Visualizing Countries Opinions of Issues Discussed at UNGA

Eyad Al Khayat

eak1u19@soton.ac.uk

I. DATA STORY SUMMARY

For the last 8 decades, delegations from each country in the world met on an annual basis at the United Nations General Assembly (UNGA) where nations discussed all the topics concerning our globe. The discussion takes the form of voting on certain resolutions where each country's delegate can vote with yes, no, or abstain from voting at all. UNGA voting dataset documents the challenges and problems that faced humanity. Moreover, it shows the position of countries towards an issue or a topic. The data story takes the reader on a personalized-interactive exploratory journey to answer the following questions:

- What were the major discussed issues in the world over the last 80 years?

- How did the reader's country vote for these issues?

- How similar were the votes of other countries to the reader's country, does trade volume affect the votes of the two countries? Is there a correlation between the two?

The story provides readers with a powerful and simple tool to proof-check their countries stance about a certain topic or issue. It is important to have such a tool that everyone can easily use to explore how and what, their countries voted without the probable, biased opinion that news agency may share. The data story will focus on 6 types of resolutions:

1. Israeli-Palestinian Conflict

2. Nuclear Weapons

3. Human Rights

4. Arms Control

5. Economic Development

6. Colonialism

The first visualization is using Treemapping method to showcase these six topics, where the size of the rectangle is proportional to the number of votes related to each topic. Then, the story prompts the user to select any country (presumably the reader's country) and show a horizontal bar chart illustrating the number of votes for each resolution type categorized by "Yes", "No", and "Abstain" votes. To answer the last question, another dataset was needed to show the total trade flow between the two countries. The Correlates of War Project Trade Data Set tracks the total bilateral trade volume between states from 1870-2014. The final chart asks the reader to select two countries and a specific period, then the story will combine the results of the two datasets and show a hybrid line and a bubble chart where the line indicates the similarity percentage between the votes of the two selected countries, and the size of the bubble shows the total amount of trade done between the two nations in that same period.

II. DATASET SUMMARY

A. UNGA Voting Data

The first dataset used was obtained from the Harvard Dataverse [1]. It contains the details of the sessions conducted at the UNGA, the votes of each country, and a summary about the resolution being voted for. The dataset covers sessions from 1946-2018. Harvard Dataverse assembled the data from multiple official UN sources and other scholars' datasets. The original dataset size is around 400MB CSV file that had 1.2 million records with 34 attributes. The following pre-processing was done on this dataset:

- Remove the records that are not related to the main 6 resolution types mentioned previously.
- Remove several attributes (columns) that adds no value for the data story.
- Reshape resolution types into one column with categorical values. They were spread across several columns and were coded with numerical values.
- Reshape vote types (yes, no, abstain) into one column with categorical values. They were also split into several columns with numerical codes.
- Add a new column related to the continent of the country.

The pre-processing of data was done using Python and Excel. This pre-processing resulted in reducing the file size of the dataset for around 50%. Moreover, it enabled the analysis of the attributes faster and easier.

B. Dyadic Trade V4 Dataset

This dataset includes the flow of imports and exports and the total trade volume in US dollars between nations from 1870 until 2014. It was collected and compiled from various reliable sources as part of the Correlates of War Project [2] [3]. The dataset contains 800k records with 23 different attributes. To reduce the dataset size and remove unnecessary attributes the following were done:

- Remove attributes that will not be used in the data story
- Convert values of total trade from negative to 0. The dataset highlighted missing data with a negative 9. This was affecting the results of the analysis and had to be changed to 0 instead.
- Remove records before 1946, as there was no UNGA session before that date.

C. Linking the Datasets

There was a similar attribute in both datasets that enabled the linkage of data. In the UNGA dataset, it was "ccode" which was using the standard country codes, and the Dyadic Trade dataset also used the same standard country codes in its attributes. This made it easy to query the two datasets in a standardized way.

III. VISUALISATIONS

The data story followed the instructional reader-driven methodology rather the traditional author-driven way of narrating a story. The story draws a clear path and structure for the user in how they should navigate the story and helps them in concluding the questions presented previously. The main colours used in the visualizations are derived from the 6 resolution types.

