Homework 03: Visualization

Visualization is a key aspect of exploring nuances of a given data but also communicating results with any readers. Many times, at a quick glance the visualization is the only aspect that someone sees and deduces the main point of the study. Therefore, it is important to design a good visualization. One way to do that is see how it holds up against five qualities of great visualizations from chapter 2 of The Truthful Art: Data, charts, and maps for communication by Alberto Cairo. In this write up, each of the five facets (truthful, functional, beautiful, insightful, enlightening) are evaluated for the following visualization. This visualization is found from a subreddit, "r/dataisbeautiful".

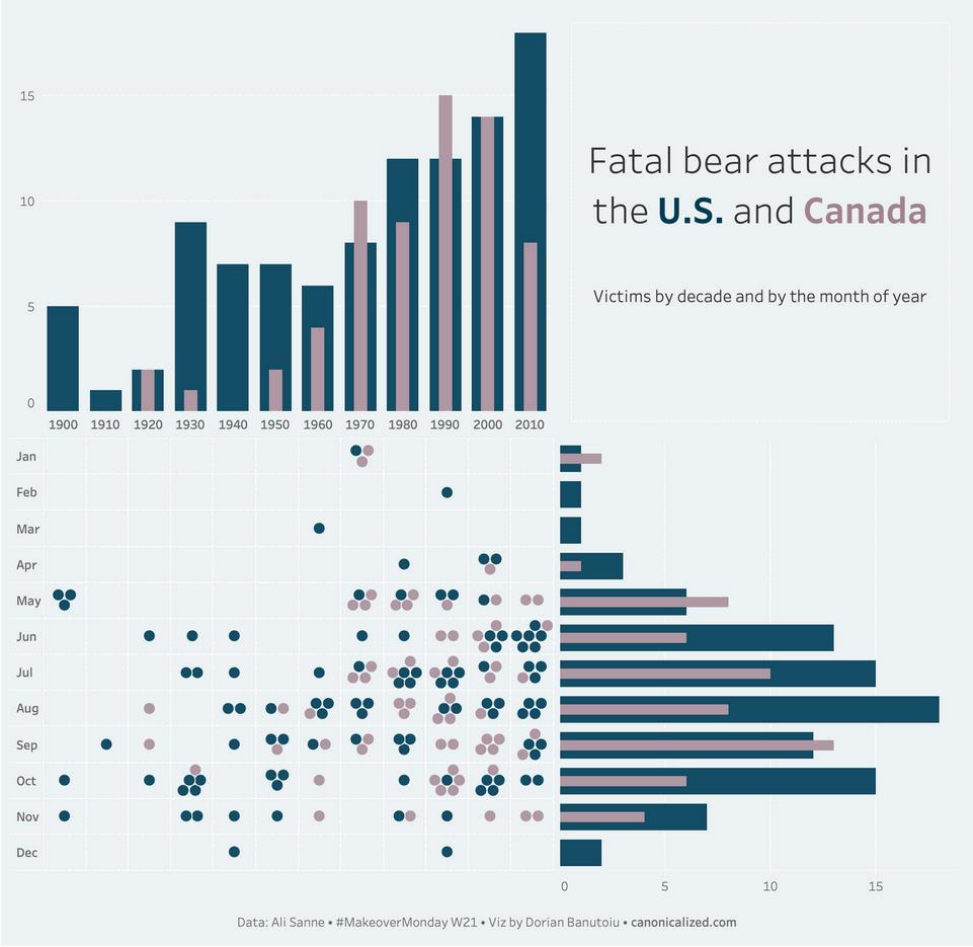

Figure 1 Visualization to Evaluate [1]

To be able to evaluate the how well was the visualization was made, the original dataset is reviewed. The csv data file, “north_american_bear_killings.csv” is found in the data folder. The data has the following information: Name of the victim, age, gender, date of the attack, month, year, type (wild/captive), location, description, type of bear, hunter, grizzly, hikers, only one killed.

Is it truthful?

Given the dataset, the visualization is truthful and presents the information in a way that is not misleading. From the graph it is clear how many people died in bear attacks in USA compared to Canada since 1900. The graph does a good job showing the breakdown of deaths per month – for each decade and cumulative. The scale used is appropriate for the graphs and it clearly represents the difference. Upon, looking at the dataset, a possible improvement would be somehow indicating if a cluster of deaths were related to one bear attack or different ones. For instance, the all three deaths in May 1901 was from a single bear attack on three children who were unfortunately attacked while picking flowers at the same time. Visualizing this distinction is important in being able to clearly gauge the number of incidents and not just the number of deaths. However, considering the title clearly states that the graph is showing the victims not incidents, it is not misleading and does not take away from the truthfulness of the graphs but the change could be an improvement.

Is it functional?

The graphs clearly meet their purpose. The title indicates that the information presented is the fatal bear attacks in US and Canada and that is very well illustrated using overlaying bar plots. The simple bar plots are a great tool for counts, and it is appropriate for interpreting the data. However, the scatterplot showing the number of deaths per month, in a given decade by country does not seem to add much value to the greater trends presented. While it is good information, it does not

seem necessary as the cumulative numbers does manage to convey the story. The scatterplot was just confusing at the first glance and required a closer read to understand the message conveyed.

Is it beautiful?

I believe that graphs are presented in an aesthetically pleasing format. The color choice is muted and goes well together. The distinction between incidents in USA versus those in Canada are clear and it was a good design to show the color distinction in the title. That is an elegant and space saving way to create a legend. The grid and axis utilize absence of lines versus dark black lines which gives the visual a clean, minimalist look. The use of complementary background color also makes it appealing. The overlaying graphs provide an easy means of comparison and clearly articulate how many bear fatalities occurred in USA versus Canada.

Is it insightful?

Yes, the visual provides the right balance of insight for a mass audience. This visual is not a part of an academic study so it is meant to provide spontaneous insight to a common reader as opposed to a graph building knowledge as typical in an academic publication or journal. On a quick glance, one can glean two major trends. First the number of bear attacks are growing over the decades and the highest bear attacks are during the summer months. It also seems like that for most decades, the number of Americans attacked is greater than number of Canadians attacked. One question that does arise is overall between America and Canada, what is the proportion of people to bear in some of the areas included. This would be an important information to be able to infer, why are more people in America attacked by bears compared to Canada. Is it because more people reside in places near bear habitat or more people visit such areas? To uncover such powerful insight, additional data could be merged to present a stronger story.

Is it enlightening?

When compared to the overall population, relatively few people are affected by this issue but understanding the trends in bear attacks can help save lives. Considering more people are visiting remote wilderness, it is crucial to gain an understanding of how big of a problem is bear attack. Both insights, bear attacks have increased over decades and that the highest attacks occur over summer, are helpful information to tourists planning outdoor vacations or rangers preparing to avoid such mishaps. While this visualization may not alter one's viewpoint on reality, it is enlightening about a problem and provides insights to understand it better and potentially devise information backed, preventative measures.

Overall, the creator of this visual has a done a good job clearly presenting facts about fatal bear attacks in North America. They have presented the information in non-misleading, aesthetically pleasing way. While there could be some improvements, such as use of additional population data source or removal of scatter plot, it is still a very well made visualization.

References

https://www.reddit.com/r/dataisbeautiful/comments/bqm1ot/fatal_bear_attacks_in_the_us_and_canada_each_dot/