Redesigning a Bad Graph

Sai Goutham Kumar Akula(G01355914)
Gowtham Mukkara(G01353489)

Abstract: Rent is one of the major expenses for any individual. So, choosing a suitable city to live plays an important role. For our redesign project we considered a graph which displays the average rents in major cities in United States of America (USA) which is visualized as a 3d graph showing the USA map with the rents information and redesigning the graph into a bar-plot and micro-maps showing the data in a more efficient way. This redesigned plot enables users to easily extract information just using the visualizations and using micro-maps they can easily compare the population of the state and the average rent.

Introduction:

Rent is one of the important factors when it comes to choosing a city to live. Almost half of central city residents in the United States are renters, as are the majority of low-income households. The rental prices for single-family homes grew an average of 7.8% in 2021[CoreLogic Single-Family Rent Index] because of the Covid-19 pandemic and this is the reason why we considered redesigning this graph(figure1) and the output will be useful for users to easily get the information about the rents in major cities.

Bad-Graph: A graph is considered to be bad if the information displayed is not engaging, inefficient or misleading, also using of 3-dimensional graphs confuses the users and makes reading data more complicated, bad graphs can also be difficult to analyze, as they can contain a lot of information or very less information in them and also bad-graphs visualizes data that are too dense, or by graphs that have too many crossings.

Good-Graph: A graph is considered to be good if it is visually appealing and easy to read. It should be properly labeled, all axes and data points should be accurately represented and correctly proportioned. The graph should also be neat and tidy, with no stray marks or errant data points. The color differentiation between different data can also be helpful for the users to easily differentiate the data and no overcrowding of information. A good graph is also clear and concise. It should present the information it is meant to convey in a way that is easy to understand, without any extraneous details or confusing data without any ambiguity.

Redesign from bad graph to good graph: Dataset from the bad graph is collected and analyzed first and using R-Studio we created bar plots to represent the data easily using different libraries. The redesign makes the graph easier to read by adding labels and spacing out the information in order and making the lines thicker and easier to follow by adding colors

The micro-maps visualization is a way to show how a particular variable changes over a geographic area. It can be used to show the distribution of a particular variable, or to show how the variable changes over time. Also, Micro-Maps is created to visually represent the state wise

population and the average rent to easily make the data available in visualizations for the users to take a decision based on population and average rent.

Bad-Graph:

