Livability of U.S. Cities

Charles Song songsc@umich.edu
Lydia Cheng lycheng@umich.edu
Alex He alexhe@umich.edu
Yifei Sun yifeisun@umich.edu
Manasa Jagadeesh jmanasa@umich.edu

Introduction

Our group project aims to evaluate the livability of cities across the United States. Choosing a place to live can be daunting, given the country's varied geographical, economic, and infrastructural landscapes. To maintain a focused scope, we have narrowed down the factors of interest to the following categories: housing, walkability, transportation, climate, and healthcare. According to the United States Census Bureau reports, individuals can expect to relocate an additional 9.1 times after turning 18, highlighting the importance of informed decision-making when selecting a place to call home. Our project aims to provide valuable insights to aid in this crucial decision. Our project is online at https://tinyurl.com/livability-cities, or here if the first link doesn't work.

Background

Existing research articles have identified numerous factors influencing housing prices and livability; however, many studies lack data visualizations. Instead, they primarily focus on developing mathematical and conceptual/qualitative models to characterize housing prices and livability. As a result, the visualizations often consist of diagrams and flowcharts that represent conceptual models rather than data-driven illustrations.

More advanced visualizations may include weighted directed graphs that depict Bayesian network models, such as the one found in a study aiming to identify features city leaders could focus on to create more livable and age-friendly communities (Brossoie & Burns, 2020). This study's diagrams showcased the relationships between key factors contributing to the development of livable communities, including the built environment (e.g., transportation, outdoor spaces, structural features), resident characteristics (e.g., age, income, health, quality of life), and neighbor relations (e.g., interactions, perceived helpfulness).

Some studies concentrating on mathematical modeling have minimal visualizations but identify the most influential factors, which can help guide our visualization efforts. For instance, a study examining housing price dynamics in 130 U.S. metropolitan areas concluded that population growth, income changes, construction costs, and interest rates are the most impactful (G. & Daniel, 2002).

The U.S. News & World Report's Best Places to Retire rankings evaluate the largest metropolitan areas in the U.S. based on six indexes: Happiness, Housing Affordability, Health Care Quality, Retiree Taxes, Desirability, and Job Market (How We Rank the Best Places to Live & Retire, 2022). Livability's Top 100 Best Places to Live ranking employs a proprietary algorithm that calculates a LivScore for each mid-sized city (population up to 500,000) using over 50 data points across eight categories (Livability: Find Your Best Place To Live, Work and Play, 2023). Weather and climate are also essential factors. Various rankings of cities with the most pleasant weather in the U.S. use a subjective definition of pleasant weather to determine the number of enjoyable days for each metropolitan and micropolitan statistical area (Kaduk, 2019).

Purpose

Our project aims to help users find the optimal place to live by considering common concerns, including:

- Housing Costs: Housing costs are a common indicator for cost of living, which is one of the key factors users listed when determining which city to live in.
- Walkability: A walkable city can significantly impact the quality of life and overall convenience.
- Traffic accidents: The frequency of traffic accidents can reflect a city's safety and transportation infrastructure.
- Weather: Climate and weather conditions are essential considerations since they can greatly influence daily life and comfort. Somewhere with large amounts of rainfall in the summer may indicate hurricanes. Somewhere with very hot summers may indicate more air conditioning expenses, while somewhere with very cold winters may indicate more heating costs.
- Life expectancy: While we could not find comprehensive healthcare accessibility data, we used life expectancy data as a proxy for overall health and well-being in a given area. Life expectancy can offer readers a comprehensive view of the overall healthcare situation in a city, as it encapsulates various factors such as access to healthcare, quality of care, and general well-being.

Our project involved analyzing specific data for each city and identifying general trends across the United States. This information will be invaluable for users who need to relocate, whether for their job or to start and raise a family: especially to analyze general trends of a geographic area and more specific details about any cities they had in mind.

