Data Visualization in Tableau

North Korea Missile Tests

Summary

The tension between the United States and North Korea have been rising in the past few months [1]. While reading news articles and searching for ideas for this project, I came across *'The James Martin Center for Nonproliferation Studies North Korea Missile Test Database'* [2], containing records for all missiles tested by North Korea that are capable of delivering a payload of at least 500 kg and a distance of at least 1300 km.

This project makes use of this dataset to display information of what I could uncover using this dataset and my reading a case study on it [3].

Links to Tableau Stories

1. **First Version**
   https://public.tableau.com/profile/pranav.suri#!/vizhome/P6_Ver1/Story1
2. **Final Version**
   https://public.tableau.com/profile/pranav.suri#!/vizhome/P6_Ver2_0/Story1

Changes Made to Dataset

The data [2] was downloaded in *.xlsx* format. I manually inspected it as the number of observations were not too high (119). The inspection has been outlined as follows:

1. Data is organized. Does not require much pre-processing.
2. For `'Landing Location'`, `'Facility Location'` attributes, some entires were very long. They were manually shortened to look nicer in the plotting.
3. `'Distance Travelled'`, `'Apogee'` data has lots of unknown values. Values noted from Google searches for the associated missile type did not closely match existing entries. Hence, they have been filled with the median-values for the associated missile type. Failed missile tests with unknown values were marked zero.
4. From `'Distance Travelled'`, `'Apogee'` values, remove 'km' to convert to numerical values for analysis.

Feedback

This section has been placed before the ‘Design’ section to create a better flow in this report i.e. by explaining the design ideas and decisions that were taken based on the feedback received.

Since, this is meant to be an explanatory data analysis i.e. to be understood by masses, asking for feedback from friends and people not from the subject seemed to be a good idea. I messaged the link of the first version of the story and noted the following feedback.

> *“Use better colors. It is a bit dull.*
> *If this was in the newspaper, I might skip it.”*

The plots in the first story were deliberately kept to default so as to get opinion of people to find a suitable color palette. Blue used earlier was too dull and hence changed/ complemented with warmer colors like orange and red.

> *“Can you show which countries do these*
> *missile-tests tests trouble?”*

The final plot section in the first story didn’t depict a plot that showed the landing position of the test missiles. The new plot added shows that most missiles end up in the Sea of Japan/East Sea.

> *“In the first graph, you show the number of*
> *launches for each type of missile. You should*
> *edit it to show the numbers for launches that*
> *passed or failed?”*

Adding another attribute/dimension to the graph helped it convey more information than before. It also complements the graph it is displayed along with.

Design

In the first visualization, the idea was to create a series of plot each depicting a particular piece of information primarily related to distance statistics and number of launches. As the the data was explored a little more, I decided to create the flow of the story as – distribution of data (based on a few attributes), missile testing under different reigns & what is the reach of the missiles.

From the beginning, I had intended 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.

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. Moreover, since different missiles travel different distances and have different apogees I encoded the missile type with color to compare missile types on the scatterplot.

To end the visualization story, the reach of the missiles and a distribution of their landing points has been plotted. This ends the story showing the 'effect' of the missile testing.

References

[1] – www.bbc.com/news/world-asia-40882877

[2] – http://www.nti.org/analysis/articles/understanding-north-koreas-missile-tests/

[3] – http://www.nti.org/analysis/articles/cns-north-korea-missile-test-database/

[4] – http://kb.tableau.com/articles/howto/plotting-geographic-data-using-custom-longitude-and-latitude

[5] – http://onlinehelp.tableau.com/current/pro/desktop/en-us/formatting_editaxes.html

[6] – https://onlinehelp.tableau.com/current/pro/desktop/en-us/publish_workbooks_tableaupublic.html

[7] – https://www.interworks.com/blog/ccapitula/2015/02/12/tableau-essentials-formatting-tips-color