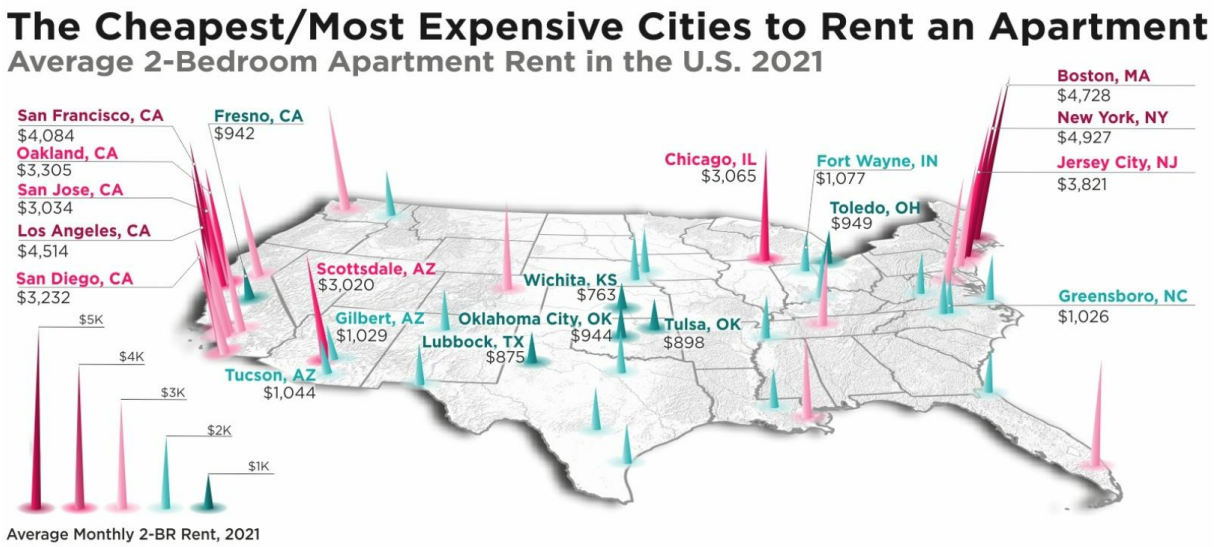

Figure1

The above graph visually represents the average rent for the most expensive and cheapest cities to live in United States of America (USA), the green bars represent the cities with the cheapest rents while pink bars represent the most expensive cities rents. The graph produced is a 3-dimensional graph which is not visually appealing and very hard to derive the information as the scale is not exactly visualized. Also, if we look at the expensive cities which are located mostly at the coastal areas appears to be clustered and it does not represent the data visually accurate. The use of 3-Dimensional graphs is not usually preferred for visualizations as there will be data manipulation with the scales and in this case the bars in the foreground obscure those in background. Also, when the graph is in 3-dimensions the designer should consider the perspective and viewpoint of the reader for correctly representing the data visually.

Good-Graphs:

After analyzing the data set, we came to know that the data has 3major columns which are Cities-States, Average Rent, Population. Using the data, we designed 2barplots which represent the expensive city average rents, cheapest city average rents and also a micro map which shows us the average rent and the population in a single visualization.

Expensive City Bar-Plot (Figure2): A bar graph breaks down the larger data into groups and represents these amounts of data by using bars of the length specified. It is something where you show distribution of data points or perform a comparison of values with all the subgroups of data selected. We used R-studio to redesign the given dataset into a bar plot where initially we imported the dataset and using the libraries ‘tidyverse’, ‘tidyr’, ‘ggplot2’ we designed a bar

plot defining the x-axis as City and the y-axis as Avg.rent and created a plot. Once the plot is generated, we also imparted the region as a fill color to represent what region in USA the city belongs to. We used different colors using the codes in R-studio to make the plot visually appealing. From this plot users will be easily able to identify the average rents for the city they are interested and know which region it is located in USA.

