Data Dive into WWII

An Interactive Scrollytelling of the Deadliest War

I. Introduction

World War II was the biggest and deadliest war in history, and it ended a mere 75 years ago. With it slowly fading into the shadows of history, it is very easy to forget the mass destruction it imposed on the global stage. But this isn't the only potential issue; time erodes all. With a historical event of this magnitude, details, temporal events and numbers matter. Most sites and mediums present surrounding information in the form a boring timeline or tabular format that does not provide enough motivation to advance through. Conversely, there's less than engaging video content that, although reveals compelling footage, lacks the statistical fortitude to truly impart the gravity of this historical event.

We aim to create a interactive scrolling story-telling visual, or *scrollytelling*, of WWII. It will be an oversimplified, sequential narration of the major events surrounding the war. It provides an interactive narrative that will begin at the intial trasngressions and agitations of the axis power, culminating to the final srrender of the Japanese Empire. Along the way, we will not only provide temporal context, but emphasize the magnitude of death, destruction, and world changing events with inline interactive statistical details. Each major battle and and turning point of the war, as told in the scrollytelling, will be bolstered by its surrounding data for deeper analysis.

To truly convey the full gravity of the most horrific war in history, we will supplement all this with BDI, or *Bayesian Death Index*. For given battles, we will supply a filterable graphic that utilizes bayesian inference to reveal the probability of survival for that theater. Whether you are interested in an allied soldier fighting against evil and tyranny, or an axis aggressor following orders, their fate is summed up in single, statistical revelation.

We wish to present history in this manner to stimulate interest, curiousity and knowledge. The traditional methods of vanilla data aggregation or documentary style mediums no longer suit the masses. We aim to create a symbiotic bridge between narration and datap because, to know of D-day is knowledge. Knowing there was only a 1 in 4 chance of walking off the beach that day, is complete dogmatic comprehension.

II. Traditional Methods and Their Shortcomings[1]

WWII is probably one of the most documented wars in history. There are no shortage of sites, references, and data stores surrounding this great war. Each method can be broken into one of three delivery paradigms – the first being traditional educational content.

Sources such as (historyextra [dane's source]) and (brittannica [Niya's source]) are examples of historical narrations. They tell us simplistic things such as, why WWII started in the first place or, the main details surrounding the war. These mediums are reflective of the antiquated style of knowledge transfer we are accustomed to. Consisting of minimal data analysis or presentation, they are a bland and to the point narration.

POV specific style tomes and published papers also fall under the first paradigm. The POV piece (Day Through German Eyes [Niya's source]) reveals the elements of history from a specific angle. This method is an interesting take, but restricts the learner to a narrow view. Finally, "The Effects of World War II on Economic Health Outcomes across Europe" is an example of published historical analysis that suffers from the same shortcomings as the aforementioned references; they're boring, non-engaging and does a disservice in conveying the magnitude of death and destruction. These mediums provide little, to no motivation to keep a learner engaged. They exclude deeper data analysis for the sake of continuity story line progression.

Figure 1: Typical depiction in a traditional historical narrative with no supplemental data.

The second paradigm are mediums that are bound in statistical fact and data dives, but they're lacking sequential context. The website (Vox [Dane's source]) for example, has provided a consolidated set of data driven maps that represent different events throughout the war. This canonical format is revealing of deeper trends, but the viewer is given fragments of background information. This relegates them to external sources to complete the narrative, and ultimately the big picture.

Finally, the third paradigm is one bounded in data and narration by nature, but lacks interactivity. An amazing data dive into WWII is done by (fallen.io [dane's source]) where they provide an sufficiently deep analysis and emphasis in the numbers that constitute everything WW2. However, the analysis is bounded by whatever content the creators have presented. There is an inability for a user to manipulate, adjust , or transform the data to satisfy a specific flavor of curiosity. The shortcomings of all 3 of these paradigms are ones we intend to rectify.

