Simona Bisiani - April 2020 - DSSS - Linköping University

"O bella ciao, bella ciao, bella ciao, ciao, ciao"

An analysis of Wikipedia´s "List of anti-war songs" and their related wars

I grew up in the North-West of Italy, in a small town where the Alps meet the Apennines, and where, now over 50 years back, my grandfather, like many grandfathers of friends of mine, fought as a Partisan in World War II, in the civil conflict that split Italians in two factions, the Resistance and the National Fascist army. Being my area quite proud of its "rebellious" nature, I was raised singing chants from the resistance, a gentle reminder of our local heritage, a political statement on a conflict that I never witnessed. It made me think: music is eternal, timeless, but it is also often timely - we make music about things as they happen, and then these songs crystallize in time. . . or do we? Within the realm of music about war, do we make music during a war or after it happens? Are some wars more sung about in popular culture? If so, is it because they were more deadly? Or because they lasted longer? In the following analysis, I use data from several Wikipedia pages on anti-war music and on related conflicts to do an exploration of the factors that affect the creation of music about war.

Data Collection

The first step involved data collection from Wikipedia List of Anti-War song. From the relevant lists, I only retained those that had a meaningful sample of songs (>1). Following some cleaning and wrangling, we got a dataframe containing 1054 songs, with information about the song title, artist, release date, and related conflict. We can start looking at the count of songs per each list in the Wikipedia page.
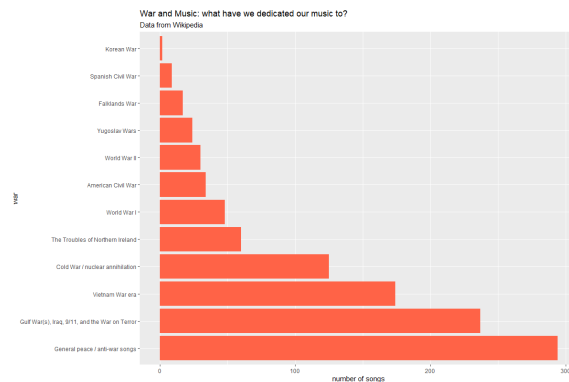

The second step in the analysis consisted in obtaining war-related information. Particularly, I was interested in the duration of the conflicts, death tolls, date of the conflicts, location, and combatants. A suitable place to obtain such data was again Wikipedia´s List of Wars by Death Toll, where I assumed I would have been able to find most of the wars in the previous dataframe. Luckily, most wars were contained in a table called "modern", thus I was able to follow a similar procedure to scrape the relevant rows and columns that interested me. For all those I was unable to find in this page, I manually googled the information and inserted into the new dataframe on information about wars. Some conflicts were not so easily coded: for instance, my anti-war songs list puts the Yugoslav War as one category, when de-facto it is comprised of several smaller conflicts, with individual dates and death tolls. As a result, I summed information of the smaller conflicts to obtain a matching category. Same was done for the sub-list "9/11, Gulf War, Iraq War and the War on Terror". I considered splitting the song list, but that would have inolved manually coding which conflict each song in the list related to, with potential overlaps. Perhaps something I could expand to with tools in text analysis? Again, some data wrangling and cleaning was required. Once I obtained my two datasets, I merged them using the war name as a key id.

Data Analysis

The analysis was run in R, and visualisations were created through several packages, from the well-known ggplot2, to streamgraph, plotly and gganimate to incorporate some interactivity on the output page. Whilst this document contains stills about the plots, feel free to check out the full interactive webpage resulting from this analysis here: http://compscjournalism.org/projects/anti_war_music/anti_war_music.html

The first plot is a plotly html embedded graph that shows the linerange of when the war took place, and distributes songs around the timeline as they got released. By hovering over the dots, information about the song is revealed.

From this graph, we can see that the patterns are quite different for each conflict. For example, the Yugoslav War had most music about it made during the years of the conflict, whereas on the opposite side of the spectrum the American Civil War became popular again in music nearly 100 years after it happened, suddenly revamping. Similarly, music about World War I and World War II has been made during, and well long after the conflicts ended, perhaps indicating them as the most "timeless" conflicts of them all. Clusters of songs can be seen around the Cold War and the Vietnam War, indicating a lot of music about them was made within the same short period of time.

Are songs made during or after a war?

