

Visual Analytics Workflow - Data Visualization

An Insight Report

CS732: Data Visualization

Tanishq Jaswani - Sri Sai Suhruth - Seethana Sreevardhan

International Institute of Information Technology, Bangalore

IMT2018077 - IMT2018012 - IMT2018068

Tanishq.Jaswani@iiitb.ac.in - SriSai.Suhruth@iiitb.ac.in - Seethana.Sreevardhan@iiitb.ac.in

Abstract—Characterizing the problem domain and understanding users’ practices and processes are recognized as important steps in order to design and validate visualization, but are often disregarded in practice, also because of their complexity. We introduce the nested workflow model for design and validation of visual analytics, aimed at providing designers with a powerful and expressive modelling tool. This model enables the description of visual analytics processes, at different design levels, in terms of tasks, data, and users, including complex workflow patterns, data and knowledge flows, and collaboration between users. We discuss its application in the following group project, demonstrating its usefulness for their design and validation.

Index Terms—Alliance, Mapping, Animation, Integration, Data distribution, Pattern discovery

I. INTRODUCTION

Collaboration in real-world analysis can be regarded as a process of working together or sharing decision-making to develop a joint strategy or answer for the given complex tasks. Since collaboration is beneficial for solving complicated tasks, domain experts often work together to solve analytical problems in a collaborative setting. However, it has been known that real-world analysts typically perform both individual and group tasks, and as a result must frequently transition between single-user and multi-user collaborative workflows during the course of their analysis.

This paper is organized as follows. First, we discuss the introduction to what exactly is visual analytics and its workflow. Further we discuss how we can integrate our individual tasks as a group. In Section 2, we represent our assigned problem statement. In Section 3, we provide an overview of our visual analytics data set and outline the methodology will be applied further. Next, in Section 4, we then describe our specific tasks performed and the visualizations presented. Finally, in Section 8, we discuss future work and conclude the paper.

II. PROBLEM STATEMENT

This assignment is about implementing a visual analytics workflow. Hence, the idea here is that team members can own different visualizations of different aspects of the datasets, and all of them must be stitched together in the workflow.

Following are the mentioned requirements to complete the problem statement.

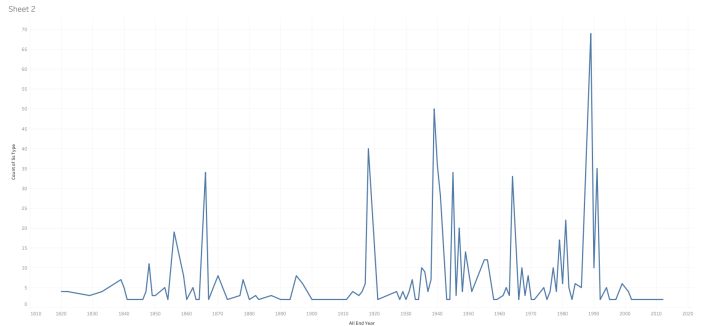


Fig. 1. Count of Ss Type vs All End Year.

1) : Since it is a visual analytics workflow, at least one instance of feedback loop is desirable, which means that the knowledge discovery from the dataset shall be used as a new dataset and a visualization/data mining cycle can be performed.

2) : The assignment implementation must start with a list of tasks you would like to perform on the dataset. Analytical tasks include tasks processing/analyzing dataset or its subsets, along with descriptive verbs such as "determine", "compare", "identify", etc., which a user should be able to do using your proposed solution. Your solution can then solve n tasks for an n-member team.

3) : The visualizations can be generated on independent tools or an integrated tool. The integrated tool can be a third-party one, e.g. Tableau, or one built by the team.

III. METHOD

A. Dataset

This data set records all formal alliances among states between 1816 and 2012, including mutual defense pacts, non-aggression treaties, and ententes. This data set is hosted by Douglas Gibler, University of Alabama.