1. Heilmeier question #2

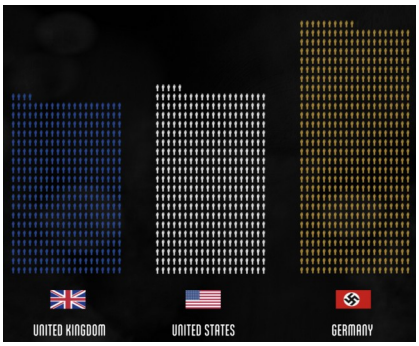

Figure 2: Graphic from Dane's source

III. *ScrollyTelling*[2]

Our efforts to generate more interest and knowledge surrounding this historical event will be instituted in the methodology of *ScrollyTelling*. This is an environment that promotes user engagement and involvement. It is a web based medium that requires the user to scroll at their pace for continuous flow and transition in a narration. Upon reaching certain points in the narration, supporting visualizations and graphics transition into view in such a way to fully supplement and complete the section being detailed. Not only will the initial revelations of these graphics default to meaningful insights, but they will provide the user with interactivity capabilites for deeper analysis.

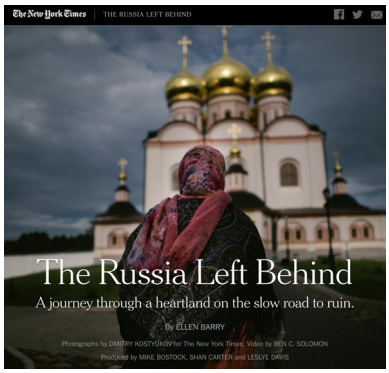

Figure 3: Example of scrollytelling from d3.js creator Mike Bostock

There are some examples of this new method in narration sytle, some of which that are masterfully executed by one of the original d3.js creators, Mike Bostock. As a developer for the New York Times, he has created *scrollytelling* pieces that transcend conventional articles. Specifically, his pieces "The Russia Left Behind" and "The New York City Marathon – A poem in sights and sounds", serve as the main inspiration behind our work.

We chose to take on this endeavor, in this format, because everyone should care about history[3]. Those who fail to know, retain and comprehend history are doomed to repeat it. It is also beneficial in that as time distances us from the actual event, we become less sensitive to it. We should never lose sight of the fact that the lives of millions were lost due to the greed, anger, and hatred of few.

IV. The Method & the Madness

Data will be scraped from various places such as (but not limited to) Wikipedia, ww2db, history channel, nationalww2museum, angel fire. Given that most of this data is already contained in tabular form on these sites, most of the data will be scraped and dumped as CSV. In the event that there is data from different sources, that can ultimately be joined in a meaningful way, database design and usage will be explored. After the data is scraped, it will be consumed by d3.js and represented as an interactive scrolling website.

The impact of our project is one that is hard to measure intrinsically[4]. The intent is not to directly affect any particular system, business, or person. It is instead, an exercise of bringing history to life with data, and giving data a contextual history. Conversely, the only apparent risks of this project is the delivery method itself. Since we aim to create an interactive scrolling visual narration, which none of us have attempted before, we expect initial hardships in execution of this content delivery style. This risk is comparatively low compared to the the payoff of a medium that teaches, inspires, elevates awareness.

Everything will be developed using Python, CSS, HTML, JS, and d3.js. We will all agree on the entire sequential narration as a whole, then divide the years of the war among each person. Assuming that each member is responsible for 1.5 years of the war, we estimate a timetable of 2 weeks for each person to source and scrape the data for their timeline, and another 2 weeks to build the site.[5]

Since the choice of tools are open source mediums, the only cost that could potentially be accrued is the server costs to host such a site. A typical low end web host on AWS costs anywhere from $3 - $10/month. Most, if not all of the data for this project is openly available data and will be scraped from multiple sites, free of charge.

In order to sufficiently gauge progression, we will adhere to the following key performance indices[6]:

KPIs			
Mid Term Check	Narration written and agreed upon	Each member has accumulated enough data to develop their data story	Data has been processed and stored in a canonical format
Final Check	Each part of the story is clear, concise and properly sequential	Data truly reflects and enhances the story being told	Scrolling interaction of the site is bug free and achieves the seamless effect of ScrollyTelling

2. Heilmeier question #1 and #3
3. Heilmeier question #4
4. Heilmeier question #5 and #6
5. Heilmeier question #7 and #8
6. Heilmeier question #9

REFERENCES

[1]
[2]

2. Heilmeier question #1 and #3
3. Heilmeier question #4
4. Heilmeier question #5 and #6
5. Heilmeier question #7 and #8
6. Heilmeier question #9