

Data Visualization of World War Battles Dataset Using Tableau

S. Rishi Karthikeya, Pavan Raj, Mrs. Radhika

Department of Artificial Intelligence and Data Science
Methodist College of Engineering and Technology
Hyderabad, India

Abstract- This paper examines WWI and WWII battlefield data to show data visualization techniques. Included in the data set is data about casualties, number of days for the battle, its location, and the outcome. Data visualization was constructed with Tableau to gain insights into the trends and patterns by region and time. The study noted Europe and Asia saw the heaviest losses, and that a number of "significant fights" played a major role in total losses. The analysis shows how visualization methods can aid in comprehending historical war data better.

Keywords- World War I, World War II, Data Visualization, Tableau, Battle Analysis, Casualties.

I. INTRODUCTION

Modern history as they caused tremendous destruction, loss of life, and had a worldwide effect. These wars had an impact on several countries, particularly those in Europe and Asia, causing significant political, social and economic changes. Discussions of such historical wars contribute to a comprehension of the magnitude and ferocity of war in different times and places.

War analysis has, in the past, relied primarily on documents, texts, and expressions to describe war. These sources are valuable information, but less effective for visualising broad trends and patterns of large datasets. This is where data visualization techniques come in handy to turn complex historic info into helpful, visual representation.

In this paper, we analyzed a dataset that includes 200 battles from each World War I & II using data visualization techniques. Crucial information like casualties, battle timings, locations, and battle outcomes are provided. Regions and time trends were used to find patterns through the use of charts, graphs, and maps created with Tableau.

The aim of the study is to gain a greater understanding of the impact and distribution of war events, by means of visual analysis.

II. LITERATURE SURVEY

Historical records and analysis of world war have been used in several studies to jump to the conclusion of a battle. Researchers have been mainly concerned with the nature of casualties, the details of battle strategies, the regional effects of war and the order of war events. Over the last few years, Tableau began to make significant gains in the historical data visualization space, which helps to make complex historical data easier to understand and more meaningful to see. Uses charts, graphs and maps to improve identification of patterns and trends and comparisons. Research has demonstrated the power of visualization for gaining understanding of large historical data sets and for effective analysis and interpretation of information on war.

One of the most frequently used datasets for the study of war, the Correlates of War (COW) dataset provides structured data on wars since 1816. It provides data about country involvement, number of deaths in wars, the length of the battles, etc., allowing the study of the development of war throughout history and the worldwide nature of wars.

Another significant dataset of interest to Armed Conflict is the Uppsala Conflict Data Program (UCDP) which is based on post-World War II armed conflicts. Another significant dataset of interest to Armed Conflict is the Uppsala Conflict Data

Program (UCDP) which is based on post-World War II armed conflicts. It includes details of battle and death fatalities and conflict intensity, useful for a comparison of today's conflicts and a knowledge of how violence has changed in recent history.

Lacina and Gleditsch (2005) have created a special dataset of battle casualties in armed conflicts. They study how to distinguish the numbers of battle victims from the overall numbers of bodies in order to facilitate more accurate analysis of the impact of the war and prevent the distortion of the war's effects.

Likewise, the Peace Research Institute Oslo (PRIO) Armed Conflict Dataset provides comprehensive information about conflicts worldwide, including numbers of casualties, location, and type of conflict. This dataset is used for exploring the space-time partitioning and intensity of wars by region.

To facilitate a consistent classification of the wars, Gleditsch and his colleagues (2002) further standardized war datasets to ensure that classification techniques are used consistently, even over periods of history stretching from World War I to World War II.

Change point analysis is used in some advanced statistical studies to detect shifts in intensity of a war over time, like Fagan et al. (2019).

They concluded that such events as World War 1 and World War II and a time period following 1950 mark main turning points in the world's pattern of conflict.

In addition, from a different perspective, Lacina, Gleditsch, and Russett (2006) determined that since World War II, the chances of dying in battle have dramatically reduced — this suggests that modern battle is more fundamentally different than the battles of previous large-scale wars.

Seven new papers discuss how to use AI-based simulation models like WarAgent (2023) to recreate war scenarios and gain insights into decision making, conflict escalation and historical war behaviours through computer simulations.