The Correlates of War Formal Alliance data set seeks to identify each formal alliance between at least two states that fall into the classes of defense pact, neutrality or non-aggression treaty, or entente agreement. A defense pact (Type I) is the highest level of military commitment, requiring alliance members to come to each other's aid militarily if

attacked by a third party. As the labels imply, neutrality and non-aggression pacts (Type II) pledge signatories to either remain neutral in case of conflict or to not use or otherwise support the use of force against the other alliance members. Finally, ententes (Type III) provide for the least commitment and obligate members to consult in times of crisis or armed attack. Each alliance classifies the highest level of military support that an alliance member pledges to another alliance member. The data set provides the type, members, and appropriate dates of activity of each identified alliance.

Version 4.1 of the alliance data expands the original Singer and Small (1966) definitions of alliance types to include neutrality and non-aggression pacts separately. Neutrality pacts are now labeled as Type 2a alliances, and Non-Aggression pacts are labeled as Type 2b alliances. Alliance terms also distinguish between neutrality and non-aggression in the dyadic data.

In addition, the fourth version now differentiates between the overall alliance type and the terms of the alliance by specific member. In the alliance member list of data, for example, both the Singer and Small (1966) alliance type and the specific terms by alliance member are included. Only the terms operational for the specific dyad are included in the dyadic and directed-dyad data.

Following is the format of the data:

1. Alliance Member Format: The file named ; alliance v4.1 by member.* ; contains the alliance dataset distributed by member, with one observation for each alliance member initiation. Note that, if a member exits an alliance and then reenters the same alliance later, then that member will be included in the alliance twice, with the appropriate dates of membership included for each observation. The file named ; alliance v4.1 by member yearly.* ; includes an observation for each alliance member year.

2. Dyadic Terms Format: The file named ; alliance v4.1 by dyad.* ; contains the alliance dataset distributed by dyad, with one observation for each dyadic alliance initiation. Note that, if a member exits an alliance and then reenters the same alliance later, then this will in ate the number of dyadic alliance initiations for each particular alliance. Alliance terms in the dyad may not be applicable to each country code. Users should consult the asymmetric variable to determine whether the alliance includes terms that only apply to certain members. Specific terms are listed in the directed-dyad format. The file named ; alliance v4.1 by dyad yearly.* ; includes an observation for each allied dyad-year.

3. Directed-Dyad Format: The file named ; alliance v4.1 by directed.* ; contains the alliance dataset distributed by directed-dyad, with one observation for each directed-dyad alliance initiation. In this format the alliance terms include the specific promises made by code1 toward code2. The file named ; alliance v4.1 by directed yearly.* ; includes an observation for each allied directed-dyad-year.

A more detailed description of the data may be found in the release notes in the ZIP file below. The ZIP files contain the following files:

1. alliance_v4.1_by_member.*
2. alliance_v4.1_by_member_yearly.*
3. alliance_v4.1_by_dyad.*
4. alliance_v4.1_by_dyad_yearly.*
5. alliance_v4.1_by_dyad_directed_dyad.*
6. alliance_v4.1_by_dyad_directed_dyad_yearly.*
7. Alliance_v4.1_ReleaseNotes.pdf

There are two ZIP files, one in .dta format, another in .csv.

IV. VISUALIZATIONS

A. Count of Ss Type vs All End Year

The Graph(Figure 1) shows the number of alliances for every year an alliance has ended. The purpose of creating this graph is to provide a better informative visualization.

B. Defense vs Mem St Year, Mem End Year

The Graph (Figure 2) shows the defence pacts of states during the start and end of an alliance. The x-axis shows the years and the y-axis shows the sum of defense levels at that year. The graph is split into two parts. The left part shows the defence levels at the start of the alliance and the right part shows the defense levels at the end of the alliance.

