# Narrative

*Casey Mallon*

*12/10/2019*

## Substantive background

NGOs like Amnesty International serve to provide information to people, states, and international organizations to advocate for a continued pledge to upholding human rights. The organization targets both state and nonstate actors aiming to promote human rights globally including, but not limited to: abolishing the death penalty, women's rights, ending torture, protection of human dignity, and the rights of refugees and asylum-seekers. This last goal is particularly relevant for the international stage today, with the world still reeling from the 2015 Migrant Crisis. Recent conservative political movements have formed as a result of the sudden shock recent migration issues have placed on the international community. Members of this coalition have drawn a suspicious eye toward NGOs, casting doubt on the reliability of their information and the relevancy of their goals.

Is Amnesty International biased in its reporting on the treatment of asylum-seekers? This project seeks to answer this question through performing regression analysis on the text of Amnesty International's articles. I hypothesize that there will be a negative relationship between particular regions and the sentiment score of an article. That is to say that Amnesty International will exhibit bias toward particular regions by using more negative diction in their articles on asylum-seekers.

## Collecting data

Amnesty International allows users to search through all of its products by product type and by keyword. I chose to filter my analysis to articles covering Asylum-seekers. Obtaining the links for all of the articles proved to be more challenging than initially expected. The nodes for the links to each article on each search page were not easily accessed through the Selector Gadget. This meant that I had to sort through using Google Chrome's "Inspect" feature for the page URLs.

Actually scraping each article URL presented similar challenges up front.T he Selector Gadget would not find the correct nodes and manually inspecting the data revealed that there were layerd nodes, so it was often hard to find the correct node that would work across articles. Even still, there were two articles that had completely different nodes than most of the data, so they were omitted from the analysis by the addition of lines 37-40 in Code 01. Eventually, I was able to get the title, date, text, and region for nearly all of the articles on the search page and create an initial dataframe to work with.

## Cleaning / pre-processing data

From there, I moved onto removing any articles for which not all of the information was available. Thanks to my screening in the collection phase, I only lost one object in this phase, going from 160 observations to 159. In the next stage I filtered out articles that had improperly coded titles; these were articles in which the nodes were incorrectly named by Amnesty International. After this, the data was reformatted to have proper column names including transforming the publication date into just the year the article was published. Additionally, the several categories of the "Region" column were renamed to reflect actual geographic region. Amnesty International occasionally chose to mark an article's "Region" as the country the article was focused on. I recoded most of these objects to be a part of already existing regions except for Afghanistan, Turkey, and the Russian Federation, which either did not neatly fit into the existing regions or could easily stand on their own given their relevance to asylum politics and/or world politics. A second csv file was then constructed from this tidied data.

## Analysis and visualization

Once the data was in a workable format, I was able to use the "Text" column to perform text analysis on each article. I first began by transforming the data into a Document Term Matrix then obtaining the sentiment score (1 or -1) for each word in each article according to the "bing" dictionary. From there I was able to create a wordcloud to visualize the words characterizing most of the articles. I was also able to create a line chart to show how the sentiment score for each region changed over time (2015-2019). I then created a series of column plots to show the breakdown of scores by region per year. Finally, I produced three models to examine the relationship between region, date, and sentiment scores. My regression results yielded no significant relationship between the factors. Thus, I have found no evidence to support that Amnesty International's coverage on asylum-seekers is unfairly negatively biased toward any particular region. However, it is interesting to note the coverage is overwhelmingly negative as a whole across regions and years.

## Future work

Future work could expand upon my code to incorporate all of the search categories and documents provided by Asylum International. While my analysis did not find evidence to support that Amnesty International's coverage could be biased, I only focused on one specific search result. An analysis encompassing multiple search categories and treating them as factors may find support for my initial hypothesis after all.