

Home Page	The Fine Print	Summary of Subreddits	Summary of Words	Overall Sentiment Breakdown	Search by word or sentiment	Scoring Sentiment	Spread of Sentiment
-----------	----------------	-----------------------	------------------	-----------------------------	-----------------------------	-------------------	---------------------

Reddit Sentiment Analysis

Alexis Campestre
Applied Data Analytics - Fall 2019
Prof. John Chandler

Click Me! When you
get bored, or need
a break from
grading!



Home Page	The Fine Print	Summary of Subreddits	Summary of Words	Overall Sentiment Breakdown	Search by word or sentiment	Scoring Sentiment	Spread of Sentiment
-----------	----------------	-----------------------	------------------	-----------------------------	-----------------------------	-------------------	---------------------

The Fine Print

My Questions:

RQ1 – Is Reddit a negative place to spend time on the internet?

RQ2 – Are some subreddits more prone to negativity or positivity?

RQ3 – Is there a significant difference in the average “Reddit Score” of a post as a result of using words that elicit positivity or negativity?

My Approach:

Reddit API

Python Script – retrieve and preliminary cleaning of data

RStudio – manipulate data and analysis (general linear model used)

NRC NLP Lexicon – word sentiment dictionary

Tableau – visualize data

My Variables:

“Rank” – assigned rank of the top 1000 posts pulled

“Title” – the actual text of the title of the post

“Number of Words” – the number of words in the title of the post

“Reddit Score” – the calculated score on reddit.com by users either ‘up-voting’ or ‘down-voting’ a post

“Subreddit” – reddit.com is broken into topics or categories, referred to as ‘Subreddits’

“Num Comments” – total number of comments for a given post

“Negative” – points assigned to every ‘negative’ word in the title of a post

“Positive” – points assigned to every ‘positive’ word in the title of a post

“Total Score” – both negative and positive points added together

“Posminusneg” – calculated difference between total positive and negative points for a post

Limitations & Assumptions:

A Note About Posts

For unknown reasons, reddit.com would not allow me to retrieve more than 1000 posts at a given time.

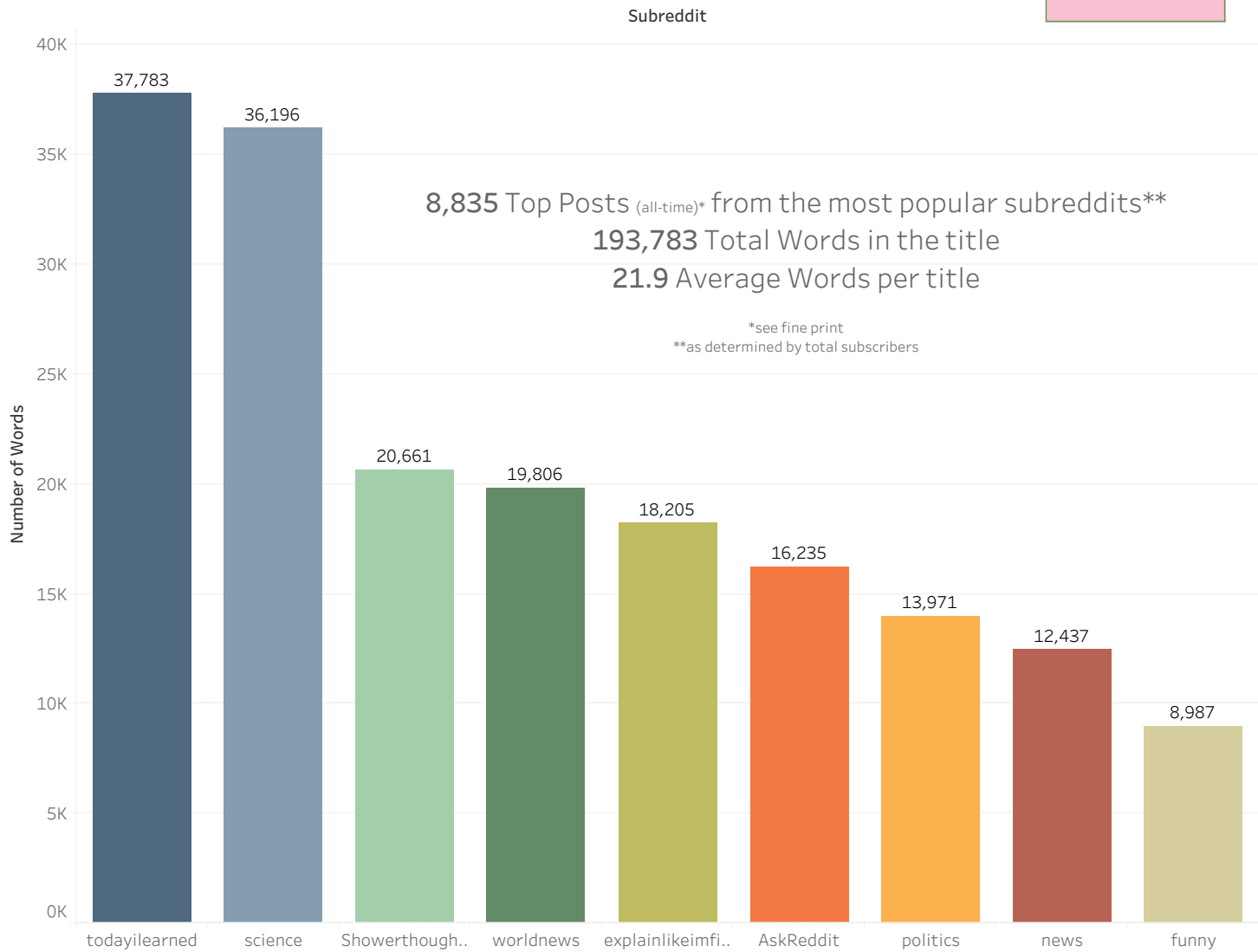
Additionally, no subreddit actually has 1000 total retrieved posts. The average is 981.67 posts/subreddit. All retrieved posts were retrieved on November 25, 2019.

