Grey's Anatomy Over the Years

By Melanie Ackerman

Abstract:

The medical drama, Grey's Anatomy, is on its 18th season and will reach its 400th episode by its conclusion in May of this year. Much has changed during that time, not only in medicine, but in what is considered socially acceptable to address on TV. The show has actually been a trail blazer in the non-medical topics addressed in its episodes, like its portrayal of LGBTQ relationships, or most recently, the unequal treatment of Black patients in health care. In this project I use natural language processing to analyze how the topics covered in Grey's Anatomy have changed over the years. Unfortunately, the aforementioned topics were too specific to be picked up by this model, but larger events (i.e. Covid) were prevalent enough to be included.

Design:

This project plots the changing focus of Grey's Anatomy episodes over the show's 18 seasons. The text was scraped from the fan website, https://greysanatomy.fandom.com/wiki/Grey%27s_Anatomy_Episodes, which includes full episode summaries for nearly every episode that has aired (a handful of episodes were dropped from the analysis as they were missing full summaries). I conducted text pre-processing and used TFIDF word vectors and a NMF model to produce 13 topics. With my knowledge of the show, I defined these categories and mapped them over time.

Data Description:

I scraped the fan website, https://greysanatomy.fandom.com/wiki/Grey%27s_Anatomy_Episodes , for episode summaries. As opposed to full episode transcripts, using summaries in theory provide enough information for topic modeling with about a third of the words included in a full transcript.

Algorithms:

I use TFIDF word vectors and an NMF model that produced 13 topics. I set max_df=0.5 and min_df=.01 to avoid common and extremely rare words from confusing the topic categories.

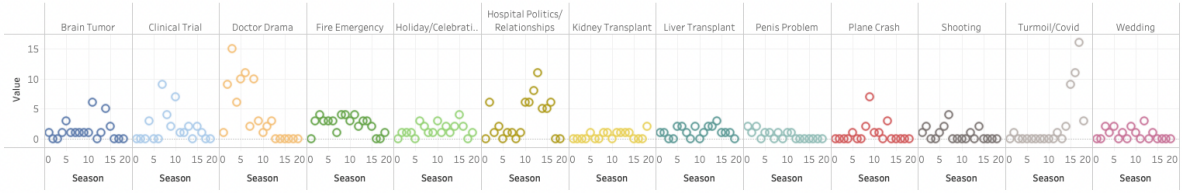
Tools:

Python reference, BeautifulSoup for web-scraping
NumPy and Pandas for data manipulation
Scikit-learn for modeling
Matplotlib and Tableau for visualization

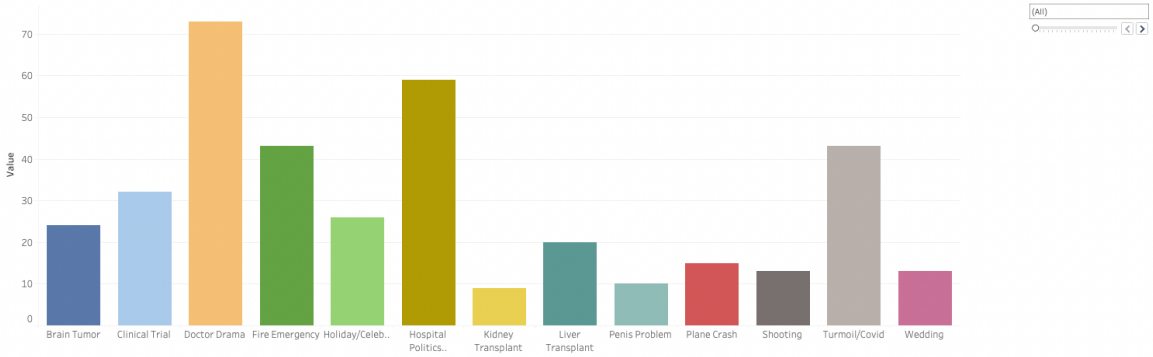
Communication:

I produced slides containing the following visuals (including a video of the interactive charts).

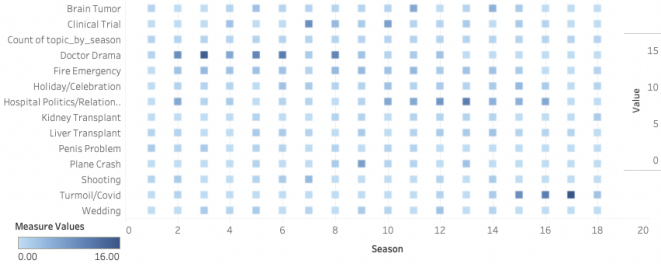
Topics Over Time



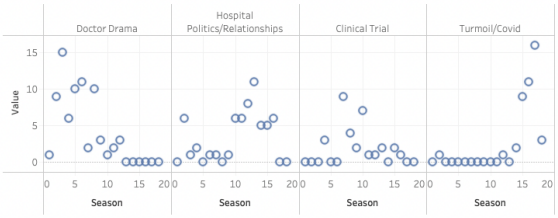
Topics by Season



All Topics and Seasons



Select Topics Over Time



Topics Overall