Dana Westley

Soltoff - DataViz

Visualization Justification

The final versions of my visualizations can be found on my Final Project repo in the R Markdown document “Final Visualizations.”

Truthful

To ensure the accuracy and transparency of the story I was telling about my data, I included in the title of the visualizations: the data being used, the content that was being conveyed, and the timeframe of the sample. It is straightforward that these visualizations are not a sample of all tweets, and that the conclusions drawn from the data should only be considered for Twitter users posting about these specific issues. By identifying the timeframe as Donald Trump’s first month in office, it not only helps the viewer to draw better informed conclusions about the data, but honestly discloses that any trends found could be due to the unique and emotional nature of information and media during this time. I believe that this does not bias the sample, and instead can be used as an explanatory factor.

Functional

I will discuss the functionality of my visualization across the two formats I used, the bar charts to display the most popular terms overall and the most popular sentiment conveying words for each group of tweets, and the line graph to make between group comparisons over

time. I believe the bar charts are extremely functional. Since we are extremely good at making comparisons of line lengths, not only are viewers able to understand how frequent each term is itself, but also to discern how much *more or less* frequent one term may appear compared to others. Additionally, the bar charts displaying sentiment word frequency have additional channels of both color and directionality. The directionality in addition to differentiating colors helps users to compare line lengths and number of lines for each sentiment, positive or negative. This helps to quickly grasp whether a specific issue had more frequent usage of positive or negative words, or if it was about even. The line chart allows viewers to see the change in net sentiment over time while also allowing them to compare overall sentiment trends across groups.

Beautiful

The intended audience for these visualizations is mainly those in the fields of academia or research and students. The minimal theme is clean and simple, and forces the focus on the data instead of potential distractors. I used the Times family of fonts because it felt like a classic academic font and was easy to read. For the net sentiment graph, I used month abbreviations on the x-axis to reduce the clutter of showing the full date in YYYY-MM-DD format. I also added more frequent interval labels to more easily identify specific dates of interest. This way, viewers could use important dates to draw conclusions about the causes of raises and dips in sentiment of tweets for each group.

Insightful

The frequency chart data could be displayed in a table, but the comparison would require more effort. By comparing line lengths, it is much faster to compare the magnitude of difference between items. For instance, in the religious tweets group, we can very quickly see that the term “Islam” is the most popular term overall, and that it occurs twice as often as nearly every other word. This comparison would be possible in a frequency table, but would require a scanning of the data and comparing individual frequency scores and mentally calculating the difference. The visual helps to eliminate the effort required on the end of the reader. Even more so than the bar charts, the line graph of change in net sentiment over time for tweets about each social issue conveys information that would be nearly impossible to display in a table format, and would be convoluted to explain entirely in words. It shows an enormous amount of data points in an easy to understand visual, and comparisons across the groups can be easily made using different line types.

Enlightening

I think these visualizations have the power to inform those interested in social justice, activism and community organizing. Going into this project, I thought that negative sentiment would be pervasive across posts about all social issues, considering that Twitter is known as a platform for complaining. As these visualizations show, that is not the case. There is a much more nuanced story to be learned about the topics and sentiment people use when posting about these causes. For instance, the relationship between other social issues people post about can help to promote coalition building and to recruit new members for different activist groups.