Program 2: Retirement Factors Visualization

Jacob Low

Abstract—When selecting a retirement location there are many factors that must be taken into consideration. Safety, health care, pollution, climate, and cost of living are a few of the many variables that must be taken into account by prospective retirees. This paper will identify the most desirable country for retirement through data analysis and web-based visualizations.

Index Terms—Retirement, data visualization, choropleth map, quality of life

1 Introduction

Many individuals base their retirement location on personal preferences such as proximity to family or community, but the following visualization method will focus on standardized attributes of many common retirement destinations across the world. For each country, a quality of life index is calculated from the 8 most important retirement factors: purchasing power, safety, health care, cost of living, property price to income ratio, traffic commute time, pollution, and climate. By visualizing the quality of life index for each country along with it's geographic location, we will be able to identify geographic trends as well as effortlessly identify the best countries to retire in.

2 Methods

An interactive 3D globe can be used to visualize these attributes. The globe is divided at the country level, where the color of each country corresponds to it's quality of life index. This granularity is effective due to the unique styles of politics, culture, and economy of each country.

This visualization is created using a data set from Numbeo.com [5], a crowd-sourced global database of quality of life data. This paper will discuss how to combine this data set with a topojson file and use it to construct a 3D web-based visualization.

The data set from Numbeo.com contains 8 retirement factors represented as indices. The *safety index*, *health care index*, and *pollution index* are calculated with data from crowd-sourced surveys. Data with a maximum age of 36 months contributes to all three of these attributes. The *climate index* is calculated with a formula that use dew point and temperature to estimate the climate likeability of a given country. The *traffic index* is calculated with a formula that use the transportation times, inefficiencies, and CO2 emissions of a given country. The *property price to income ratios* is calculated as the ratio of median property prices to median familial disposable income. The *purchasing power index* and *cost of living index* uses grocery, restaurant, transportation and utilities prices relative to New York City.

All of these indices are taken into account when estimating the quality of life of each country. However it should be noted that complete data sets are available for only 86 countries. This is acceptable for this visualization because all of the most popular retirement destinations are captured by the data set.

The topoJSON file used in this visualization is taken from an example of Versor Dragging [4] by Mike Bostock on the D3 website. The file contains condensed geometry of every country with a 1:110m small scale. The countries are stored in an array of geometries where each object in the array contains the geometry of each country and a list of properties. If a country is a single land mass it is stored as a polygon, and if a country contains multiple land masses it is stored as a multipolygon.

Although a GeoJSON file could have been used for this project, the TopoJSON file format is more efficient in term of run-time and storage space since it merges land boundaries together automatically. In almost every geographical visualization boarders are shared between the geometries. In the GeoJSON file format, these boarders are repeatedly stored in each geometry, but in the topoJSON file format, these shared boarders are stored once and referenced by each geometry, decreasing the amount of space required to store the shapes.

2.1 Data Extraction and Organization with TopoJSON

The following method to combine the Numbeo.com data set into a single topoJSON file uses the TSV file containing the quality of life indices and the topoJSON file containing geographic shape data for each country. The creation of this file can be divided into three steps: data extraction, data analysis, and data combination.

2.1.1 Data Extraction

The data available on Numbeo.com can be converted into a Tab Separated Value file by simply copying the table into an excel spreadsheet and saving the file in the desired format. The TSV file is essentially a table where each row represents a different country and each column pertains to a different attribute. The attributes that are relevant to this analysis are the 8 retirement factors mentioned above as well as the new 'quality of life' attribute.

2.1.2 Data Analysis

The quality of life index available in the Numbeo.com data set is an estimation of overall quality of life of a specific country using the 8 retirement factors: Q: quality of life index, PP: purchasing power index, R: housing to income ratio, COL: cost of living index, S: safety index, H: health care index, T: traffic index, POL: pollution index, CLI: climate index.

$$Q = max(0, \frac{PP}{2.5} - R - \frac{COL}{10.0} + \frac{S}{2.0} + \frac{H}{2.5} - \frac{T}{2.0} - \frac{POL * 2.0}{3.0} + \frac{CLI}{3.0}) \quad (1)$$

A high index score does not necessarily increase the quality of life index. In fact, a country with a lower pollution index, house price to income ratio, traffic index, and cost of living index is better in terms of quality of life. On the other hand, a country with a higher purchasing power index, safety index, health care index, and climate index is also better in terms of quality of life. When calculating the quality of life index, where a higher score represents a more desirable retirement destination, these inverse scales must be taken into account by adding beneficial attributes (PP, S, H, CLI), and subtracting undesirable attributes (R, COL, POL, T).

Furthermore, each index has a different range of possible values (see Figure 1). In order to handle the variety of ranges, each attribute is scaled up or down. For the purpose of this visualization, I assume each attribute has an impact on the overall quality of life. However, attributes can be excluded from the calculation by scaling the attribute term by 0 or exaggerated by scaling the attribute term by a non-zero constant.