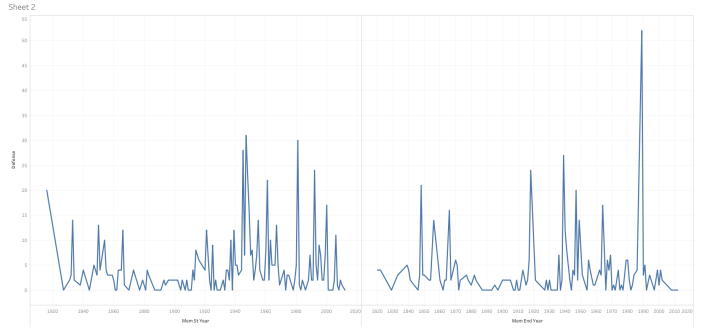


Fig. 2. Defense vs Mem St Year, Mem End Year

C. Count of Ss Type vs State Name

The Bar graph (Figure 3) shows total number of each alliance type each country has. Each country has a bar which shows the total number of alliances it has and each bar has different colors depicting different types of alliances a country has and the total number of each alliance.

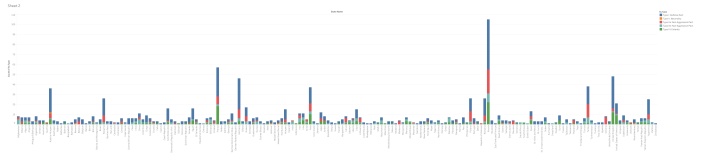


Fig. 3. Count of Ss Type vs State Name

D. Ss Type, All End Year vs All St Year

The graph (Figure 4) shows the start and of all types of alliances. The x-axis shows the years and the y-axis shows the end year and is split into 5 different alliance types.

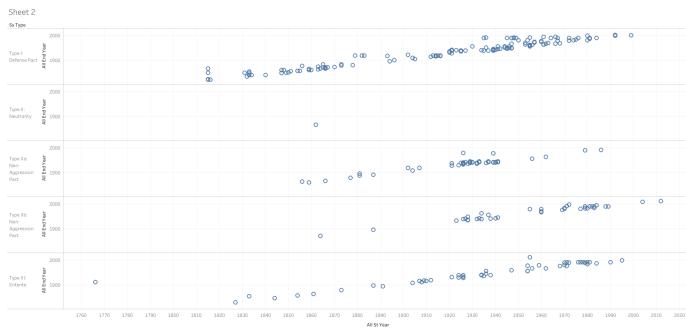


Fig. 4. Ss Type, All End Year vs All St Year

E. Defense, Neutrality vs State Name

This graph(Figure 5) shows the countries with defence and neutrality pact. The x-axis shows the pairs of countries and they y-axis shows the number of pacts. The y-axis is split into two parts one for each type of pact.