Posts are filtered for “Top” with the time range of “All-Time”

The scoring algorithm for reddit.com has changed in recent years. This was not taken into consideration.

BY NO MEANS IS THIS A COMPLETE ANALYSIS OF ALL OF REDDIT

Total Words in each Subreddit



Average Number of Words by Subreddit, Ranked

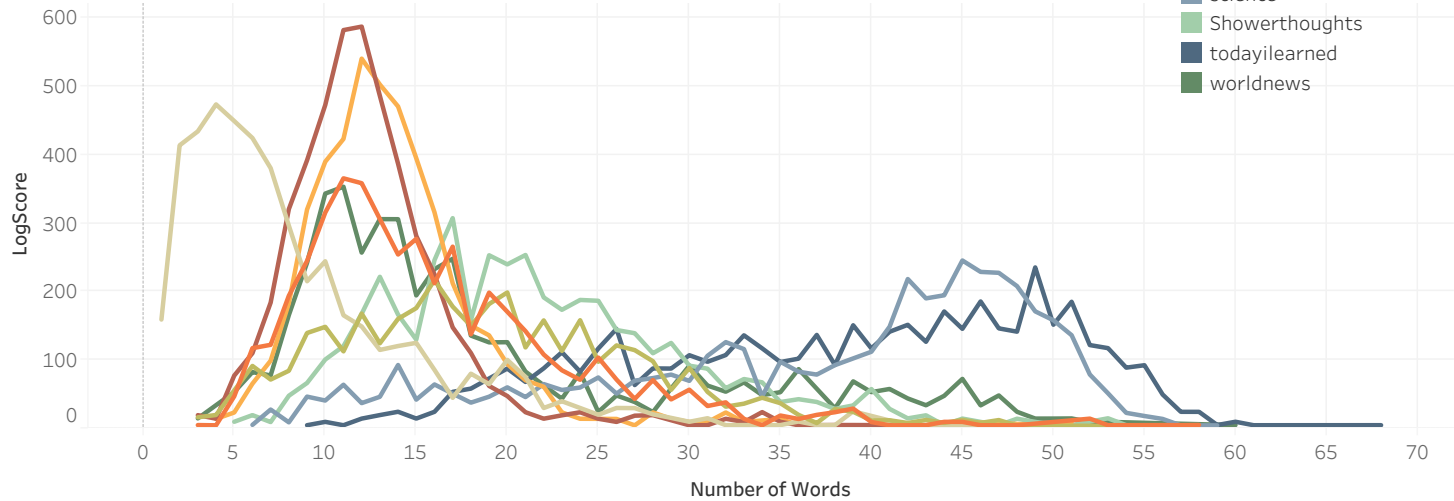
Subreddit	Rank (Tenths)									
	1-100	101-200	201-300	301-400	401-500	501-600	601-700	701-800	801-900	901-999
todayilearned	41.20	39.30	39.73	37.85	37.15	38.06	38.92	36.48	36.06	36.76
science	35.94	37.11	37.86	37.55	38.93	37.62	35.09	37.78	35.21	31.73
Showerthoughts	23.73	23.12	21.30	21.75	22.44	21.66	21.25	21.61	19.37	22.57
worldnews	18.83	20.38	20.18	19.29	19.72	18.74	19.86	18.41	20.58	22.75
explainlikeimfive	21.46	22.22	20.59	18.58	19.72	17.81	17.47	18.99	18.09	16.56
AskReddit	20.65	18.54	15.17	14.70	16.67	15.49	15.38	17.20	15.42	14.75
politics	14.52	14.44	13.73	13.79	14.57	14.89	13.77	12.71	13.71	13.72
news	11.47	13.11	13.07	12.44	12.75	12.48	12.91	12.19	12.44	12.65
funny	10.60	10.38	9.35	8.15	8.42	7.89	9.24	10.18	8.70	7.82

- Subreddit
- ☒ AskReddit
- ☒ explainlikeimfive
- ☒ funny
- ☒ news
- ☒ politics
- ☒ science
- ☒ Showerthoughts
- ☒ todayilearned
- ☒ worldnews

