

Graph Redesign Project

Introduction

When the information being shown is unclear and the reader cannot understand it, the graph is said to be a bad graph. Basically, it presents the information in a way that deceives anyone viewing the graph and causes them to draw false conclusions. In this report we will redesign a graph showing the credit card debt of USA by state in years 2019 and 2020, and the percentage change in debt. We first made a scatterplot using ggplot() showing two graphs, one for the year 2019 and another for 2020 to understand the data, then we made a micromap using micromapST() showing the data state-wise and finally we redesigned it into an interactive barplot using girafe().

Bad Graph

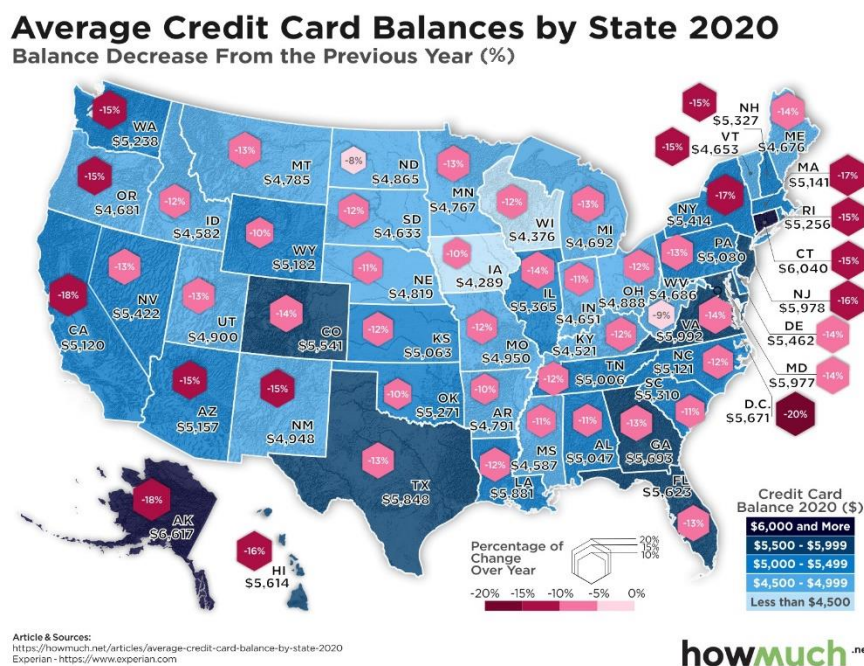


Figure 1: Bad Graph

This graph shows the average credit card balance by state in year 2020 and percentage decrease from year 2019. As we can see in the graph, information shown in large states like Texas is understandable but in the right corner of the graph the data is too crowded with values and the states are extending by using lines to depict information.

The disadvantages/mistakes in this graph are:

1. The values and change in percentage are not visually clear and understandable for most of the graph.
2. There is no scale, grid or legend for the representation which makes it difficult for new readers to understand the data.
3. The graph does not show accurate representation of data by state and it does not show the previous year balance values.

Redesign Goals

We will redesign this graph, focusing on all the errors and displaying values and percentages in a clear manner by mapping them on a scatterplot, moving on to a micromap, and finally to an interactive bar plot that when hovered, displays the state's name, the percentage change in the balance, and the balance in the years 2019 and 2020.

Redesigning the bad graph

1. Refining the dataset

The values of credit card balance in year 2019 and 2020 are in thousands which can crowd the scale of the graph, so we refine the dataset by dividing those values by 1000 so that the scale can look clear for the reader. We also added columns for state names abbreviation and tooltip_text for interactive bar plot.

