Michael Wornow and Theo Walker 4/3/18

US World 34: Project Proposal

**Leveraging Data to Uncover Novel Insights on the Civil War**

Though many of the major themes and arguments presented in this course during lecture and section seem intuitive and readily apparent, the nuanced nature of these claims can be difficult to empirically substantiate. With this data-driven project, we hope to analyze data relating to the Civil War to uncover evidence for interesting insights related to the themes of this course. Hopefully, we will then be able to definitely prove or disprove some of these claims, and offer support for several new novel narratives based on data. The main questions we are seeking to answer are thus:

1) How does data verify the narratives of US World 34?

2) What novel insights on the Civil War can be made using historical/modern datasets?

Within these higher-level questions, there are many potential sub-questions and data sets that we can explore. First, we will attempt to investigate the themes of selective remembrance and historical re-interpretations of the meaning of the Civil War by using Twitter. We will run a sentiment analysis of tweets to determine the context in which the Civil War is being referenced by modern commentators, and measure the positivity or negativity of words used to describe it. We could then plot this data on a map, graph trends in sentiments over time (potentially pre- and post-Trump), and compare it to other data sources to see if there is a correlation between how one remembers the Civil War and other factors like political affiliation, socioeconomics, etc. This would allow us to make useful inferences such as how the meaning of the Civil War has been recently re-interpreted and how citizens today draw parallels between it and current events.

Another possible direction is to analyze primary sources (such as letters) from the Civil War period itself from the War of the Rebellion archive. We could then run a sentiment analysis on this data set like we did with modern tweets, or alternatively analyze the frequency of certain terms and phrases and see how they changed throughout the war. For instance, one idea we discussed in lecture is that those involved in the fighting became more pragmatic and less spiritual over time. We could explore this effect by documenting how the frequency of the words “God,” “prayer,” and other religious phrases changed in our data as a function of time.

Although these are the data sets we are most strongly considering, these are just a few of the many promising ones we have discovered. Other datasets include information on black sailors in the Union army, military enlistment info and battle data from both sides, Census data from 1850-70, and economic statistics. We could plot this information on an interactive map to allow users to fully explore this data and gain a greater understanding of it.

The project will likely consist of three main stages of progress:

1) Determine desired data sources and mine data

2) Perform regression analysis and various calculations

3) Create medium through which results can be presented in a clean fashion

For the last stage, we will likely create a website on which users can interact with our data and findings, as well as be connected to our sources so that they can explore them further.

Michael will primarily be responsible for data mining whereas Theo will run the analysis of that data. We will work collaboratively on designing the end presentation, which will likely entail designing and programming an interactive website.

Guys this is a fantastic idea/proposal. You have identified how the workload will be evenly split and there is enough here to equal a final project. Regarding the final product you mention an end presentation and a website. How do you envision that presentation? Will the website include a summary of your findings? I think it is important to have a place where you can state your findings/conclusions and I would imagine this would be on the website.

As I said in my comments, I don’t really know much about data analysis, but once you decide on a dataset I can help you with finding the right source materials. Based on my suggestions, I have included some source materials below. I am really excited to see the final product here, and if you have any questions along the way please do not hesitate to reach out.

**Suggested Sources**

Charlottesville

[New York Times: A Guide to the Charlottesville Aftermath](https://www.nytimes.com/2017/08/13/us/charlottesville-virginia-overview.html)

[Washington Post: Recounting a day of rage, hate, violence and death](https://www.washingtonpost.com/graphics/2017/local/charlottesville-timeline/?utm_term=.40967c675417)

Charleston Massacre

[GQ Magazine: A Most American Terrorist: The Making of Dylann Roof](https://www.gq.com/story/dylann-roof-making-of-an-american-terrorist)

[New York Times: Nine Killed in Shooting at Black Church in Charleston](https://www.nytimes.com/2015/06/18/us/church-attacked-in-charleston-south-carolina.html)

New Orleans Confederate Statue

[NPR: With Lee Statue's Removal, Another Battle Of New Orleans Comes To A Close](https://www.npr.org/2017/05/20/529232823/with-lee-statues-removal-another-battle-of-new-orleans-comes-to-a-close)

[Time Magazine: New Orleans Mayor on Removing Confederate Monuments](http://time.com/4790674/mitch-landrieu-new-orleans-confederate-monuments-speech/)