III. FIGURES AND TABLES

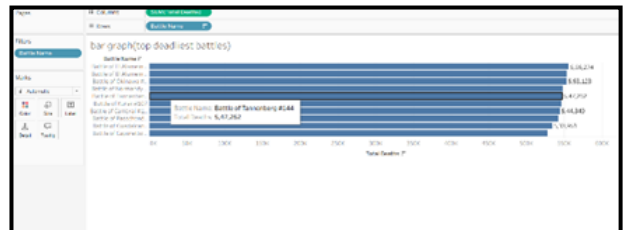


Fig1 : Top Ten Deadliest Battles in Terms of Total Number of Deaths

On the chart (Figure 1), are the 10 most deadly battles by number of dead. During the analysis of all the battles, largest number of casualties was registered in El Alamein. The numbers of casualties were also high on other major battles like Okinawa, Normandy and Tannenberg. El Alamein is still on top, though the numbers of deaths are diffused over the remaining engagements, indicating that there were some similar engagements with similar casualties among them. The overall figure indicates the extent of the human casualties inflicted by warfare and underscores the loss of life due to several battles.

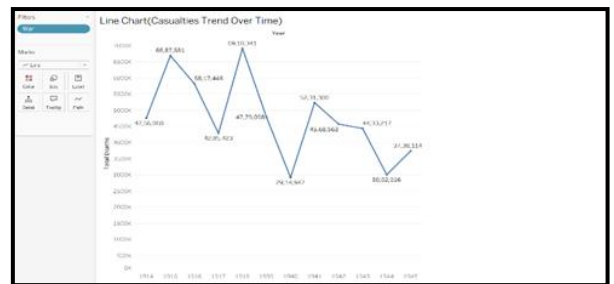


Fig. 2. Casualty Trend Over Time

The figure at change in casualties from year to year is displayed in figure 2. The graph has considerable fluctuations according to years, the years 1915 and 1918 are particularly marked. These are times of great bloodshed and mass murder. The numbers of casualties were smaller in the 1940s, suggesting a short-term dip in the number of war fatalities after 1918. There's another sudden jump about the year of 1941 though, that's a foretaste of a re-flashing of massive warfare. Generally it shows there are differences in casualty

figures from one year to the other, with certain years having very high casualty numbers.



Fig. 3. Duration vs Total Deaths

This is shown in the scatter graph in figure 3. The scatter graph does not display any obvious association between the length of time a battle takes and the number of casualties in that battle. A few battles were of several minutes duration, had much higher casualties involving Brenner guns, and some battles were longer, but had relatively low casualties. The practice of the last few wars has shown that ground factors other than duration played a large part in ratings of casualties, as at Kursk. This means that many things like battle severity, military strategy and technological development were responsible for the dimensions of battle casualties.

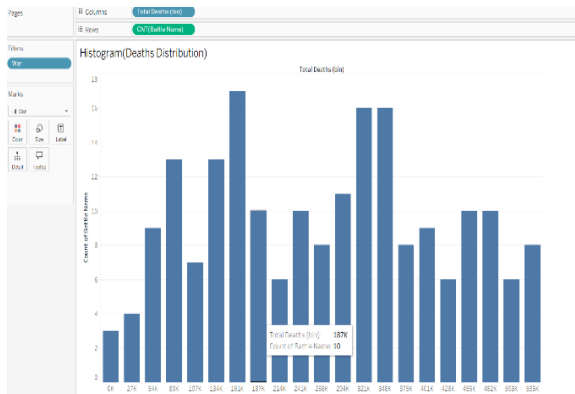


Fig 4: Deaths Distribution

The histogram shows the distribution of casualties for the overall number of casualties per battle for the world wars. The x values indicate the different ranges for the number of deaths, while the y values indicates how many battles are in each such range. The histogram shows that the deaths in most of the battles were evenly spread across a variety of ranges with

many battles having a high number of deaths, but there are only a few battles with high numbers of deaths.

