DD2 WW2
Data Dive into WW2
An Interactive storytelling and analysis of survival chance.

1.) What are you trying to do?
Ww2 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's 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, the details matter, temporal events matter, catalysts matter. Most historical sites present surrounding information in the form of a boring timeline that most people have no motivation to scroll through, or less than engaging videos to sit through. Because of this, we aim to create a storytelling visual of WW2. It will be an "oversimplified" interactive narrative that will begin at the initial transgressions and agitations of the axis powers, culminating to the final surrender of the Japanese Empire. Along the way, we will not only provide sequential context, but emphasize magnitude, death and destruction with interactive visuals. Each major battle and turning point of the war, as told in the narration, will be bolstered by its underlying statistics and data. To truly reveal the gravity of the most horrific war in history, we will reveal what your chances of survival would have been for various theaters of war, whether you were an allied soldier fighting against tyranny, or an axis aggressor, following orders. Knowing of D-day is knowledge, knowing you only had a 1 and 4 chance of walking off that beach, is complete dogmatic comprehension.

2.) WW2 is probably the most documented historical event of the great wars in history. There are many sources from books, websites, history channels, and videos the describe events during the war. However, most sources focus on one of two sides; narration or data. Places like Wikipedia are a great place for specific data pertaining to certain events in the war. But it fails to give a complete sequential storytelling of the war. You have to click from link to link, putting the pieces together. And with so many little battles in the war, it is hard to follow the most important ones that primarily contributed to the flow of the entire war campaign. This is true of many other statistical sites. On the other side of the spectrum is sources such as , YouTube, History Channel, Smithsonian, etc. that have great sources of storytelling and narration. But these sources typically spew general data and high level statistics for the sake of continuity in their narration. You have no ability to comb through the data in the context of the narration. Simply hearing facts along the lines of "100,000 soldiers died in this battle" really does a disservice to highlighting the true impact of each battle. If you want a contiguous narration with immersive data, you'd have to pause the narration and check other sources for the full battle details. Even then, most sites provide this data in a basic tabular format. There is no ability to pivot, reshape, or transform the data in a meaningful way outside of general contexts.

3.) We aim to merge the two (data & narration) by taking the user on an end to end narration of WW2, and along the way we will supply the data inline. That way there will be no break in continuity, the visuals will bolster the story, and the story provides context to the data. It is a full compliment structure. We will start with the initial acts of aggression (Hitler rebuilding his army and air force in violation of the Treaty of Versailles, Japan invading China to grow their "Sphere of Influence). From there we will allow the story to flow, following the most crucial components of the main campaign. For each major event, we will provide the surrounding data specific to that event and present it in a format such that the user can interact with it and manipulate it to reveal deeper insights. We will also provide for certain events in the narration, a "Chance of survival" graphic. This graphic will be the results of a Bayesian inference method, highlighting just how likely each soldier was marching to their death (e.g. given that the battle is Dunkirk, and you are a British soldier on the beach, your chances of survival are 28%). A user will be able to filter, for given battles, specifics such as which power a soldier was

fighting for, what country their military belong to, if they were ground troops or pilot, if they were a pilot – what type of plane did they fly. From this filtering, we will deduce the exact chance of survival (or inversely, death) of a given soldier. We aim to achieve this interactive storytelling via the same methods Mike Bostock, co-creator of d3.js achieves his engaging stories with the New York Times. Examples of this are:
http://www.nytimes.com/newsgraphics/2013/10/13/russia/index.html

https://www.nytimes.com/interactive/2014/10/28/sports/marathon-videos.html?mtrref=undefined&assetType=REGIWALL&mtrref=www.nytimes.com&gwh=F106F0DE4573825F25F5E052C7E5EAA8&gwt=pay&assetType=REGIWALL

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.

4.) Everyone should care. Those who fail to know 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.

5.) The impact of our project is one that is hard to measure intrinsically. The intent of this project 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 home.

6.) The only immediately apparent risks of this project is the delivery method. We aim to create an interactive scrolling visual narration. Given that none of us have done this type of thing before, we expect initial hardships in execution of the medium such that the overall desired effect is achieved. The payoff is a medium that teaches, inspires, elevates awareness.

7.) The project will be developed using CSS, HTML, JS, and d3.js. Since these 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. We can scrape this data from multiple sites, free of charge.

8.) We will divide the history of the war among each team member. We will all agree on the entire sequential narration as a whole, then divide it. 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.

9.) Midterm exams for success are:
- a.) Has the narration been written and agreed upon
- b.) Has each member accumulated enough data to develop their data story.

Final exams for success are:
- a.) Is each part of the story clear, concise and properly sequential
- b.) Does the data provided truly reflect and enhance the story being told

c.) Does the scrolling interaction of the site achieve the seamless combination of data and storytelling in a truly symbiotic way