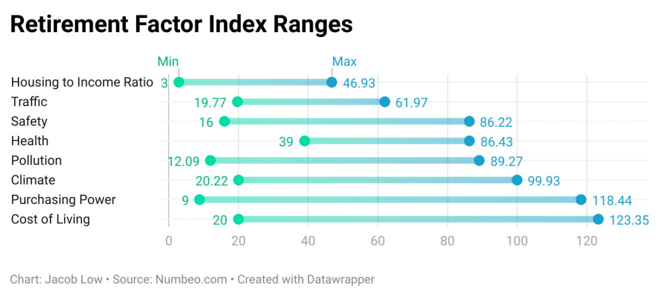

Fig. 1. Range plot visualizing the maximum and minimum values of each index used in quality of life calculation.

2.1.3 Data Combination

The quality of life TSV file can now be combined with the topoJSON file using Data-Driven Documents (D3), a JavaScript library for analyzing and presenting data on the web. The goal is to add each of the 9 TSV attributes for each country to the properties of it's corresponding geometry in the topoJSON file. In order to achieve this, both files are loaded into JSON and TSV objects, which allows for the data to be easily accessed and manipulated in the same environment. We then iterate through each TSV object and find it's corresponding geometry within the JSON object. After locating the matching objects, each TSV attribute is added as a new property to the JSON geometry. Once all TSV objects have been processed, the new topoJSON file can be used to create the 3D globe visualization.

2.2 3D Globe Creation

The following 3D globe visualization is built off of an excellent example by Jorin Vogel [2].

Since we are using topoJSON instead of a GeoJSON, the geometries must be converted into a collection of features in order to draw them using D3. These geometries are still stored in 2D space, since each country is an array of longitude and latitude coordinates connected by line segments. In order to map these 2D shapes onto a 3D globe, we have to use the azimuthal orthographic projection supported in D3. This projection converts the 2D coordinates into 3D coordinates on the surface of a sphere.

The color of each country depends on it's quality of life index stored within the properties of the geometry. The 'properties' object is a dictionary of attributes corresponding to the country. A linear color scale from black to green takes the quality of life index as input and calculates the green component of the RGB string used to fill each country. After each country has been filled we have a static 3D globe visualization (See Figure 2). Since data for only 86 countries was available in the Numbeo.com data set, many countries are colored grey to signify that no data exists.

Fig. 2. Static quality of life 3D globe visualization.

The legend (See Figure 3) in the top right corner of the visualization displays 9 labeled boxes filled with colors corresponding to equidistant values within the quality of life range. This tool adds meaning to the different colors in the visualization.

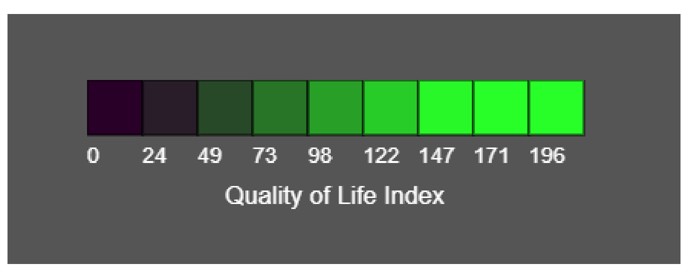

Fig. 3. Quality of life color legend for 3D globe visualization.

In order to make this visualization interactive, I expanded upon the features within Vogel's example. The starting visualization supported automatic rotation, mouse controlled movement, and mouse hover country labels.

I kept the automatic rotation, converting the static 3D globe visualization into a constantly revolving globe. Even if there is no interaction with the visualization, the globe will display all the information in a cyclical pattern.

I also chose to keep the mouse controlled movement capabilities, allowing the user to control the rotation of the globe independent of the automatic rotation. Click and drag actions on the canvas are translated from Cartesian positions into rotation variables that update the orientation of the globe.

Finally, I converted the country label that appears when the mouse is hovering a specific country into a drill down display. Without the drill down display, the quality of life index is the only attribute presented in the visualization. However, the quality of life index is a combination of the 8 retirement factors discussed earlier. I modified the hover functionality to display the country name, quality of life index, safety index, health care index, cost of living index, traffic index, pollution index, and climate index. The drill down display allows the user to see the specific retirement factors that contribute to the overall quality of life score. Additionally, when hovered with the mouse, the country changes to a bright red color to signify that the country is being selected (See Figure 4).

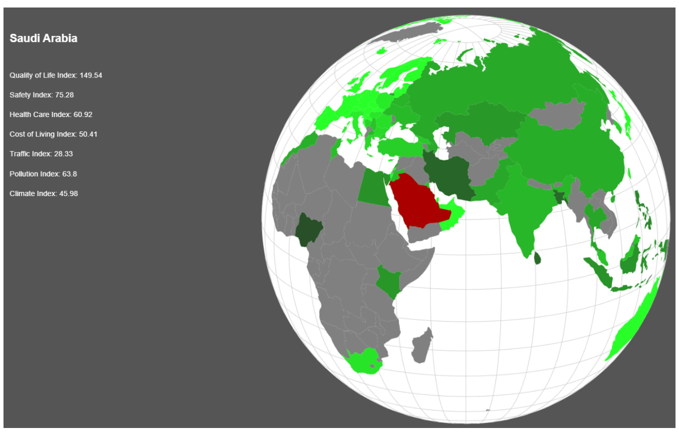

Fig. 4. 3D globe visualization with drill down display and country selection.

