysis. This time, we plotted the median housing units for each state against walkability index values at various walkability levels.

Across different walkability index values, we encountered a familiar sight—an alignment of medians that closely hugged each other within a fixed spectrum, but this time the spectrum became little wider. This striking similarity underscores the stability of housing units for the majority of states, reaffirming the notion that deviations from this trend are indeed noteworthy outliers. The trends mentioned could be found in Figure 15.

4.4.4 Employment Percentage Across States

To visualize this crucial aspect, we employed a dot plot, presented as Figure 17. This dot plot offers a compelling snapshot of employment distribution, allowing us to discern patterns and variations among states. As we continue our exploration, the insights gleaned from this visualization will undoubtedly contribute to a more comprehensive understanding of the complex interplay between employment, housing, and walkability, thus further enriching our investigation into urbanization dynamics.

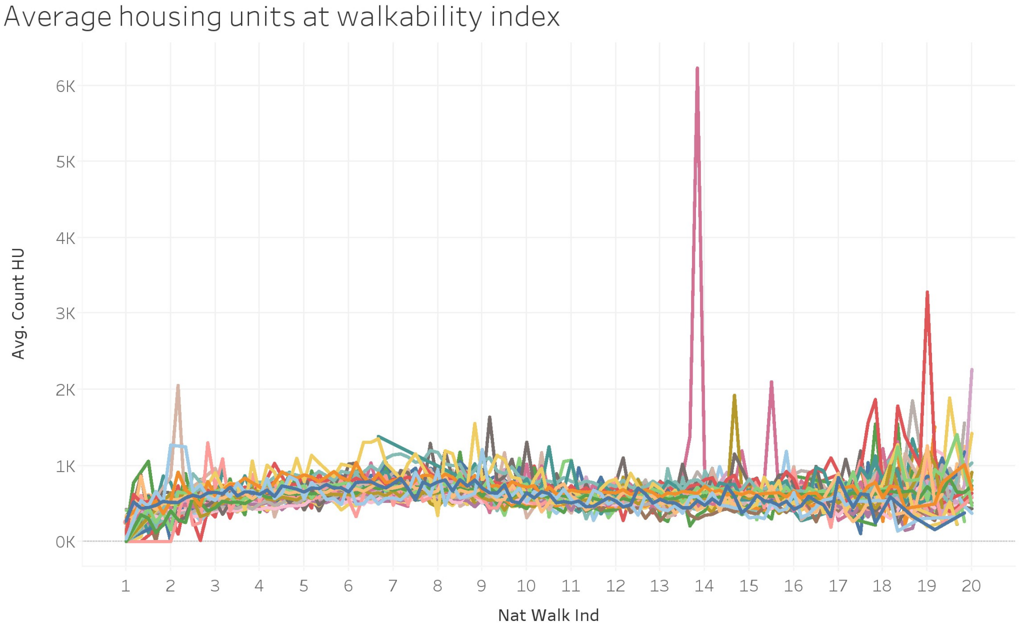

Figure 14: Line Graph

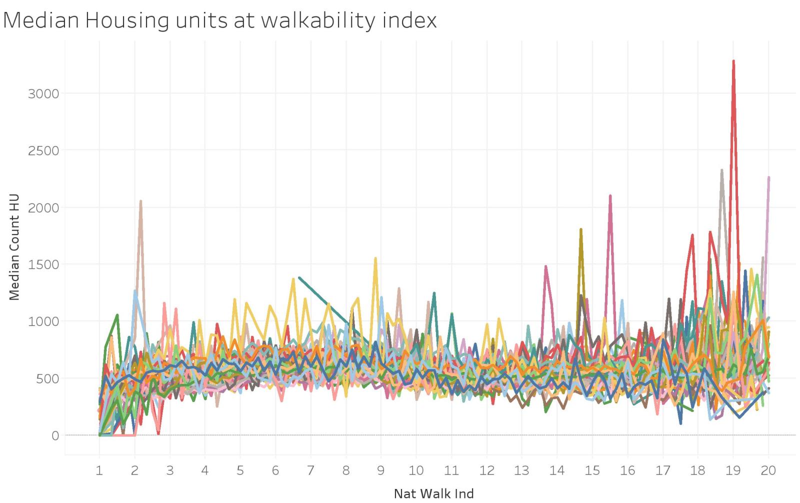

Figure 15: Line Graph

Figure 16: Block-wise Employment mix plotted on the USA Map

4.4.5 Most Walkable States

