Measuring Anti-China Sentiment using the Wesleyan Media Project data

Video to audio to text Conversion:

Using python, we convert every video file (each video file consists of one political advertisement) to audio file, from ".wmv" to ".wav" format using "ffmpeg" and "moviepy" which are video to audio converters that are installable as packages in python. Each audio file is then transcribed to a text output in python. The transcription process is done using IBM's Watson Speech to Text API, a machine learning tool that transcribes audio to text. Further documentation about the IBM speech to text service can be reached at: https://cloud.ibm.com/docs/speech-to-text?topic=speech-to-text-about.

Each raw text output is saved under the variable "text" to the matching video advertisement in the Wesleyan Media Project data.

Variables

Note: "average" is used for most of these variables because we divide the variable by the number of observations (advertisements) for each dma in order to account for the variation in the number of observations (advertisements) per dma.

avg_AFINN_negcount_2006, avg_AFINN_negcount_2010, avg_AFINN_negcount_2012, avg_AFINN_negcount_2014, avg_AFINN_negcount_2016, avg_AFINN_negcount_2018:
Average number of AFINN dictionary's negative words mentioned in each advertisement per dma for the indicated year.

avg_HuLiu_negcount_2006, avg_AFINN_negcount_2010, avg_AFINN_negcount_2012, avg_AFINN_negcount_2014, avg_AFINN_negcount_2016, avg_AFINN_negcount_2018:
Average number of Hu&Liu dictionary's negative words mentioned in each advertisement per dma for the indicated year.

avg_count_neg_emotion_2006, avg_count_neg_emotion_2010, avg_count_neg_emotion_2012, avg_count_neg_emotion_2014, avg_count_neg_emotion_2016, avg_count_neg_emotion_2018:
Average number of negative emotions shown in each advertisement per dma for the indicated year. The count of negative emotions were calculated by adding up the three variables, "fear", "anger", and "sadness" from the Wesleyan Media Project Data. According to the Wesleyan Media Project Data codebook, the three variables "fear", "anger", and "sadness" are coded as follows:

"Please assess the intent of the ad maker (not your own personal reaction) below. Does the ad make an appeal to the following emotions?
0 No
1 Some appeal
2 Strong appeal
***fear**: Does the ad make an appeal to fear?*
***anger**: Does the ad make an appeal to fear?*
***sadness**: Does the ad make an appeal to fear?"*

Since each ad has a score of 0, 1, or 2 for each of the variables "fear," "anger", and "sadness," the "**avg_count_neg_emotion_2006"** variable counts as 1 if only one of the variables indicates as 1 or 2, 2 if two of the variables indicates as 1 or 2, and 3 if three of the variables indicates as 1 or 2.

avg_AFINN_2006, avg_AFINN_2010, avg_AFINN_2012, avg_AFINN_2014, avg_AFINN_2016, avg_AFINN_2018:
Average sentiment score using the AFINN dictionary for each advertisement per dma for the indicated year. We use the difference metric for the sentiment score, which adds up the sentiment value of the words that are mentioned in each ads. The sentiment value are pre-coded by the dictionary.

avg_HuLiu_2006, avg_HuLiu_2010, avg_HuLiu_2012, avg_HuLiu_2014, avg_HuLiu_2016, avg_HuLiu_2018:
Average sentiment score using the HuLiu dictionary for each advertisement per dma for the indicated year. We use the difference metric for the sentiment score, which adds up the sentiment value of the words that are mentioned in each ads. The sentiment value are pre-coded by the dictionary.

count_china_2006, count_china_2010, count_china_2012, count_china_2014, count_china_2016, count_china_2018:
Number of ads that are coded as 1 for "issue 65" per dma for the indicated year. According to the Wesleyan Media Project Data codebook, "issue 65" is coded as follows:

"Please assess the intent of the ad maker (not your own personal reaction) below. Does the ad make an appeal to the following emotions?
0 No
1 Some appeal
2 Strong appeal

issue65: *China"*