

HW6-Text Mining

Richard Barad, Dave Drennan, Jarred Randall

2023-12-11

Contents

Setup	2
Data	2
Corpus	2
Frequency Word Cloud	2
Sentiment Analysis	3
Plotting NRC Lexicon Sentiment Analysis	3
Comparing Plots of Different Lexicons	3
Positive, Neutral, and Negative Sentiment Counts by Lexicon	7

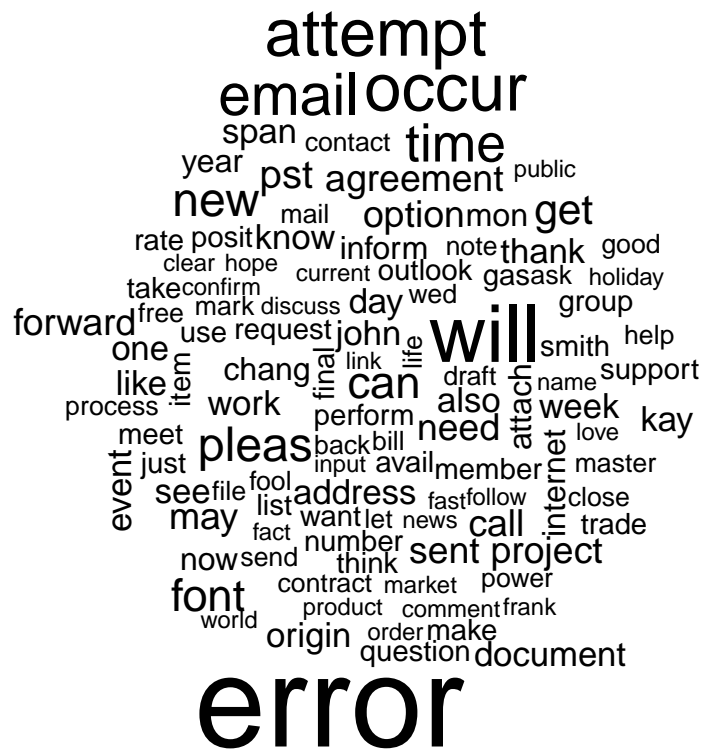
Setup

Data

Corpus

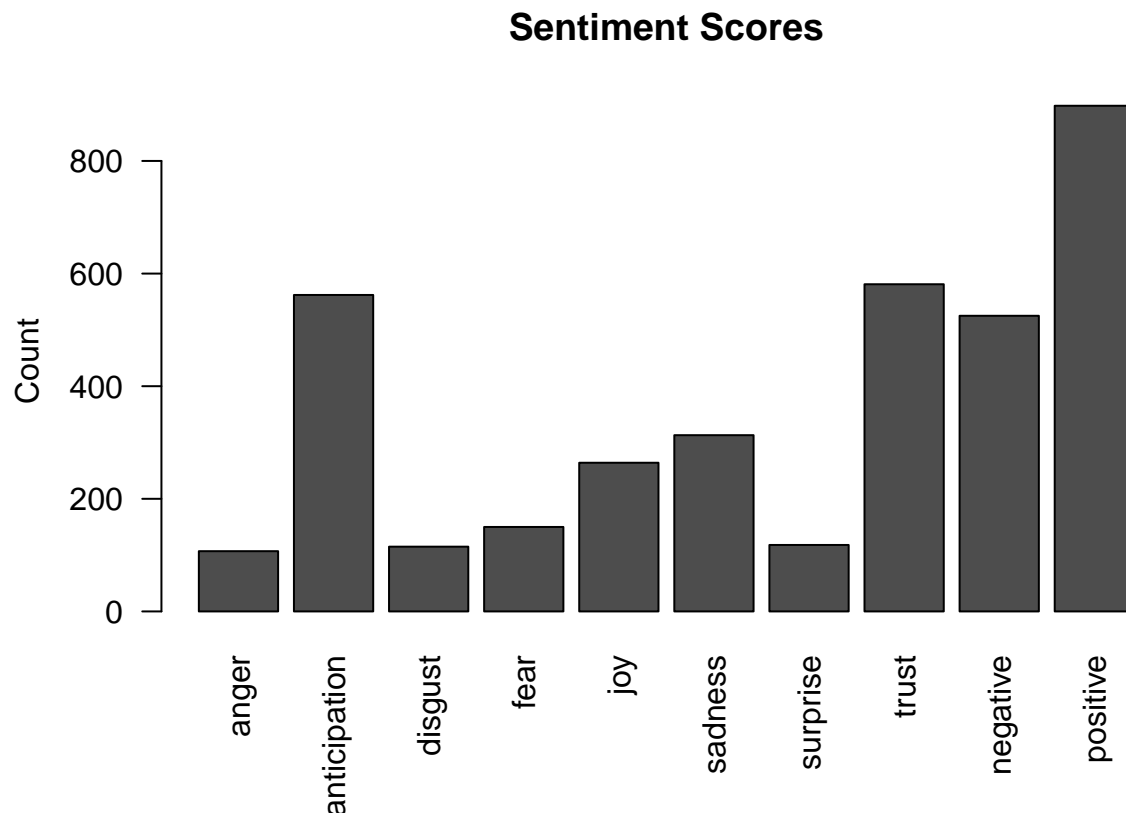
Corpus Text Cleaning

Frequency Word Cloud



Sentiment Analysis

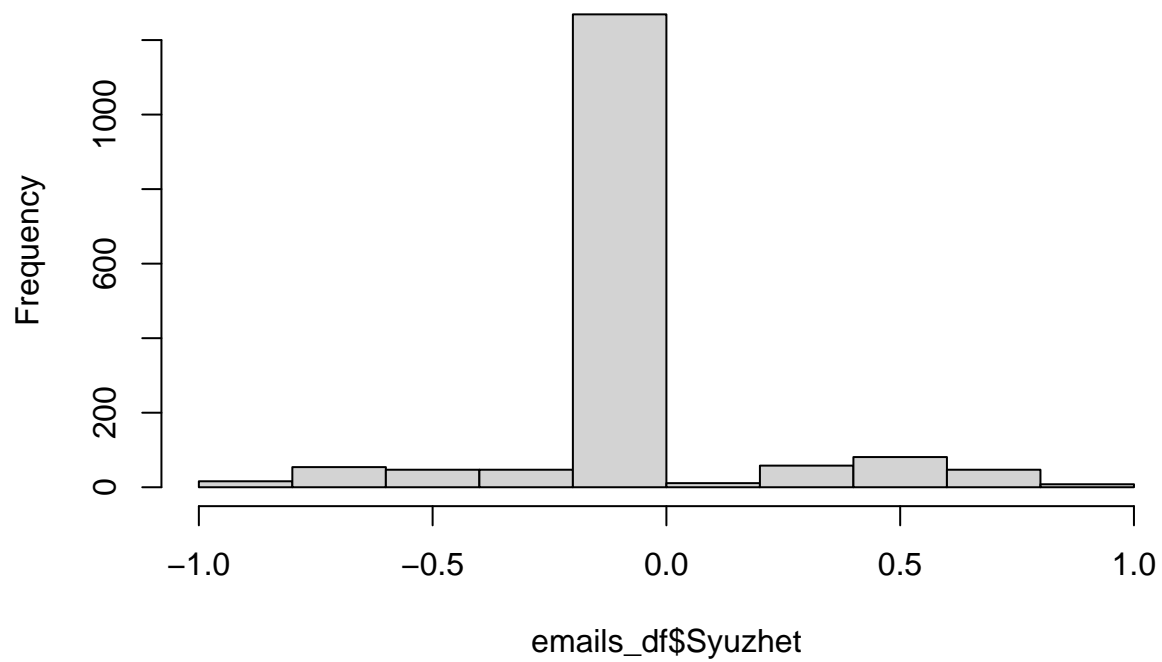
Plotting NRC Lexicon Sentiment Analysis

