Finding the Perfect City: Data Analysis
Website: Quality of Life

Yasmine Aitouny, Daniel Carmona, Terisha Kolencherry, Pooja Nagrecha

Overview

The purpose of our project was to visualize data pulled from an API in ways that allow for a given user to explore different metrics and compare metrics between cities. The Teleport API gave us an abundant list of cities including scores for the following: Housing, Cost of Living, Safety, Economy, Education, Business Freedom, etc. The categorical data were scored from 1-10 by inhabitants of any given city. The higher the score, the better. The scores help a user decide on where to live based on factors that are important to them. Our data analysis we performed includes many charts to visualize the scores in a meaningful way and help choose your perfect city.

Data and Model

The data source we handled was the Teleport API. It is a database that focuses on the quality of life for various cities around the world and pulls data from different sources such as the World Bank, World Health Organization, United Nations, Without Borders, Air BnB, and more. During data exploration, we realized that the API was structured like a Russian nesting doll with many layers to unravel. The extraction required plenty of JSON traversing. Due to the complexity of API calls needed, we decided to refrain from calling the API on the live website and we instead opted to create a CSV based on the API calls and used this CSV as our basis for the visualizations.

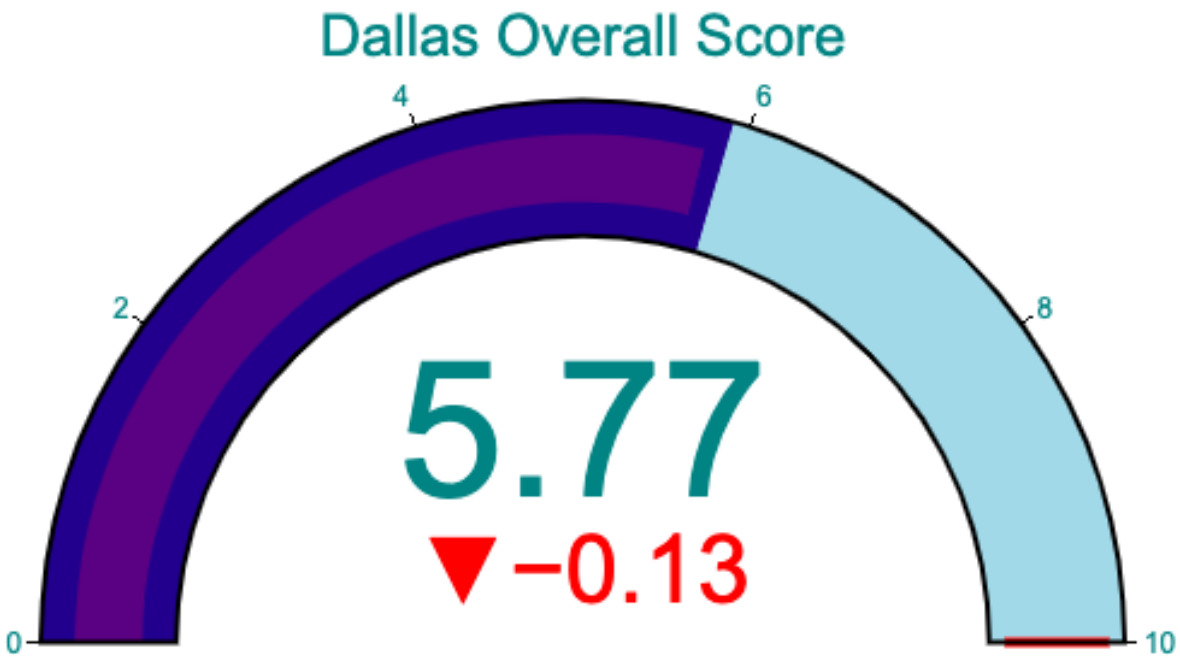
Extraction Process

The Teleport API had some limitations in the amount of cities listed. Because of this, we could not just use any random list of cities for our API calls. In order to find the total list of available cities, we performed a scrape of the main Teleport website to extrapolate the city names. With the list of cities, we then had to find the geolocation data (lat, long). To accomplish this, we performed calls to the google places API and mapped the data to the city name which allowed us to generate a leaflet map. Further cleaning of the data was needed however, as some of the city names were mapped to the incorrect city (e.g. Birmingham, AL instead of Birmingham, UK). The cities that were incorrectly mapped were corrected manually. The remaining categorical data was extracted from the Teleport API through the various JSON paths. For the leaflet map, the CSV was converted to geoJSON for ease of use. We used the convertcsv.com website to accomplish this.

