Data Appendix:

episode1_clean

- This is a cleaned data file of Episode 1, The Phantom Menace, movie script
- The cleaning code turns the original SW01_txt data file into one column containing every word in the script, cleaned of any punctuation (excluding spaces), multiple spaces, capitalization, numbers, and screenplay markers
    - Each new line in the original script was replaced with a space in order to put the entire new cleaned script into one column that we could run our analysis on
    - Unit of Observation: text from script of Episode 1
- On this cleaned file we performed both the VADER sentiment analysis and NRC lexicon analysis

episode2_clean

- This is a cleaned data file of Episode 2, Attack of The Clones, movie script
- The cleaning code turns the original SW02_txt data file into one column containing every word in the script, cleaned of any punctuation (excluding spaces), multiple spaces, capitalization, numbers, and screenplay markers
    - Each new line in the original script was replaced with a space in order to put the entire new cleaned script into one column that we could run our analysis on
    - Unit of Observation: text from script of Episode 2
- On this cleaned file we performed both the VADER sentiment analysis and NRC lexicon analysis

episode3_clean

- This is a cleaned data file of Episode 3, Revenge of the Sith, movie script
- The cleaning code turns the original SW_txt data file into one column containing every word in the script, cleaned of any punctuation (excluding spaces), multiple spaces, capitalization, numbers, and screenplay markers
    - Each new line in the original script was replaced with a space in order to put the entire new cleaned script into one column that we could run our analysis on
    - Unit of Observation: text from script of Episode 3
- On this cleaned file we performed both the VADER sentiment analysis and NRC lexicon analysis

episode4_clean

- This is a cleaned data file of Episode 4, A New Hope, movie script
- The cleaning code turns the original SW04_txt data file into one column containing every word in the script, cleaned of any punctuation (excluding spaces), multiple spaces, capitalization, numbers, and screenplay markers
    - Each new line in the original script was replaced with a space in order to put the entire new cleaned script into one column that we could run our analysis on
    - Unit of Observation: text from script of Episode 4
- On this cleaned file we performed both the VADER sentiment analysis and NRC lexicon analysis

episode5_clean

- This is a cleaned data file of Episode 5, The Empire Strikes Back, movie script

- The cleaning code turns the original SW05_txt data file into one column containing every word in the script, cleaned of any punctuation (excluding spaces), multiple spaces, capitalization, numbers, and screenplay markers
    - Each new line in the original script was replaced with a space in order to put the entire new cleaned script into one column that we could run our analysis on
    - Unit of Observation: text from script of Episode 5
- On this cleaned file we performed both the VADER sentiment analysis and NRC lexicon analysis

episode6_clean

- This is a cleaned data file of Episode 6, The Return of the Jedi, movie script
- The cleaning code turns the original SW06_txt data file into one column containing every word in the script, cleaned of any punctuation (excluding spaces), multiple spaces, capitalization, numbers, and screenplay markers
    - Each new line in the original script was replaced with a space in order to put the entire new cleaned script into one column that we could run our analysis on
    - Unit of Observation: text from script of Episode 6
- On this cleaned file we performed both the VADER sentiment analysis and NRC lexicon analysis

episode7_clean

- This is a cleaned data file of Episode 7, The Force Awakens, movie script
- The cleaning code turns the original episode7.txt data file into one column containing every word in the script, cleaned of any punctuation (excluding spaces), multiple spaces, capitalization, numbers, and screenplay markers
    - Each new line in the original script was replaced with a space in order to put the entire new cleaned script into one column that we could run our analysis on
    - Unit of Observation: text from script of Episode 7
- On this cleaned file we performed both the VADER sentiment analysis and NRC lexicon analysis