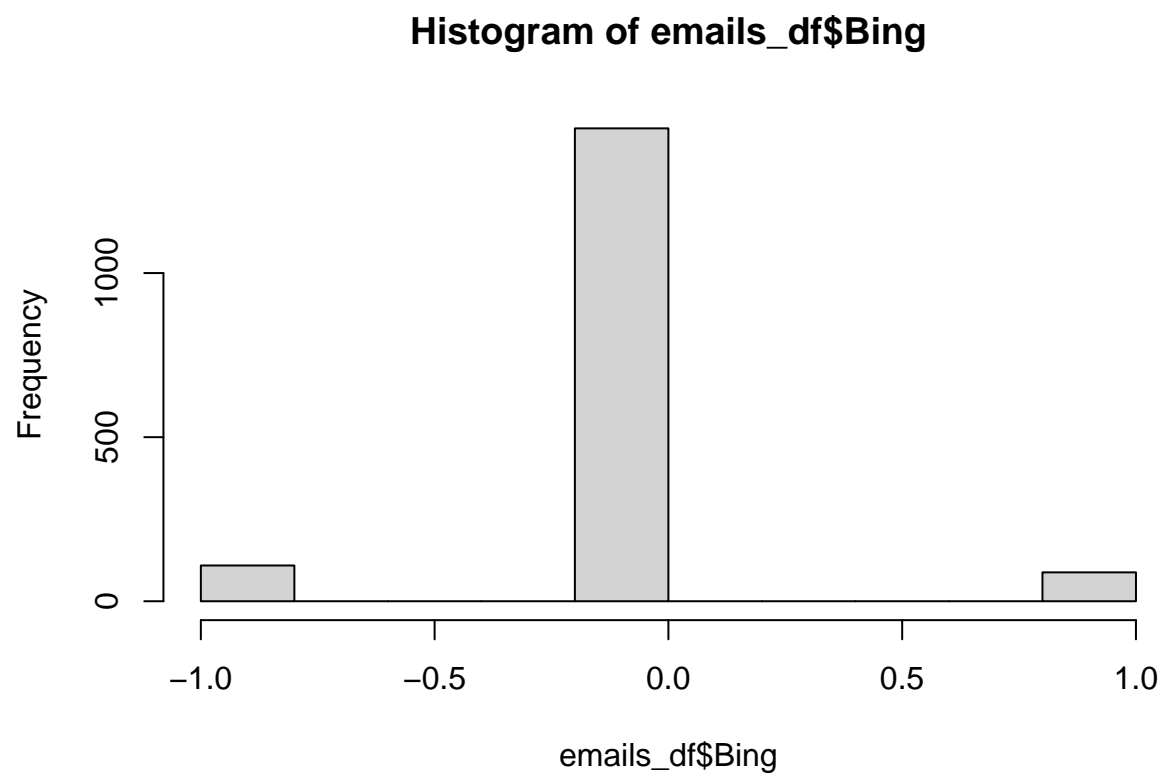


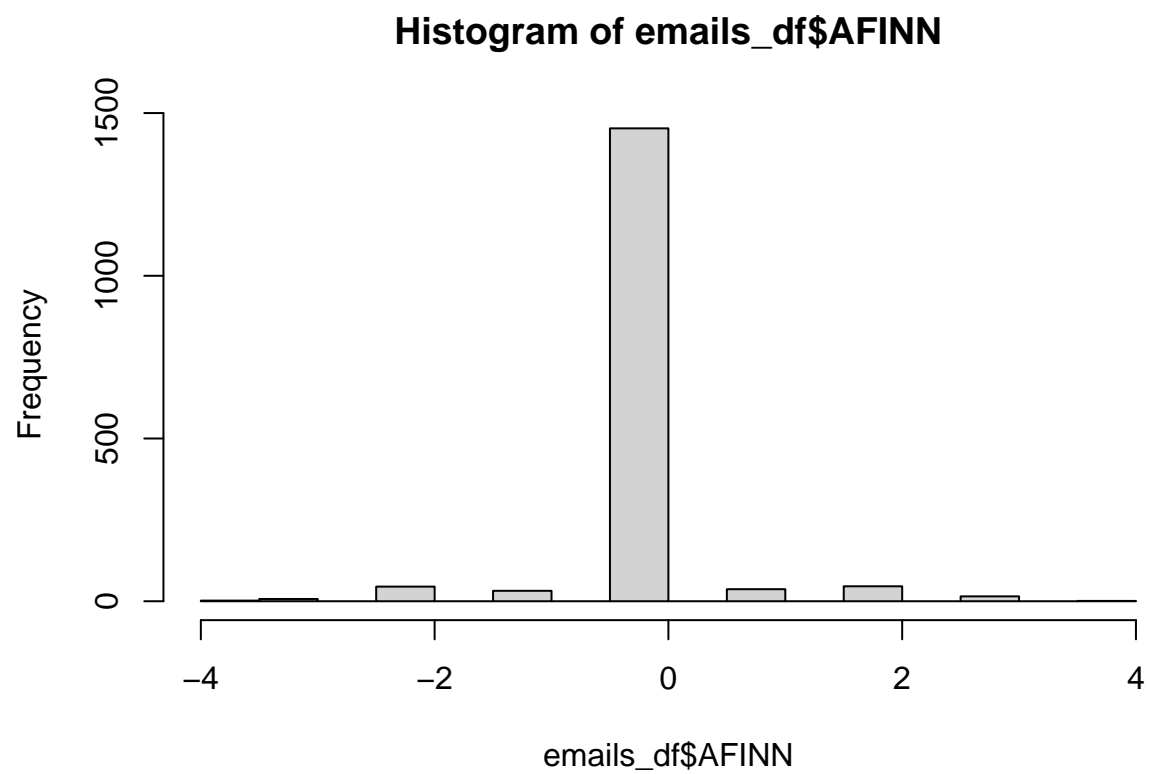
Comparing Plots of Different Lexicons

Uses NRC, AFINN, Bing, and Syuzhet lexicons

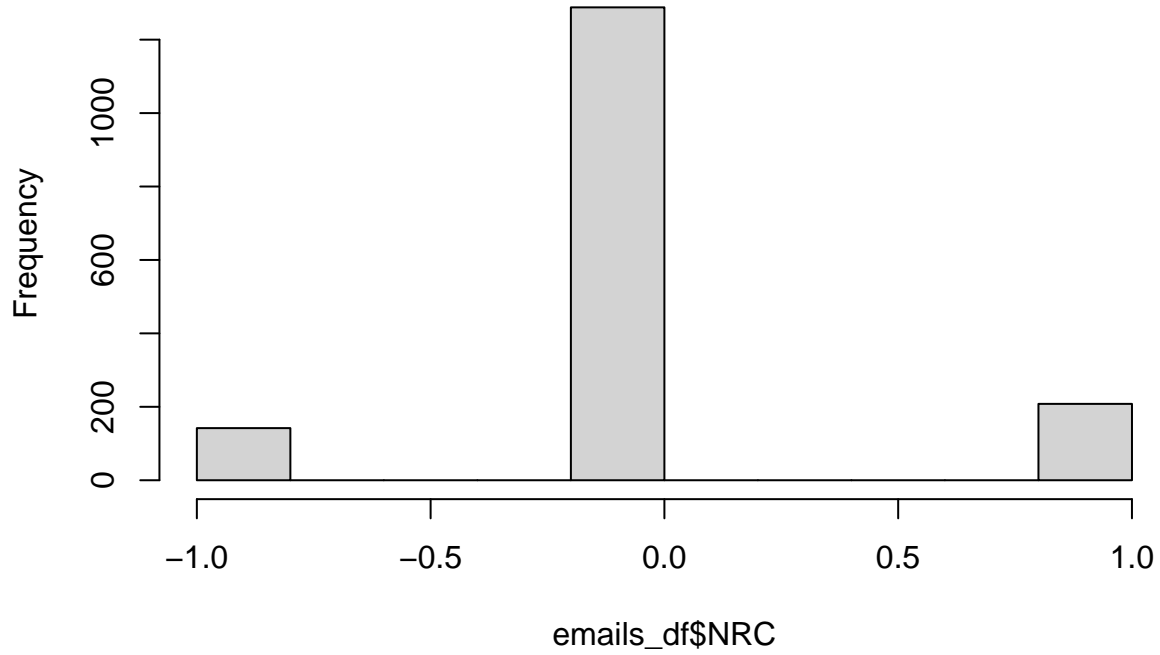
Histogram of emails_df\$Syuzhet







Histogram of emails_df\$NRC



Positive, Neutral, and Negative Sentiment Counts by Lexicon

	Syuzhet	Bing	AFINN	NRC
-1	169	109	86	142
0	1264	1441	1453	1288
1	205	88	99	208

	Syuzhet	Bing	AFINN	NRC
-1	0.1031746	0.0665446	0.0525031	0.0866911
0	0.7716728	0.8797314	0.8870574	0.7863248
1	0.1251526	0.0537241	0.0604396	0.1269841