My Website:

https://morganntaylor.github.io/dh-topic-models/

I chose to look at articles that pertained to Health from 1972-2012. I first selected to look for items particularly relating to nutrition within the Environmental Health Perspectives Journal. When we first started this, I learned that you could not put in keywords since we are using topic modeling. The results I received were different than I expected, but they still fell along the same guidelines of public health.

When I first got my journals, I received way too many. In fact, there are so many journals that the R program crashed on me numerous times during class. I was able to narrow down the number of corpuses during class to 1,000. That way my computer was able to hand the process better. I was told that I have an older Mac Air, which could also have slowed the process down too. Our instructor for the day rewrote a code for me and was able to select 1,000 journals for me to go through and topic model. I chose to do 25 topics because I felt like anything more was too repetitive, and anything less seemed to miss important topics.

Well, I think that the first thing that I discovered by doing topic modeling on the journals from JSTOR, was that I was actually getting more environmental factors than I had originally hoped for. I wanted to stay more on the side of Nutrition since that is my career interest. However, I also am aware of public policy and environmental health so the topics made sense. I do think that a possible argument for these topics is figuring out how bad external factors play within a human being. Wether that is the amount of allergens in the air, or the amount of pollution people may inhale from living in a large city, it all poses the same

question about wether or not it is healthy. The 25 topics that I found all surround the big question and that is what is the environment that we live in doing to our health. We can see this with some topics mentioning women's health, air quality, ocean pollution. All of these are external factors that are having an effect. The data and such on these topics can be found within these journals themselves. Another thing to note is that we can see trends through topic modeling. For my subject, we are able to see that the number of data points are high, meaning that experiments have a certain validity to them. Also, we can see that there are more than just one environmental factor that can contribute to our health. Finally, we can also see all of the history behind what people thought was safe and then what is not considered safe.

From the readings, I would think that the pros to topic modeling is that it is rather easy to do and that it is a unique way to present the data. Previously, we talked about tidy data, and for me, seeing the data neatly arranged in a graph form is the cleanest. I know that other authors suggested that in a chart is the tidiest, leaving one data set for each column and row. And that can be seen for the topic models. However, the way they are presented is more appealing, at least to me. Another positive about topic modeling is just the process and what we get out of doing it. Programs and code are able to see words and align them up, figuring out the word frequency and then using them to form a group, all labeled by a topic. It is pretty neat that coding and programing is able to turn a bunch of words into 25 perfect circles, containing a group of words found within 1,000 journals.

However, there are definitely some cons. For instance, there are some topics that are not completely repeated, but they are very similar. Also, some groups do not make complete sense. One of mine for example with the data points, seem like they would matter to someone doing research about previously research

done. But for me, it kind of skewed my topics a bit. Besides that, I have not read many other cons. It is definitely an interesting tool to use, saving people countless hours of going through 1,000 journals in order to make a connection between all of them.