Exercise 7B Report: Spanish Speakers in the United States

Motivation

The provided visualization named Distribution of Spanish Speakers in the US is characterized by a stationary map of the United States with state borders and coloring aimed at showing differences in Spanish Speakers as a percentage of state population across the country. While the use of a map is a good approach, the execution of the map by the original author betrays the principles of effectiveness and expressiveness. In particular, the main channel (percentage of Spanish Speakers) is encoded by partially filling states with colors. The amount of coloring should reflect the percentage of speakers, but too many colors are used thus making it hard to locate on the map. Additionally, many of the states with very lower percentages are all marked with a barely visible coloring line which makes it very hard to distinguish differences. In particular, a larger state will gain significance on the map and thus be more noticeable, but many smaller states like New York having much higher percentages are barely noticeable. A larger state with a smaller percentage of Spanish speakers will thus appear to have more speakers than a smaller state with a higher percentage. Finally, the choice of background color (black) does not make for a pleasant experience to the viewer. Thus the purpose of the new visualization is to better encode the same information.

Problem Statement / Task Description

The dataset provided consists of a simple spreadsheet with states and percentage of Spanish speakers. This has been augmented with geographic region (e.g. Southeast) and state code, both derived from the state attribute. The author leveraged the geographic features to create two maps and then arrange/group by in different ways to provide insights that could not be easily derived by the original visualization. The visualization is a story made up by four idioms. The story's main task is to highlight state and regional differences in terms of Spanish speakers as a percentage of state and region population.

Visualization

The visualization consists of four elements: an interactive map of the United Sates characterized by state borders; an interactive bar chart sorted by highest to lowest state in terms of percentage of Spanish speakers; a second interactive map defining regional differences; and finally a plot of regional differences across the country displaying key regional statistics.

The most important idiom is the map of the fifty states. Each state is clearly labeled by state code in order to keep the amount of text low while clearly identifying each. Additionally, the each state code is accompanied by the actual percentage of Spanish speakers. The color used by the map is a red-gold palette where gold represents lower presence and red signifies higher percentages. The legend accompanying the map clearly defines the color spectrum along with minimum and maximum percentage values. Each state can be highlighted to show the actual state name, or selected to hide other states and just leave the selection on the map.

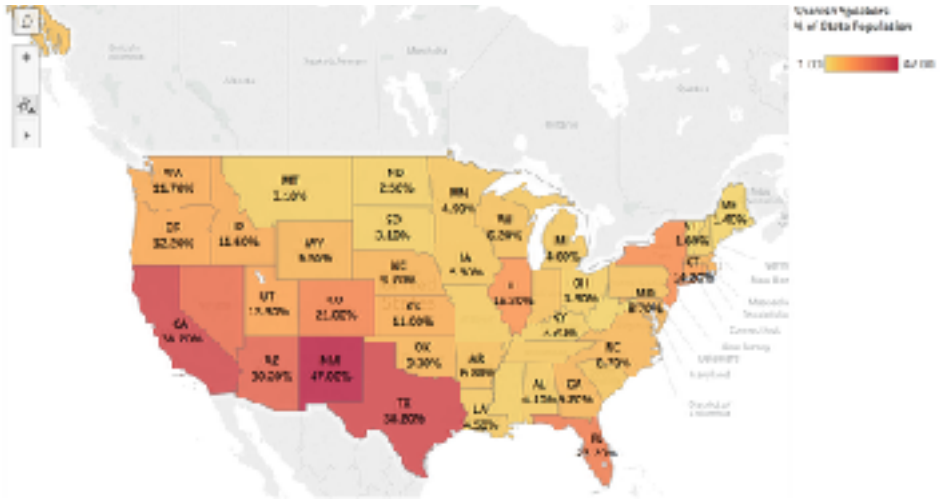

Figure 1: Map of the United States

The interactive bar chart presents the same information as the map above but sorts states by percentages in order to show maximum, minimum, and contextualize smaller states which might be harder to seen the map if the viewer does not zoom. This chart increases the effectiveness of the story as it allows for easy comparison of all states across a common objective numeric scale. The color palette and its color spectrum are exactly the same as in the map to allow for easy recognition of state percentages.