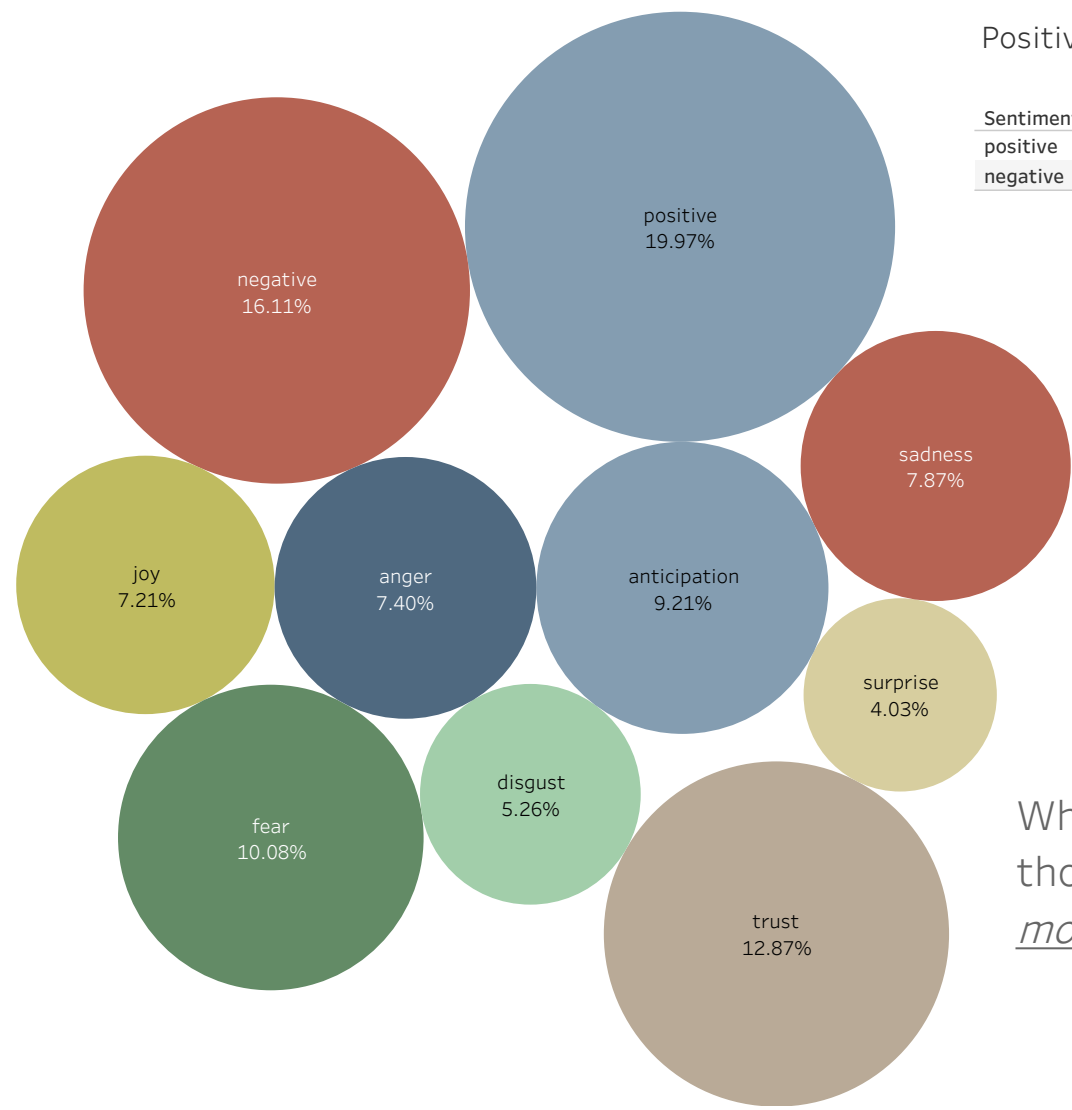
INSIGHTS

There seems to be a sweet spot for average words in a "Title" of a post, varying by subreddit. Generally speaking, /r/funny does not desire much text, whereas, /r/todayilearned and /r/science/ have over 10 words more, on average, than any other subreddit.

Log of Reddit Score (not sentiment) by Number of Words



Sentiment Overall



Positivity vs. Negativity

Sentiment	% of Total N along Senti..	N
positive	55.34%	10,160
negative	44.66%	8,198

Who would of thought Reddit is mostly positive?*

The Fine Print	Summary of Subreddits	Summary of Words	Overall Sentiment Breakdown	Search by word or sentiment	Scoring Sentiment	Spread of Sentiment	/r/Showerthoughts	/r/News
----------------	-----------------------	------------------	-----------------------------	-----------------------------	-------------------	---------------------	-------------------	---------

