STAT-515-MID TERM REDESIGN PROJECT REPORT (GROUP 12)

HOW LONG CAN YOU LIVE ON 1 MILLION DOLLARS IN EVERY STATE?

Introduction

The gap in the Cost of Living Indexes (2017) among the various US States and its possible influence on savings are revealed by newly available data. In the States of Iowa and Nebraska in particular, it may be challenging, if not impossible, to reap the rewards of years of laborious work without a comprehensive savings plan. A senior citizen's savings of $1 million in Hawaii can only last around 13 years due to the high Cost of Living Index (166.4), whereas it can last almost twice as long in Mississippi and Arkansas due to the relatively low Cost of Living Indexes (85.5 and 87.4, respectively).

Objective

The ultimate aim of this project is to make a graph easily understandable for both people who knew Analytics as well as for people who are unaware of it. Making graphs simple so that anyone viewing the plot could understand and analyze the content.

Bad graph

A bad graph is a graph which does not contain a clear representation of data and has poor mapping of values. Sometimes it is due to the misleading choice of color to represent the data or poor alignment of values or sometimes it is due to the choice of using the graph type for example like pie charts. The bad graph we have chosen does not give a complete picture of the content. The choice of colors too does not speak much of the necessary idea. Also, detailed and correct values should be shown in the graph which is clearly missing here.

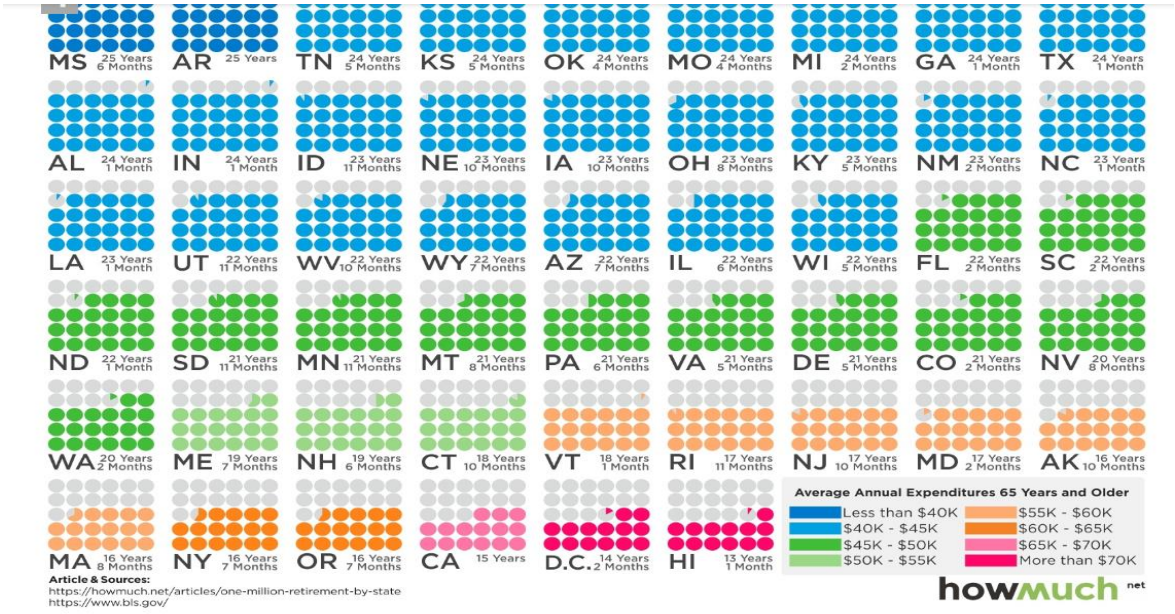

Fig-1: Snippet of bad graph

Packages used

library(ggplot2)

library(plotly)

library(usmap)

library(dplyr)

library(maps)

library(micromapST)

About Dataset

The dataset which we chose had 7 columns in it, namely Rank, Code, State, Cost of living, Annual expenditure, One Million would last for (years) and Years and Months. Rank is based on the annual expenditure. The code and states tell the list of states in the United States. The Annual expenditure tells the amount spent in each state annually. Also, the last 2 columns summarize how many years and months it takes to spend $1 million after retirement. With this dataset, we can categorize different states into different categories.

Dot plot

We began with the simplest chart, which is dot plot. It was interactive. It suggests the name of state and the ranking when we hover over it. With that, we were able to identify the most and least expensive States. Though it gave us some knowledge, we still felt that the graph could be made better.