3 Results

The 3D globe interactive visualization can be used to identify which countries across the world have the most desirable retirement advantages. According to this visualization, Switzerland, Denmark, Netherlands, Finland, Australia, Iceland, Germany, Austria, New Zealand, and Norway are the 10 countries with the highest quality of life indices respectively. But what differentiates these countries so desirable?

Stacked bar graphs containing the 8 retirement factors for each country can be used to select a country that best suits a prospective retiree's personal preferences. I was hoping to add another dynamic element to the D3 visualization that would display the stacked bar chart for each country when hovered, but static visualizations will suffice for the purpose of this paper (See Figures 5 and 6).

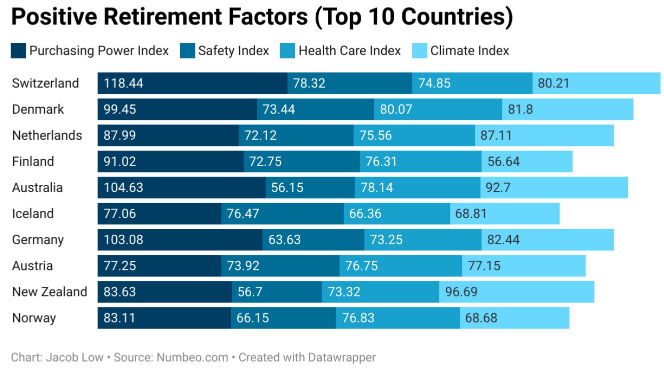

Fig. 5. Stacked bar graph detailing the desirable retirement factors of the top 10 countries.

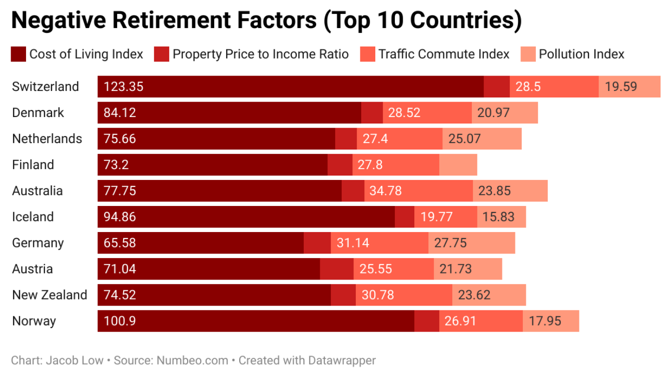

Fig. 6. Stacked bar graph detailing the undesirable retirement factors of the top 10 countries.

The stacked bar charts are separated by their contribution to the quality of life score. The attributes that increase the quality of life score are visualized in Figure 5, and the attributes that decrease the quality of life score are visualized in Figure 6.

According to the visualization, Switzerland is the best country to retire in. Out of the top 10 countries, Switzerland has most of the highest positive retirement factor indices. Surprisingly, even though Switzerland has the highest quality of life score, it also has a cost of living index far greater than any of the other top 10 countries. From this analysis of the 3D globe visualization along with the stacked bar charts, we can conclude that Switzerland is an excellent retirement destination suitable to retirees that can afford the exceedingly high cost of living. Similar analysis can be done on the remaining countries.

4 Limitations

This visualization method is limited to the scope of the data. Measurement for only 86 countries were used in the creation of this visualization, and many portions of the 3D globe are greyed out. In future exploration, a larger data set capturing data from as many countries as possible should be used.

Additionally, the formula for calculating the quality of life for each country was provided by Numbeo.com. In future exploration, it would be useful to dynamically modify the weights of each retirement factor in real time and design the visualization automatically update the quality of life score for each country.

5 User Interface

This visualization is web-based and runs on a HTML web page using JavaScript. All of the libraries used are imported via URL links to files hosted on the internet, so no additional software download is necessary. In order to run the visualization on your local computer, download the source code from the GitHub repository and launch the HTML file with a python server.

References

[1] "The Data Visualization Catalogue," The Data Visualisation Catalogue. [Online]. Available: https://datavizcatalogue.com/index.html. [Accessed: 30-May-2022].

[2] J. Vogel, "Create a 3D globe with d3.js," Create a 3D globe with D3.js –, 02-Feb-2017. [Online]. Available: https://jorin.me/d3-canvas-globe-hover/. [Accessed: 30-May-2022].

[3] M. Bostock, "Data-driven documents," D3.js. [Online]. Available: https://d3js.org/. [Accessed: 30-May-2022].

[4] M. Bostock, "Versor dragging," Observable, 25-Oct-2020. [Online]. Available: https://observablehq.com/@d3/versor-dragging. [Accessed: 30-May-2022].

[5] "Quality of Life," Cost of Living. [Online]. Available: https://www.numbeo.com/quality-of-life/rankings$_b y_c ountry.jsp.[Accessed : 30 - May - 2022]$.