**INTRODUCTION**

We are interested in finding the differences in speech between high peace and low peace countries. We have expanded the work from the previous teams on two more robust datasets, **TWITTER** and **NEWS**, and several different metrics with more diverse approaches.

**COUNTRY SELECTION**

To define High-Peace and Low-Peace Countries, we referred to 5 indexes: GPI, PPI, HDI, WHI, and FSI for 137 countries. We define our peace index using PCA and K-means, and chose 16 countries for the analysis.

<table>
<thead>
<tr>
<th>HIGH PEACE</th>
<th>LOW PEACE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada, UK, Finland, Norway, Ireland, France, Australia, Singapore</td>
<td>India, Iran, Nigeria, Uganda, Gambia, Libya, Pakistan, Zimbabwe</td>
</tr>
</tbody>
</table>

**METHOD**

**TWITTER**

- Remove links, country names, convert contractions, emojis, apply lemmatization
- Get word clusters using GloVe embedding and find word topics using KMeans
- Apply BERT Model (BertForSequence Classification and BERTweet)

**NEWS**

- Remove links, country names, email, numbers, punctuations, convert contractions, apply lemmatization
- Get word clusters using GloVe embedding and find word topics using KMeans
- Classify top words for peace and non-peace group based document types with the highest word occurrence
- Extract feature importance using model coefficients/attention scores

**RESULT**

Words used in high peace countries are more diverse compared with those in low peace countries

- Keywords have different interests in topics
  - High and low peace countries use different sets of words focusing on different topics
  - High peace countries talk about more topics than low peace countries
  - High peace countries discuss more about topics in daily lives, such as festivals, films, and feelings, whereas low peace countries focus more on political subjects