Chart Building

In order to build our visualizations, we used various javascript libraries - Jquery, D3, Leaflet, and Plotly. We built four types of charts - a set of gauge charts made using Plotly, a radar chart made using D3, a circular bar plot made using D3, and a cluster map made using Leaflet. The charts were

separated into two categories: charts by indicator and charts by city. The charts by indicator category showcased the values for all 256 cities in our dataset for any given indicator and included the cluster map and circular bar plot. The charts by city category showed all indicator values for a given city and included the gauge chart and radar chart. Screenshots of each chart are included below:



Dallas

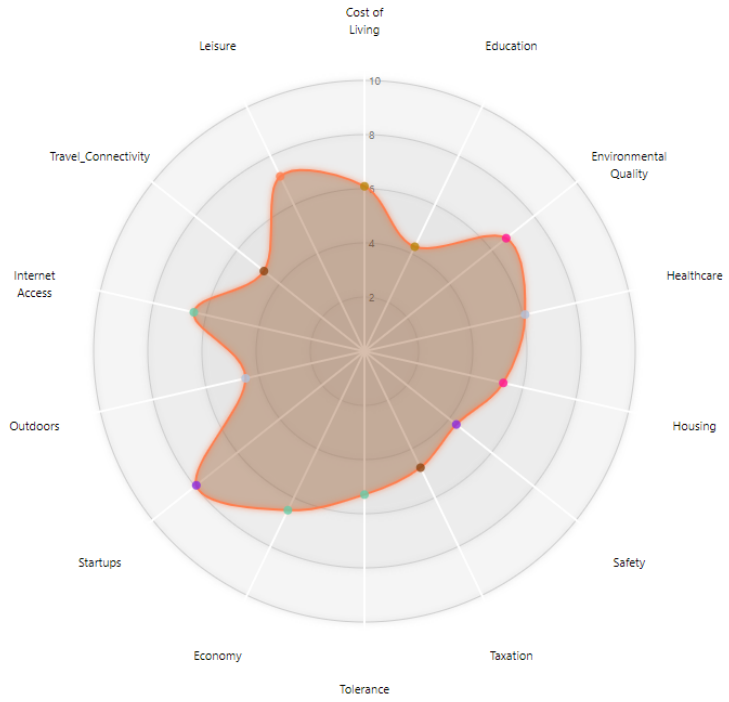
This is a visual representation of Dallas scores, the city's average score is 5.73/10