The first colour for the Israeli-Palestinian Conflict is a combination between the main colours of the flags of both countries, blue, and red, resulting in a purple colour. The second for the nuclear weapons topic was yellow, as it is always related to nuclear issues and biological weapons. The third one, representing human rights, was taken from the United Nations flag. The fourth is linked with Arms Control, is a light brown colour, as most people remember the famous AK47 brown colour when thinking about weapons! Light green for Economic Development, as this colour conveys growth and prosperity. For Colonialism, when searching for this term on Google Images, a shade of orange was always in the results, thus it was chosen. The final colours shades were generated using an online tool[1] to make sure that they fit together and are visually appealing.

A. Treemap

1) Description

The first tab of the visualization illustrates the top controversial resolution types during the whole period. Showing the percentage of the number of votes for each resolution. On the right of each story tab, as shown in Fig 1, a short analysis is given to the reader highlighting the main trends in the story.

1 https://coolors.co/
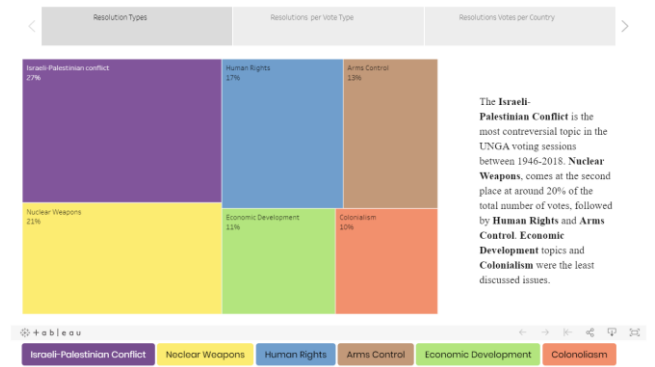
Figure 1 Treemap, story tab 1. Top controversial Issues

The second story tab of the visual shows the percentage of vote types for each resolution taking the reader one step deeper into the story while explaining key trends on the right.
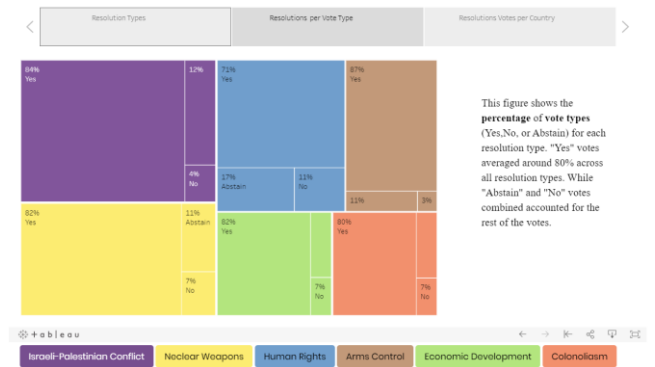
Figure 2 Treemap, story tab 2. Resolutions per vote type.

The last tab of the story, shown in Fig 3, is showcasing the total number of countries' votes for each resolution structured as yes, no, and abstain. It enables the reader to have an overview of how countries voted for each resolution.
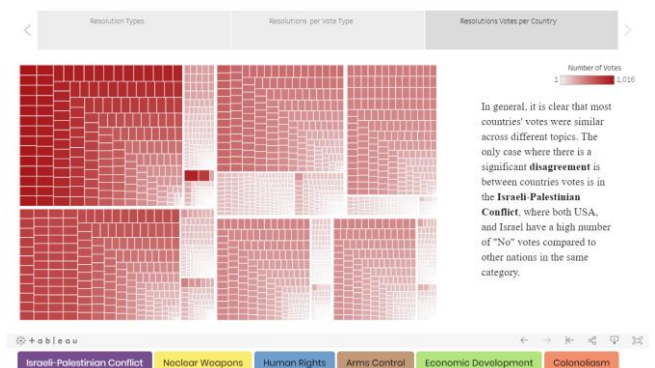
Figure 3 Treemap, story tab 3. Resolutions votes per country

On the bottom of the visual, the six resolution types are presented acting as a legend for the treemap.