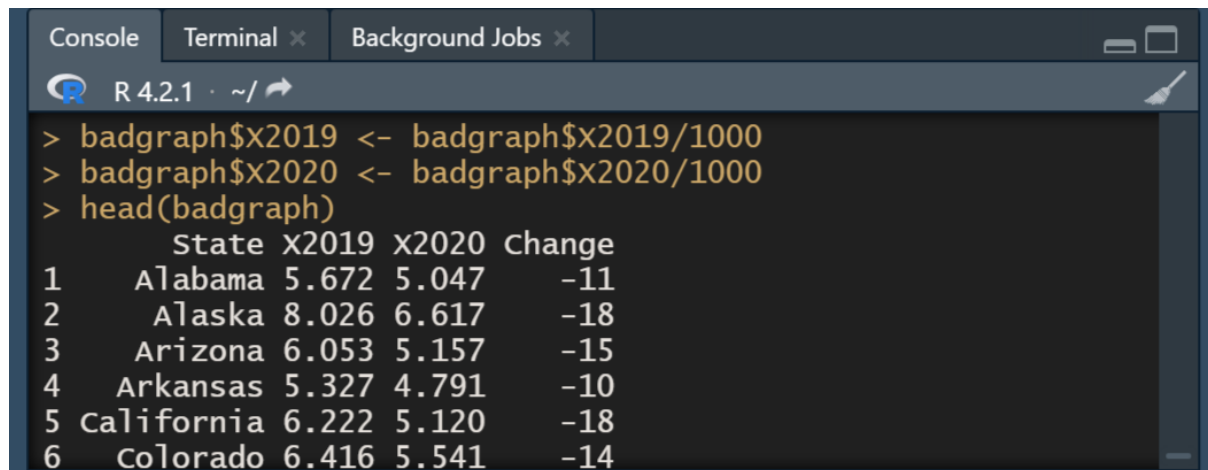


Figure 2: Refined Dataset

2. Designing a scatterplot

We first plotted a scatterplot by combining two graphs for year 2019 and 2020 by state to understand the values of the dataset and how they changed over the year.

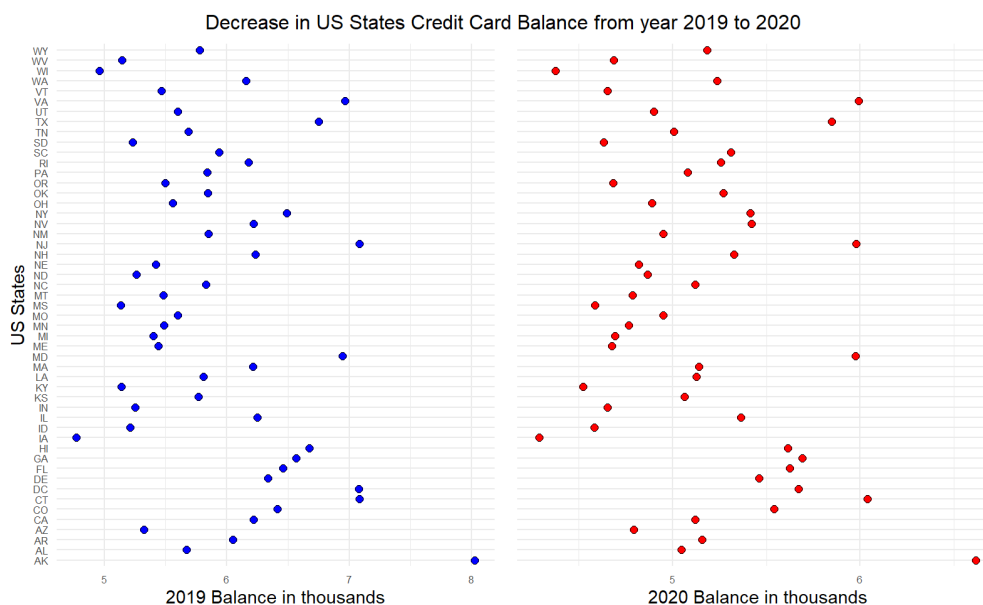


Figure 3: Scatterplot of dataset showing the balance for year 2019 and 2020 state-wise

This graph was used to understand the values in each state and change in balances. We used `ggplot()`, `geom_point()` and `grid.arrange()` functions to make this scatterplot.

