

## Milestone 2

Dash Chin

9/24/2020

Github Repo Link: [https://github.com/dashchin/Milestone\\_2.git](https://github.com/dashchin/Milestone_2.git)

The first GOV 50 final project I looked at is called “Behind Billboard: Exploring the Audio Features of Popular Music” by Parker Mas. I chose this project because it did extensive data gathering (American pop music since 1968) and is tangentially related to projects I have in mind right now (this project is different, however, because it is looking at features of songs I don’t intend to explore). I think this project models general change very well. The author made several predictive models not just in different styles, but also with different time frames. For example, the “Audio Features Over Time” section features graphs where you can choose to view changes over the course of a decade or a year. I think this a thoughtful inclusion. However, I think there may be some flaws within this project. I admire the intention behind the functionality under “Artist and Song Data”, where a user can search for the attribute values of any artist’s most or least popular songs. There do seem to be some bugs. For example, if I search “Daft Punk” only two songs are modeled. I think the data set may be limited to songs that reached the Billboard top 10, but their biggest song ever, “Get Lucky” is missing. For artists where there are 10 songs, issues seem to persist. For example, “Billie Jean” is designated one of Michael Jackson’s top 10 least popular songs, when it is definitely his most popular song. For Earth Wind & Fire, “Fantasy” is ranked as both a least popular song and a most popular song for the same attribute. Overall, I think this searching idea is great, but might need some double-checking for functionality. Lastly, I would have `fct_reorder()`’d these graphs for aesthetics.

The second GOV 50 final project I took a look at is called “The Language of Emojis” by Linda Qin. The data set is undoubtedly smaller than the previous project as it is based on 45 responses from Harvard students (many of which are from GOV 1005). The author made some thoughtful considerations. For example, much of the data was collected around the time of Harvard’s COVID-19 evacuation. Therefore, there was a concern that emoji use might have been heavily affected by the sudden removal of Harvard students. I think this is a great example of taking a real-world event and narrowing the data set accordingly. To keep the data more objective, the author excluded all responses from late March onward, which I think is a good call. Besides the great foresight, I think this project also does a nice job with presenting a variety of graphs, modeling emoji use with concentrations, gender, secondaries, and more. I also think this project does a better job with aesthetics, as all of the graphs are easily readable. That being said, I do think there is room for improvement with “The Emoji Network” section. The author made a network where one can drag the vertices to see how use of a certain emoji might be linked to use of another. However, I think the graph is hard to read, and once you try to drag one vertex to another, it is impossible to see label of the other vertices. Therefore, it is hard to know what you’re comparing your emoji with. I am unfamiliar with the arguments and parameters with this type of graph, but I think I would try to make this graphic much more navigable.