Data from Wikipedia
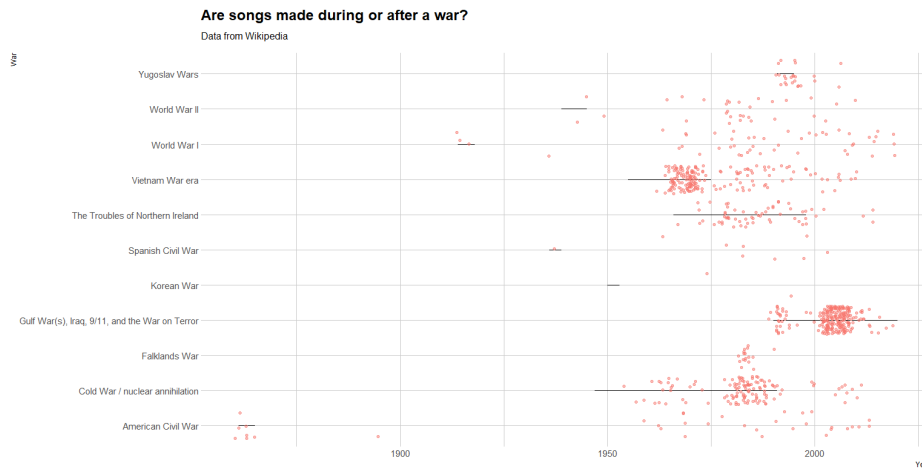
The second plot shows statistical correlation between number of songs per conflict and the deadliness of the conflict. Using linear regression, I model the relationship between those two variables and notice that they are negatively correlated. However, the dataset is a good example of when not to run a statistical analysis: the sample size is not large enough, and the conflicts are hardly comparable. The Cold War, for example, has hardly any deaths (when not counting its proxy wars, or when counting its proxy wars music as separated data points), whereas World War II had an incredible amount of death however, not as much music was made for it as for other conflicts. Another question we have to ask ourselves is, is our data source reliable enough to return a good estimate of the relationship between deadliness and popularity in music? From a quick glance of the Wikipedia song list, I noticed most songs were in English, made by American or British artists. Certainly, when looking within-combatant nations, the dataset would grow larger of international music. Finally, the dataset focuses heavily on modern music, just as it focuses heavily on modern conflicts, yet the singular American Civil War is included. How much can we trust the data? It is Wikipedia, after all. Despite its lack of purpose, I thought it was interesting to point out the limitations of my dataset in those regards.

Correlation of war death toll and number of songs per war
Data from Wikipedia
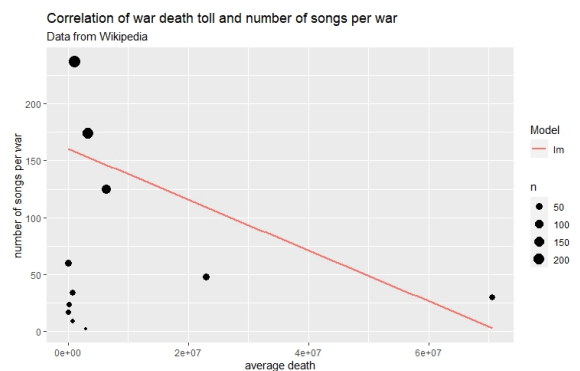
Finally, I have included some interactive visualizations of the music-making overtime. Please note you can find the interactive versions at the aforementioned link. The first one is a plot of how many songs were made yearly per each conflict. The plot was made using R´s interactive package streamgraph. It shows that some conflicts have dominated some decades more than others: particularly the Vietnam War, the Cold War and the more recent Iraq War, Gulf War, War on Terror and 9/11. Following is a plot just showing the cumulative version of this plot, again visualised through streamgraph. I here included name labels of the conflicts to provide an indication of when the conflicts happened. It shows mainly that most music included in the Wikipedia list is released from the 1960s onwards, despite some conflicts taking place much earlier. Going back to our original question, is music made during or after a war? If we were to really trust our data, we would conclude that there has been a boom in anti-war music-making following the sixties, and we could speculate that perhaps jazz, blues and classical music were unsuited genres to speak about war? Could the rise of rock, folk, punk, disco, hip-hop and pop have lead to more topical music, more political music? I guess I will leave that to my next project on text analysis.
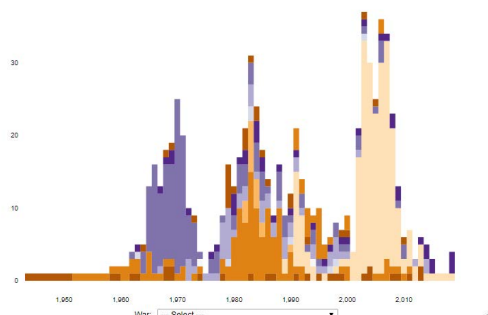
I have added a section on automation at the end of the R script, where I scrape 133 Wikipedia pages on songs by music themes, which I can use to expand this research in the future.