In the context of the dot plot analysis shown in Figure 18a., which focuses on five specific states - California, Washington DC, Florida, Massachusetts, and Nevada, an intriguing narrative emerges, which primarily revolves around the "Above Average" walkable category. Remarkably, in these states, a significant majority, surpassing 50% of their populations, reside in areas boasting an above-average walkability index. Conversely, a mere fraction, less than 15%, of these states' populations find themselves in the "Least Walkable" category.

Upon further analysis, a noteworthy pattern emerges, one that establishes a link between the prevalence of above-average walkability and higher employment rates in these states. This connection underscores the intricate interplay between walkability, employment opportunities, and the path to urbanization. As we continue to dissect these findings, we inch closer to unraveling the nuanced dynamics of urban development within these regions.

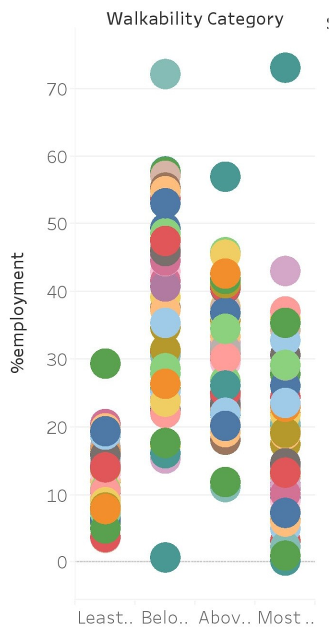

Figure 17: Dot Plot

4.4.6 Least Walkable States

In the dot plot shown in Figure 18b., our attention is drawn to a distinctive pattern emerging in the states of Arkansas, Maine, Mississippi, Montana, and West Virginia. Figure 5 (below) vividly showcases a significant concentration of data points within the "Least Employment" and "Below Average" categories for these states.

Urban areas tend to offer a higher abundance of employment opportunities, often leading to the development of vibrant, closely-knit markets. This, in turn, fosters a self-reinforcing cycle of improved employment prospects, creating a dynamic environment where job growth continually fuels further economic opportunities.

