INFORMATION VISUALIZATION: VISUALIZING WARS

KU Leuven, Leuven, Belgium
Luca Van Deun, Max Dekoninck, Vital D'haveloose
Master Computer Science
E-mail: luca.vandeun@student.kuleuven.be, max.dekoninck@student.kuleuven.be,
vital.dhaveloose@student.kuleuven.be

Abstract

In this paper we describe a visualization of the intensity of inter-country wars. We present the dataset, the calculation of our intensity metric and an overview of the visualization. Related work will be analyzed and the most critical design decisions are discussed.

Keywords Information visualization; design; war; conflict.

ACM Classification Keywords H.5.0. Information interfaces and presentation (e.g., HCI): General.

I. INTRODUCTION

The subject of 'wars' is interesting, because wars have both a geographical and a temporal component as well as other quantitative and qualitative information. We narrowed down the topic to only inter-country wars in the last three centuries. Everyone knows the two World Wars, but there are many conflicts that affected millions of lives that not as many people have heard of. The classic mantra of historians applies: by looking at the past we can better understand the present.

II. GOALS AND TARGET AUDIENCE

GOALS

It is impossible to convey the horrors of a war with a simple visualization, but by comparing different conflicts with each other, we can put them into perspective. The goal of our visualization is to make it easy to compare different wars, not only based on the number of casualties, but also based on their intensity (a metric that is further explained in the Dataset section). Users will also be able to situate the different conflicts in space and time, with a timeline and a world map. We want to expand people's knowledge about wars and make them look at them from different angles than solely the total number of casualties, as this can help their understanding of the conflict. We can be satisfied if someone walks away from our visualization having found a link between two seemingly unrelated wars.

TARGET AUDIENCE

The visualization should appeal to anyone who is interested in history and wants to know more about (the evolution of) wars throughout the last few centuries. Therefore, the visualization should not include any history-specific jargon.

III. DATASET

Originally, it was unclear if the focus should be solely on complete wars or if individual battles should also be included. It became clear that focusing on individual battles would cause problems in terms of data, since they often are not well documented. It is also difficult to determine when armed combat becomes a battle, since unfortunately that's a daily occurrence in most wars.

In order to create the visualization, the following data is required for each conflict that we want to visualise:
- Name of the conflict
- Start date
- End date
- Countries involved
- Number of affected people
- Number of casualties
- Brief summary of the conflict

The Correlates of War website [6] offers spreadsheets with the conflict names, dates, involved countries and number of casualties. Only the inter-country conflicts were chosen out of personal preference. The short summaries were looked up manually on online encyclopedias like Wikipedia [2] for better accuracy, since the names of the wars in the database often aren't the same as the titles of the Wikipedia pages. If no corresponding Wikipedia pages could be found, alternative online sources were sought. The wars range from 1823 up until 2003. There were a few obvious typographical errors in the dates of the conflicts, like end dates smaller than start dates, that we corrected manually. The dates were also changed into the mm-dd-yyyy format for better browser support.

Figure 1 shows the histogram of the number of casualties in the wars of our dataset. We notice that

most wars have between 10^3 and 10^5 casualties. The number of casualties can give a general idea of the scale of a war, but we also want to be able to convey the intensity of a war. Figure 2 shows a scatterplot with the number of casualties (logarithmic) and also the conflict duration. Here we see that longer conflicts tend to have more casualties, but the amount of casualties for shorter conflicts can vary a lot more. This is where an 'intensity' metric could help better distinguish these conflicts, by taking more variables into account. Sometimes 'intensity' is synonymous with a metric used to categorize different conflicts based on the number of combat deaths [12]. We will be giving it a new meaning (see below).
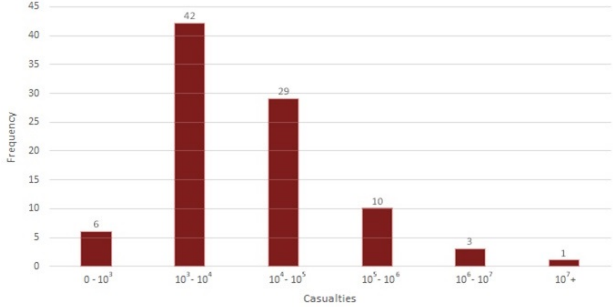