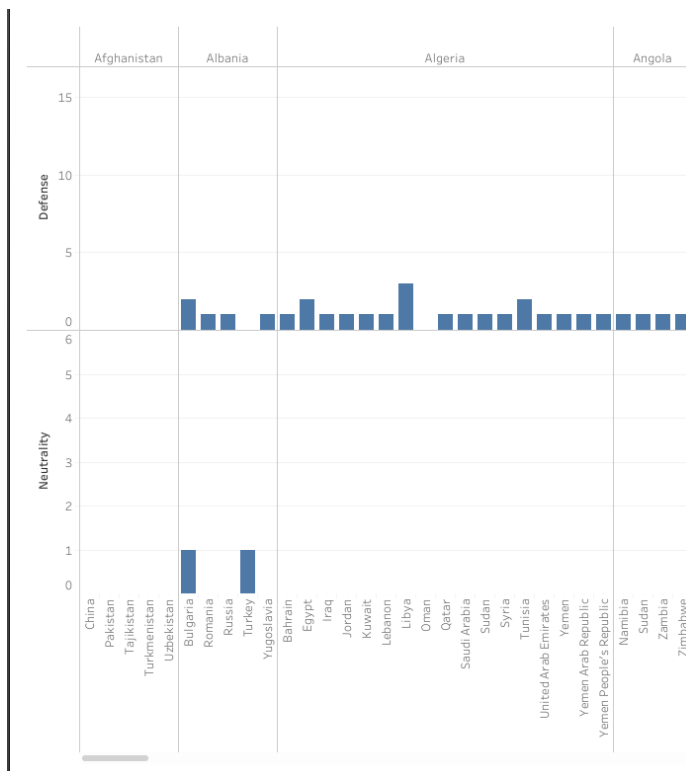


Fig. 5. Defense, Neutrality vs State Name

F. Treemap of Countries and Defense,Entente

The plot(Figure 6) is a Treemap which explains about the number of defense pacts and entente pacts that each country has. In Figure 6, Defence is of the Mark "Size" and Entente is of the Mark "Color". For example, United States of America has 81 defense packs and 115 entente pacts. The purpose of creating this graph is to provide a better informative visualization.

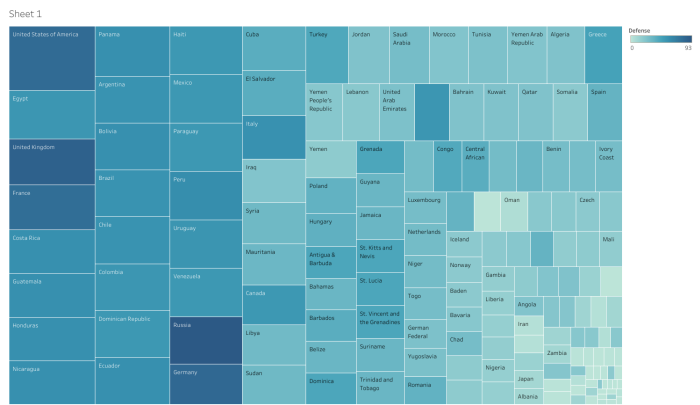


Fig. 6. Treemap of Countries and Defense

G. Packed bubbles of Countries and Nonaggression

The plot(Figure 7) is a Packed bubble plot which explains the number of nonaggression pacts that each country has. For example, Iraq has 52 nonaggression pacts. The purpose of creating this graph is to provide a better informative visualization.

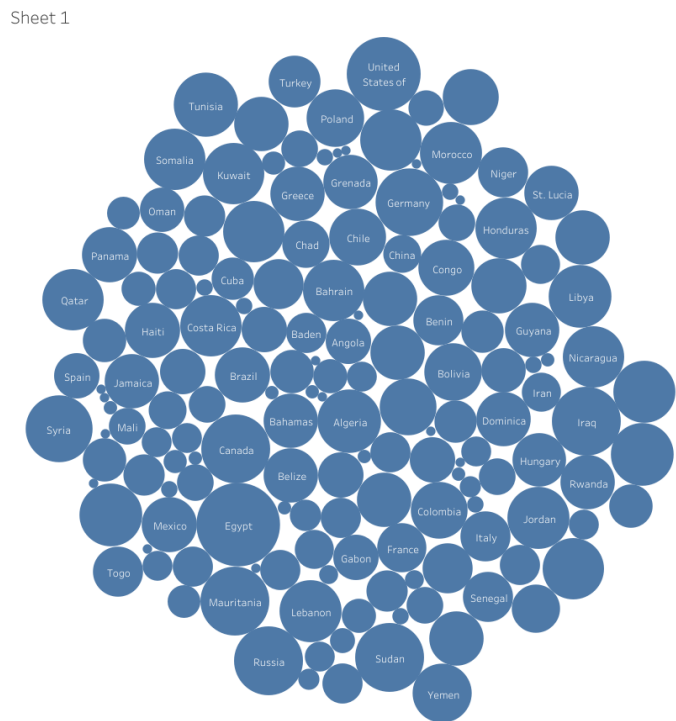


Fig. 7. Packed bubbles of Countries and Nonaggression

H. Geographical Maps of Countries and it's Defense

The plot(Figure 8) is a Geographical Map plot which explains the number of defense pacts that each country has. For example, Russia has 93 defense pacts. The purpose of creating this graph is to provide a better informative visualization.