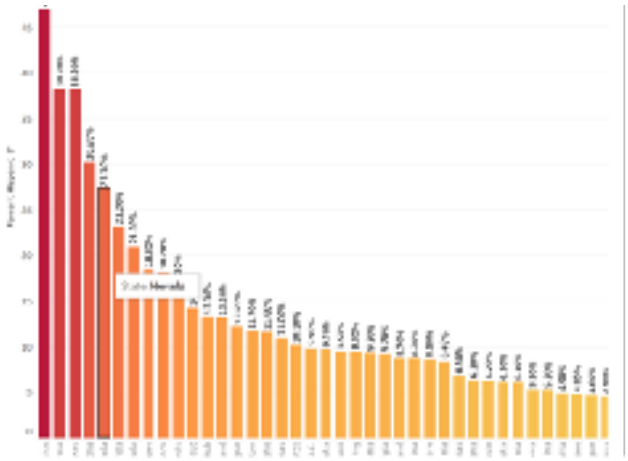

Figure 2: Interactive Bar-chart Sorted

The third idiom consists in a map similar to the first one in terms of color (the same palette is used) and visual impact (the same dimension and similar interactivity tools are used). However, the focus is on the five regions of the US: West, Midwest, Southwest, Southeast and Northeast. At any given point the viewer can interact with the map to select regions among these five from the tooltip. When a selection is made, only states belonging to that particular region are shown. All states are labeled with state code and share the same color which represents the regional average of Spanish Speakers calculated as the average of the percentages (this is not a weighted average by each state's population). The viewer can hover over each state to view the states full name, region, and four key statistics: state percentage of Hispanic Speakers (same as in the previous two idioms), regional average, regional maximum, and regional minimum.

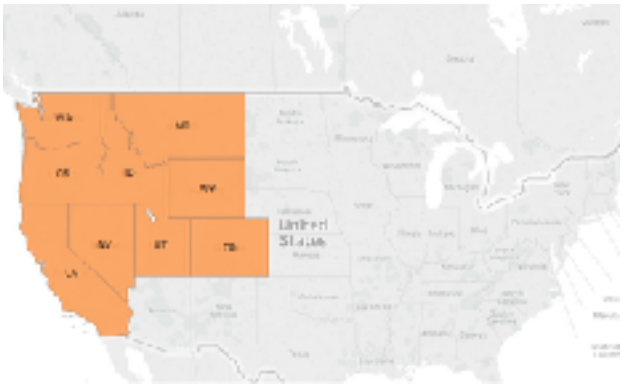

Figure 3: Regions Interactive Map

Finally, the fourth idiom consists of a plot demonstrating regional differences across regions (x-axis) and in terms of Average Spanish Speakers for the region (y-axis). Additionally, the same red-gold palette is used to color each circle on the plot; the palette uses the same color spectrum as the previous idioms. Upon hovering over each of the five circles (Southwest is not shown here because its value is higher in the plot) the viewer will be presented with the same derived statistics shown in the second map above.

Figure 4: Regions Plot

The story as a whole is expressive and effective. Each idiom encodes the main statistic (percentage of Spanish speakers) along an objective scale and allows for objective judgements. States are grouped both into regions and ordered in terms of scale. States and regions with higher percentages pop-out in red (channel salience), while states with lower percentages are also clearly visible despite their lower importance remaining intuitive to the human eye. Maps include interactive tooltip features allowing to zoom into areas of the United States and visualization information for smaller states. States are clearly labeled but the use of state codes to initially represent them on the map allows to decrease the amount of text on the map, thus making the states easily separable.

Conclusion

The visualization accomplishes the task of showing state differences in terms of Spanish Speakers as a percentage of state population across the United States.