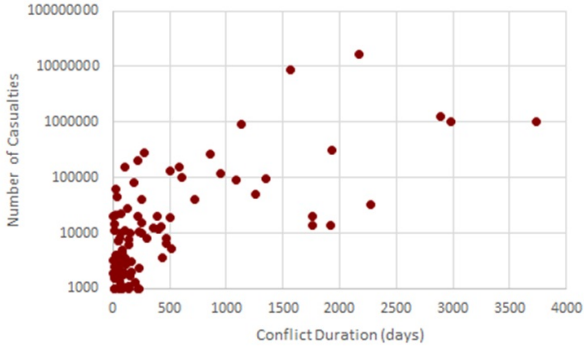Figure 1 – Histogram of the number of casualties in the wars of our dataset.

Figure 2 – Scatterplot of the dataset displaying the number of casualties (logarithmically) and conflict duration.

The number of affected people is necessary to calculate our intensity metric. We used a separate dataset in order to obtain these data. The Gapminder website [1] has a very extensive spreadsheet that contains the population for every country going back to the year 1700. There are a lot of missing data points prior to 1950 however, which will be linearly interpolated with the previous and next available data points if needed. Nonzero cells in this dataset are shown as dots in Figure 3, with larger countries usually having more complete data.

All the data is collected into one online spreadsheet that our visualization can read.
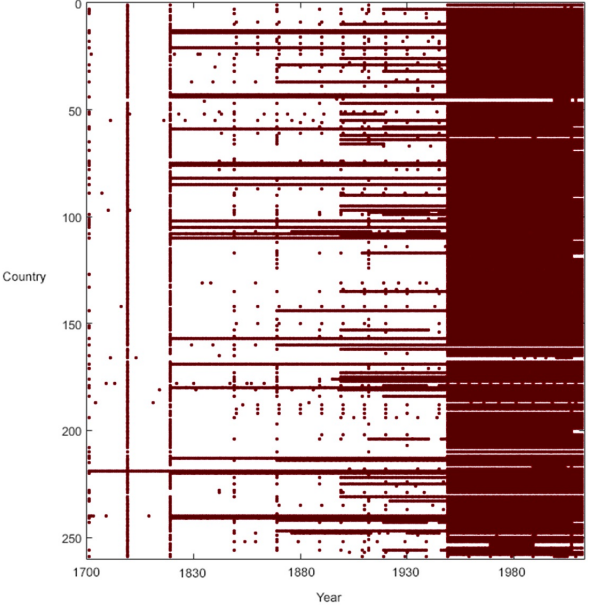

Figure 3 – Nonzero entries of the population dataset.

Intensity Metric

We measure the intensity of a war by calculating the average daily chance of dying for someone involved in the war. A short war with few people involved and a lot of casualties will score higher on this metric than a longer war with more people involved and a lower number of casualties. In order to calculate this percentage, Formula 1 is used. The duration of the war is expressed in number of days and the number of people involved is the sum of the population sizes of the countries involved at the start of the war.

$$dailyChanceOfDying = \frac{\#Victims}{\#Involved \cdot duration} \quad (1)$$

The 'average daily chance of dying' denotes the chance that a person involved in the war is likely to die on any given day.

There are some considerations with our metric. The change in population size from the start of the war until the end is not taken into account. If, for example, the population of a country involved in a war would grow, the average daily chance of dying would slightly decrease, since the number of people involved increases. On the other side, the people that die during the war are no longer part of the population at the end of it. Therefore, only looking at the population at the end of the war, or taking an average between the start and the end would also be somewhat arbitrary, since the victims should obviously be part of the number of people involved. We also don't take into account that certain countries might not be involved from the very start of the war.

