```
--ALL COMMANDS USED WITH EXPLANATIONS

--do these commands from sql editor before transferring to linux fs

--remove columns
alter table May2015
drop column created_utc, link_id, score_hidden, author_flair_css_class,
author_flair_text,
removal_reason, archived, retrieved_on, edited, controversiality,
parent_id;


--remove non default subreddits excluding announcements
delete from May2015
where not subreddit='Art'
and not subreddit='AskReddit'
and not subreddit='askscience'
and not subreddit='aww'
and not subreddit='blog'
and not subreddit='books'
and not subreddit='creepy'
and not subreddit='dataisbeautiful'
and not subreddit='DIY'
and not subreddit='Documentaries'
and not subreddit='EarthPorn'
and not subreddit='explainlikeimfive'
and not subreddit='food'
and not subreddit='funny'
and not subreddit='Futurology'
and not subreddit='gadgets'
and not subreddit='gaming'
and not subreddit='GetMotivated'
and not subreddit='gifs'
and not subreddit='history'
and not subreddit='IAmA'
and not subreddit='InternetIsBeautiful'
and not subreddit='Jokes'
and not subreddit='LifeProTips'
and not subreddit='listentothis'
and not subreddit='mildlyinteresting'
```

```
and not subreddit='movies'
and not subreddit='Music'
and not subreddit='news'
and not subreddit='nosleep'
and not subreddit='nottheonion'
and not subreddit='OldSchoolCool'
and not subreddit='personalfinance'
and not subreddit='philosophy'
and not subreddit='photoshopbattles'
and not subreddit='pics'
and not subreddit='science'
and not subreddit='Showerthoughts'
and not subreddit='space'
and not subreddit='sports'
and not subreddit='television'
and not subreddit='tifu'
and not subreddit='todayilearned'
and not subreddit='UpliftingNews'
and not subreddit='videos'
and not subreddit='worldnews';


--check count should be 46
select count(distinct subreddit) from May2015;


--optional double check
select distinct subreddit from May2015;


--remove all rows where comment body is [deleted]
delete from May2015 where body='[deleted]';



--upload file to linux server
scp may_2015_trim.csv dkang10@129.150.64.74:/home/dkang10


--create hdfs directory
```

```
hdfs dfs -mkdir group3
hdfs dfs -mkdir group3/dataset

--check directory was made
hdfs dfs -ls group3

--put file in hdfs
hdfs dfs -put may_2015_trim.csv group3/dataset

--check file in hdfs
hdfs dfs -ls group3/dataset

--change permissions of new directory
hdfs dfs -chmod -R o+w .


--create external table from csv file in hdfs
create external table may_2015 (
    id string,
    subreddit_id string,
    subreddit string,
    name string,
    author string,
    ups int,
    downs int,
    score int,
    gilted int,
    distinguished string,
    body string
)
row format delimited
fields terminated by '|'
lines terminated by '\n'
location '/user/dkang10/group3/dataset'
tblproperties('skip.header.line.count'='1');

--check new table
show tables;
describe may_2015;
select count(*) from may_2015;
```

```
--clean table from csv and text box errors
create table may_2015_a as
select id, subreddit_id, subreddit, name, author, ups, downs, score,
gilted, distinguished, body from may_2015
where subreddit='Art'
or subreddit='AskReddit'
or subreddit='askscience'
or subreddit='aww'
or subreddit='blog'
or subreddit='books'
or subreddit='creepy'
or subreddit='dataisbeautiful'
or subreddit='DIY'
or subreddit='Documentaries'
or subreddit='EarthPorn'
or subreddit='explainlikeimfive'
or subreddit='food'
or subreddit='funny'
or subreddit='Futurology'
or subreddit='gadgets'
or subreddit='gaming'
or subreddit='GetMotivated'
or subreddit='gifs'
or subreddit='history'
or subreddit='IAmA'
or subreddit='InternetIsBeautiful'
or subreddit='Jokes'
or subreddit='LifeProTips'
or subreddit='listentothis'
or subreddit='mildlyinteresting'
or subreddit='movies'
or subreddit='Music'
or subreddit='news'
or subreddit='nosleep'
or subreddit='nottheonion'
or subreddit='OldSchoolCool'
or subreddit='personalfinance'
or subreddit='philosophy'
```

```
or subreddit='photoshopbattles'
or subreddit='pics'
or subreddit='science'
or subreddit='Showerthoughts'
or subreddit='space'
or subreddit='sports'
or subreddit='television'
or subreddit='tifu'
or subreddit='todayilearned'
or subreddit='UpliftingNews'
or subreddit='videos'
or subreddit='worldnews';



--general analysis begins~~~~~~~~~~~~~~~~~~~~

--query comments with highest upvotes across default subreddits
select subreddit, ups, downs, score, body from may_2015_a
order by score desc limit 25;
```

subreddit	ups	downs	score	body
AskRedditx	6761	0	6761	Then you got yourself a one night standoff.
AskRedditx	6736	0	6736	"If it's untreated mood...
IAmA	6713	0	6713	"Okay, here's one.
AskRedditx	5992	0	5992	"I have family who sailed around the world. One day in the North Atlantic, their sailboat was going over some GIGANTIC swells. They didn't have breaks at the top, so it was safe, but the boat was rising and falling way beyond the neutral.