2) Justification

Treemap visualisation was used as it is the best method of visualising structural data that may have several main branches and sub-branches. This is exactly the case for

this question, as the first story showed the 6 types of resolutions, then it branched into resolutions per vote types, and finally resolutions per vote type per country. Drilling down into the data easily while being able to distinguish the categories quickly.

3) Narrative Design Patterns
Two main design patterns were followed in this visualisation. The first one is the "Compare" [4] principle that uses treemap, and the second one is the "Repetition" [] pattern where the same visual was shown when progressing with the story and at the same time going one step deeper until reaching the last one.

4) Strengths and Weaknesses
The visualisation compares the 6 types of resolutions and how countries voted for each one. The reader can smoothly switch between the story tabs and read the narrative supported with each story to have a summary of what the story is. The colours and the legends in the bottom help the reader to identify the types of resolutions comfortably. Hovering the treemap will show additional information related to that tile. The main disadvantage is that the reader needs to gauge the different sizes of rectangles that are presented. It also takes more time for non-numerical readers to understand this type of visualisation from first glance.

5) Improvements
Increasing the spaces between the rectangles (resolution types) would be a positive addition as the reader will be able to identify categorize easily.

B. Word cloud

1) Description
The second function of the tab showing the resolution types is that readers can click on the buttons to be presented with a short description of the resolution. This will enable the reader to know more about the resolution type. In addition to the short description, a word cloud summarizing the important keywords that were written in the text of the resolution. This word cloud is covering all resolutions for the selected topic from 1946 until 2018.

Figure 4 Word cloud summarizing the Israeli-Palestinian Resolutions from 1946-2018

2) Justification
Word clouds are a one of the simplest methods to visualize text data. They show the frequency of words in a text. Reading the resolutions text one by one for all years is difficult and time consuming. This way the reader can spot a keyword from figure and search further.

3) Narrative Design Patterns
This visualisation pushes the reader to find and explore the keywords in the word cloud, making it follow the model of "Users-find-themselves" [4]. When exploring the word cloud the reader will know more about the types of resolutions discussed under this category, which is contributing to understanding the whole data story in general.

4) Strengths and Weaknesses
A simple way to grab the attention of the reader to common words mentioned in the resolutions. One drawback of the word cloud presented is that the colours have no meaning.

5) Improvements
Even though a lot of time was spent in identifying the jargon words from the text there are still words that could be cleaned to have a better cloud.

C. Horizontal Bar Chart

1) Description
The bar chart shows the 6 types of resolutions ordered from Israeli-Palestinian Conflict, at the top, until the Colonialism, at the bottom. The horizontal bars indicate the total number of votes. The legends on the top of the chart can be clicked to filter the graph.

Figure 5 Horizontal bar chart, Votes of Country by Resolution Type

2) Justification
Bar charts are the best for comparing multiple categorical variables at the same time. The reason behind using

horizontal bar chart instead of the regular, vertical, chart, is that the items that are compared are not ordinal in nature. As this book suggests [5], vertical bar charts are best used for time-series comparisons, and horizontal charts are better used for categorical values with no importance of order.

3) Narrative Design Patterns
"Exploration" [4] principle was applied here by giving the readers the ability to freely explore the votes of different countries by selecting a country from the dropdown list. Readers will be able to make their conclusions and see how different countries voted.

4) Strengths and Weaknesses
Engaging and interactive chart with visuals to link resolution types in the mind of the reader with icons. The negative thing is that the types of resolution names were abbreviated as the full names can not fit with the icon. However, the table on the right with the explanation of each icon is helpful.

5) Improvements
Finding a way to fit the full name of resolution type in the Y-axis without affecting the icon. Increasing the space between each resolution type.