Data

For our project, we utilized five datasets: housing costs, walkability, traffic accidents, weather, and life expectancy. Below are the descriptions of each dataset and the values we extracted:

- Housing costs: The Zillow Home Value Index (ZHVI) measures the typical home value and market changes across a given region and housing type. The ZHVI is calculated using various data sources, including public records and user-submitted data. It includes ZHVI values for homes across the United States, aggregated by region or city and seasonally adjusted. Regarding the specific dataset chosen for the project, we chose the ZHVI value across all home types, including single-family residences, condos, and co-ops, smoothed and seasonally adjusted for each city. To further narrow down the dataset, we limited the data to only that of 2022.
- Traffic accidents: Available on Kaggle, A Countrywide Traffic Accident Dataset is a collection of traffic accident data in the contiguous United States that contains over 2.8 million records of individual traffic accidents from February 2016 to December 2021. The dataset provides a range of metrics, including location (coordinates and city), time, the severity of accidents, and weather and road conditions at the time of the accident. Other metrics the dataset provides include the type of accident, the number of people involved, and the distance to the nearest airport. We used Cartographic Boundary Files and 500 Cities shapefile data, as well as the US Zipcodes to County State to FIPS Crosswalk dataset, to aggregate geospatial walkability and traffic datasets at the city level and adjusted them by city area. This process required the use of the GeoPandas library in Python.
- Walkability: The Walkability Index dataset, provided by the EPA, is based on a combination of national datasets that provide information on transportation, demographic, and land-use characteristics. It covers ~1,000 Core-Based Statistical Areas (CBSAs) in the United States for 2019. The index is calculated from a range of metrics related to walkability, such as access to public transportation, pedestrian infrastructure, land-use mix, and population density. For each city, we simply extracted the index value for the CBSA containing the city. We used Cartographic Boundary Files shapefile data to determine which CBSA each city belonged in.
- Weather: We used Comparative Climatic Data from the National Center for Environmental Information of NOAA. It includes the maximum, minimum, and mean temperature data (in Fahrenheit) and precipitation data (in inches) for each month over 30 years of 280 weather stations in the US. The normal precipitation is the arithmetic mean for each month over the 30-year period and includes the liquid water equivalent of snowfall.
- Life expectancy: We obtained life expectancy data from the National Center for Health Statistics/USALEEP for the year 2015. The dataset is accessible through the City Health Dashboard. After cleaning, the dataset includes 833 cities, with life expectancy values ranging from 68.9 to 85.5 years.

We utilized Python and the Pandas library to process and clean all datasets. Our project focuses on city-level data, so we ensured that all entries in the datasets corresponded to this level of granularity. We employed the "relation" function in Tableau to connect different datasets. We retained the "State" and "City" columns in each dataset, which served as the keys for combining datasets and obtaining geographic coordinates. The final joined dataset contained ~545 cities.

Methods

To initially split the work, we delegated one dataset to each person. Each person processed data and created visualizations for their dataset. The initial guidelines were simply that each dataset needs a 2-letter State column and a City name column. We conducted meetings whenever someone encountered difficulties with data processing, especially for the aggregation operations using shapefile data.

For the weather data, additional meetings were conducted to determine the best data source to use and how to join the data by the nearest weather station. The selected weather dataset is by station, so we extracted the coordinates of each city and station, determined the nearest station to each city using nearest neighbors, and used the nearest station's data for each city.

Once each dataset's visualization was complete, we joined the data and associated visualizations into a single Tableau dashboard. Since the data has been joined at this point, we also regression plots to highlight potentially useful trends that may assist users with creating conclusions or making decisions about the optimal cities to live in.

Results

We created a dashboard that presented each dataset as a point map covering the Contiguous United States (Fig. 1). We also included a few regression plots to help the user understand certain trends (Fig. 2). When selecting a city on click within any of the graphs, the same city will be highlighted within all other graphs, except within the weather station point map. For this, the nearest station or city would be highlighted. Unfortunately, all stations within the same state are highlighted despite the fact that stations data is only joined by coordinates. This seems to be a quirk with Tableau.

We received user feedback that suggested areas for improvement in our data visualizations. To enhance the user experience and understanding, we will implement the following changes:

1. Legends: We will ensure that each chart includes a clear and explicit legend, making it easier for users to interpret the data.
2. Organization: We will improve the organization of the charts, presenting them logically and coherently to facilitate user comprehension.
3. Descriptions: To guide users through the visualizations, we will add brief descriptions to each graph, explaining its purpose and relevance.
4. Units and Clarity: We will ensure that all graphs specify the units being used and provide any necessary explanations to prevent confusion or misinterpretation.

