Assignment 1

Team project: Topic Modeling and Sentiment Analysis using Text Data

1. We wanted to explore the relationship between health crises and the prevalence of fake news in media. We started by gathering online data of news articles and cleaned data according to "fakeness" and "realness"
2. Our hypotheses are:
    a. Health misinformation exhibits distinct thematic patterns, with certain words and topics recurring more frequently than others.
    b. Fake health information is disseminated through specific channels and platforms, contributing to its persistence and amplification.
3. Data is collected through online data organizations and stored using R and Python if needed.
4. Topic modeling and sentiment analysis were methods that stood out to us because they are intuitive and can be easily tracked alongside societal context. Lining up data with events would make sentiment analysis easier.
5. Kendall: Construction of research topic, reviewing literature to see what strategies have been employed in similar studies and comparing their results to ours
    a. Sharon: Molding research idea into a question, data collection from online organizations, cleaning data and sorting it based on "realness" and "fakeness"
    b. Mason: Molding research idea into question, topic modeling and sentiment analysis of text data using R, creating visuals of results