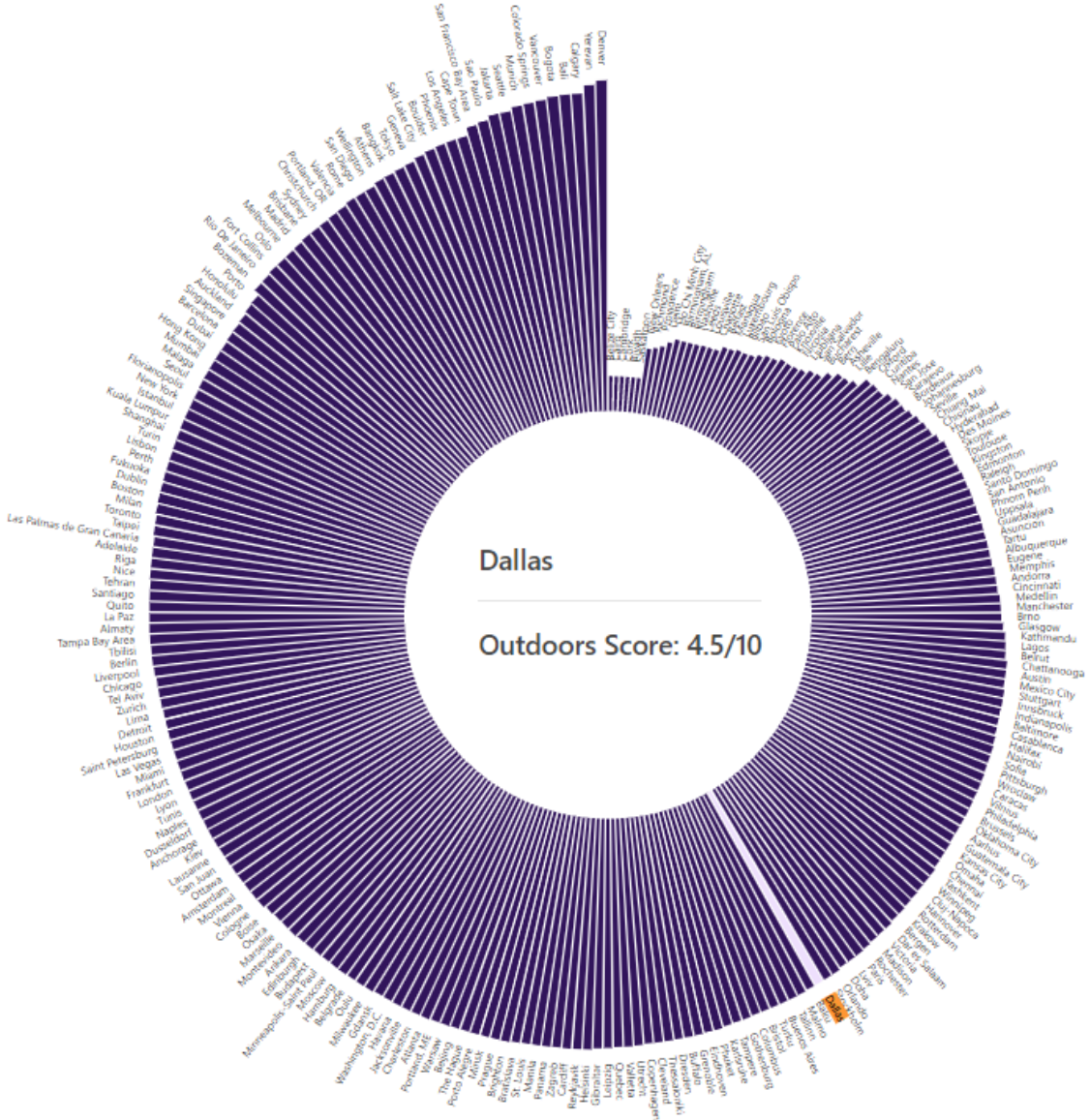
Choose city

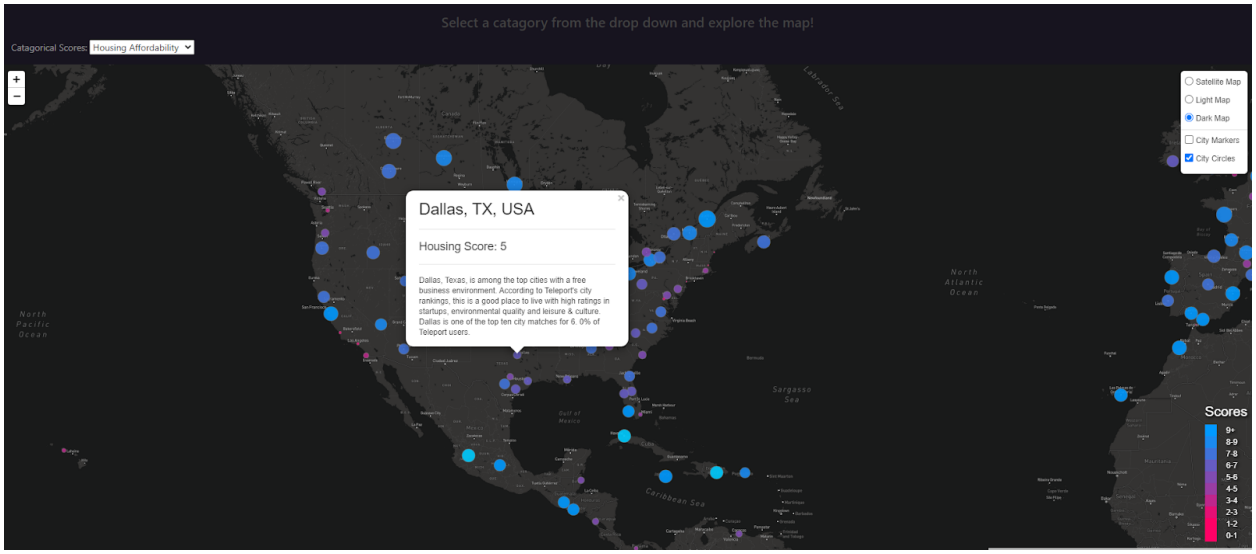
Dallas

City Description

City: Dallas
City_Summary: Dallas, Texas, is among the top cities with a free business environment. According to our city rankings, this is a good place to live with high ratings in startups, environmental quality and leisure & culture. Dallas is one of the top ten city matches for 6. 0% of Teleport users.
Address: Dallas, TX, USA
Latitude: 32.78
Longitude: -96.8
Housing: 5.26
Cost_of_Living: 6.09
Startups: 7.94
Venture_Capital: 4.92
Travel_Connectivity: 4.75
Commute: 4.47
Business_Freedom: 8.67
Safety: 4.34
Healthcare: 6.09
Education: 4.28
Environmental_Quality: 6.69
Economy: 6.51
Taxation: 4.77
Internet_Access: 6.46
Leisure: 7.17
Tolerance: 5.29
Outdoors: 4.5




Dallas
Outdoors Score: 4.5/10



An area of interest during the chart building process was adapting code from charts we found online to our dataset. For the radar and circular charts, we were inspired by the charts below and sourced code from websites to build the visualizations. In some cases, adapting code meant updating versions of D3, and in others it meant going through and commenting out portions in order to understand what portion of the code did what to the SVG.

Pain Points

One of the pain points we encountered was extracting geo-location data. Since some cities had duplicate names, but were located in different countries, we had to go in and manually update information.

Another one of the paint points we encountered was aggregating our individual visualizations. Our group decided to have separate pages for each visualization since each of the visualizations had different dropdowns. The downside of this approach was that we needed to pay more attention to detail across each page to make it cohesive and consistent. An example of this attention to detail is the styling of the navigation bar. At one point, while the content of our bar was the same, the style differed between pages. Additionally, some elements of our visualizations didn't transfer over well to the hosted page. An example of this lack of transferability would be the tool tip in the circular barplot, which had to be manually adjusted and ultimately was off-center for our live demo.

Data Takeaways

Much can be gleaned from the analysis of the quality of life charts. The three highest scoring cities from the dataset were Singapore in the first place, followed by Munich and Toronto. While, not one city in the United States was in the top 20 cities for quality of life. In contrast, the US performed best