Words by Sentiment

Sentiment

positive

negative

trust

fear

anticipation

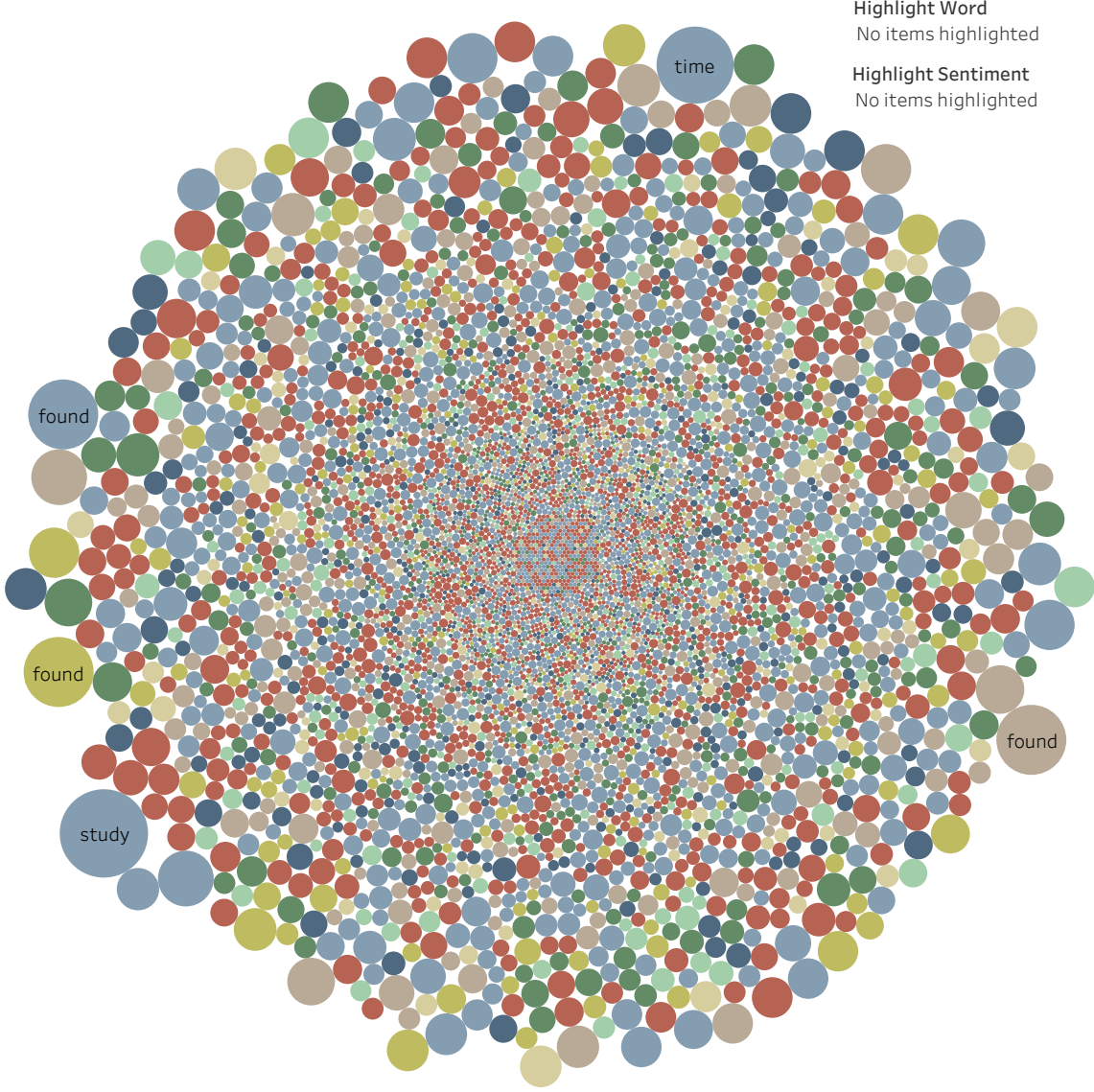
sadness

anger

joy

disgust

surprise



Highlight Word

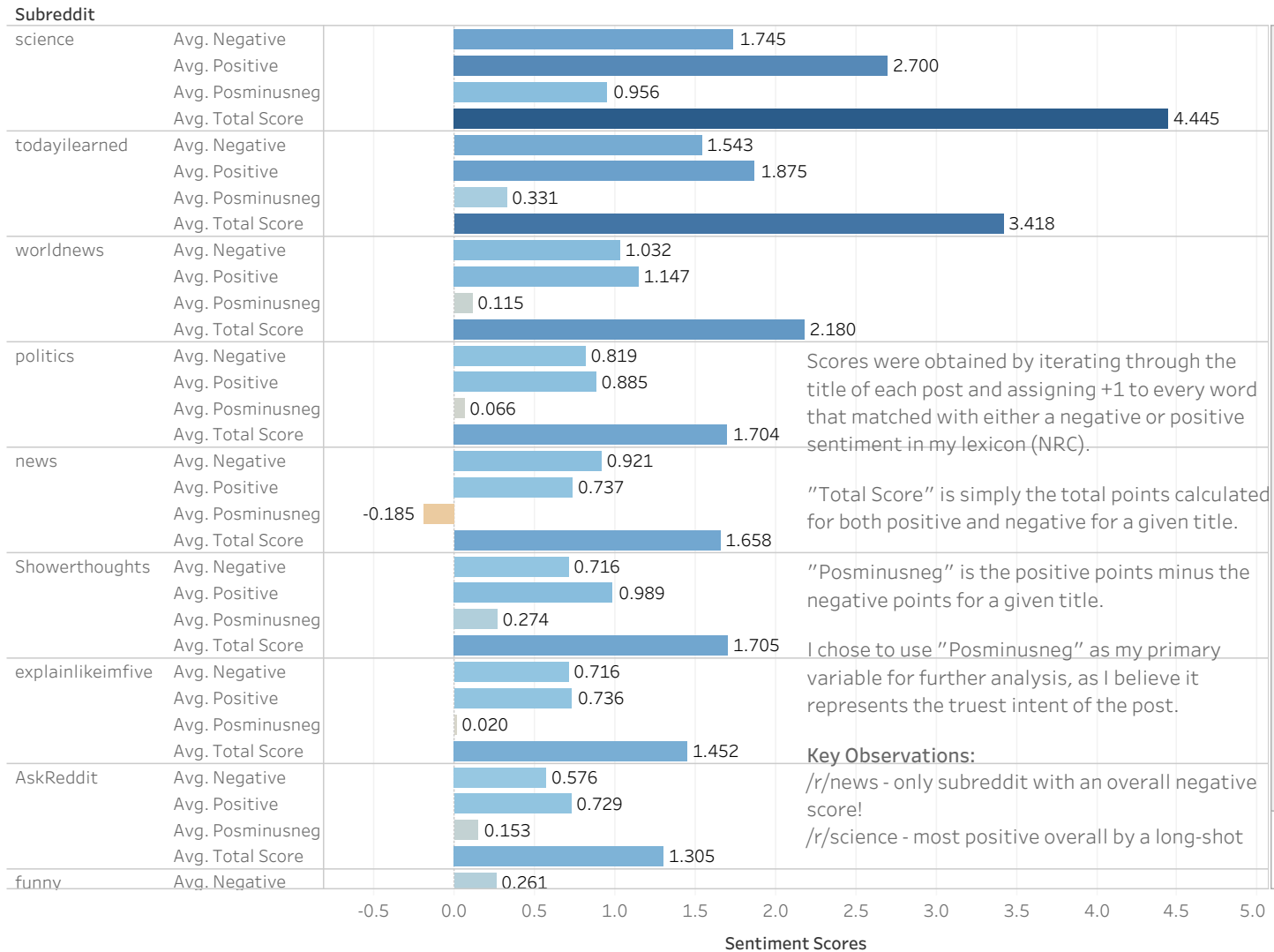