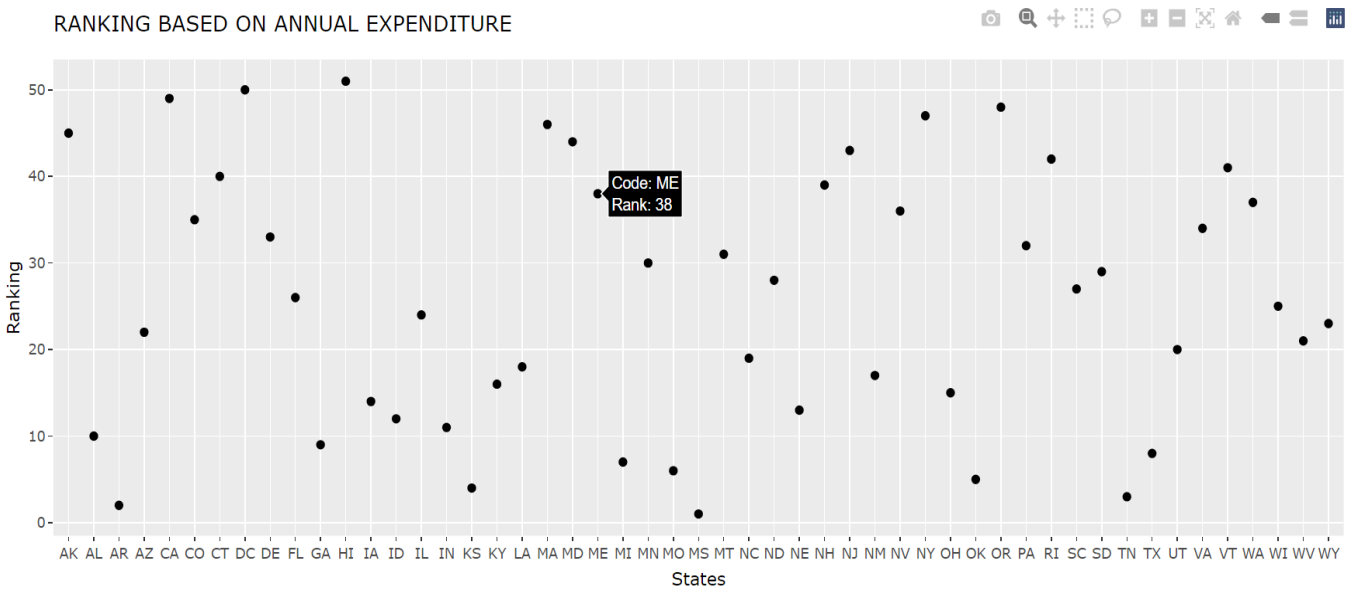

Fig-2: Dot Plot

The graph gives the idea about the expenses to live in different states respectively.

Bar chart-

We then decided to plot a bar chart. The bar chart was made between the States and the years it takes to spend a million dollars. It was made interactive. This bar chart helped us by giving a better idea about the plot. To get a further better understanding, we sorted the bar chart based on the values.

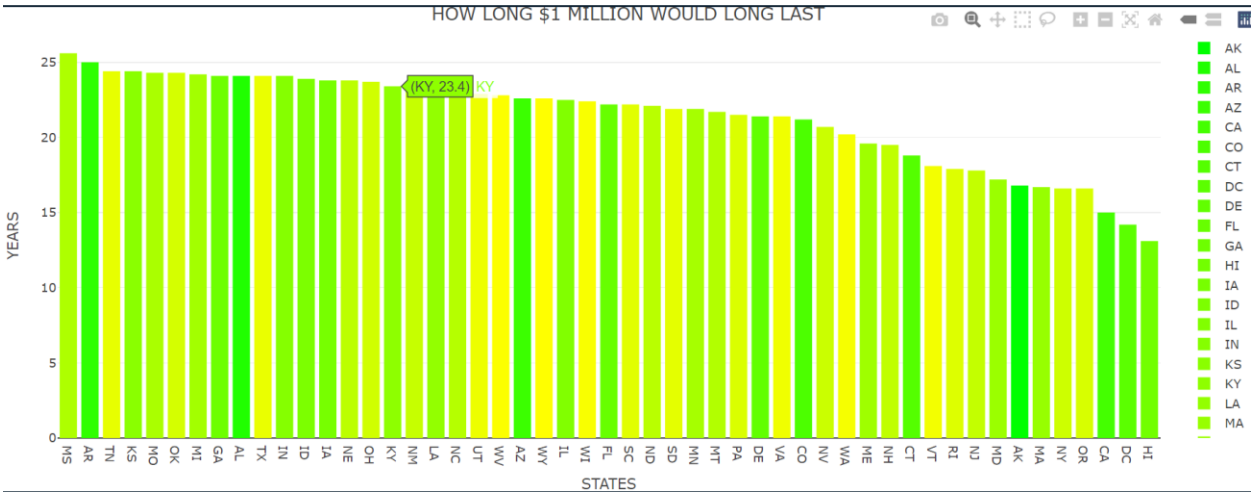

Fig-3: Bar chart

From this bar chart, we can get a clear understanding of the States where $1 million could be spent. Even though this bar chart gave a good image, we still wanted to make it even better.