AskRedditx	5927	0	5927	"My boss is this giddy Asian lady (I work at a restaurant) and one day somehow we are on the subject of racial slurs. I am black and tell her I have been called ""nigger"" before in a fight. She then asks me what the ""Asian equivalent of 'N'"" would be and I tell her maybe ""chink"".
AskRedditx	5849	0	5849	thanks man and also fuck you man
AskRedditx	5845	0	5845	"this happened to med school. it was taking the history of a guy in clinic and I asked about his past medical problems, including if he had had any heart attacks.
AskRedditx	5776	0	5776	We had an office picture day. At the time, I was his sole employee.
AskRedditx	5767	0	5767	OP are you alright? I can send you food if you'd like.
AskRedditx	5762	0	5762	So she's pregnant huh
AskRedditx	5741	0	5741	"Once as a tired medical resident I was called to the ER to admit someone at like 3am. This bonehead had gall bladder removal a week ago and now had a surgical-site wound infection. I asked if they'd taken their post-op antibiotics they were prescribed, and they weren't sure. I was getting more and more frustrated with this dumbass preventing my sleep when I decided to use a ""pregnant pause"" interview technique, and just shut up. This usually results in either awkward silence and the patient saying ""Uhh WTF doc"" or awkward silence followed by some useful deep revelation.
AskRedditx	5720	0	5720	"My Dad would not let me take a shit after dinner. I had to wait until he went to bed or sometime in the middle of the night. He claimed I was wasting food by shitting it out right after eating. He was fucking retarded.
AskRedditx	5710	0	5710	Porum based website moderators.
AskRedditx	5699	0	5699	It would be the inventor of punctuation, as he would have died after reading the title of this post.
AskRedditx	5673	0	5673	¯_(ツ)_/¯
AskRedditx	5668	0	5668	"Boss would always shout out the same thing as he walked out for his lunch break
funny	5642	0	5642	HEY! THAT'S ME!!! https://twitter.com/RyanMirvis/status/595392230463905793
AskRedditx	5641	0	5641	"Mr clean magic erasers.
AskRedditx	5570	0	5570	Being fucking late to every fucking thing fucking ever.
AskRedditx	5564	0	5564	I sometimes wake up around 4am to drive around or shop for groceries. The world is so much better without other people in it.
AskRedditx	5562	0	5562	knowing me procrastinate till the last day and hate my life on that day
AskRedditx	5545	0	5545	Dicks!
AskRedditx	5535	0	5535	"I got a call from my father's wife that he was on his deathbed and needed me to come out right away. I couldn't afford the tickets at the time (over $1000) and made a Facebook post about it.
AskRedditx	5531	0	5531	"Walking into a grocery store and casually strolling past the security guard after not buying anything.
AskRedditx	5527	0	5527	If everyone could stop pissing on the floor in front of the urinal THAT WOULD BE FUCKING GREAT

```
--observation: many top upvoted comments are from askreddit

--query number of comments per subreddit
select subreddit, count(score) as count from may_2015_a
group by subreddit order by count desc;
```

subreddit	count
AskReddit	3950729
funny	695296
pics	568131
videos	515397
todayilearned	489510
news	483832
worldnews	445479
movies	357404
gaming	329248
gifs	229450
explainlikeimfive	195420
Showerthoughts	162825
IAmA	147176
tifu	145783
aww	142269
personalfinance	110905
mildlyinteresting	109951
Music	97177
nottheonion	85618
television	81860
books	68073
Futurology	66264
LifeProTips	58613
Jokes	56189
science	55016
food	51488
creepy	50601
sports	45969
DIY	42090
dataisbeautiful	34786
OldSchoolCool	32328
askscience	30946
space	29858
Documentaries	27148
photoshopbattles	26340
nosleep	25333
InternetIsBeautiful	25018
gadgets	22146
EarthPorn	21302
history	20357
UpliftingNews	17873
GetMotivated	16873
Art	16147
philosophy	15265
blog	13377
listentothis	12541

```
--observation: askreddit has over 5x the largest comment participation
over the next largest subreddit

--query average upvote per comment per subreddit
select subreddit, avg(score) as avg_score from may_2015_a
group by subreddit order by avg_score desc;
```

subreddit	avg_score
photoshopbattles	28.870652999240697
IAmA	20.451065391096375
gifs	14.995219001961212
AskReddit	13.244326806521025
blog	12.784555580473947
videos	12.618364096026946
tifu	12.233778972856918
pics	12.200760035977618
funny	12.02165408689249
todayilearned	11.632318032318032
mildlyinteresting	10.766304990404816
askscience	10.712531506495186
aww	10.437755238316148
television	10.226349865624236
nottheonion	10.202761101637506
history	10.075944392592229
Jokes	10.00553489117087
movies	9.800259650143815
gaming	9.618026533190786
Showerthoughts	9.467372946414862
UpliftingNews	9.334303138812734
sports	9.28895560051339
science	8.736840191944161
news	8.645775806478282
worldnews	7.852819100339185
OldSchoolCool	7.5560504825538235
creepy	7.5380130827454
space	7.515640699310068