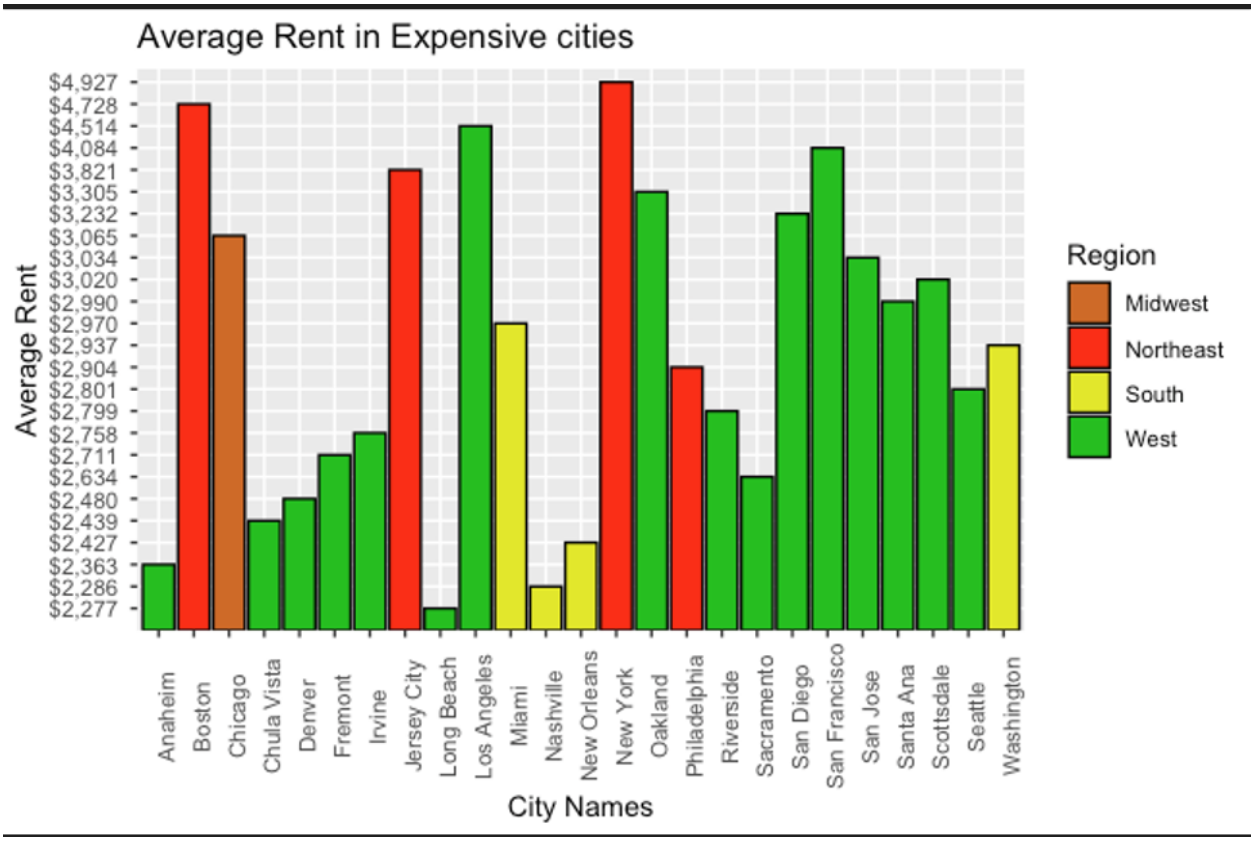

Figure2

Cheapest City Bar-Plot (Figure3): Bar plots are used to compare the data of one specified segment of data to the other with a visual understanding. When the changes are larger the bar plot looks more effective than a usual change in the data. Using R-studio a bar plot is designed by importing the data and defining the aesthetics as x = City, y=Avg.rent which will give us a bar plot when we define the geom_bar using the libraries, once a bar plot is created we defined the fill = Region to represent the region of the city. Labels were added using the labs function and also colors of the bars were changed using the scale_fill_manual function by defining the specific colors. The plot created is much easier to understand than the original visualization as users from any walk of life can easily understand and derive the data without any effort. The grid line provided in the background will also help to determine the average rents which is represented on the Y-axis.