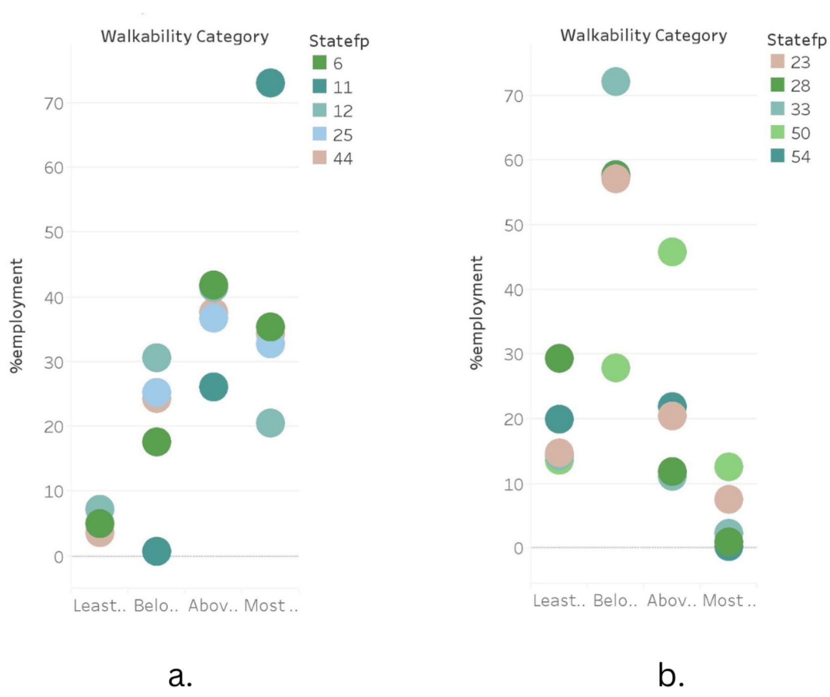

Figure 18: Most and least walkable states

4.4.7 Key Results

- **Consistency in Housing Units**: Our exploration revealed a remarkable consistency in housing units values across most states, with only a few outliers. This suggests that housing conditions remain relatively uniform, irrespective of other factors, in the majority of regions.
- States with higher walkability indices tended to exhibit greater employment opportunities, while lower walkability states often had less robust employment prospects. This correlation between walkability and employment levels is indicative of a relationship intertwined with urbanization and overall standards of living.
- In states where walkability is more pronounced, residents benefit from easier access to amenities, job centers, and transportation hubs, fostering increased employment prospects. Conversely, in areas with lower walkability, limited access to these essential components can hinder employment opportunities and, consequently, contribute to lower urbanization rates and potentially lower living standards.
- **Urbanization Implications**: While not directly concluding urbanization, these inferences collectively paint a picture of urbanization as a multifaceted phenomenon. It is not solely driven by employment opportunities or housing conditions but involves complex interactions among various factors, including walkability. These insights encourage further investigation into the nuanced dynamics that contribute to the evolution of urban landscapes.

4.5 The Conclusion

For all three factors: Population, Street Intersection Density, and Housing & Employment, there was a correlation with Walkability Index as shown by the different visualizations. In general, an increase in these variables showed an increase in Walkability Index. This trend matches the formula used to calculate Walkability Index, explained in the section 2.2.

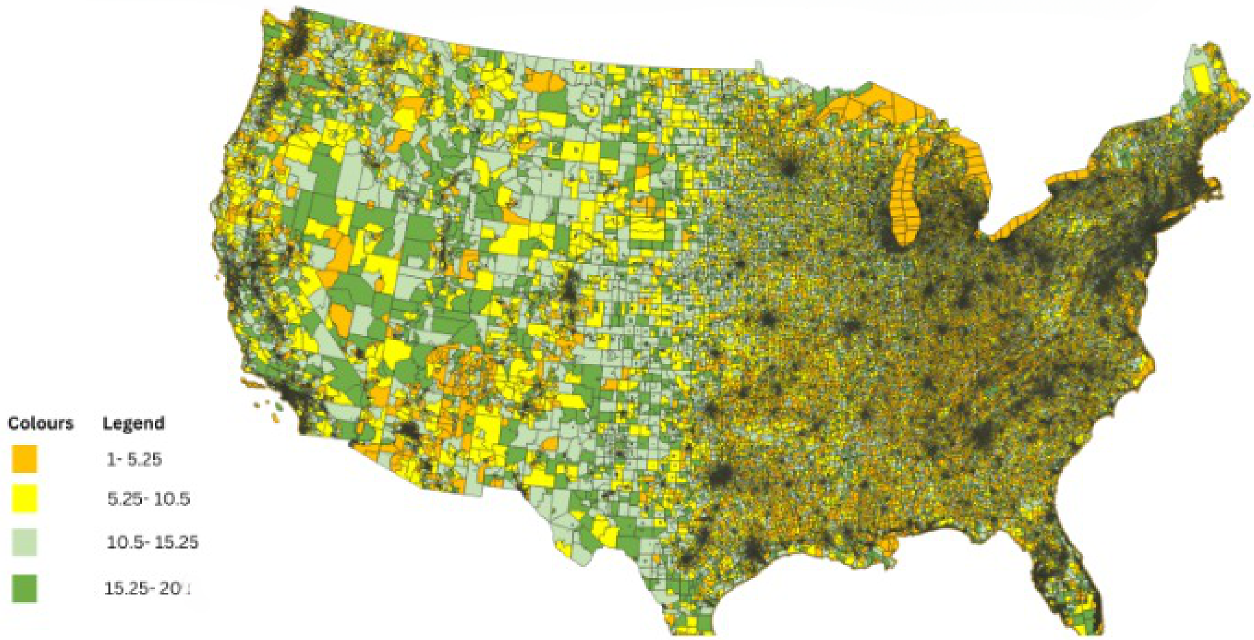

Figure 19: Block-wise Household & Employment mix plotted on the USA Map

We also noticed that higher population and street intersection denity was found in urban areas, whereas rural areas had lower population and lower street intersection densities. However, in the case of employment and housing, the correlation was not as strong.