D. Hybrid Line and Bubble Chart

1) Description
When the selects the two countries and press the "Visualize!". A function will calculate the percentage of similar votes on resolutions between the two selected countries for each year. Another function will query the Dyadic Trade dataset and aggregate the total trade volume for the two countries for each year. The generated graph will first show a line representing votes similarity per year. Then on each node of the line chart, a bubble is placed showing the total trade flow between the two countries in that year. On the right of the charts are additional insights. The first one is the average voting similarity over the period. The second one the average trade values and the last most important one is Pearson's correlation coefficient that checks the linear relationship between two variables.
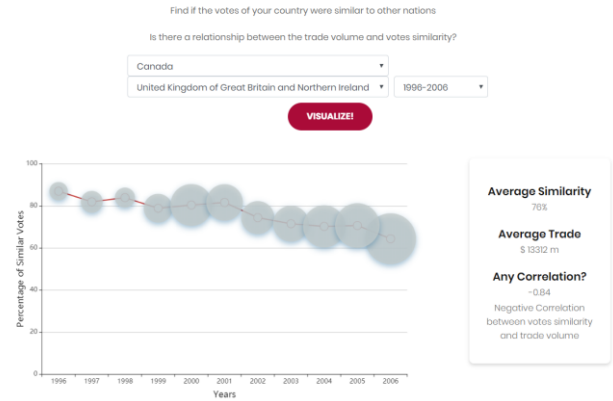

Figure 6 Hybrid line and bubble chart showing voting similarity and trade flow in million USD between two countries.

2) Justification
Bubble charts are the best way to visualize 3 variables at the same time. It was used here to show how the similarity percentage is changing over the years with the trade flow of the countries. To make it easier to read the trend of similarity, a line chart was incorporated.

3) Narrative Design Patterns
"Users-find-themselves" [4] is used by supporting the user to deduce the conclusion of the chart by reading the correlation value and the insights on the right of the chart. "Explore" pattern is also used where the reader can select any two countries of interest to them.

4) Strengths and Weaknesses
The clear trend that the line chart makes with similarity percentage, combined with the trade flow bubble and the insights on the right makes it easy for the reader to deduce results related to the original question. Bubble size for similar values may seem confusing, thus when the reader hovers over the bubble it shows the total trade value correctly.

5) Improvements
Using a better bubble sizing function and generate a legend that shows the bubble size value.

IV. Conclusion

Several people and organizations have analysed the UNGA voting data before. In the work done by [6] and [7], the visualizations that answer some of the questions presented here, are not interactive and can be hard for a reader to find the information related to his country. This data provided the interactive element that was missing from the previous projects and explored how trade flows of countries might affect countries votes similarities. All the functions related to filtering data, and dealing with interactive requests are hosted on Azure Functions and written in Python.

REFERENCES

[1] Erik Voeten "Data and Analyses of Voting in the UN General Assembly" Routledge Handbook of International Organization, edited by Bob Reinalda (published May 27, 2013). Available at SSRN: http://ssrn.com/abstract=2111149

[2] Barbieri, Katherine and Omar M. G. Omar Keshk. 2016. Correlates of War Project Trade Data Set Codebook, Version 4.0. Online: http://correlatesofwar.org.

[3] Barbieri, Katherine, Omar M. G. Keshk, and Brian Pollins. 2009. "TRADING DATA: Evaluating our Assumptions and Coding Rules." Conflict Management and Peace Science. 26(5): 471-491.

[4] B. Bach, M. Stefaner, J. Boy, S. Drucker, L. Bartram, J. Wood, P. Ciuccarelli, Y. Engehardt, U. Köppen, and B Tversky. "Narrative design patterns for data-driven storytelling." In Data-Driven Storytelling, N. H. Riche, C. Hurter, N. Diakopoulos, and S. Carpendale, Eds. CRC Press, USA, 2018, ch. 5, pp. 107–134

[5] P. Milne, "Presentation Graphics for Engineering, Science and Business," 2015, p. 31.

[6] U. S. a. O. Z. Mohammed Haddad, "How has my country voted at the UN?," aljazeera, [Online]. Available: https://interactive.aljazeera.com/aje/2019/how-has-my-country-voted-at-unga/. [Accessed 10 1 2020].

[7] Y. Fu, "Data Analysis: Who Votes with China, and Who Votes with the US and Europe at the UN?," [Online]. Available: https://yiqinfu.github.io/posts/united-nations-general-assembly/. [Accessed 10 1 2020].