Micromap

A micromap is basically a set of organized small maps that show the geographical patterns in a highlighted manner. It is used to identify the spatial content, make comparisons between the attributes given in a dataset and to explore the dataset in a geospatial context.

A linked micromap for the dataset Can you retire on $1 Million dollars is plotted and you can see that the comparison is made between the attributes Cost of Living Index in dollars and timeline for $1 million will last in years. There are 4 columns in which the first two represent the US states and corresponding name of each state and the second two columns represent the dot plot for the Cost-of-living Index and how many years will the $1 million dollars last. The analysis between the values for each state has been done and it is observed that the state Mississippi has a cost-of-living index of 85.5 and the $1 million dollar will last for 25 years. This the best state to get retired and settle because the cost of living is lower, and the years value is higher. The state which has the highest cost of living index is Hawaii which is around 166.4 dollars, and the $1 million dollars will last for only 13 years. Therefore, Hawaii is not the ideal state to get retired and settle down. The mapcum data frame is used which gives the outline for each state. South Carolina is the median state that has the cost-of-living index of 98.6 dollars and $1 million years will last for 22.2 years.

The states above South Carolina have lowest cost of living and highest years for $1million dollars and the states that lie below South Carolina have the highest cost of living index and lowest years of $1million dollars to last.

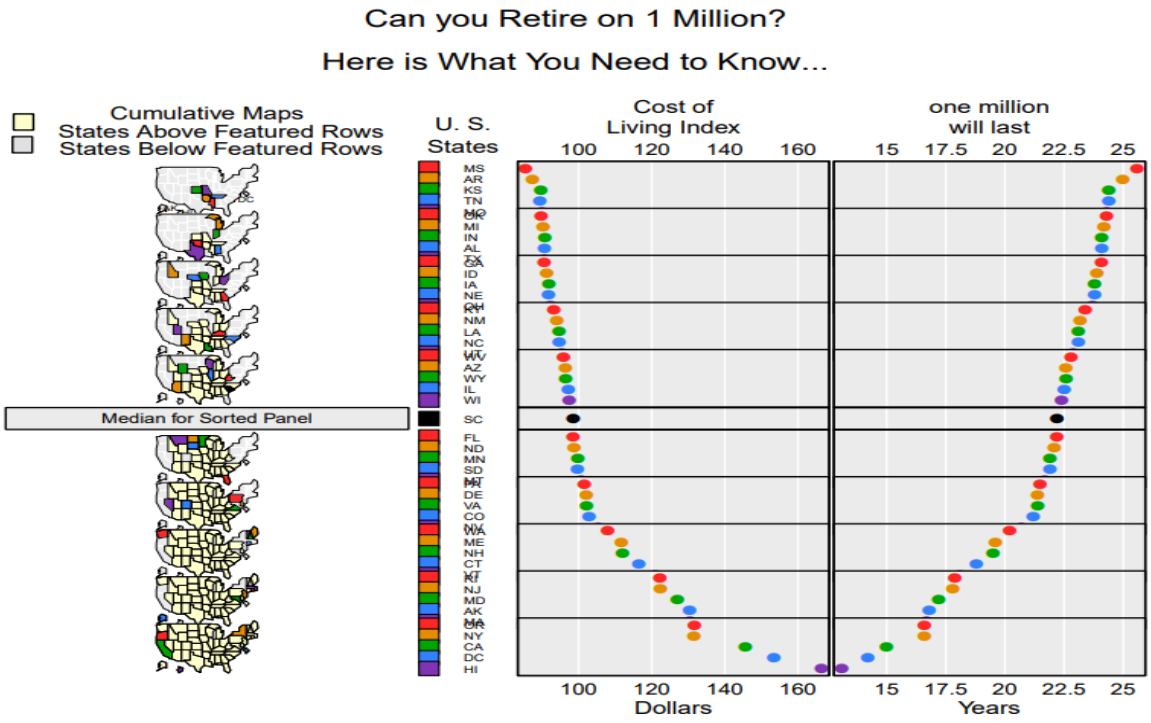

Fig-4: Micromap

Final Plot

To redesign the bad graph, As a final approach, we have used an interactive user-friendly plot. We can easily hover to find the annual cost of living, rank, state Year and month. These criteria are based on how the 1 million will last after retirement. In bad-graph colors and comparisons made based on Annual expenditure, dots are added based on year and month which is countable and the user finds it difficult to interpret once seen. We have redesigned by using an us-map with a representation of values even easier just by hovering gives the entire data which is easy to interpret for a layman as well for an analyst. So, highly interactive with visualization. Code Plotly package highly helped us to plot the map. To be more interactive, we have added the hover data represented using the hover information argument. Additionally, add trace argument helps in poly visualization.

