

Title Phrase Comparison in the Politics Subreddit

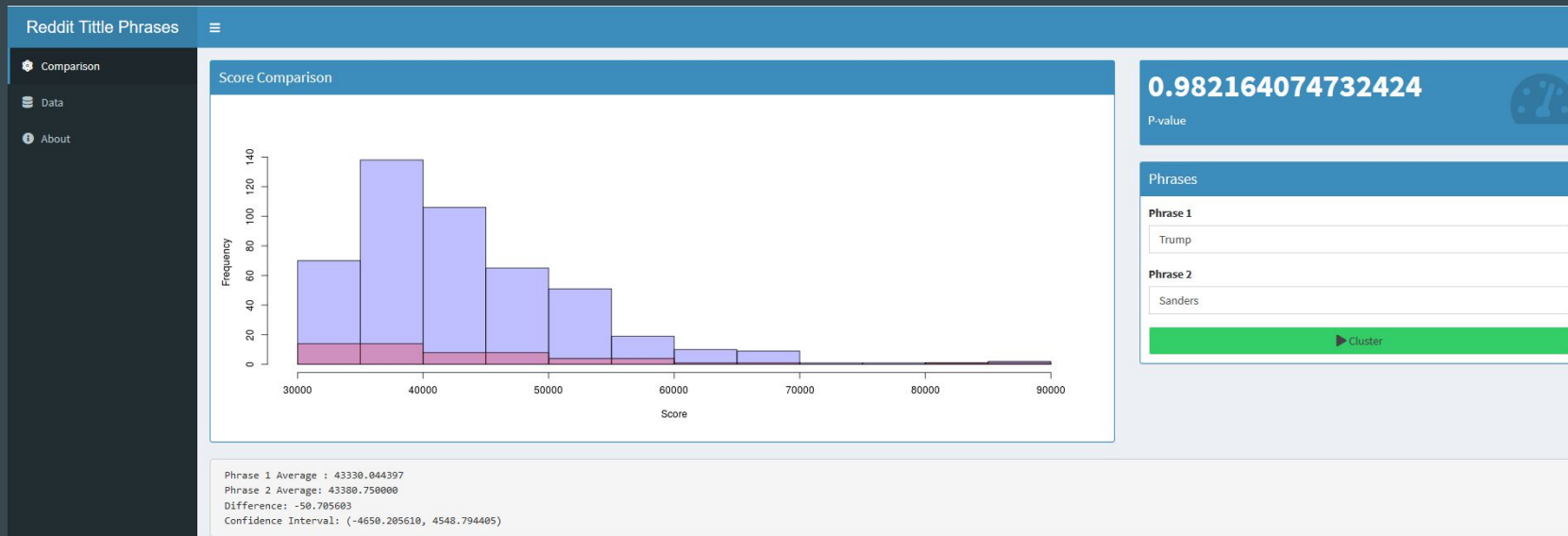
...

Cole Sanders

Summary

- Data is collected using the praw library and is collected from the top post from the politics subreddit
- We want to compare scores for different phrases that can be used in titles
- The model is a simple t-test comparison to see what the average difference between scores for each phrase
- Histograms for each score are plotted together to visualize spread of data
- Shiny App deployed to ShinyApps.io at https://cole-stats.shinyapps.io/Reddit-Tittle_Phrases/
- Github <https://github.com/cgsanders/STAT418-Final>

Sample Screen Shot



Data

X	Type	author	date	domain	glited	M	ratio	num_comments	num_controversy	score	title	upvote_ratio	url
0	Top	OrangePinehead	2018-11-07 01:00:55	berkley.com	1	0.0000	0.000	2771	15	101980	Vin DeLu, clerk who refused to sign marriage licenses for gay couples, leads to Democrat	0.83	https://www.berkley.com/news/politics/government/jarvisde1111145.html
1	Top	Bloodbath McGrath	2018-11-10 23:20:44	washingtonpost.com	4	Symbolic	False	8518	6	80624	Ivanka Trump used a personal email account to send hundreds of emails about government business last year	0.83	https://www.washingtonpost.com/politics/ivanka-trump-used-a-personal-email-account-to-send-hundreds-of-emails-about-government-business-last-year/2018/11/10/5C61a5f0-470c-11e8-b081-114742c0147c_story.html
2	Top	Eurymenides	2018-08-12 01:21:08	nytimes.com	30	aOfing	False	7079	11	87732	F.B.I. Opened inquiry into whether Trump files secretly working on behalf of Russia	0.83	https://www.nytimes.com/2018/08/12/us/politics/fbi-trump-russia-inquiry.html

About this app

This app looks at the top 25,000 post in the Politics subreddit on Reddit and allows you to compare phrases to see the average score they receive.

Details

We uses a separate process to join query Reddit's public API to collect titles and scores for the top submission on the politics subreddit. We use this process to collect, structure, and save the data. We then uses this data in our shiny app to allow users to compare different phrases to uses in their titles. A histogram overlaying the distribution of scores is shown for both.

