**Reflective evaluation of my visualisation**

During the data analysis, I struggled to find a multidimensional dataset which was interesting to visualise, eventually I chose a dataset involving historic volcanic eruptions from the Global Volcanism Program at the National Museum of Natural History. I was aware that due to technological advances the dataset might be biased, I tried to limit this slightly by only choosing eruptions from the last 1000 years. Showing an increasing number of volcanic eruptions in the more recent years may not always reflect the full picture, despite this I believed it would be an interesting and unique data set, I also chose total population and land area from world bank datasets to give more context to my overall message.

Overall, I wanted to look at the volcanic eruption threat to the human population nearby, although this is just a potential threat as the volcanoes were not categorised by active volcanoes or dormant, this made me doubt the usefulness of the visualisation, but I believed there was still enough interesting and useful information to be created from the dataset. I learnt that the datasets also had huge discrepancies, when I was cleaning the data and created the percentages, many countries were showing over 100% of the population living near a volcano, I concluded to drop the outliers and plot only between 0-100% otherwise the visualisation would not be valid. I believe world bank data was collected in 2020 and the volcanic population data in 2022 but one data set must be flawed.

During the design process, I wanted to learn more python for future data analysis, I researched the best visualisation libraries, matplotlib, seaborn, plotly and holoviews. Ultimately I chose Hvplot from the Holoviews family as this provides an interactive hover mechanic to add context to the visualisation without complicated the graphic too much. I really enjoyed using this library as it was modern, intuitive, and simple. The trade off was that I initially wanted to use a bubble plot but this library does not have a way to plot that efficiently, along with that the documentation is still a work in progress so learning the customising the plot was quite difficult at times. I concluded that the loss of advanced customisations such as annotations was worth it for the amount od extra context and data the hover adds.

The first graph I plotted was a bubble plot in matplotlib to show the percentage of the population living within 10km of a volcano, the substantial number of countries along with the lack of time series data or time progression made me believe this was not the optimal choice. Initially I wanted to create more complex and intricate designs using hexbins, scatter graphs etc, but the most clear and concise way to visualise this was a series of horizontal bar charts looking at percentage of the population living 10km/30km/100km away from a volcano. This helped add context and massively change perspective based on the land area of the countries. The size of each bar and order them from top to bottom creates an easily readable and memorable graphic.

I believe my graphic could benefit from adding annotations at the end of the bars to give the percentage in a more visual and obvious method, I couldn’t find a way to do this in hvplot, also the hover numbers could benefit from being formatted with commas to make it more readable. I would have also liked more customisation over padding around the graph, title and interact buttons. Adding these to a dashboard in future could help visualise the picture better.

I believe my graphic supports Tufte’s principles and represents a good graphic, it displays a large number of ideas, in the shortest time using the least amount of ink (Interaction-design, 2020). My visualisation communicated complex ideas with clarity, precision and efficiency while telling the truth about the data. It also supports graphical integrity by showing the true data with no variation in design and no distortion. The lie factor is 1:1, the bars are represented by percentage which are proportionate with the size in the graphic. Bertin also highlighted that relative size is the most widely useful and easily perceived visual variable, no matter what information is visualized (Sluis, K. 2016).

My visualisation also supports a high data to ink ratio with additional hover context not shown in the graphic. One critique could be, is the land area relevant to the volcanic eruption data and it is not present in the graphic visually, in this instance I would have preferred to use a bubble plot to show the size based on land so I added land area data as context information so the user can assess the number of eruptions compared to land mass. Tufte offers the idea that borders, backgrounds, use of 3D, etc. may do nothing but serve to distract the user from the information itself (Interaction-design, 2020). I took this idea and I believe in this instance it was correct to not distract the data.

I believe my colour map contributes to the visualisation positively, creating an aesthetically pleasing and memorable visualisation which also add information, perspective, and context. The colours are chosen to be more colourblind friendly using pastel shaded colours instead of plain solid defined colours such as red, blue, green. The colours also have some familiarities behind them such as a shade of red for Asia relating to the common prominent colour on the flags. Bertin also describes that a good visualisation should answer an elementary, intermediate and an overall question. I believe there are many angles and questions that you could cipher from my visualisation such as, how many volcanic eruptions happened in Japan over the last 1000 years, Which continent has the most volcanic eruptions in the last 1000 years or Does higher land area always correlate with more eruptions.

Finally, Junk Charts describes a great visualisation should encompass an interesting question with relevant data in combination with a visual which addresses the question in a clear, concise manner (Junk Charts, 2014). I believe the weakest part of my visualisation is the combination of two different sources of data which did not align for the population statistic, some countries clearly had the wrong number in one of the datasets. Alongside this, it was hard to create a clear overall message which could be useful to the user, I could have chosen a more basic message, just looking at volcanic eruptions over the last 10,000 years but I believed this added no useful information. Looking at threat was a more interesting message in my opinion, but it does not correlate with proximity to volcanoes as they can be dormant. Combining the two graphics I believe the information is interesting, but I couldn’t definitively answer threat levels as there are too many other factors at play that I didn’t have data for.

Another problem was adding continent data, I added this manually but there is a lot of conflict over what continent countries belong too, this is qualitative categorical data so to simplify it, I chose where I though the country was most associated with. In the case of Russia, it is part of Asia and Europe, the land mass covers a huge area, but as I believe it’s often associated with Europe and a large amount of the population live In Moscow which is considered in Europe. I could have included some extra detailed continents such as Eurasia, but I chose to keep it simplified for the user.

In conclusion, I believe the visualisation includes interactive and useful data which is clear and concise for my intended audience which is the average person looking to enjoy an interactive aesthetically pleasing visualisation about volcanos. Despite flawed data and the improvements that can be made, I’m very happy with the visualisation and data analysis of these datasets.

**References**

Interaction Design. 2020. Guidelines for Good Visual Information Representations. Available at: https://www.interaction-design.org/literature/article/guidelines-for-good-visual-information-representations [Accessed: 4 April 2022].

Junk Charts. 2014. Junk Charts Trifecta Checkup: The Definitive Guide. Available at: https://junkcharts.typepad.com/junk\_charts/junk-charts-trifecta-checkup-the-definitive-guide.html [Accessed: 4 April 2022]

Sluis, K. 2016. Before Tufte, there was Bertin. Available at: https://karlsluis.medium.com/before-tufte-there-was-bertin-63af71ceaa62 [Accessed: 4 April 2022].