LifeProTips	7.431747223312234
dataisbeautiful	7.275829356637728
explainlikeimfive	7.175202128748337
books	6.844534543798569
Music	6.628296819206191
InternetIsBeautiful	6.358901590854584
gadgets	5.821502754447756
DIY	5.697481587075314
nosleep	5.049145383491888
Futurology	5.023255462996499
food	4.90186062771908
EarthPorn	4.591493756454793
GetMotivated	4.5320926924672555
Documentaries	4.314756151466038
personalfinance	4.169063613002119
Art	4.092772651266489
listentothis	2.459532732636951
philosophy	2.2024238453979694

```
--observation: surprisingly askreddit is in 4th place, maybe due to sheer
comment volume and not enough upvote distribution
--photoshopbattles comments are each tailor made and more impactful
--IAmA comments consist of questions asked to high profile people who only
respond within a pre determined time frame
--gifs has more upvote participation likely due to its lesser comment
participation so it has less throwaway comments

--see most common 2 word combinations
```

```
select explode(ngrams(sentences(lower(body)), 2, 50)) as two from
may_2015_a;
```

```
+----------------------------------------------------+--+
|                        two                         |
+----------------------------------------------------+--+
| {"ngram":["in","the"],"estfrequency":638462.0}     |
| {"ngram":["of","the"],"estfrequency":631859.0}     |
| {"ngram":["to","be"],"estfrequency":366549.0}      |
| {"ngram":["i","was"],"estfrequency":314278.0}      |
| {"ngram":["on","the"],"estfrequency":295589.0}     |
| {"ngram":["to","the"],"estfrequency":290884.0}     |
| {"ngram":["it","was"],"estfrequency":279669.0}     |
| {"ngram":["if","you"],"estfrequency":271564.0}     |
| {"ngram":["i","don't"],"estfrequency":268100.0}    |
| {"ngram":["is","a"],"estfrequency":261561.0}       |
| {"ngram":["in","a"],"estfrequency":230547.0}       |
| {"ngram":["i","think"],"estfrequency":229899.0}    |
| {"ngram":["this","is"],"estfrequency":229047.0}    |
| {"ngram":["for","the"],"estfrequency":228932.0}    |
| {"ngram":["and","i"],"estfrequency":216135.0}      |
| {"ngram":["i","have"],"estfrequency":211305.0}     |
| {"ngram":["it","is"],"estfrequency":202114.0}      |
| {"ngram":["a","lot"],"estfrequency":192877.0}      |
| {"ngram":["have","a"],"estfrequency":190666.0}     |
| {"ngram":["the","same"],"estfrequency":187752.0}   |
| {"ngram":["and","the"],"estfrequency":185324.0}    |
| {"ngram":["but","i"],"estfrequency":181902.0}      |
| {"ngram":["have","to"],"estfrequency":172076.0}    |
| {"ngram":["for","a"],"estfrequency":170837.0}      |
| {"ngram":["is","the"],"estfrequency":165347.0}     |
| {"ngram":["you","can"],"estfrequency":164771.0}    |
| {"ngram":["with","the"],"estfrequency":164459.0}   |
| {"ngram":["at","the"],"estfrequency":162088.0}     |
| {"ngram":["of","a"],"estfrequency":161078.0}       |
| {"ngram":["would","be"],"estfrequency":156869.0}   |
| {"ngram":["i","am"],"estfrequency":155192.0}       |
| {"ngram":["going","to"],"estfrequency":154916.0}   |
| {"ngram":["as","a"],"estfrequency":153556.0}       |
| {"ngram":["to","get"],"estfrequency":153375.0}     |
| {"ngram":["was","a"],"estfrequency":152216.0}      |
| {"ngram":["you","are"],"estfrequency":151630.0}    |
| {"ngram":["want","to"],"estfrequency":151152.0}    |
| {"ngram":["to","do"],"estfrequency":145731.0}      |
| {"ngram":["out","of"],"estfrequency":145619.0}     |
| {"ngram":["with","a"],"estfrequency":145107.0}     |
| {"ngram":["when","i"],"estfrequency":142073.0}     |
| {"ngram":["one","of"],"estfrequency":141489.0}     |
| {"ngram":["you","have"],"estfrequency":134646.0}   |
| {"ngram":["they","are"],"estfrequency":131709.0}   |
| {"ngram":["i","would"],"estfrequency":130755.0}    |
| {"ngram":["all","the"],"estfrequency":128557.0}    |
| {"ngram":["lot","of"],"estfrequency":127845.0}     |
| {"ngram":["from","the"],"estfrequency":123960.0}   |
| {"ngram":["that","the"],"estfrequency":121784.0}   |
| {"ngram":["be","a"],"estfrequency":120029.0}       |
+----------------------------------------------------+--+
```

```
--results are mostly prepositions, articles and generally results of
english grammar
```

```
--see most common 3 word combinations
select explode(ngrams(sentences(lower(body)), 3, 50)) as three from
may_2015_a;
```

```
+------------------------------------------------------------------------+--+
|                                  three                                 |
+------------------------------------------------------------------------+--+
| {"ngram":["a","lot","of"],"estfrequency":124501.0}                     |
| {"ngram":["one","of","the"],"estfrequency":64822.0}                    |
| {"ngram":["https","www.youtube.com","watch"],"estfrequency":60975.0}   |
| {"ngram":["i","don't","know"],"estfrequency":54975.0}                  |
| {"ngram":["i","don't","think"],"estfrequency":49913.0}                 |
