Create a Tableau Project: Write-up

By: Quentin Ikeno

Summary

With the increasing tensions between North Korea and the United States, doing some sort of data analysis involving North Korea had piqued my interest. I was able to find the *The James Martin Center for Nonproliferation Studies North Korea Missile Test Database*, containing records for all missiles tested by North Korea capable of delivering a payload of at least 500 kg a distance of at least 1300 km. Using this dataset I wanted to see what I could find out about North Korea's missile programs. To aid in the investigation, Tableau was used to create visualizations of the data.

Story

The visualizations are collected into a Tableau story. Below are links to the first version of and final version of the story:

- [Version 1](#)
- [Final Version](#)

Design

For the first visualization in the story, I wanted to show the number of missiles tests over time, so I decided to use a bar graph, showing the amount of tests for each year. I then considered making a pie chart showing how the number of successful tests compared to the number of failures, however I ended up showing the successes and failures by turning the first graph into a stacked bar graph.

Next, I wanted to look at what type of missile was launched the most. So, I created a bar chart of missile types and sorted the bars in descending order to make comparison of the bars sizes easier.

From the beginning I knew I wanted to make some sort of map for the story. Thus, I decided to make a map showing every missile testing facility in North Korea. To compare how often each facility was used, the number of launches for each facility was encoded by bubble size.

I next wanted to compare which facilities were used over time. To do this I settled on using a map which would use animation to show the different testing facilities used for each year. Again, like the first map the number of launches at each facility was encoded using bubble size.

I was interested in seeing the distribution of distances travelled for all the missiles in the dataset; particularly which distance was the most frequent. Therefore, I decided to make a histogram to look at the distribution of distances.

After thinking about distances I wanted to compare distances and apogee (i.e. the highest altitude in the missile's trajectory) simultaneously, so I made a scatterplot for apogee and distance travelled. I put apogee on the vertical axis since it is a height and distance travelled on the horizontal axis since it was a horizontal measurement. Since different missiles travel different distances and have different apogees I encoded the missile type with color to compare missile types on the scatterplot.

For the final visualization, to compare the frequency of each missile landing location I decided to make a word cloud. The frequency of each word is shown by the size of the word in the cloud and to make it easier to discern each landing location, I used a different color for each location.

Feedback

After making the first iteration of the story and before receiving any outside feedback, I realized that there were parts of the data that needed to be cleaned that I had initially missed. First, there was quite a bit of missing data for apogee and distance travelled. To reconcile this problem, I decided to impute the data by taking the median apogee and distance travelled for each missile type and replacing the missing data with the median distance and apogee for each missile type. Second, after looking at the word cloud, I saw there was one location named "330km east of Hachinohe and 4000 km out into the Pacific Ocean." I decided this was a little too detailed and that changing the landing location to "Pacific Ocean" would suffice.

The next thing I noticed, was that there were some legends and axes labels that I had forgotten to change from the default labels. I then changed them to provide a more accurate description.

The first piece of feedback I received was that I should change the color of the bubbles in the map legend from gray to green, to reflect the green bubbles on the map. Unfortunately, I couldn't find any information on how to change the legend bubble colors, so I had to leave them as they were. A change that I did make to the map was to add a highlight action, so that if you want to look at a specific location you can hover over it with your mouse and all the other locations will fade.

The next two pieces of feedback were for the scatterplot of distance and apogee. It was suggested to change the circles I was using from open circles to filled in circles to make the points easier to see. Next, I was told that it would be good to see what years each missile was launched, so I tried to include text to indicate the launch year for each point.

Of the three people that I asked to look at my story, there was a unanimous agreement that the word cloud was ambiguous. It was difficult to compare the word sizes and the different sizes and colors made it more distracting than informative. One person suggested I make a pie chart instead, however I am not too keen on pie charts, so we compromised and settled on using a bar chart instead. I found this to be the best solution as it provides concrete numbers

and information, unlike a word cloud, where one has to guess what a word's size means. Overall, I feel most people would find the bar chart easier to interpret.

Conclusion

A lot of important changes have been made from the first version of my story to the final version. The most important change in my opinion was changing the word cloud to a bar chart. In retrospect, using the word cloud was to distracting whereas the bar chart was simpler and unambiguous. Overall I think that my story is much stronger now than how it started. All the feedback received have made the visualizations clearer and more informative.

References

http://www.nti.org/analysis/articles/cns-north-korea-missile-test-database/
https://en.wikipedia.org/wiki/International_Space_Station
http://www.bbc.com/news/world-asia-40489750