P values

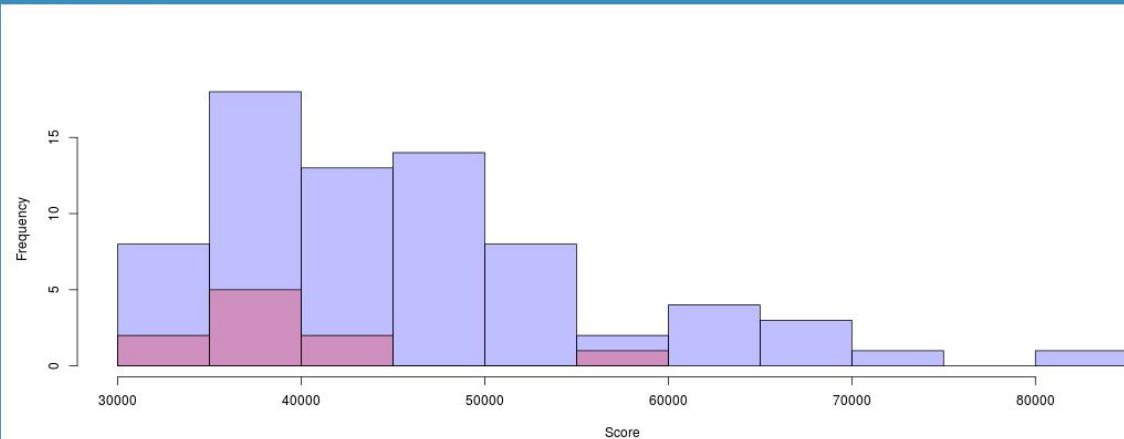
In the top left corner a P-value is shown to allow you to gauge the statistical significance of the difference. Below the graph estimates and confidence intervals are given.

Creator

Cole Sanders

Example 1: President vs Supreme Court

Score Comparison



Phrase 1 Average : 45776.069444
Phrase 2 Average: 39386.600000
Difference: 6389.469444
Confidence Interval: (837.719805, 11941.219084)

0.0268439706126743

P-value



Phrases

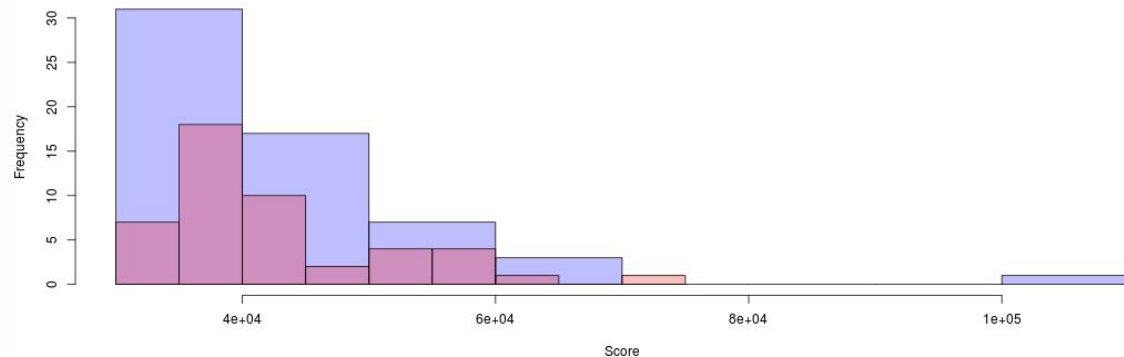
Phrase 1

Phrase 2

▶ Cluster

Example 2: Democrat vs Republican

Score Comparison



Phrase 1 Average : 42949.372881
Phrase 2 Average: 42355.702128
Difference: 593.670754
Confidence Interval: (-3301.528349, 4488.869856)

0.763074432459747

P-value



Phrases

Phrase 1

Democrat

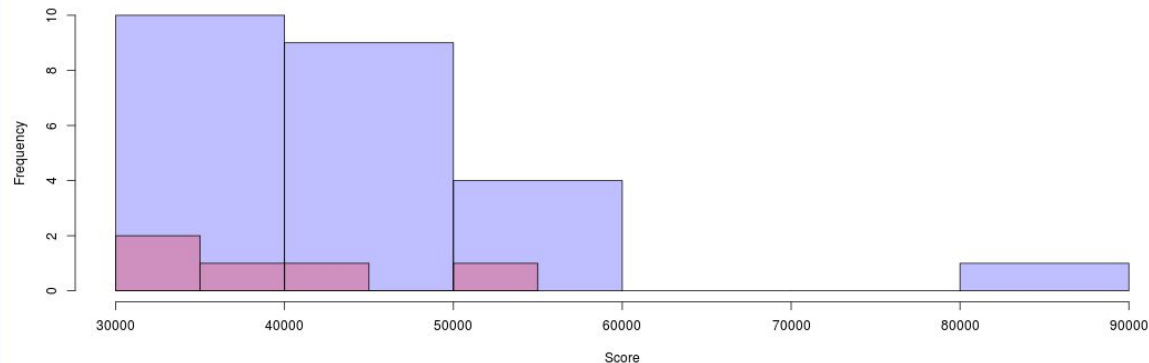
Phrase 2

Republican

Cluster

Example 3: McConnell vs Schumer

Score Comparison



Phrase 1 Average : 44106.708333
Phrase 2 Average: 40600.400000
Difference: 3506.308333
Confidence Interval: (-5271.515128, 12284.131794)

0.38701254890906

P-value



Phrases

Phrase 1

McConnell

Phrase 2

Schumer

▶ Cluster