| {"ngram":["when","i","was"],"estfrequency":46081.0}                    |
| {"ngram":["i","have","a"],"estfrequency":42840.0}                      |
| {"ngram":["to","be","a"],"estfrequency":41652.0}                       |
| {"ngram":["it","was","a"],"estfrequency":37688.0}                      |
| {"ngram":["be","able","to"],"estfrequency":34437.0}                    |
| {"ngram":["the","fact","that"],"estfrequency":33301.0}                 |
| {"ngram":["you","have","to"],"estfrequency":32770.0}                   |
| {"ngram":["http","www.reddit.com","r"],"estfrequency":31748.0}         |
| {"ngram":["it","would","be"],"estfrequency":31583.0}                   |
| {"ngram":["to","do","with"],"estfrequency":29045.0}                    |
| {"ngram":["out","of","the"],"estfrequency":28921.0}                    |
| {"ngram":["there","is","a"],"estfrequency":27769.0}                    |
| {"ngram":["all","the","time"],"estfrequency":27009.0}                  |
| {"ngram":["you","want","to"],"estfrequency":26900.0}                   |
| {"ngram":["i","want","to"],"estfrequency":26614.0}                     |
| {"ngram":["this","is","a"],"estfrequency":25542.0}                     |
| {"ngram":["the","rest","of"],"estfrequency":24800.0}                   |
| {"ngram":["http","en.wikipedia.org","wiki"],"estfrequency":24770.0}    |
| {"ngram":["going","to","be"],"estfrequency":24628.0}                   |
| {"ngram":["submission","has","been"],"estfrequency":24589.0}           |
| {"ngram":["this","is","the"],"estfrequency":24311.0}                   |
| {"ngram":["i","feel","like"],"estfrequency":24197.0}                   |
| {"ngram":["i","used","to"],"estfrequency":23514.0}                     |
| {"ngram":["a","bunch","of"],"estfrequency":22526.0}                    |
| {"ngram":["the","end","of"],"estfrequency":22162.0}                    |
| {"ngram":["i","have","to"],"estfrequency":22152.0}                     |
| {"ngram":["there","is","no"],"estfrequency":21997.0}                   |
| {"ngram":["thank","you","for"],"estfrequency":21573.0}                 |
| {"ngram":["i","had","a"],"estfrequency":21565.0}                       |
| {"ngram":["most","of","the"],"estfrequency":21157.0}                   |
| {"ngram":["you","have","a"],"estfrequency":20844.0}                    |
| {"ngram":["been","automatically","removed"],"estfrequency":20496.0}    |
| {"ngram":["the","same","thing"],"estfrequency":20450.0}                |
| {"ngram":["one","of","my"],"estfrequency":19983.0}                     |
| {"ngram":["if","you","are"],"estfrequency":19603.0}                    |
| {"ngram":["part","of","the"],"estfrequency":19450.0}                   |
| {"ngram":["has","been","automatically"],"estfrequency":19202.0}        |
| {"ngram":["if","you","have"],"estfrequency":19122.0}                   |
| {"ngram":["have","to","be"],"estfrequency":18888.0}                    |
| {"ngram":["i","am","a"],"estfrequency":18806.0}                        |
| {"ngram":["the","first","time"],"estfrequency":18763.0}                |
| {"ngram":["don't","want","to"],"estfrequency":18540.0}                 |
| {"ngram":["in","the","us"],"estfrequency":18169.0}                     |
| {"ngram":["its","entirety","before"],"estfrequency":18057.0}           |
| {"ngram":["at","the","end"],"estfrequency":17901.0}                    |
+------------------------------------------------------------------------+--+
```

```
--top phrase by almost 2x occurrence is "a lot of"
--youtube links are also the third most common

--see most common 4 word combinations
select explode(ngrams(sentences(lower(body)), 4, 50)) as four from
may_2015_a;
```

four
{"ngram":["has","been","automatically","removed"],"estfrequency":20468.0}
{"ngram":["in","its","entirety","before"],"estfrequency":19807.0}
{"ngram":["this","message","in","its"],"estfrequency":18590.0}
{"ngram":["read","this","message","in"],"estfrequency":18573.0}
{"ngram":["entirety","before","taking","action"],"estfrequency":18461.0}
{"ngram":["its","entirety","before","taking"],"estfrequency":18461.0}
{"ngram":["message","in","its","entirety"],"estfrequency":18461.0}
{"ngram":["please","read","this","message"],"estfrequency":18461.0}
{"ngram":["your","submission","has","been"],"estfrequency":16849.0}
{"ngram":["a","lot","of","people"],"estfrequency":15437.0}
{"ngram":["submission","has","been","automatically"],"estfrequency":15090.0}
{"ngram":["been","automatically","removed","because"],"estfrequency":14523.0}
{"ngram":["http","www.reddit.com","r","askreddit"],"estfrequency":13536.0}
{"ngram":["r","askreddit","wiki","index"],"estfrequency":13326.0}
{"ngram":["keep","in","mind","that"],"estfrequency":12649.0}
{"ngram":["the","rest","of","the"],"estfrequency":12216.0}
{"ngram":["be","removed","http","www.reddit.com"],"estfrequency":12151.0}
{"ngram":["mind","that","the","op"],"estfrequency":11962.0}
