# Sentiment analysis on Game of Thrones

**Names:** Laura Bundgaard Jørgensen, Amanda Mondrup, Sarah Astrup, Cecilie Hvenegaard Jensen

## **Date: 12/5-2025**

## **URL to Github code** (*where is the code for your visual)*: <https://github.com/Digital-Methods-HASS/AU754423_Amanda_Mondrup> - **The file is in the folder “Portfolie”, and it is called “SentimentAnalysis****”**

## **Visualisation:** (*include a legible and meaningful graphic with clear labels and no scientific notation*)

## Et billede, der indeholder skærmbillede, diagram, Farverigt, linje/række Indhold genereret af kunstig intelligens kan være forkert.

## **Significance**: *Explain in 250 words what your visualisation (sentiment chart, map, or colonial legacy timeseries or regression chart) represents and what are its implications vis-a-vis your [historical] research question? Also, consider briefly what your visual does not show.*

The visualisation is a sentiment chart from doing a sentiment analysis on the book “Game of Thrones: A Song of Ice and Fire”. It represents the most common words from the book, after removing stopwords such as “the” “and” and so on with the tidyr (Wickham H, Vaughan D, Girlich M, 2024) antijoin function. These common words are sorted into different sentiment categories based on the pre-built NRC lexicon (Dr. Saif M. Mohammad, Dr. Peter Turney). The NRC lexicon puts the words in the sentiment categories that they are most associated with. It also shows how many times each word is used.

This shows us the most common words from the book, such as lord, time, bran, king etc. It also shows us which sentiment(s) these words are associated with based on the NRC lexicon.

One word can appear in several categories. For example, we can see that the NRC lexicon associates the word lord with four different sentiments: disgust, positive, negative and trust. This means, that we should be careful when using a sentiment analysis to see whether a text is mostly positive or negative, as a word can be associated with both positive and negative sentiments, and thus possibly skew the chart.

As the word lord has such a high frequency, it can also be harder to properly sense the frequency of the other common words, as lord changes the scale so dramatically. (from 100’s to 1000)

We also need to be aware that when using a prebuilt lexicon and removing stopwords, we also risk removing words that could be relevant, and we should check which words are excluded.