Figure 4 shows the scatterplot of the chosen dataset displaying our metric for intensity versus conflict duration. The majority of the wars are found near the origin and along the axes. Longer wars tend to have lower average daily chance of dying, while shorter wars are more likely to have higher intensity. Figure 5 shows the same data but with a logarithmic scale for the vertical axis. If we compare Figure 2 with Figure 5 we can clearly see a better spread in the data achieved by our intensity metric.
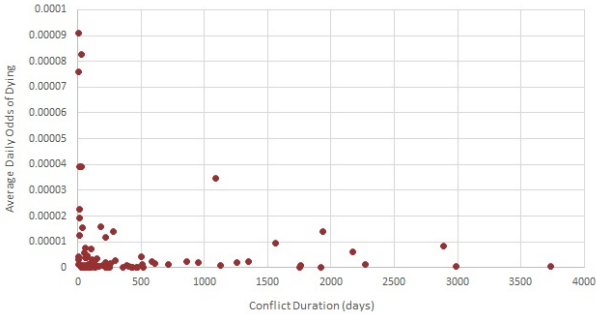

Figure 4 – Scatterplot of the dataset displaying our metric (linearly) and conflict duration.
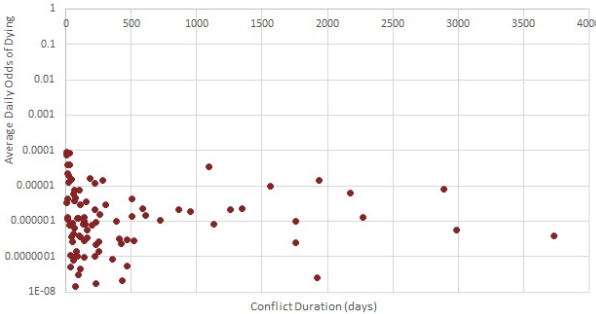

Figure 5 – Scatterplot of the dataset displaying our metric (logarithmically) and conflict duration.

Since the average daily odds of dying result in very small numbers, we decided to find an alternative way of representing these numbers to the users, to make them more tangible. We chose to represent these chances by comparing them to the number of consecutive heads thrown with a coin. These numbers range from 14 to 26, with the lower number being a higher average daily chance of dying, since the odds of throwing 14 consecutive heads is a lot higher than throwing 26 consecutive heads. These are still the odds of dying on one day, not during the entire conflict. The analogy would be to wake up every day during a war and to flip a coin to find out if you would live to fight another day. Figure 6 shows the data with this new representation and how the distribution will look in the final visualization. The horizontal axis now shows their start date instead of their duration.
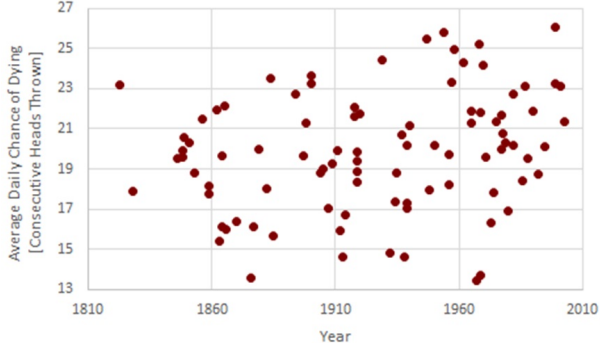

Figure 6 – Scatterplot of the dataset displaying our metric with the coin flip representation and the conflict start date.

IV. VISUALIZATION OVERVIEW

Our visualization is a combination of multiple sections, as shown in Figure 8. Each section adds a new layer of information in order to fit as much relevant data as possible in limited screen space.

Figure 7 – The Korean War represented on our timeline graph.

Timeline graph

The main graph shows a timeline on the horizontal axis and the average daily chance of dying on the vertical axis using a logarithmic scale. Figure 4 already showed the average daily chance of dying linearly on the vertical axis, which results in large gaps or overlapping data. As illustrated by Figure 5, a logarithmic scale spreads the data much better.