{"ngram":["replies","that","are","jokes"],"estfrequency":11929.0}
{"ngram":["at","the","end","of"],"estfrequency":11845.0}
{"ngram":["therefore","any","replies","that"],"estfrequency":11587.0}
{"ngram":["that","the","op","of"],"estfrequency":11134.0}
{"ngram":["has","chosen","to","mark"],"estfrequency":10795.0}
{"ngram":["chosen","to","mark","this"],"estfrequency":10787.0}
{"ngram":["in","mind","that","the"],"estfrequency":10741.0}
{"ngram":["of","this","thread","has"],"estfrequency":10741.0}
{"ngram":["with","the","serious","replies"],"estfrequency":10698.0}
{"ngram":["or","are","otherwise","non-contributory"],"estfrequency":10681.0}
{"ngram":["this","post","with","the"],"estfrequency":10643.0}
{"ngram":["mark","this","post","with"],"estfrequency":10538.0}
{"ngram":["only","tag","therefore","any"],"estfrequency":10537.0}
{"ngram":["that","are","jokes","puns"],"estfrequency":10535.0}
{"ngram":["please","keep","in","mind"],"estfrequency":10476.0}
{"ngram":["thread","has","chosen","to"],"estfrequency":10435.0}
{"ngram":["the","op","of","this"],"estfrequency":10411.0}
{"ngram":["www.reddit.com","r","askreddit","wiki"],"estfrequency":10291.0}
{"ngram":["this","thread","has","chosen"],"estfrequency":10198.0}
{"ngram":["post","with","the","serious"],"estfrequency":10185.0}
{"ngram":["removed","http","www.reddit.com","r"],"estfrequency":10149.0}
{"ngram":["the","end","of","the"],"estfrequency":10108.0}
{"ngram":["puns","off-topic","or","are"],"estfrequency":10099.0}
{"ngram":["are","otherwise","non-contributory","will"],"estfrequency":10005.0}
{"ngram":["i","thought","it","was"],"estfrequency":9996.0}
{"ngram":["i","am","a","beautiful"],"estfrequency":9922.0}
{"ngram":["to","mark","this","post"],"estfrequency":9895.0}
{"ngram":["am","a","beautiful","person"],"estfrequency":9798.0}
{"ngram":["askreddit","wiki","index","wiki"],"estfrequency":9663.0}
{"ngram":["serious","replies","only","tag"],"estfrequency":9655.0}
{"ngram":["the","serious","replies","only"],"estfrequency":9655.0}
{"ngram":["replies","only","tag","therefore"],"estfrequency":9654.0}

```
--the most frequent longer phrases appear to be administrative comments..


--see most common 5 word combinations
```

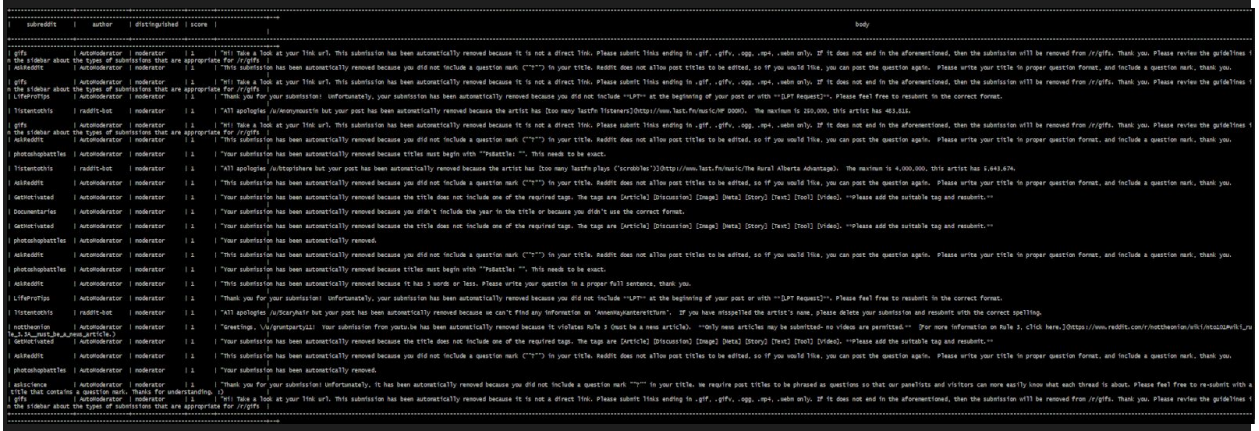
```
select explode(ngrams(sentences(lower(body)), 5, 50)) as five from
may_2015_a;
```

```
+----------------------------------------------------------------------------------------------------+--+
|                                                five                                                |  |
+----------------------------------------------------------------------------------------------------+--+
| {"ngram":["its","entirety","before","taking","action"],"estfrequency":18765.0}                     |  |
| {"ngram":["please","read","this","message","in"],"estfrequency":18765.0}                           |  |
| {"ngram":["in","its","entirety","before","taking"],"estfrequency":18694.0}                         |  |
| {"ngram":["message","in","its","entirety","before"],"estfrequency":18694.0}                        |  |
| {"ngram":["read","this","message","in","its"],"estfrequency":18694.0}                              |  |
| {"ngram":["this","message","in","its","entirety"],"estfrequency":18694.0}                          |  |
| {"ngram":["submission","has","been","automatically","removed"],"estfrequency":15323.0}             |  |
| {"ngram":["has","been","automatically","removed","because"],"estfrequency":15239.0}                |  |
| {"ngram":["has","chosen","to","mark","this"],"estfrequency":13188.0}                               |  |
| {"ngram":["in","mind","that","the","op"],"estfrequency":12365.0}                                   |  |
| {"ngram":["only","tag","therefore","any","replies"],"estfrequency":12084.0}                        |  |
| {"ngram":["serious","replies","only","tag","therefore"],"estfrequency":11388.0}                    |  |
| {"ngram":["non-contributory","will","be","removed","http"],"estfrequency":11304.0}                 |  |
| {"ngram":["this","post","with","the","serious"],"estfrequency":11219.0}                            |  |
| {"ngram":["keep","in","mind","that","the"],"estfrequency":11185.0}                                 |  |
| {"ngram":["that","are","jokes","puns","off-topic"],"estfrequency":11082.0}                         |  |
| {"ngram":["replies","only","tag","therefore","any"],"estfrequency":11058.0}                        |  |
| {"ngram":["or","are","otherwise","non-contributory","will"],"estfrequency":11057.0}                |  |
| {"ngram":["please","keep","in","mind","that"],"estfrequency":11040.0}                              |  |
| {"ngram":["are","otherwise","non-contributory","will","be"],"estfrequency":11000.0}                |  |
| {"ngram":["with","the","serious","replies","only"],"estfrequency":10995.0}                         |  |