But in this graph, we cannot clearly understand what data is being visualised, hence we further redesign the dataset to make it more clear to the reader.

3.Designing a micromap

By using `micromapST()` we plotted the dataset into a micro map which shows the percentage change in credit card balance and the change in amount from 2019 to 2020. We chose a cumulative type of map representation, represented the percentage change by dot plot, change in balance amount by arrow plot and specified clear labels for each column. We sorted data based on percentage change in balance of each state from highest to lowest.

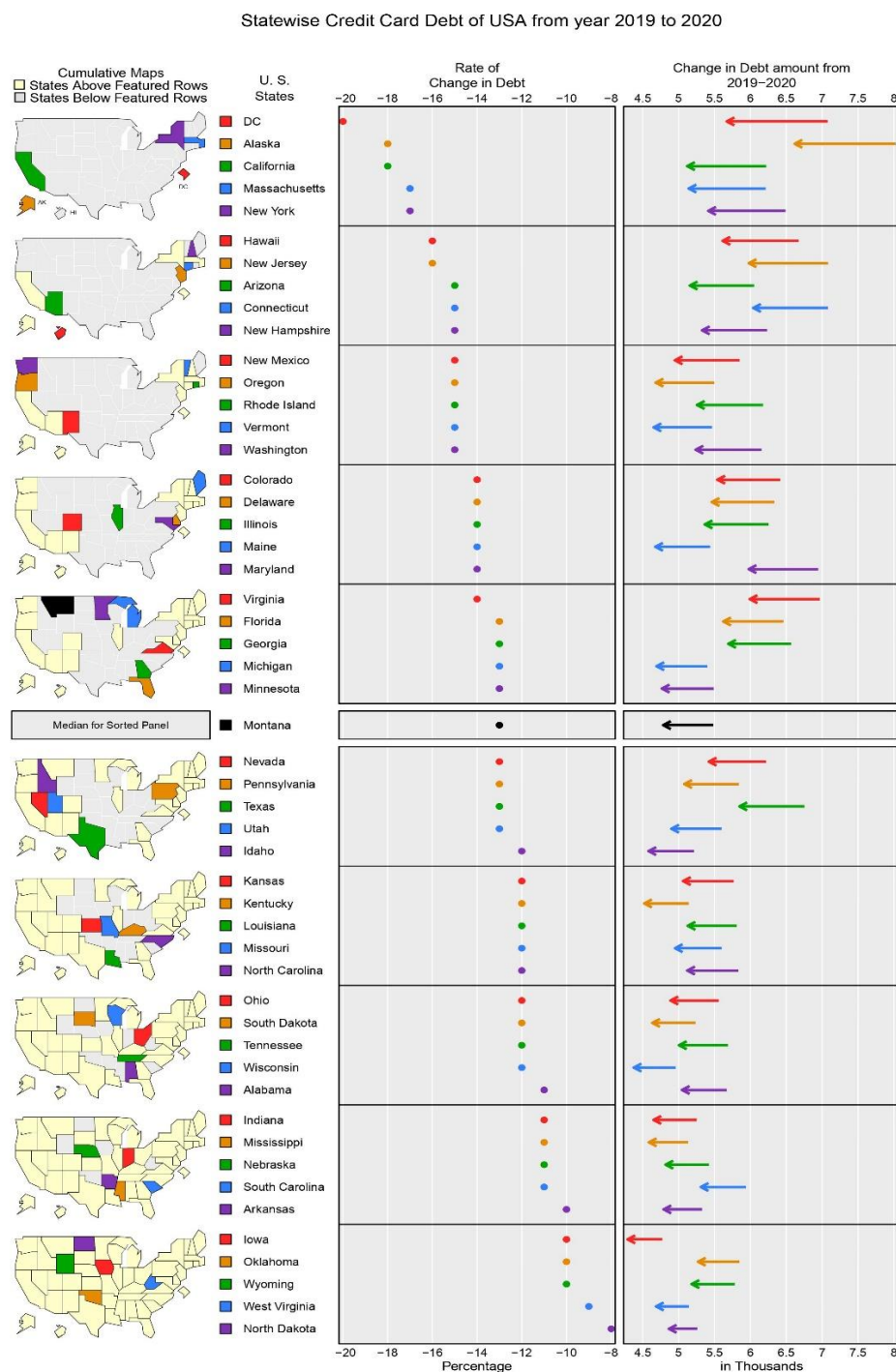


Figure 5: Redesigned Micromap

4.Designing an interactive bar plot using ggplot2 and ggiraph

By using ggplot() we created two interactive bar plots using geom_bar_interactive() function and displayed year 2019 and 2020 balances. By using girafe() we then interlink both the bar plots and display them side by side for clear understanding to the reader.

```
girafe(code = print(bplot1+bplot2), width_svg = 8, height_svg = 5) %>%  
  girafe_options(opts_hover(css = "fill:orange;"))
```