By implementing these changes, we can create more effective and user-friendly data visualizations, allowing users to understand the information better and make informed decisions based on our findings.

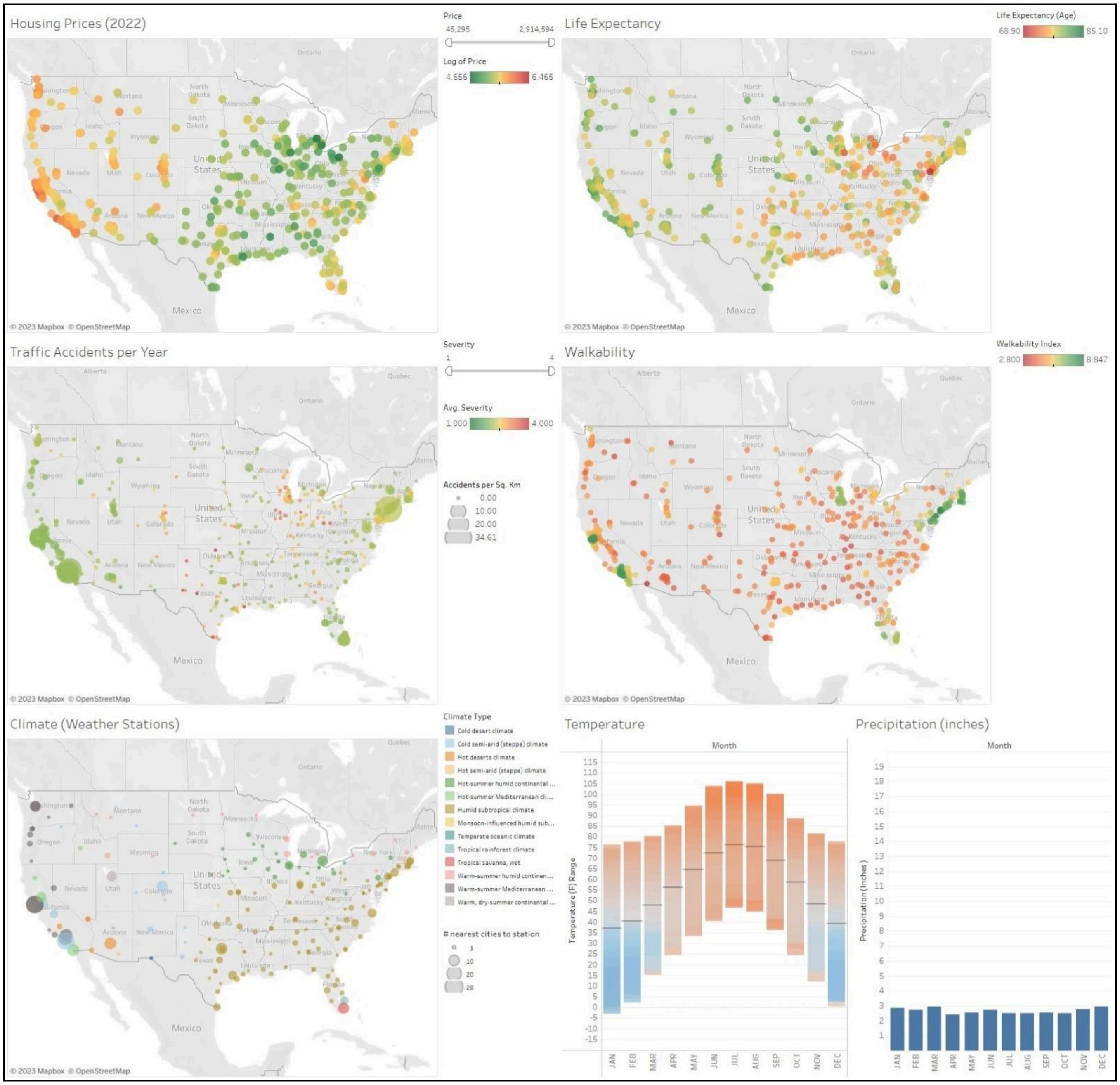

Figure 1: Point maps and charts for weather (temperature and precipitation).