| {"ngram":["puns","off-topic","or","are","otherwise"],"estfrequency":10937.0}                       |  |
| {"ngram":["removed","http","www.reddit.com","r","askreddit"],"estfrequency":10934.0}               |  |
| {"ngram":["will","be","removed","http","www.reddit.com"],"estfrequency":10934.0}                   |  |
| {"ngram":["thread","has","chosen","to","mark"],"estfrequency":10873.0}                             |  |
| {"ngram":["chosen","to","mark","this","post"],"estfrequency":10846.0}                              |  |
| {"ngram":["op","of","this","thread","has"],"estfrequency":10833.0}                                 |  |
| {"ngram":["jokes","puns","off-topic","or","are"],"estfrequency":10813.0}                           |  |
| {"ngram":["any","replies","that","are","jokes"],"estfrequency":10812.0}                            |  |
| {"ngram":["are","jokes","puns","off-topic","or"],"estfrequency":10812.0}                           |  |
| {"ngram":["tag","therefore","any","replies","that"],"estfrequency":10811.0}                        |  |
| {"ngram":["therefore","any","replies","that","are"],"estfrequency":10811.0}                        |  |
| {"ngram":["that","the","op","of","this"],"estfrequency":10805.0}                                   |  |
| {"ngram":["mind","that","the","op","of"],"estfrequency":10765.0}                                   |  |
| {"ngram":["to","mark","this","post","with"],"estfrequency":10714.0}                                |  |
| {"ngram":["http","www.reddit.com","r","askreddit","wiki"],"estfrequency":10677.0}                  |  |
| {"ngram":["of","this","thread","has","chosen"],"estfrequency":10670.0}                             |  |
| {"ngram":["this","thread","has","chosen","to"],"estfrequency":10670.0}                             |  |
| {"ngram":["the","serious","replies","only","tag"],"estfrequency":10622.0}                          |  |
| {"ngram":["mark","this","post","with","the"],"estfrequency":10621.0}                               |  |
| {"ngram":["otherwise","non-contributory","will","be","removed"],"estfrequency":10618.0}            |  |
| {"ngram":["the","op","of","this","thread"],"estfrequency":10520.0}                                 |  |
| {"ngram":["post","with","the","serious","replies"],"estfrequency":10474.0}                         |  |
| {"ngram":["www.reddit.com","r","askreddit","wiki","index"],"estfrequency":10466.0}                 |  |
| {"ngram":["r","askreddit","wiki","index","wiki"],"estfrequency":10424.0}                           |  |
| {"ngram":["replies","that","are","jokes","puns"],"estfrequency":10405.0}                           |  |
| {"ngram":["off-topic","or","are","otherwise","non-contributory"],"estfrequency":10380.0}           |  |
| {"ngram":["be","removed","http","www.reddit.com","r"],"estfrequency":10377.0}                      |  |
| {"ngram":["i","am","a","beautiful","person"],"estfrequency":9837.0}                                |  |
| {"ngram":["ↄ","ↄ","ↄ","ↄ","ↄ"],"estfrequency":7809.0}                                             |  |
+----------------------------------------------------------------------------------------------------+--+
```

```
--same situation here..



--see subreddit source and full comment of most frequent 4phrase
select subreddit, author, distinguished, score, body from may_2015_a
```



```sql
where lower(body) like '%has been automatically removed%' limit 25;
```

```
+------------------+---------------+---------------+-------+-----------------------------------------------------------------------------------------------
--------------------------------------------------------------------+--+
|    subreddit     |     author    | distinguished | score |                                                              body
                                                                    |
+------------------+---------------+---------------+-------+-----------------------------------------------------------------------------------------------
--------------------------------------------------------------------+--+
| gifs             | AutoModerator | moderator     | 1     | "Hi! Take a look at your link url. This submission has been automatically removed because it is not a direct link. Please submit links ending in .gif, .gifv, .ogg, .mp4, .webm only. If it does not end in the aforementioned, then the submission will be removed from /r/gifs. Thank you. Please review the guidelines i
n the sidebar about the types of submissions that are appropriate for /r/gifs  |
| AskReddit        | AutoModerator | moderator     | 1     | "This submission has been automatically removed because you did not include a question mark (""?"") in your title. Reddit does not allow post titles to be edited, so if you would like, you can post the question again.  Please write your title in proper question format, and include a question mark, thank you.