No items highlighted

Highlight Sentiment

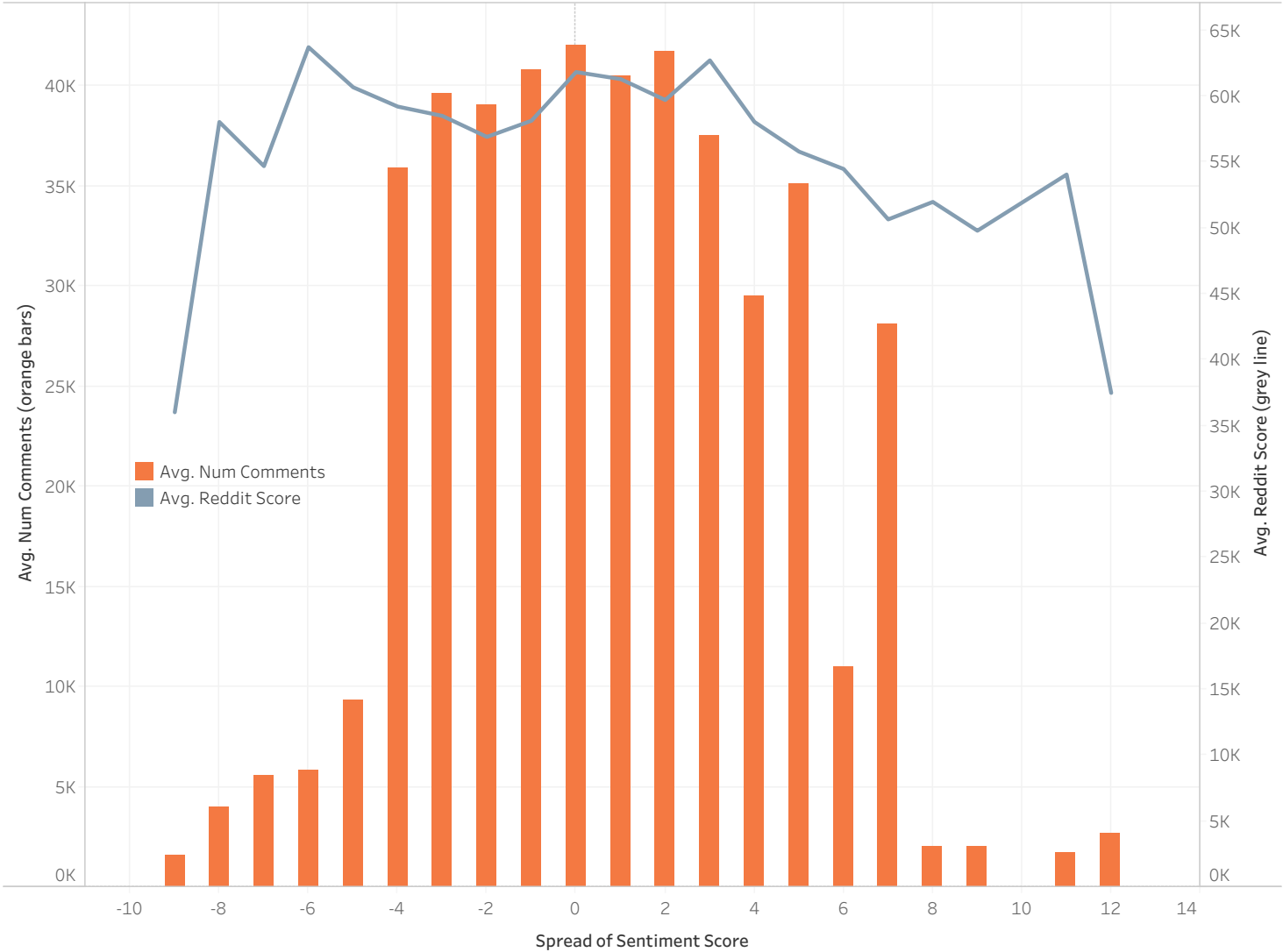
No items highlighted

Summary of Subreddits	Summary of Words	Overall Sentiment Breakdown	Search by word or sentiment	Scoring Sentiment	Spread of Sentiment	/r/Showerthoughts	/r/News	Conclusion
-----------------------	------------------	-----------------------------	-----------------------------	-------------------	---------------------	-------------------	---------	------------