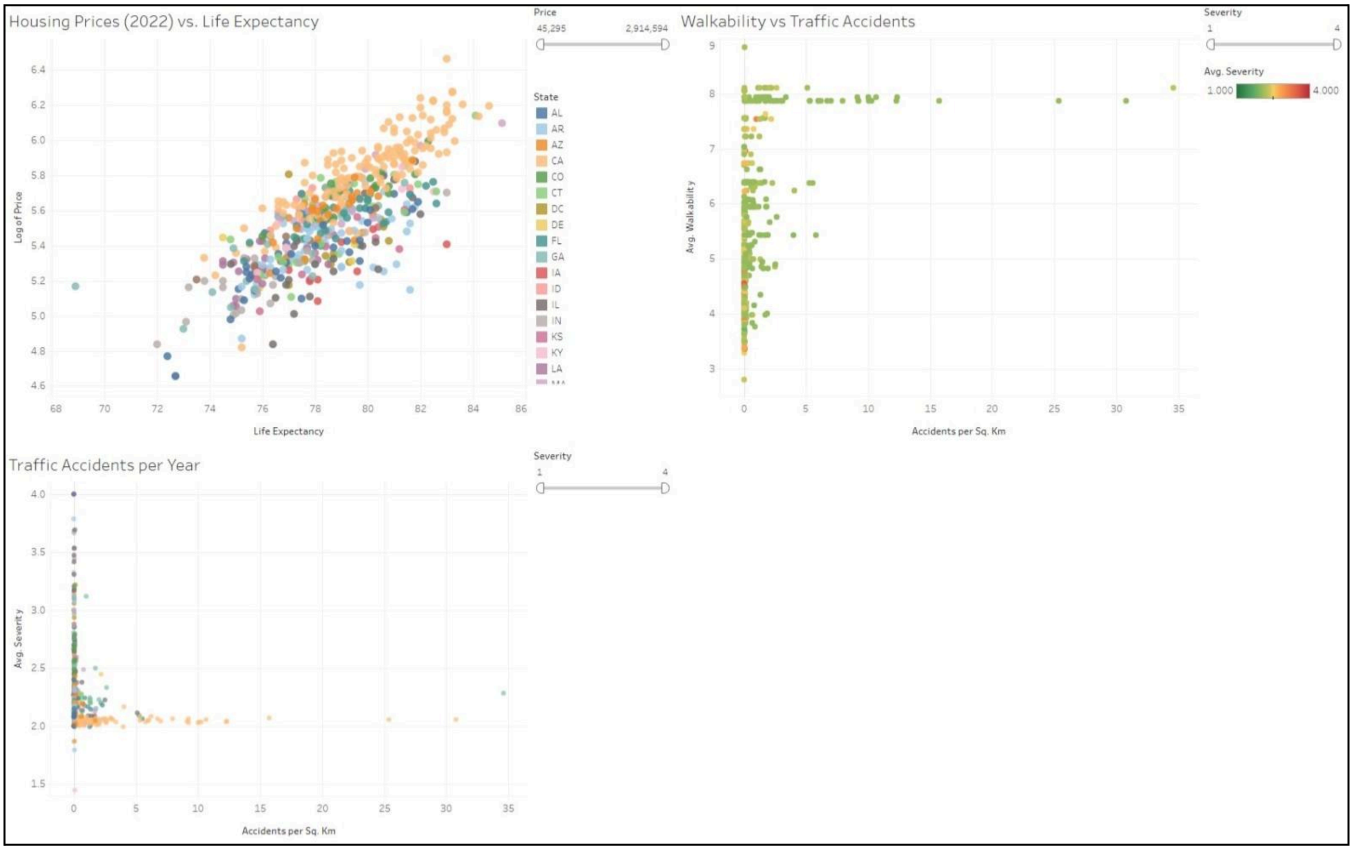

Figure 2: Regression plots.