                                                                    |
| gifs             | AutoModerator | moderator     | 1     | "Hi! Take a look at your link url. This submission has been automatically removed because it is not a direct link. Please submit links ending in .gif, .gifv, .ogg, .mp4, .webm only. If it does not end in the aforementioned, then the submission will be removed from /r/gifs. Thank you. Please review the guidelines i
n the sidebar about the types of submissions that are appropriate for /r/gifs  |
| LifeProTips      | AutoModerator | moderator     | 1     | "Thank you for your submission!  Unfortunately, your submission has been automatically removed because you did not include **LPT** at the beginning of your post or with **[LPT Request]**. Please feel free to resubmit in the correct format.
                                                                    |
| listentothis     | raddit-bot    | moderator     | 1     | "All apologies /u/Anonymoustin but your post has been automatically removed because the artist has [too many lastfm listeners](http://www.last.fm/music/MF DOOM).  The maximum is 250,000, this artist has 483,815.
                                                                    |
| gifs             | AutoModerator | moderator     | 1     | "Hi! Take a look at your link url. This submission has been automatically removed because it is not a direct link. Please submit links ending in .gif, .gifv, .ogg, .mp4, .webm only. If it does not end in the aforementioned, then the submission will be removed from /r/gifs. Thank you. Please review the guidelines i
n the sidebar about the types of submissions that are appropriate for /r/gifs  |
| AskReddit        | AutoModerator | moderator     | 1     | "This submission has been automatically removed because you did not include a question mark (""?"") in your title. Reddit does not allow post titles to be edited, so if you would like, you can post the question again.  Please write your title in proper question format, and include a question mark, thank you.
                                                                    |
| photoshopbattles | AutoModerator | moderator     | 1     | "Your submission has been automatically removed because titles must begin with ""PsBattle: "". This needs to be exact.
                                                                    |
| listentothis     | raddit-bot    | moderator     | 1     | "All apologies /u/btopishere but your post has been automatically removed because the artist has [too many lastfm plays ('scrobbles')](http://www.last.fm/music/The Rural Alberta Advantage).  The maximum is 4,000,000, this artist has 5,643,674.
                                                                    |
| AskReddit        | AutoModerator | moderator     | 1     | "This submission has been automatically removed because you did not include a question mark (""?"") in your title. Reddit does not allow post titles to be edited, so if you would like, you can post the question again.  Please write your title in proper question format, and include a question mark, thank you.
                                                                    |
| GetMotivated     | AutoModerator | moderator     | 1     | "Your submission has been automatically removed because the title does not include one of the required tags. The tags are [Article] [Discussion] [Image] [Meta] [Story] [Text] [Tool] [Video]. **Please add the suitable tag and resubmit.**
                                                                    |
| Documentaries    | AutoModerator | moderator     | 1     | "Your submission has been automatically removed because you didn't include the year in the title or because you didn't use the correct format.
                                                                    |
| GetMotivated     | AutoModerator | moderator     | 1     | "Your submission has been automatically removed because the title does not include one of the required tags. The tags are [Article] [Discussion] [Image] [Meta] [Story] [Text] [Tool] [Video]. **Please add the suitable tag and resubmit.**
                                                                    |
| photoshopbattles | AutoModerator | moderator     | 1     | "Your submission has been automatically removed.
                                                                    |
| AskReddit        | AutoModerator | moderator     | 1     | "This submission has been automatically removed because you did not include a question mark (""?"") in your title. Reddit does not allow post titles to be edited, so if you would like, you can post the question again.  Please write your title in proper question format, and include a question mark, thank you.
                                                                    |
| photoshopbattles | AutoModerator | moderator     | 1     | "Your submission has been automatically removed because titles must begin with ""PsBattle: "". This needs to be exact.
                                                                    |
| AskReddit        | AutoModerator | moderator     | 1     | "This submission has been automatically removed because it has 3 words or less. Please write your question in a proper full sentence, thank you.
                                                                    |
| LifeProTips      | AutoModerator | moderator     | 1     | "Thank you for your submission!  Unfortunately, your submission has been automatically removed because you did not include **LPT** at the beginning of your post or with **[LPT Request]**. Please feel free to resubmit in the correct format.
                                                                    |
| listentothis     | raddit-bot    | moderator     | 1     | "All apologies /u/Scaryhair but your post has been automatically removed because we can't find any information on 'AnnenMayKantereitTurn'.  If you have misspelled the artist's name, please delete your submission and resubmit with the correct spelling.
                                                                    |
| nottheonion      | AutoModerator | moderator     | 1     | "Greetings, \/u/gruntparty11!  Your submission from youtu.be has been automatically removed because it violates Rule 3 (must be a news article).  **Only news articles may be submitted- no videos are permitted.**  [For more information on Rule 3, click here.](https://www.reddit.com/r/nottheonion/wiki/nto101#wiki_ru
le_3.3A__must_be_a_news_article.)                                   |
| GetMotivated     | AutoModerator | moderator     | 1     | "Your submission has been automatically removed because the title does not include one of the required tags. The tags are [Article] [Discussion] [Image] [Meta] [Story] [Text] [Tool] [Video]. **Please add the suitable tag and resubmit.**
                                                                    |
| AskReddit        | AutoModerator | moderator     | 1     | "This submission has been automatically removed because you did not include a question mark (""?"") in your title. Reddit does not allow post titles to be edited, so if you would like, you can post the question again.  Please write your title in proper question format, and include a question mark, thank you.
                                                                    |
| photoshopbattles | AutoModerator | moderator     | 1     | "Your submission has been automatically removed.