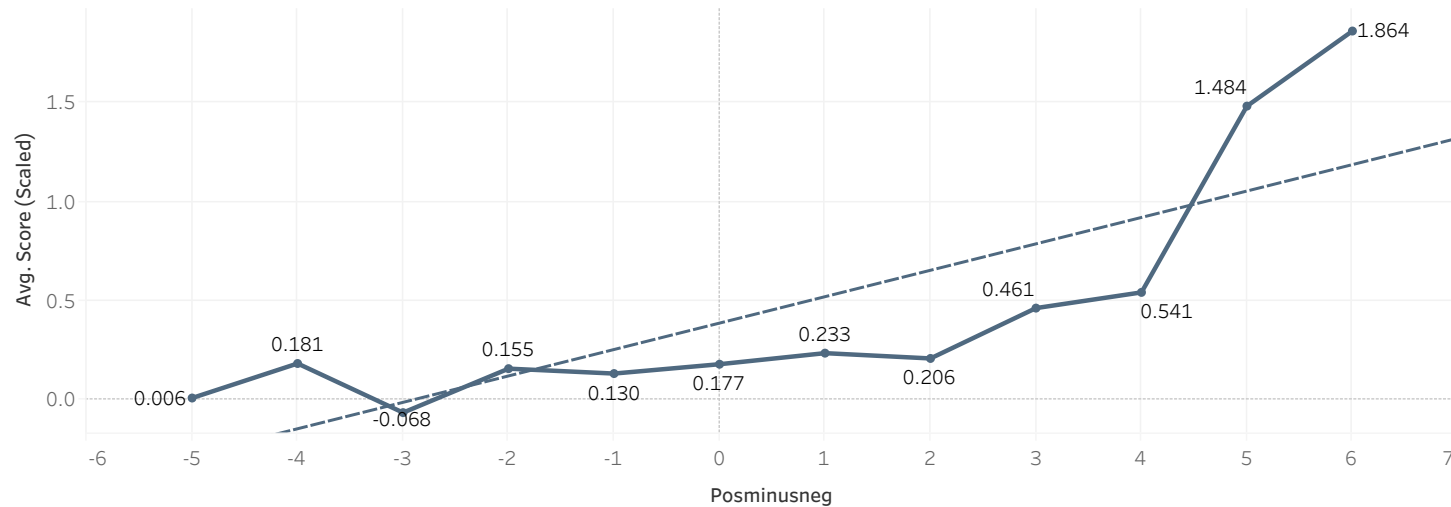
Scored Sentiment by Subreddit



Spread of Sentiment Score



A Closer Look at /r/Showerthoughts



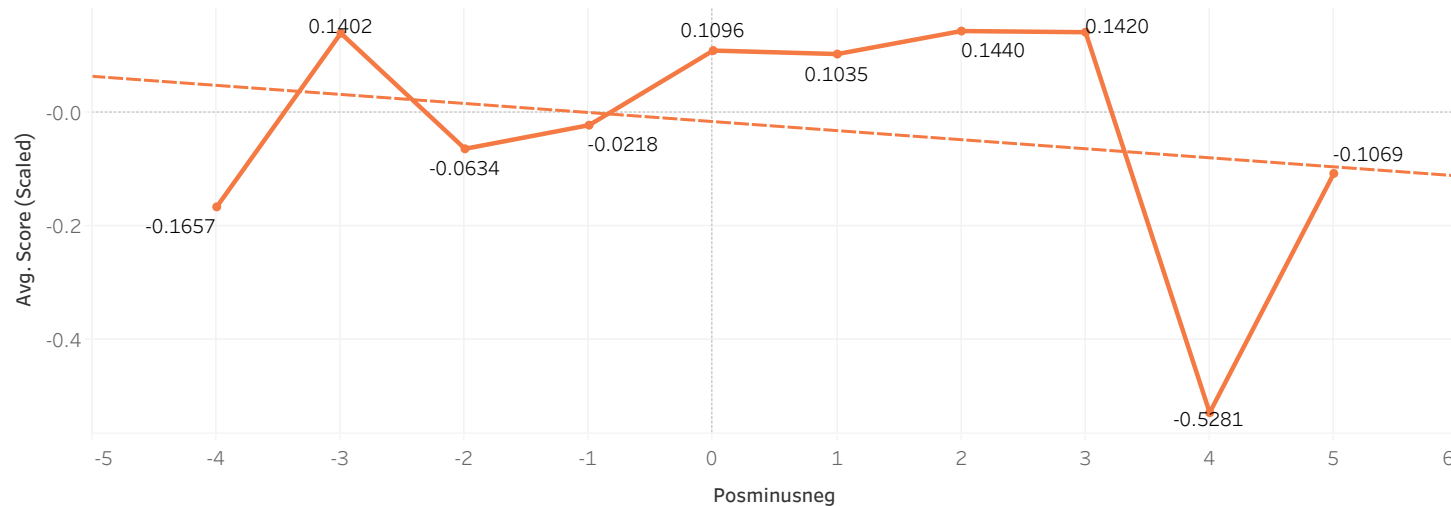
For every **net additional point** in score ("posminusneg"), we can expect to see an **increase** in "Reddit Score" of **1675.18 pts** (p-value < .05), for a post submitted to /r/Showerthoughts, all else being equal.

No one likes to be bummed out in the shower?

Model: score ~ posminusneg * subreddit

Coefficients:	Estimate	Std. Error	t value	Pr(> t)
posminusneg	281.30	454.35	0.619	0.5359
subredditAskReddit	54767.34	524.71	104.377	<2e-16 ***
subredditexplainlikeimfive	11667.74	532.70	21.903	<2e-16 ***
subredditfunny	101214.27	539.62	187.567	<2e-16 ***
subredditnews	62212.74	524.94	118.513	<2e-16 ***
subredditpolitics	50903.99	518.20	98.232	<2e-16 ***
subredditscience	45306.49	553.63	81.835	<2e-16 ***
subredditshowerthoughts	65450.97	542.91	120.556	<2e-16 ***
subreddittodayilearned	76540.25	526.36	145.415	<2e-16 ***
subredditworldnews	72047.74	519.35	138.728	<2e-16 ***
posminusneg:subredditexplainlikeimfive	-36.16	601.00	-0.060	0.9520
posminusneg:subredditfunny	438.55	712.73	0.615	0.5384
posminusneg:subredditnews	789.01	608.97	1.296	0.1951
posminusneg:subredditpolitics	-326.30	612.50	-0.533	0.5942
posminusneg:subredditscience	-171.01	496.45	-0.344	0.7305
posminusneg:subredditshowerthoughts	1393.88	604.87	2.304	0.0212 *
posminusneg:subreddittodayilearned	-841.30	518.22	-1.623	0.1045
posminusneg:subredditworldnews	-175.83	557.19	-0.316	0.7523

A Closer Look at /r/News



For every **additional point** in sentiment score ("Negative"), we can expect to see an **increase** in "Reddit Score" of **1,657.74 pts** (p-value < .05), for a post submitted to /r/news, all else being equal.

Negativity sells news?