Figure 6: Using girafe() to interlink two bar plots

Now the resulted bar plot is

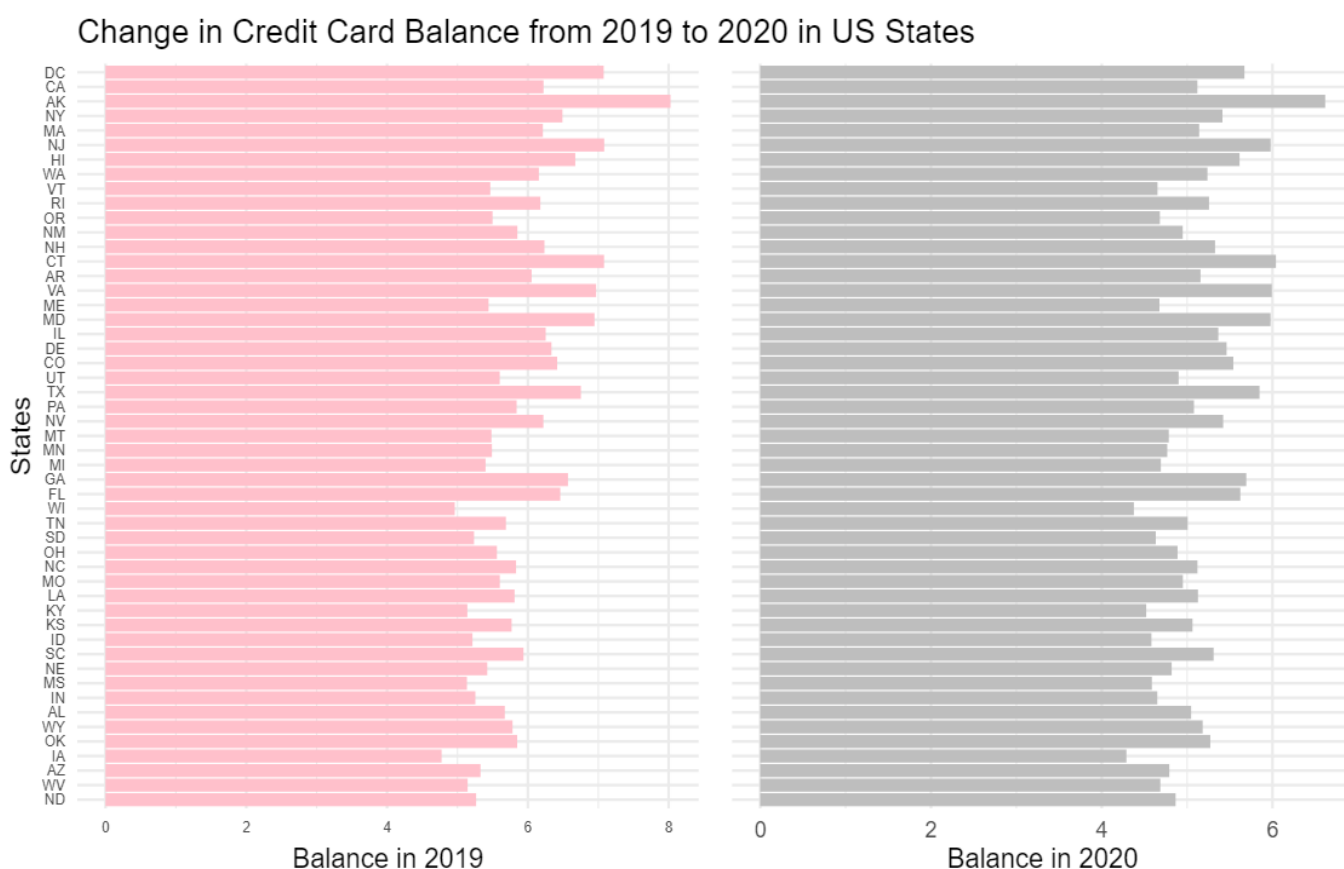


Figure 7: Interactive bar plot using ggplot(), geom_bar_interactive() and girafe()

We sorted the barplot in the same way as the micromap (Highest to lowest change in balance) so that the reader can understand both the plots similarly.

Redesign Analysis

Compared to the bad graph in Figure 1, the redesigned graph in Figure 5 and 7 are more clear, understandable to the user and is much more interesting and attractive. In the first scatterplot (Figure 3) we cannot clearly understand the data being visualised, though it is correctly plotted according to the scale it does not attract the reader to visualise the graph. However, after redesigning it into a micro map each state has a specific colour and a legend showing the names of each state and percentage of change and actually, how much value has changed over the year. In the interactive bar plot we further increase the visualisation

by clearly showing details when hovered on the bars of the plot, this visualisation engages the reader even more and creates a clear understanding of the data being visualized.

Challenges

1.The first and foremost challenge was finding a graph which was visualised poorly along with a dataset. We either found just the poor graph or just the data set with a lot of noise.

2.After finding the graph with data, we had to try a lot of ways to clearly plot the graph and make it visually clear and attractive along with information for the reader.

3.For the scatterplot we could not join the plots using `facet_grid()` so we had to plot two different graphs and then join them using `grid.arrange()` and creating the interlinked interactive barplot was not functional on `ggplotly()` so we had to create and link the two graphs using `girafe()` which was the final challenge in redesigning the graph.