2. Table Captions

Table I: Dataset Attributes

S.no	Field Name	Data type	Description
1	Battle Name	String	Name of the battle
2	Year	Integer	Year in which battle occurred
3	War	String	Type of war (WWI/ WWII)
5	Region	String	Region of battle
6	Location	String	Place where battle occurred
7	Duration	Integer	Duration of battle in days
8	Total Casualties	Integer	Total number of deaths
9	county	string	Level of casualties (High/Medium/Low)

Level of casualties (High/Medium/Low)

V. METHODOLOGY

The approach in this study was to gather, preprocess, analyze and visualize historical data of World War battle events using Tableau. The dataset used for the analysis consisted of World War 1 and World War 2 battle records from Kaggle, with a total number of around 200 records. Portable dataset included crucial data including battle name, year, war type, region, location, battle length, total casualties, casualty type.

The first step was to thoroughly review and clean the data in the dataset. Wherever appropriate, missing values were found and corrected. Entries that did not match and formatting errors were to be eliminated in order to enhance the reliability of the analysis. The numerical and categorical data types were also correctly formatted for import into Tableau.

Following the pre-processing of data, it was then uploaded to Tableau for visualization and analysis. Various visualization

techniques were used to investigate patterns and trends in the data. The most deadly battles were identified using bar charts to identify difference in the total number of casualties. Line graphs were used to compare the number of people who were killed in various years. To determine if there was a relationship between battle length and overall casualties, scatter plots were used and to explore the distribution of casualties in battles, histograms were used.

These visualisations were then examined to find out regional effects, battle intensity differences and battle trends over time. The greatest number of large engagements and losses were observed in Europe and Asia. Tableau dashboards enabled large-scale war data to be easy to interpret and made meaning of the vast historical information.

VI. CONCLUSION

Tableau data visualization techniques were used to analyze the data set, which includes information on 200 battles and events during World War II. The key learnings gained were the number of casualties, the geographic distribution of epicentres of battles, and the distribution of the timeline of war events. Tableau enabled the visualization of tons of historical data in meaningful and understandable ways. The results showed that not all battles had the same number of casualties. Many of the deaths were in just a few major battles, suggesting that there were some battles that were much more violent than others. The trend changed irregularly with some years showing higher peaks and sharp fluctuations reflecting the years when there were higher levels of war intensity. The results were in line with the historical timeline of wars. A correlation between the number of casualties and the duration of the battle was also analyzed, but no clear direct correlation was found. This means that battles didn't necessarily determine casualties based on how long they lasted, but also in terms of military strategy, technology and army size. Thereafter, the data was analyzed regionally, revealing that Europe and Asia were the regions most affected, underlining the central role of the two world wars in them. Finally, it was shown that data visualization is a useful method to analyze historical data. Visual analysis can be used to communicate complicated information in a clear and organized way, such that trends, patterns, and insights that are not clear using traditional analytical techniques can be identified. There are other factors that could influence the

results, and more sophisticated analysis methods could be used in the future.

Acknowledgment

S. Rishi Karthikeya and authors would like to thank miss Sana Mateen, for her direction and support with Data Visualization Course and along this project work. S. Rishi Karthikeya and classmates at AI&DS Department for their help and cooperation. Thanks to Kaggle for dataset and Tableau public for visualization tools to carry out this project.

REFERENCES

1. Encyclopaedia Britannica. "World War II Overview." 2026. Available: Britannica – World War II Overview
2. History.com. "World War II History and Facts." 2025. Available: History.com – World War II History and Facts
3. History.com Editors. "World War II Battles Timeline." 2025. Available: History.com – World War II Battles Timeline
4. United States Holocaust Memorial Museum. "World War II Overview." 2024. Available: USHMM – World War II Overview
5. [Our World in Data. "War and Peace Data." 2023. Available: Our World in Data – War and Peace
6. Google Dataset Search. "Historical War Datasets." 2024. Available: Google Dataset Search
7. Tableau Documentation. "Tableau Documentation." 2024.
8. BBC History. "World War Two." 2024. Available: BBC History – World War Two
9. Kaggle. "World War Datasets." 2024. Available: Kaggle – World War Datasets
10. United Nations. "Peace and Security Reports." 2023. Available: United Nations – Peace and Security