Wars are represented on this timeline graph as horizontal lines, starting at their start date and stopping at their end date, as shown in Figure 7. The timeline graph adds the dimension of time, which is very important for visualising historic events. The vertical axis introduces our metric, which should be one of the main focuses when looking at the visualization.
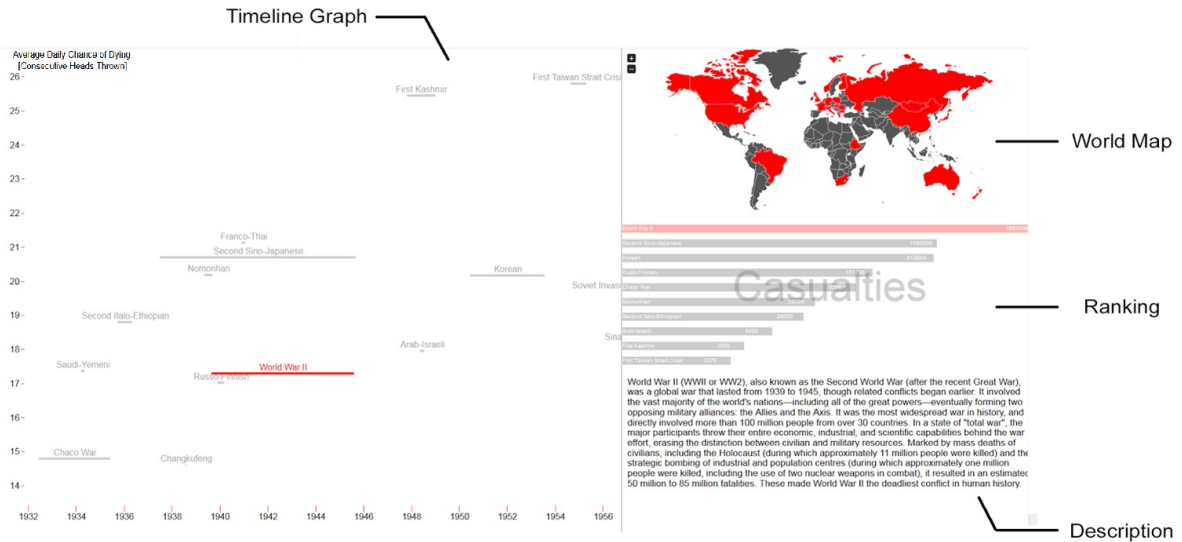
3

Figure 8 – An overview of the visualization, which is divided into multiple sections.

Map

A world map on the top right shows the countries involved in a certain war. It adds a geographical element to the visualisation, which is an important aspect of armed conflict.

Ranking

A ranking shows ten wars visible in the graph sorted by the number of casualties in descending order. We still wanted to show the number of casualties for each war, but didn't want to overload the timeline graph with more information.

Description

To lead users to more information about the selected war, we added a brief summary for each war and a link to a source with more information. By adding a description, users don't have to manually look for one and can learn more about wars right away. It helps our goal of expanding people's knowledge about wars.

More details can be found in the Visualization and Interaction section VI.

V. RELATED WORK

In our visualization, we use multiple sections to show related information. These complementary sections help users discover otherwise hidden relations in the data as explained by Baldonado et al. [9]. For example by adding the world map section you can discover that Japan hasn't been involved in any wars over the last few decades, while the USA have been very active in recent years. By partitioning the data into multiple views, they become more manageable to understand and compare. Each section has a format that is appropriate for the type of information it contains. The sections are always consistent, as they dynamically change with the changes applied to the timeline graph. The importance of coordination between multiple views is also pointed out by Aigner et al. [7].

Our timeline uses linear time intervals, since that is how people perceive time. Users can interact with the timeline graph by zooming in and out or by panning left and right, which was also used by Omar et al. [8]. Being able to manipulate this temporal granularity can lead to different conclusions, as stressed by Shabtai et al. [16]. For example the Football War and the Yom Kippur War seem to have the same duration on the timeline, but by zooming in you find out that the Yom Kippur War actually lasted more than four times as long as the Football War. We preferred this approach over other interaction methods like timeboxes [13] or brushing techniques [10] as the pan and zoom is more intuitive on a first try, which is important for a visualization that the user might only briefly interact with. If our visualization was aimed at experts, looking into these other interaction approaches that offer more functionality would be recommended. Another benefit of allowing users to zoom in on the timeline is that this reduces overlap of the different data points [11]. In addition, we slightly reduced the opacity of the data points, which

won't avoid overlap but helps reveal underlying points [17].