Since two of the factors (Population and Street Intersection) which contribute to walkability index also contribute to urbanization of a location, there is a noticeable correlation between Walkability Index and Urbanization. In general we can claim the following:

*"The Walkability Index is mostly **higher** for urban states, as compared to rural states in the USA"*

5 Some important indexes and legends

Here are some common legends and indexes which are used in all the visualizations:

5.1 Categories of area on the basis of walkability index

Walkability Category	Walkability Ranges (in minutes)
Least Walkable	0 - 5.75
Below Average Walkable	5.76 - 10.5
Above Average Walkable	10.51 - 15.25
Most Walkable	15.26-20

Table 1: Walkability Categories and Ranges

5.2 Correlation of state codes of states along with their color codes as used in further graphs

There are 50 states in the United States and the District of Columbia, making a total of 51 entries in this table.

State Code	State Name
01	Alabama
02	Alaska
04	Arizona
05	Arkansas
06	California
08	Colorado
09	Connecticut
10	Delaware
11	District of Columbia
12	Florida
13	Georgia
15	Hawaii
16	Idaho
17	Illinois
18	Indiana
19	Iowa
20	Kansas

State Code	State Name
21	Kentucky
22	Louisiana
23	Maine
24	Maryland
25	Massachusetts
26	Michigan
27	Minnesota
28	Mississippi
29	Missouri
30	Montana
31	Nebraska
32	Nevada
33	New Hampshire
34	New Jersey
35	New Mexico
36	New York
37	North Carolina

State Code	State Name
38	North Dakota
39	Ohio
40	Oklahoma
41	Oregon
42	Pennsylvania
44	Rhode Island
45	South Carolina
46	South Dakota
47	Tennessee
48	Texas
49	Utah
50	Vermont
51	Virginia
53	Washington
54	West Virginia
55	Wisconsin
56	Wyoming

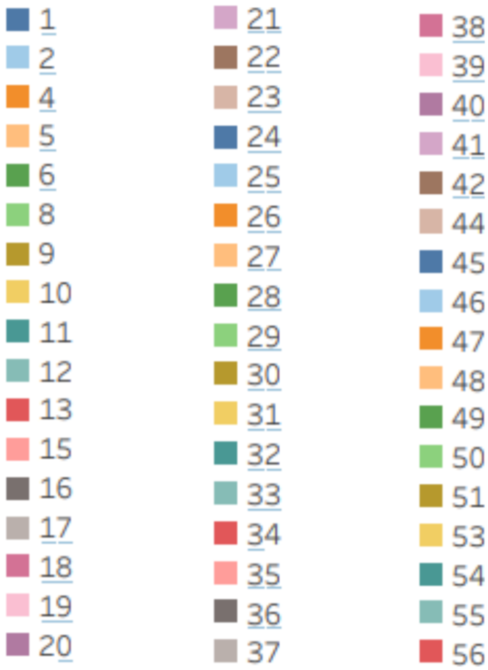

Figure 20: Colour codes of the states

6 References

- Wikipedia
- Data Source
- Video

7 Contributions

7.1 Preprocessing:

- Removing Null Values in Dataset - Tanmay Jain
- Writing Python Script to convert FIPS codes - Adithya Sunil
- Using block census API to conver FIPS codes - Raj Bunsha

7.2 Visualizations:

- Walkability and Population - Raj Bunsha
- Walkability and Street Intersection Density - Adithya Sunil
- Walkability and Housing & Employment - Tanmay Jain

7.3 Report Writing:

Each member of the group was also responsible for writing their tasks in the report, which includes the preprocessing and the visualizations.