**Data Exemplar**

**Exemplar 1 – US News Articles**

**Data collected by:** Erin M. Buchanan &William E. Padfield

This dataset consists of gathered text from four notable US news sources compiled into a spreadsheet for further processing and analysis. The sources included in this research include: *The New York Times*, *National Public Radio (NPR)*, *Fox News*, and *Breitbart*. The researchers specifically scraped political news coverage and commentary from the denoted politics pages on each website, as political coverage was suspected to include discussion of moral topics. The dataset includes the website link for each article (**Url**), examples of the processed data (**Processed**), and word counts from the overall article (**WordCount**), and each of the Moral Foundations areas created by Graham, Haidt and Nosek (2009; **HarmSum, FairSum, IngroupSum, AuthoritySum, and PuritySum**).

**Moral Foundations Dictionary**

The Moral Foundations Dictionary includes lists of words that are associated with each of the moral foundation areas. To create the **Sum** columns described in the exemplar dataset, you would need to know the list of words for foundation area. The table below includes the words used for each area in their unstemmed (i.e., full word) format. In the analysis, you would process these words like the text sources, so that you could match the root word stems in the processed text to the word stems for the dictionary. For example, you would count the number of times that *abus* appeared in the text source for the Harm category, as *abus* is the processed text for *abuse*. The number of times these words appeared would be included in the **Sum** columns in the dataset, and you would create a sum of the number of times all the words from each category appeared.

Table 1. *Moral Foundations Dictionary Word Lists*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Harm | Fair | Ingroup | Authority | Purity |
| abuse | balance | collect | abide | abstain |
| attack | bias | community | authority | adultery |
| benefit | constant | deceive | class | church |
| care | discriminate | desert | command | clean |
| cruel | equal | family | control | dirt |
| crush | equate | fellow | defect | disease |
| damage | even | foreign | defer | disgust |
| danger | exclude | group | defy | gross |
| defend | fair | individual | desert | innocent |
| destroy | favor | member | duty | modest |
| fight | honest | nation | faith | preserve |
| guard | impartial | side | father | promiscuous |
| harm | just | together | honor | pure |
| hurt | justify | trait | law | right |
| kill | prefer | unite | lead | ruin |
| preserve | prejudice |  | legal | sacred |
| protect | reason |  | mother | sick |
| ruin | right |  | obey | sin |
| safe | tolerate |  | oppose | trash |
| shelter |  |  | order | whole |
| spurn |  |  | permit |  |
| stomp |  |  | position |  |
| suffer |  |  | preserve |  |
| sympathy |  |  | protest |  |
| violent |  |  | refuse |  |
| war |  |  | respect |  |
|  |  |  | revere |  |
|  |  |  | serve |  |
|  |  |  | tradition |  |
|  |  |  | trait |  |

**Exemplar 2 - Immigration and Political Party**

**Data collected by:** William E. Padfield

The dataset includes thirty congressional speeches covering the nature of immigration given to the House and Senate in more recent (2017 and later) months. These speeches were coded by political party, Republican or Democrat, to allow for a similar moral analysis as described above, as political lean is expected to influence speech and writing patterns.

This dataset would allow you to apply Moral Foundations Theory to current political trending topics to determine if there are differences in moral speech across political party. Immigration has always been a contested topic and with the current political climate, this topic appeared to be an apt area to explore speech and discourse patterns for distinctions between Republican and Democratic Congress people. Given the research on Moral Foundations Theory, we might expect to find that Democratic speakers to use more individualising foundations of *harm/care* and *fairness/reciprocity*, while the Republican speakers to use more *binding* foundations such as *ingroup/loyalty*, *authority/respect,* and *purity/sanctity*.

The dataset provided gives you an opportunity to try a word frequency analysis to determine if these differences exist between parties. The **Source** column includes the political party of the original speaker, and the **Url**column includes a link to the Library of Congress website where the data was found. The **Text** column is the unprocessed data from the Library of Congress that you would use to start your analysis.