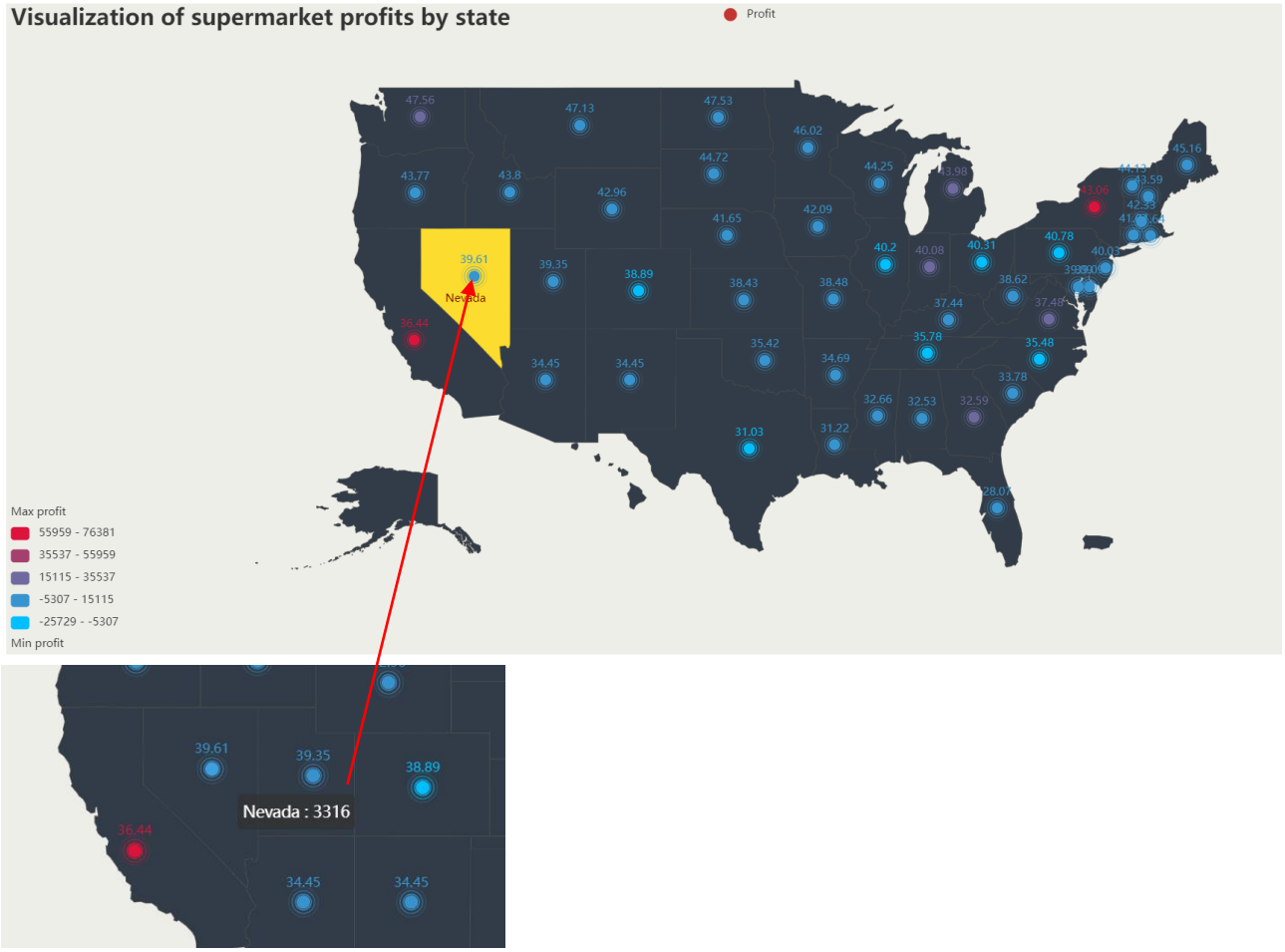
INFSCI 2415 INFORMATION VISUALIZATION Final Report

Yunpeng Guo

Github link to the project: github.com/Anton0830/INFSCI-2415



Visualization of supermarket profits by state

- **Data source:**

https://community.tableau.com/s/question/0D54T00000CWeX8SAL/sample-superstore-sales-excelxls

The excel file contains 9995 records of supermarket orders that were placed between 3/1/2014 and 30/12/2017 in different states of the US. Since we want to visualize the total profit of each state during this period, the `profit` was grouped by `state` and the accumulated profit of each state was calculated. Additionally, the coordinate of each state was provided to label the geographic location.

- **Explanation of visualization components**

As shown in the figure above, a United States map is presented. Data points with different colors at the central of each state represents the accumulated profit of supermarket sales during the period where these data were collected. The number above the data point represents the latitude of the state, which was automatically generated when the map was plotted. I tried reading the documentation of

pyecharts but did not find a method to remove it yet. This number could be a bit confusing and replace it with the profit value of each state could be much better.

Since this is an Effect Scatter plot made by pyecharts, it also provides some interactive features (link: https://infsci-2415-demo.glitch.me/). If hovered the mouse on a state, the background color will change and the name of the state will be showed. At the bottom left corner, a color bar is split piecewise to show the relationship between the colors and different profit ranges. The lowest profit is represented by blue, and as the profit becomes higher, it gradually turns into red. If hovered the mouse on a data point, it will show the name of the state and the corresponding profit value (shown as the smaller figure above).

The piecewise color bar also supports data filtering by clicking the color block. For example, if clicking the blue block, all the data points with profits range from -25729 to -5307 would be invisible, and click again they could be viewed again.

- **Findings and insights based on the visualization**

By filtering the data, it can be seen from the figure that California and New York had the top 2 sales profit during this period, which was significantly higher than other states since there was no state with an accumulated profit within the next lower range, which was 35537 to 55959. The profit of the most states was lower than 15115, and there were 7 states facing heavy losses that were greater than 5 thousand units.

- **Importance of this visualization**

This effect scatter plot provides a first glance at the overall profit and loss conditions of each state during the specific time period. By clicking the color block and filtering the data points, it provides an interactive way of analyzing data, such as finding the states with the largest profits and the states that were facing huge losses. With the initial grasp of the overall situation, further analysis could be taken by drilling down the sales data