The mantra "Overview first, zoom and filter, then details-on-demand" [17] was used when determining the placement of the different sections (see Section VII Placement). While making our visualization we adopted the principle of having a high data-ink ratio [18]. The data speaks for itself and redundant non-data-ink, like the axes on the timeline, was removed to keep the visualization clean and simple.

A similar visualization is 'Poppy Field' [4], shown in Figure 9. It shows the wars from the 20th century onwards represented as poppies (a type of flower that is the symbol of the First World War). The horizontal axis also represents time, while the vertical axis shows the duration of the different wars on a logarithmic scale. The surface area of the flowers represents the number of casualties. The user can filter by continent, choose ranges for the number of casualties and choose a start and end year. Clicking on a flower shows some information about the war, including the duration, the number of fatalities, the continents where it took place, the participants and a link to the data source and some additional notes. One critique that can be given on this visualization is that it uses a surface area to express the number of casualties in each war. Humans are not very accurate when it comes to comparing surface areas [15]. Thus, it is difficult to compare the number of casualties for certain wars with this visualisation.

VI.   VISUALIZATION AND INTERACTION

The user can control the timeline of the main graph by clicking and dragging it left or right and by zooming in or out by scrolling. This functionality is made clear by the instructions given when opening the visualization and is reinforced by the cursor changing its appearance when hovering over certain interactive elements.

The different sections are all interconnected and interact with each other. Clicking on a conflict on the timeline graph will highlight all the involved countries on the world map. The conflict will also be highlighted in the ranking and the description will change to the one of the selected war. Selecting a conflict in the ranking will result in the same behavior as selecting it on the timeline. A country can be selected on the world map and all the conflicts that the selected country was part of will be highlighted on the timeline, as well as on the ranking.
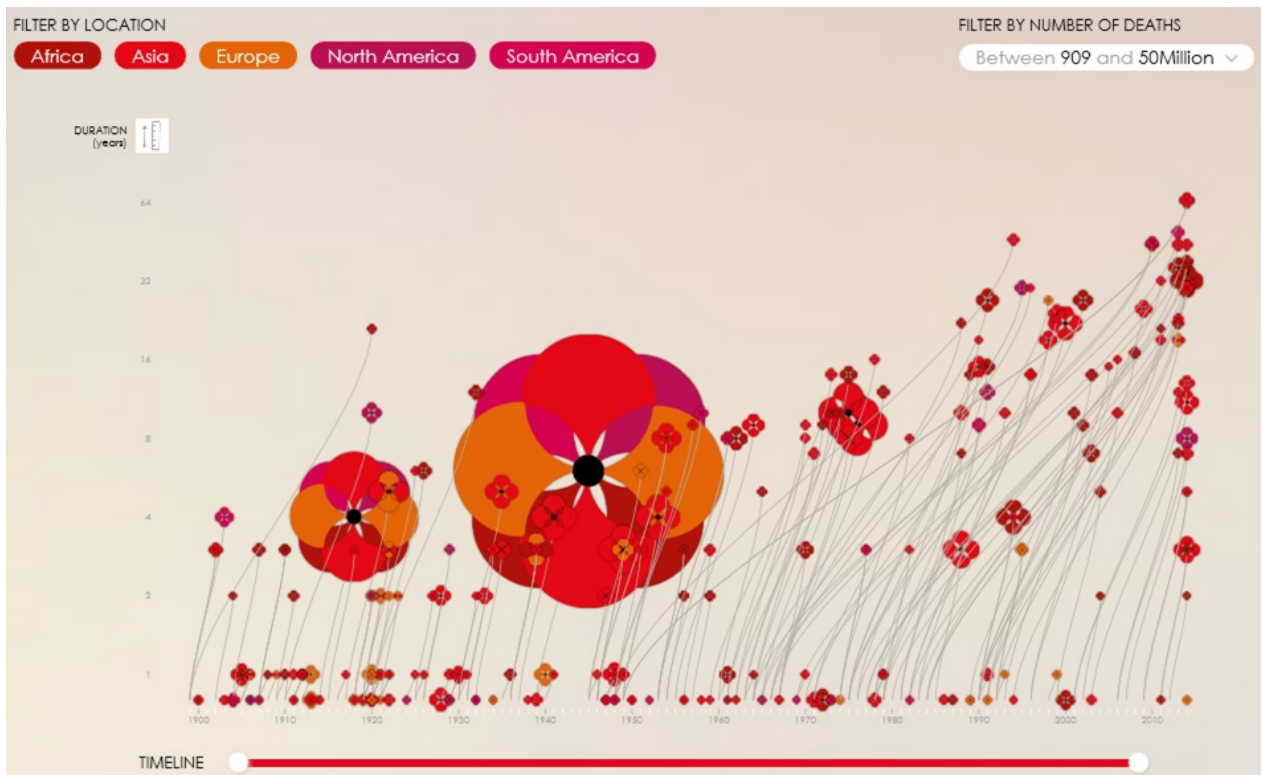