Model: score ~ negative * subreddit

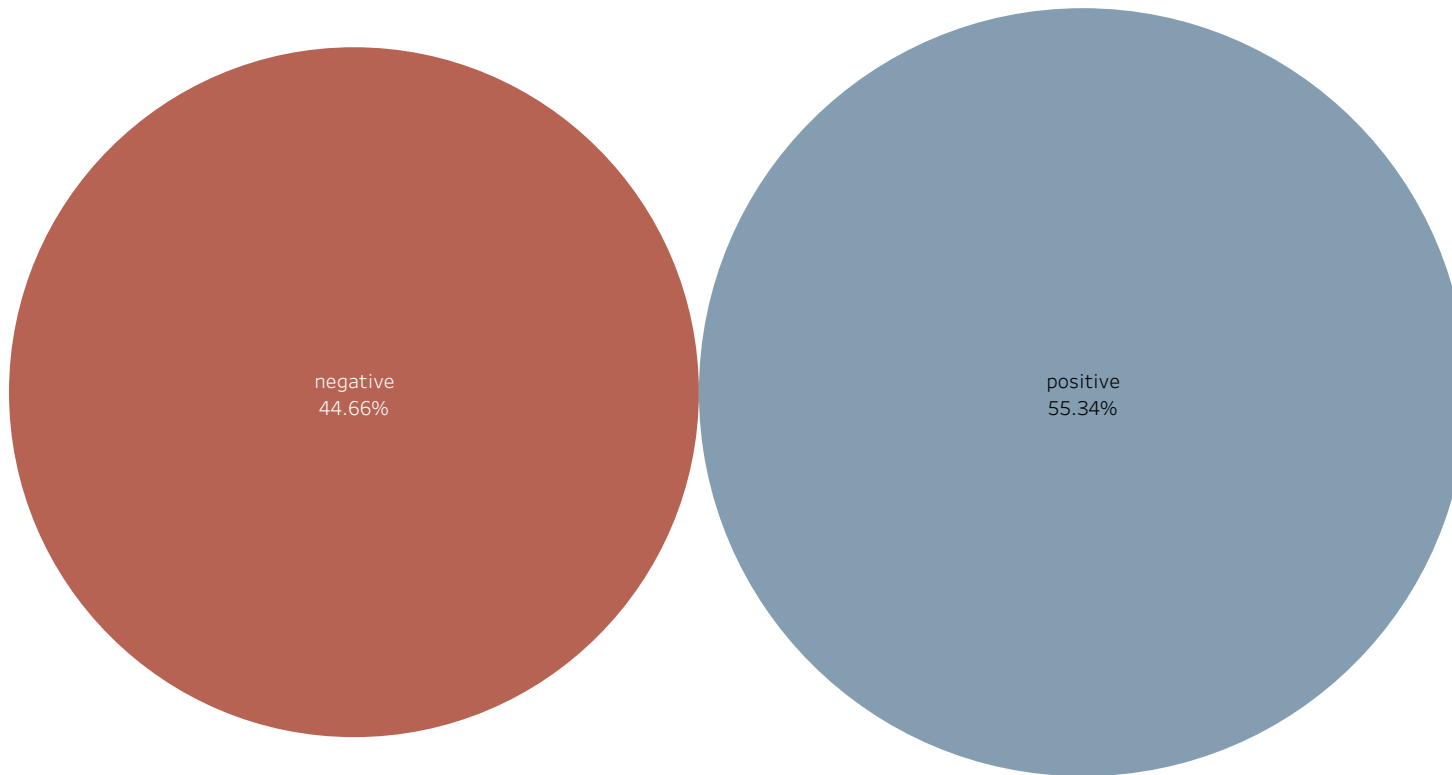
Coefficients	Estimate	Std. Error	t value	Pr(> t)
negative	214.98	679.81	0.316	0.7518
subredditAskReddit	54686.38	651.38	83.955	<2e-16 ***
subredditexplainlikeimfive	11494.59	668.87	17.185	<2e-16 ***
subredditfunny	100819.20	581.05	173.512	<2e-16 ***
subredditnews	63542.36	722.31	87.971	<2e-16 ***
subredditpolitics	50594.65	707.11	71.552	<2e-16 ***
subredditscience	45269.47	731.00	61.928	<2e-16 ***
subredditshowerthoughts	66479.87	674.75	98.525	<2e-16 ***
subreddittodayilearned	75034.62	739.90	101.412	<2e-16 ***
subredditworldnews	72137.87	690.87	104.416	<2e-16 ***
negative:subredditexplainlikeimfive	33.82	883.76	0.038	0.9695
negative:subredditfunny	2022.10	1201.76	1.683	0.0925 .
negative:subredditnews	-1872.72	870.88	-2.150	0.0316 *
negative:subredditpolitics	159.18	898.92	0.177	0.8595
negative:subredditscience	-133.36	740.89	-0.180	0.8572
negative:subredditshowerthoughts	-1011.83	893.52	-1.132	0.2575
negative:subreddittodayilearned	640.31	760.53	0.842	0.3998
negative:subredditworldnews	-290.53	811.21	-0.358	0.7202

Overall Sentiment Breakdown	Search by word or sentiment	Scoring Sentiment	Spread of Sentiment	/r/Showerthoughts	/r/News	Conclusion	Get My Code!
-----------------------------	-----------------------------	-------------------	---------------------	-------------------	---------	------------	--------------

So what's the verdict? Is Reddit a positive or negative place on the internet?

From a high level, it appears that Reddit.com is indeed a more positive than negative place on the internet.

Of course, this is only taking into consideration the 9 subreddits and the top 1000 posts of all-time for those subreddits.



Overall Sentiment Breakdown	Search by word or sentiment	Scoring Sentiment	Spread of Sentiment	/r/Showerthoughts	/r/News	Conclusion	Get My Code!
-----------------------------	-----------------------------	-------------------	---------------------	-------------------	---------	------------	--------------