Fig-5: Plot

Projecting Top 4 countries with High Rank and High Cost of Living. In this part, We have represented the Top four Countries which are ranked High based on costs of living. In this section, we have represented the top four countries in terms of living costs. Plotly and us map packages were powerful enough to highlight a particular number of areas to be plotted and visualized. Initially, there was no abbreviation for states, and they were not listed in ascending order.

The Us map package's state pop and order arguments were useful in correcting the state pop and order argument.

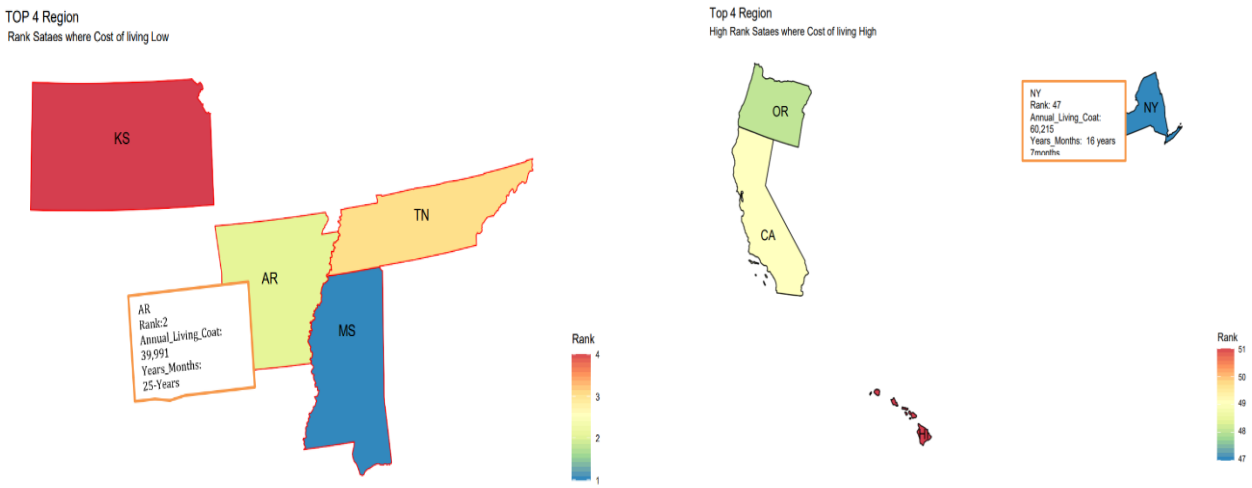

Fig-6: Maps showing top 4 countries having High and Low Rank based on Cost of Living

Changes Compared to Bad-Graph

The bad graph which we took was not satisfactory. Also, the colors didn't give complete understanding of the data. We redefined the same dataset as a interactive plot and a micro map. Upon hovering our final plot, we could get the number of years one could stay with $1 Million is respective states. We feel this graph is more self-explanatory and gives better understanding. Also with the maps, we could compare the expenses in neighboring states and visualize in a much better way.

Conclusion

The analysis of how $1 million dollars could last for how many years in each US state on an annual expenditure of 65 years and above is made with the help of using the R software. The implementation of the statistical methods, the improvement of graphs is made from a dot graph to US plot graph. So, the comparison of each state becomes clearer and easier to predict which state would be suitable to get retired and which state has the highest years of $1 million-dollar lasting period based upon the cost-of-living index. Mississippi would be the best state to get retired and Hawaii would be the worst. From this we can conclude that the use of good statistical functions and use of well-versed software like R can bring out a good graph.

Citations

1. Bar chart [2022] cran.r-project.org/web/packages/lessR/vignettes/BarChart.html [Accessed on 2/10/2022]
2. Interactive plots [2022] paldhous.github.io/ucb/2018/dataviz/week13.html [Accessed on 1/10/2022]
3. Plotly[2022] plotly.com/r/choropleth-maps/[Accessed 5/10/2022]
4. The plot_map() Retrieved from the link: stackoverflow.com/questions/68181278/adding-labels-to-plotly-map-created-using-plot-geo [Accessed 3/10/2022]
5. Visualization [2022] www.jotform.com/blog/bad-data-visualization/ [Accessed on 3/10/2022]
6. Good graphs [2022] scc.ms.unimelb.edu.au/resources/data-visualisation-and-exploration/data-visualisation [Accessed on 2/10/2022]
7. Referred class notes and presentations.