Figure 9 – The Poppy Field visualization, image used with the permission from the author.

The ranking section adjusts according to what wars are visible in the timeline graph. If the user zooms in on the timeline graph, there will be fewer wars visible. The ranking section only ranks the ten wars with the most casualties that are currently visible on the timeline graph, therefore zooming in or moving through the timeline dynamically changes what wars are being ranked.

In order to select smaller countries, the user can zoom in on the world map. A tooltip shows the name of the country that is currently being hovered over on the world map.

VII. IMPORTANT DESIGN DECISIONS

One of the first decisions that had to be made was choosing on what aspect of wars we would focus. Given the broad nature of the topic, narrowing it down makes the problem more concrete and reduces the amount of data we have to look at. As mentioned in the introduction, we chose to forego individual battles and focus on full wars. Battles aren't always very distinct and finding data like the number of casualties for a specific battle is not always easy. Inter-country conflicts were chosen over intra-country conflicts out of personal preference. Intra-country wars could be added later to expand the visualization.

Originally, the number of casualties was shown on the main graph, as depicted in Figure 10. The surface area of a circle indicated the number of casualties. The main problem with this approach is the same that we brought up in the critique of the Poppy Field visualization in the Related Work subsection V. It is difficult to compare surface areas. Overlapping circles also make it more difficult to distinguish what circle represents which war. This first attempt tried to show too much information in too little space. Splitting up the main graph in multiple sections helped to improve the clarity of the visualization.

A subsequent version of the visualization already adopted the idea of showing multiple graphs. The timeline graph was the same as the current timeline graph, but there were two additional graphs that supported it. An upper graph also showed a timeline on the horizontal axis and showed the number of casualties on the vertical axis. A graph to the right of the timeline showed the average daily chance of dying on the vertical axis and the number of casualties on the horizontal axis. One of the problems with this approach is that we were showing the same data multiple times with only small differences in representation. The number of casualties were shown in both supporting graphs. This is something that we can't afford to do if we want to show as much data as possible in the little screen space we have available.

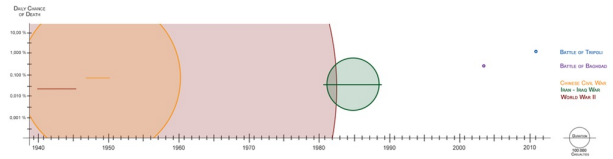Showing the same data twice can also cause unnecessary confusion.

Figure 10 – The initial design of the visualization used the surface area of circles to indicate the number of casualties.

Our last important decision was choosing how to display the information of the average daily chance of dying on the vertical axis of the timeline graph. Displaying the chance as a number or percentage resulted in very small numbers with negative exponents, making it difficult for users to compare those numbers or relate to them. That's why we chose to compare those chance to throwing consecutive series of heads with a coin. The average daily odds of dying are now equal to flipping a series of coins, which is something much more tangible. We went with flipping coins over rolling dice, since the resulting numbers are more fine-grained with coins. The chances lie between 14 and 26 consecutive heads thrown, where the lower number indicates higher odds and therefore a more intense war.