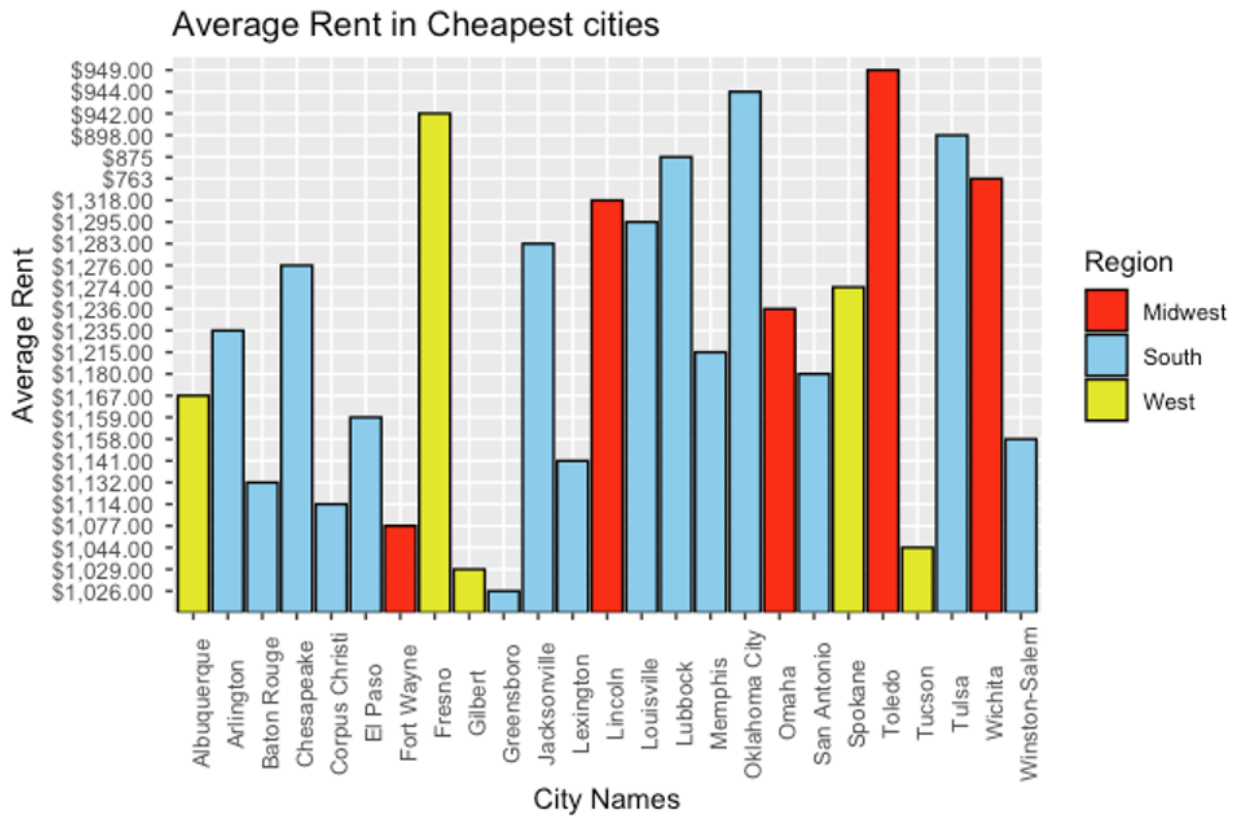

Figure3

Micro Map (Figure 4):A Micro map visualizes the whole data from the data set in a single picture. To display the statistical summaries associated with the whole units or the polygons. Displaying both the statistical and geographic distributions by linking the summaries for a better understanding. Uncovering and displaying the patterns and their associations with respect to the environmental and demographic factors to link the substituents. A linked micro map graphical design used distinct colors for its statistical elements that need to be categorized.

The micro map we created represents the variations in average rent with respect to population and the population is shown as dots while the average rent is shown with arrows. The map is created using the R-studio and “micromapST” library which enables us to define the parameters for the columns and the labels and once these are defined a data frame is created to store these variables and using the micromapST the map is generated using the data frame and the title for the micro map is also given. Using this micro map users can easily find both the population factor and the average rent in that city to choose a city of their choice.

Conclusion: To conclude the bar plots and micro maps are user friendly to convey information in the simplest form possible and from the observations the coastal cities are highly expensive compared to the central based states in the USA and most of the cities in the California are observed to be highly expensive compared to the other states. Thus, using the redesigned graphs users have a source to compare the rents in different cities also taking population into consideration.

