COMP40020: Human Language Technologies  
Assignment 1

* Investigating language styles from data on Reddit
* Build a corpus from one or more subreddits
* Pre-processing steps that should be used.
* Can you infer the subject of a subreddit from its most common words
* Common n grams capture popular jokes
* Text lexically diverse

**Research Question:**

Is New Zealand really that much more happy than Ireland? A sentiment and lexical analysis of comments of r/NewZealand and r/Ireland

**Methods**

**Results**

**Discussion**

**Bibliography**