Section Placement

In order to determine the optimal placement of the different sections, there were a lot of factors to take into account. The main focus of the visualization should be on the timeline, since that includes our metric. It's the graph that the users should be interacting with the most. The timeline graph should have as much vertical space as possible to prevent overlap of data points. The ranking section should also have a lot of vertical space in order to show as many conflicts as possible.

Figure 11 shows the different layouts that we considered. The layout at the top leaves a lot of vertical space for the ranking, but sacrifices space from the timeline graph. The ranking isn't always full, leaving quite a bit of unused space in the bottom right corner of the visualization. The layout in the middle leaves a lot of vertical space for the ranking while still leaving a lot of vertical space for the timeline graph. The timeline is placed in the middle to grab the user's attention. What we don't like about this layout is that the world map has to be small and is used in a vertical space, while it is better suited to fit into a horizontal space. The description also doesn't always fill up the entire space on the left. The layout on the bottom of Figure 11 is the one we eventually chose. The timeline graph has a lot of vertical and horizontal space

and is placed on the left, since English speaking users tend to start reading from the top left to the bottom right [14]. The ranking section is shorter than in the other layouts, but the three big sections on the right are given ample horizontal space. Looking at the final visualization from left to right and top to bottom: we get an overview of the timeline graph first, we can zoom and filter by interacting with the timeline and finally we can get more details-on-demand by using the world map or the ranking to highlight conflicts on the timeline graph and by reading the description. This coincides with Shneiderman's mantra of "Overview first, zoom and filter, then details-on-demand" [17].
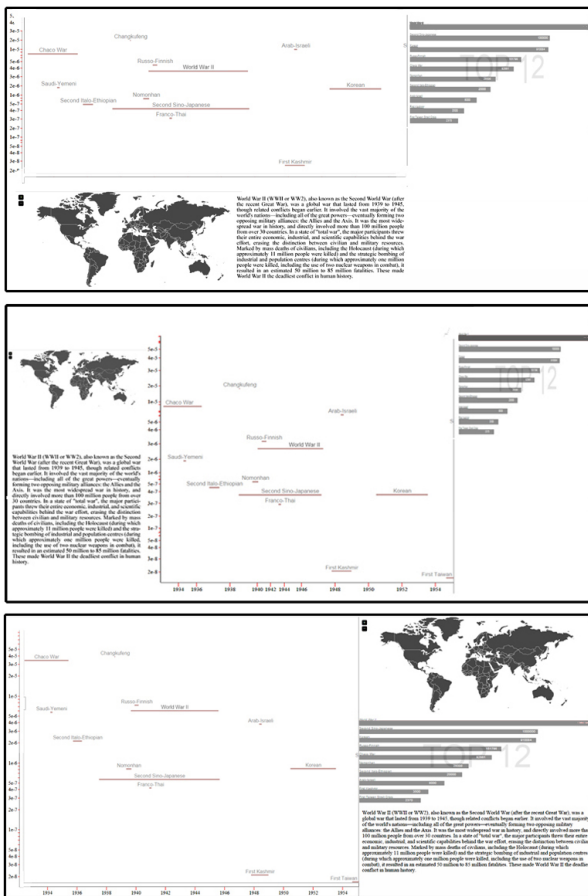

Figure 11 – The different layouts we considered, with the final choice as the bottom image.

VIII. TAKEAWAYS

Creating this visualization taught us that there are many factors to consider when making design decisions. For example trying to add more information can be detrimental if it will decrease the clarity and readability of the visualization, like we learned with our early mockup. Luckily there are some rules of thumb that can help with this process. Concepts like the data-ink ratio [18] or the "Overview first, zoom and filter, then details-on-demand" [17] mantra can guide those design decisions.