Discussion

In our analysis and presentation of the data, we encountered several challenges and limitations. Some areas for improvement and potential future work include:

1. Interface Improvements: Enhance the interface for accessing data about a selected city to make it more user-friendly and informative.
2. Dataset availability and usability were limiting factors. For example, we could obtain data on many cities for housing prices. Meanwhile, we opted to use a 280-station dataset for weather since the other weather datasets we found required far more data processing.
3. Geospatial Data Challenges: Address the unique challenges associated with working with geospatial data and location boundaries to improve data accuracy and representation. This was especially the case for the walkability data. We opted to use CBSAs to determine the values of each city and utilized a makeshift method to adjust values by CBSA area. We can increase accuracy by using each State's Census Tracts shapefile (also available in Cartographic Boundary Files), which is of higher resolution since CBSAs each have multiple census tracts. However, we wanted to avoid the

possibility that individual cities may span multiple Census Tracts. Again, dataset usability and complexity significantly affected how much we could accomplish.

4. Tableau Limitations: Explore alternative visualization tools to overcome Tableau's customizability limitations and create more tailored and interactive visualizations. We could consider using Altair and Streamlit in Python to do these visualizations to overcome the limitations.

Future work could also involve expanding our project in various ways:

1. Regression Plots: Add more regression plots to analyze further the relationships between different factors affecting city livability. We plan to do a multi-variable linear regression to explore the relationships between different variables.
2. Detailed City Information: Create a dedicated section to display comprehensive information for a selected city, allowing users to understand the city better. For example, we can insert a table showing the value of different metrics of that city into the dashboard. When a city is selected, the corresponding city information is shown in the table. Currently, we only highlight the city across all charts, which is not as useful.
3. Additional Datasets: Incorporate more datasets, such as public transportation, quality of local schools, safety, and other relevant factors, to provide a more holistic assessment of city livability. This process can be implemented based on user feedback and opinion collection, in which we can decide the most important factors people care about when considering livability.
4. International Expansion: Extend the project to include cities in other countries, broadening its applicability and usefulness for users considering international relocation. However, we could encounter more inconsistency in the data when looking for more global datasets.
5. Central Visualization: Develop a central visualization that displays the best cities to live in based on user preferences, making it easier for users to find their ideal location. This user preferences input can be combined with an Analytic Hierarchy Process to get a final output of ranked cities based on livability and preferences.
6. Insights and Conclusions: Provide more insights and conclusions based on the results of these visualizations. Choose a way to measure and quantify the results. For example, we can find a way to measure the livability of a climate based on temperature and precipitation, then give a final index representing its livability. Is there a particular region with a very livable climate? How about housing prices? Is there a region with a well-balanced climate and low housing prices? We could consider assigning different weights to different factors using an Analytic Hierarchy Process and use the weights and indices to get a final result of a city's livability.

By addressing these challenges and exploring future work possibilities, we can continue to enhance our project and provide users with valuable insights to help them make informed decisions about where to live.

References

Brossoie, N., & Burns, D. (2020). What Makes a City a Good Place to Live and Grow Old? Journal of Applied Gerontology, 073346482095714. https://doi.org/10.1177/0733464820957149

G., D. J., & Daniel, W. (2002). The Dynamics of Metropolitan Housing Prices. Journal of Real Estate Research, 23(1-2), 29–46. https://doi.org/10.1080/10835547.2002.12091069

How We Rank the Best Places to Live & Retire. (2022, November 1). U.S. News & World Report. https://realestate.usnews.com/places/methodology

Kaduk, T. (2019). Taras Kaduk: Where are the places with the best (and the worst) weather in the United States? Taraskaduk.com. https://taraskaduk.com/posts/2019-02-18-weather/

Livability: Find Your Best Place To Live, Work and Play. (2023). Livability.com; Livability - Journal Communications, Inc. https://livability.com/

National Center for Health Statistics. U.S. Small-Area Life Expectancy Estimates Project - USALEEP: Life Expectancy Estimates File for United States, 2010-2015. 2020; https://www.cdc.gov/nchs/nvss/usaleep/usaleep.html. Accessed April 8, 2020.