in economic growth, startups and education. Alternatively, Cities in developing countries scored well in cost of living and housing while scoring very poorly in education and environmental quality. Moreover, there is a slight correlation between indicators, as cost of living score increases, environmental quality score declines. Cities with high education scores have highest scores in business freedom. Education, environmental quality and economy affect the housing and cost of living score and vice versa. Cities with geographical features such as mountains and beaches tend to score higher in outdoors as expected. Major metropolitan cities scored high in Leisure and Culture. Examples include: New York, Paris, LA, San Francisco, Moscow, Paris, Tokyo etc. Additionally, clusters in Europe, Africa, East Coast and West Coast of the United States have similar scores, and based on the radar chart, these clusters also have similar characteristics.

Limitations and Looking Forward

One of the limitations of this dataset would be the lack of transparency when it comes to how scores are calculated. While the website itself does list some of the factors taken into account when calculating scores, some aspects are still proprietary. The methodology is important because it impacts the trends that a user can take away from our visualizations. For example, major metropolitan cities scored higher in Leisure and Culture, but an argument can be made that other cities that might not be major tourist attractions have more culture to experience. Of course, some of this fluidity is unavoidable since the API is trying to quantify things that are based on opinion.

If our group were to do another iteration of this project, we would likely try to incorporate some of the other data, such as average salary. This information wasn't included in this version of the project since the API endpoint was different from the scores that we pulled.

In the future, our group would have likely sat down and sketched what each page of the site should look like in terms of if there's going to be a header for each page, if there's a card for analysis for each graph, etc. and then make a bare-bones outline of what needs to be consistent across each page. This approach would have allowed for individual creativity on the page, while maintaining some consistency across the site. While our visualizations were our main focus, building some very basic and cohesive HTML/CSS could have eased the second half of the project. Additionally, our group would have set aside a whole day for deployment and debugging. The great part about the structure of this project was that we were able to effectively divide up the work so that everyone had something to do. The trade-off was that everyone codes slightly differently, and so last-minute debugging each others' code proved to be more time consuming than originally expected.