

Research Questions

- How does stress levels differ among subreddits?
- How can we predict stress level given words in the data and lexical features?
- Is there an association between stress-related data and subreddits?

Dataset Overview

- Dataset consists of training data with 2838 posts, test data with 715 posts
- Sampled from 11 subreddits

Getting the Labels Distribution

```
label_counts <- reddit_stress_data %>%  
  group_by(label) %>%  
  count()  
plot_ly(label_counts, x = ~label, y = ~n, type = "bar")
```