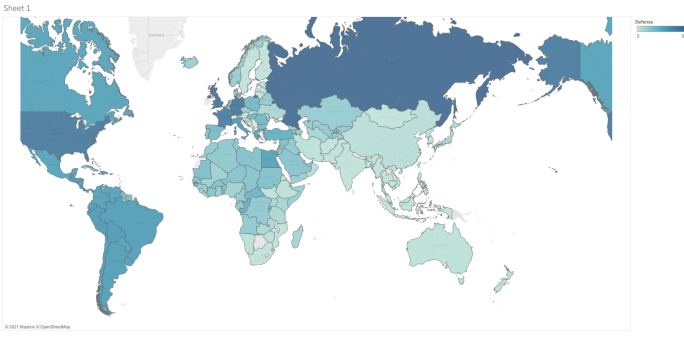


Fig. 8. Geographical Maps of Countries and it's Defense

I. Geographical Maps of Countries and it's Entente

The plot(Figure 9) is a Geographical Map plot which explains the number of entente pacts that each country has. For example, Brazil has 73 entente pacts. The purpose of creating this graph is to provide a better informative visualization.

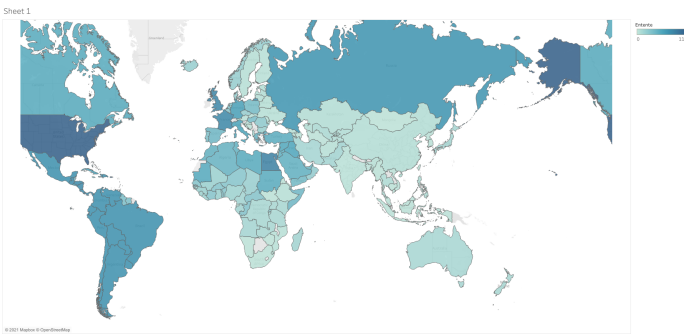


Fig. 9. Geographical Maps of Countries and it's Entente

J. Feedback loops and Nonaggression

A feedback loop means that the knowledge discovery from the dataset shall be used as a new dataset and a visualization can be performed. In Figure 7, if the size of the country is more, then there are more number of nonaggression treaties which roughly should mean that the country is safe from foreign country attacks. Figure 10 shows the number of nonaggression treaties that Canada has with other countries. There is a very low chance of foreign countries that are too far from the host country to be violating the nonaggression treaty as majority of the violations occur through land based and it isn't possible for countries that are too far away from the host country to attack it. So, in most of the cases of violation of nonaggression treaty, it would be done by the neighbouring countries. So, we felt there should be a new visualization to show the nonaggression treaties of a country with other Neighbouring countries so as to get a better idea on the nonaggression treaty. In Figure 11, We compared the nonaggression treaties of Egypt with other neighbouring countries. Egypt has 3 neighbouring countries(i.e, Sudan, Libya and Israel) and is having a nonaggression treaty with all of them. Therefore, we can say that Egypt has a very low

chance of being attacked by the foreign countries. In Figure 12, We compared the nonaggression treaties of Turkey with other neighbouring countries. According to Figure 7, Turkey has a high number of nonaggression treaties with other countries which should indicate that Turkey has a low chance of getting attacked. But, only 3(Iraq, Bulgaria and Greece) out of 8(Iraq, Bulgaria, Greece, Azerbaijan, Armenia, Georgia, Iran, Syria) neighbouring countries have a nonaggression treaty with Turkey indicating that Turkey has a higher chance of being attacked. Therefore, we realized the importance of a feedback loop as we were able to gain new insights of the visualizations.