We also learned that trying to make the data fit the visualization can be a lot more difficult than finding a visualization that already fits the data. A lot of manipulation had to be done in order to fit all our data nicely onto the timeline, while there might have been a simpler solution. That's why it's so important to use real data in early mockups, so there won't be any unpleasant surprises later on.

IX. TECHNOLOGY

The entire dataset was kept on a Google Sheet so it could be edited by all team members and easily be loaded. We chose Google Sheet because it is easily maintainable and it also avoids the having to set up a database.

JavaScript was a great tool to create the visualization with, but a lot of effort had to be put into keeping the code readable and understandable so we could still adjust it later.

D3.js was used for drawing certain elements of the visualization. Particularly helpful were the support for adding scaled axes and for drawing elements with ".data" in combination with anonymous functions to determine the size and location of elements.

Last but not least jQuery was used to add interactivity to the visualization (more specifically linking mouse clicks to the appropriate functions) and for managing the different classes for the CSS code.

X. FUTURE WORK

The current visualization is a good starting point, but it can certainly be expanded with more data. We chose to focus on inter-country wars, but intra-country wars could also be added to our dataset. Other features that could be added to the visualization are different options to sort the ranking, allowing the users to choose according to what metric the wars should be sorted. A dropdown list could also be added to the timeline graph allowing users to choose a different metric for the vertical axis.

In the visualization countries are highlighted on the world map when selecting a war, but they are all highlighted in the same color. Using different colors would allow us to color the countries according to what side of the conflict they were on, which adds a lot of value for larger wars where multiple countries were involved.

XI. CONCLUSION

In this paper, we discussed our process for creating a visualization about wars, from gathering the data to choosing the final layout. We introduced our own metric for calculating the intensity of a war, so

as to not solely focus on the casualties and to look at the data from a different angle. To make these chances more tangible, we compared them to flipping a consecutive number of heads with a coin. Multiple alternatives were presented during the design of the visualization and our decisions were explained.

XII. ACKNOWLEDGEMENTS

REFERENCES

[1] Gapminder. URL `http://www.gapminder.org/data/`.

[2] Wikipedia. URL `http://www.wikipedia.org`.

[3] D3plus. URL `http://d3plus.org/`.

[4] Poppy field. URL `http://www.poppyfield.org/`.

[5] Jvectormap. URL `http://jvectormap.com/`.

[6] The correlates of war project, 2000. URL `http://www.correlatesofwar.org/`.

[7] W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski. Visualizing time-oriented data—a systematic view. *Computers & Graphics*, 31(3):401–409, 2007.

[8] O. Alonso, R. Baeza-Yates, and M. Gertz. Exploratory search using timelines. In *SIGCHI 2007 Workshop on Exploratory Search and HCI Workshop*, number 1, pages 1–4, 2007.

[9] M. Q. W. Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Advanced Visual Interfaces*, pages 110–119, 2000.

[10] H. Doleisch, M. Gasser, and H. Hauser. Interactive feature specification for focus+ context visualization of complex simulation data. In *VisSym*, volume 3, pages 239–248, 2003.

[11] G. Ellis and A. Dix. A taxonomy of clutter reduction for information visualisation. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1216–1223, 2007.

[12] N. P. Gleditsch, P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand. Armed conflict 1946-2001: A new dataset. *Journal of peace research*, 39(5):615–637, 2002.

[13] H. Hochheiser and B. Shneiderman. Dynamic query tools for time series data sets: timebox widgets for interactive exploration. *Information Visualization*, 3(1):1–18, 2004.

[14] W. Huang and P. Eades. How people read graphs. In *proceedings of the 2005 Asia-Pacific symposium on Information visualisation-Volume 45*, pages 51–58. Australian Computer Society, Inc., 2005.

[15] J. Mackinlay. Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog)*, 5(2):110–141, 1986.

[16] A. Shabtai, D. Klimov, Y. Shahar, and Y. Elovici. An intelligent, interactive tool for exploration and visualization of time-oriented security data. In *Proceedings of the 3rd international workshop on Visualization for computer security*, pages 15–22. ACM, 2006.

[17] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.

[18] E. Tufle. The visual display of quantitative information. *CT Graphics, Cheshire*, 1983.