Comments on the graph

Every state is seeing a decrease in average credit card balance. In a single year, District of Columbia decreased 20% of their credit card balance, followed by Alaska and California with 18% each. There are two possible causes for the reduction in credit card balances. First, COVID-19, which has forced companies to close and discouraged customers from leaving their homes which continued to have an impact on the economy. As more people stay at home and avoid traveling, they are reducing their expenditures, hence the decrease in balance. Second, because of unemployment benefits provided by government, people are now making more money than they were making previously which might lead to decrease in credit card balance.

Conclusion

The bad graph was redesigned into a good graph in 3 stages. First by creating a scatterplot to understand the balance values for each state and then plotting a micromap to understand the change in balance from year 2019 to 2020 and then finally plotting an interactive bar plot to show the values and percentage change in balance clearly when hovered over the bars on the graph. The four guidelines to create a good graph have been strictly followed in this redesigning process. The packages used to redesign the bad graph were `ggplot2`, `micromapST()`, `tidyverse`, `gridExtra()` and `ggiraph`.

Learnings

We learnt about the principles for plotting a good graph which makes it visually clear and more informative for the reader. We will make use of the packages learnt in this project for future works and projects.

References:

Irena (2021, Feb 08). Visualizing the Sharp Decline in Credit Card Debt Around the U.S:
<https://howmuch.net/articles/average-credit-card-balance-by-state-2020>