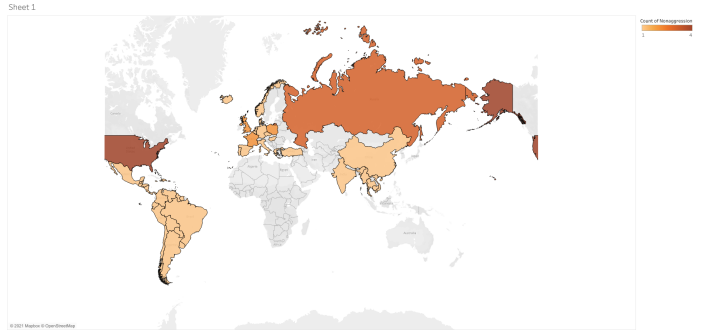


Fig. 10. Nonaggression treaties of Canada

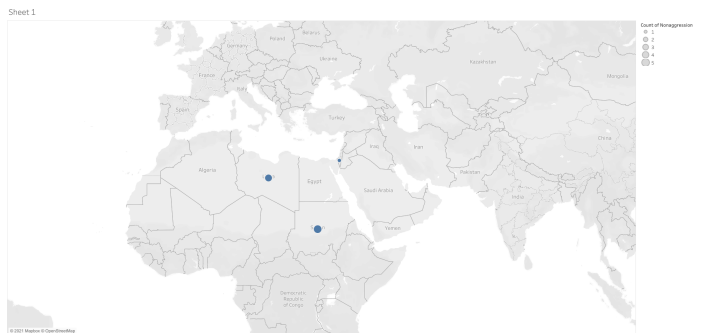


Fig. 11. Nonaggression treaties of Egypt with it's neighbouring countries

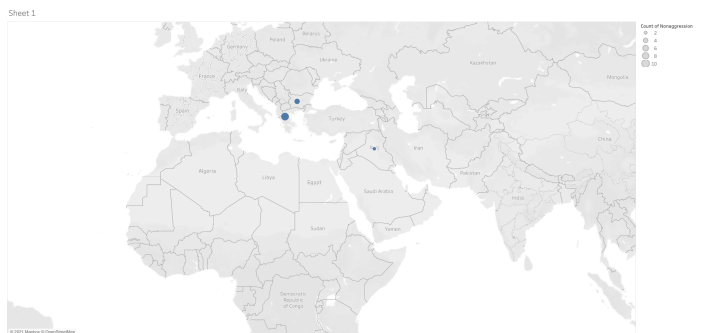


Fig. 12. Nonaggression treaties of Turkey with it's neighbouring countries

V. CONCLUSION

Although many useful visual analysis applications have been developed to assist users in understanding complicated relationships in large data sets, they are mostly limited desktop applications designed for single users. Collaborative visual analytics environments have also been developed, which allow users to work together to solve complex analytical problems. However, on their own, neither of these two modalities entirely reflects the continuous analytic processes carried out by real world experts.

In this report we have successfully generated some beautiful visualizations. Unlike the other assignments dealing with large dataset we had an appropriate size data this time which helped us to visualize better and infer a lot of information. Working on this final project made us realise the importance of data visualization with it's integrated applications. During these times of pandemic, it is very crucial to effectively visualise data so that we can make meaningful inferences from them and take action accordingly. These results will provide guidelines for designing visual analytics systems that accurately reflect the analytical processes carried out by real-world experts. Overall it was a fun activity to do and complete the task within the deadline.

VI. CONTRIBUTION

At the beginning, 3 of us analyzed all the columns of the dataset so as to get a better insight on how to visualize the data. There were 6 csv files in the dataset, the files with yearly columns had an extra column which shows the year values as to when the general population got to know about the pacts. Sreevardhan analyzed on all the columns like defense, entente, neutrality etc. and tried to interpret on how the data can be visualized. Tanishq focused on the start and the end year of the pacts and tried to find data visualizations as a geographical map plot. Suhruth focused on analyzing the pacts between all the countries with the type as well as the count, trying to map the data as a form of geological map plots and finding out the instances for feedback loops.