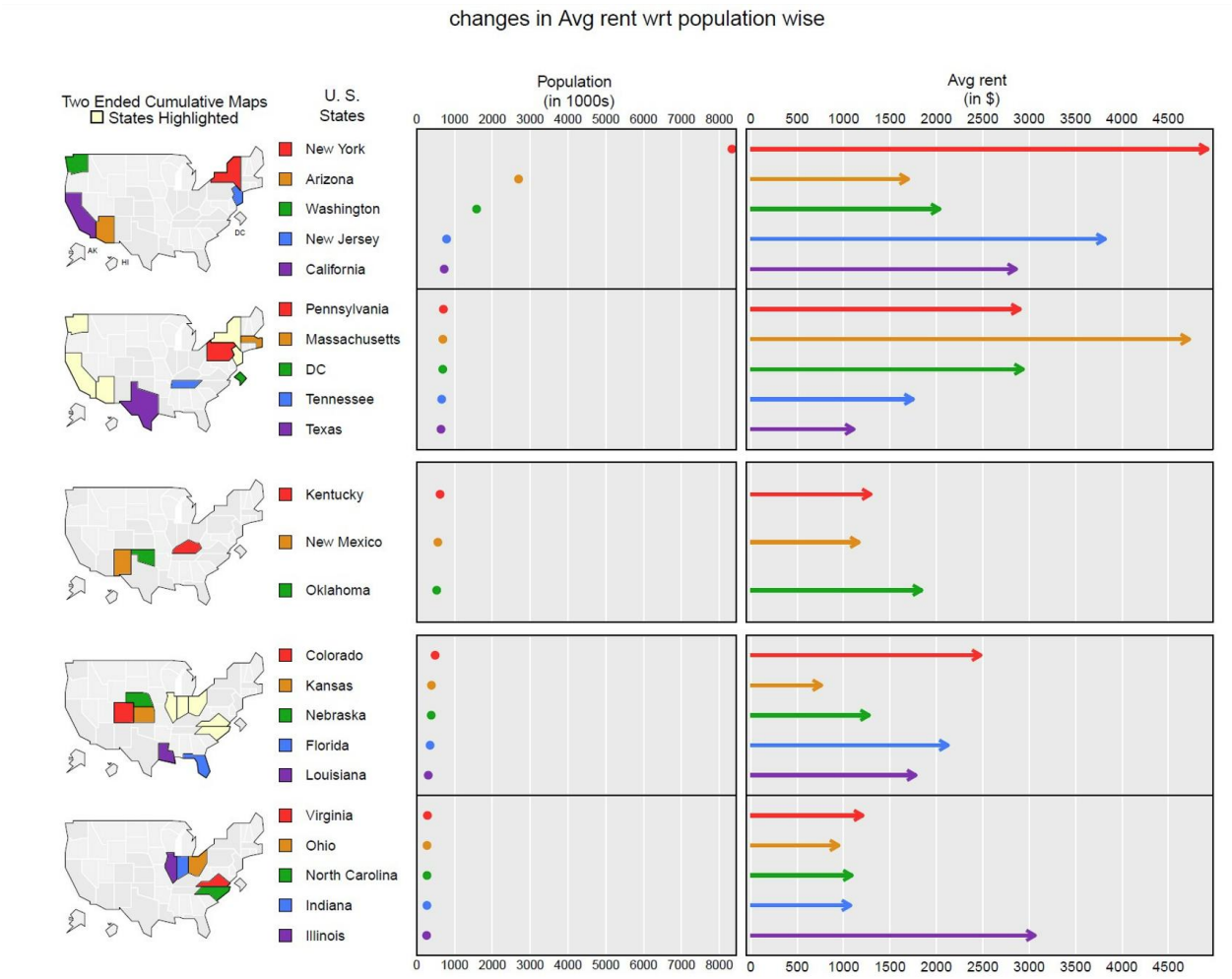

Figure4

REFERENCES:

1. 'The Cheapest and Most Expensive Cities to Rent an Apartment in America'. HowMuch, https://howmuch.net/articles/the-cheapest-and-most-expensive-cities-to-rent-an-apartment-us-2021. Accessed 9 Mar. 2022.
2. Tidyverse package: Kuk, John, et al. 'The COVID-19 Pandemic and the Rental Market: Evidence From Craigslist'. American Behavioral Scientist, vol. 65, no. 12, Nov. 2021, pp. 1623–48. DOI.org (Crossref), https://doi.org/10.1177/00027642211003149.
3. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." Journal of Open Source Software, 4(43), 1686. doi: 10.21105/joss.01686.
4. Micromapst package: Daniel B. Carr and Linda Williams Pickle (2010), Visualizing Data Patterns with Micromaps, Linda Williams Pickle, James B. Pearson, Jr. and Daniel B. Carr (2014), micromapST: Exploring and Communicating Geospatial Patterns in U. S. State Data
5. Ggplot2 Package: Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.