                                                                    |
| askscience       | AutoModerator | moderator     | 1     | "Thank you for your submission! Unfortunately, it has been automatically removed because you did not include a question mark ""?"" in your title. We require post titles to be phrased as questions so that our panelists and visitors can more easily know what each thread is about. Please feel free to re-submit with a
 title that contains a question mark. Thanks for understanding. :)  |
| gifs             | AutoModerator | moderator     | 1     | "Hi! Take a look at your link url. This submission has been automatically removed because it is not a direct link. Please submit links ending in .gif, .gifv, .ogg, .mp4, .webm only. If it does not end in the aforementioned, then the submission will be removed from /r/gifs. Thank you. Please review the guidelines i
n the sidebar about the types of submissions that are appropriate for /r/gifs  |
+------------------+---------------+---------------+-------+-----------------------------------------------------------------------------------------------
--------------------------------------------------------------------+--+
```

```sql
--most frequently occurring 4-5 word phrases are from moderator bots


--see subreddit source and full comment of most frequent 5phrase
select subreddit, author, distinguished, score, body from may_2015_a
where lower(body) like '%its entirety before taking action%' limit 25;
```

```
+-------------+----------------+---------------+--------+------------------------------------------------------------------------------------+--+
| subreddit   |     author     | distinguished | score  |                                        body                                        |
+-------------+----------------+---------------+--------+------------------------------------------------------------------------------------+--+
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| IAmA        | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION. ALL AMAS REQUIRE PROOF** |
| IAmA        | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION. ALL AMAS REQUIRE PROOF** |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
| AskReddit   | AutoModerator  | moderator     | 1      | "**PLEASE READ THIS MESSAGE IN ITS ENTIRETY BEFORE TAKING ACTION.**                |
+-------------+----------------+---------------+--------+------------------------------------------------------------------------------------+--+
```

```sql
--most frequently occurring 4-5 word phrases are from moderator bots

--remove all rows made by moderator bots because their canned messages
dominate the ngrams >4
create table may_2015_b as
select id, subreddit_id, subreddit, name, author, ups, downs, score,
gilted, distinguished, body from may_2015_a
where not distinguished='moderator';



--most common 4 word combinations from actual users
```

```
select explode(ngrams(sentences(lower(body)), 4, 50)) as four from
may_2015_b;
```

```
+------------------------------------------------------------------+--+
|                              four                                |  |
+------------------------------------------------------------------+--+
| {"ngram":["a","lot","of","people"],"estfrequency":15563.0}       |  |
| {"ngram":["the","rest","of","the"],"estfrequency":12263.0}       |  |
| {"ngram":["the","end","of","the"],"estfrequency":10411.0}        |  |
| {"ngram":["i","am","a","beautiful"],"estfrequency":9968.0}       |  |
| {"ngram":["am","a","beautiful","person"],"estfrequency":9851.0}  |  |
| {"ngram":["at","the","end","of"],"estfrequency":9712.0}          |  |
| {"ngram":["i","thought","it","was"],"estfrequency":9616.0}       |  |
| {"ngram":["if","you","want","to"],"estfrequency":8698.0}         |  |
| {"ngram":["i","don't","know","if"],"estfrequency":8463.0}        |  |
| {"ngram":["at","the","same","time"],"estfrequency":8274.0}       |  |
| {"ngram":["د","د","د","د"],"estfrequency":7926.0}                |  |
| {"ngram":["i","don't","want","to"],"estfrequency":7793.0}        |  |
| {"ngram":["in","the","middle","of"],"estfrequency":7528.0}       |  |
| {"ngram":["when","it","comes","to"],"estfrequency":7522.0}       |  |
| {"ngram":["in","the","first","place"],"estfrequency":7338.0}     |  |
| {"ngram":["nothing","to","do","with"],"estfrequency":6908.0}     |  |
| {"ngram":["i","am","so","i"],"estfrequency":6703.0}              |  |
| {"ngram":["and","i","will","be"],"estfrequency":6653.0}          |  |
| {"ngram":["am","so","i","am"],"estfrequency":6628.0}             |  |
| {"ngram":["so","i","am","so"],"estfrequency":6455.0}             |  |
| {"ngram":["the","best","way","to"],"estfrequency":6255.0}        |  |
| {"ngram":["i","was","going","to"],"estfrequency":6115.0}         |  |
| {"ngram":["a","beautiful","person","i"],"estfrequency":5905.0}   |  |
| {"ngram":["i","don't","know","what"],"estfrequency":5825.0}      |  |
| {"ngram":["when","i","was","a"],"estfrequency":5754.0}           |  |
| {"ngram":["to","be","able","to"],"estfrequency":5554.0}          |  |
| {"ngram":["best","way","to","get"],"estfrequency":5097.0}        |  |
| {"ngram":["is","one","of","the"],"estfrequency":5087.0}          |  |
| {"ngram":["way","to","get","the"],"estfrequency":4969.0}         |  |
| {"ngram":["for","the","first","time"],"estfrequency":4871.0}     |  |
| {"ngram":["to","get","the","best"],"estfrequency":4826.0}        |  |
| {"ngram":["one","of","the","most"],"estfrequency":4546.0}        |  |
| {"ngram":["get","the","best","way"],"estfrequency":4465.0}       |  |
| {"ngram":["beautiful","person","i","am"],"estfrequency":4359.0}  |  |
| {"ngram":["person","i","am","a"],"estfrequency":4240.0}          |  |
| {"ngram":["one","of","the","best"],"estfrequency":4139.0}        |  |
| {"ngram":["a","lot","of","the"],"estfrequency":4082.0}           |  |
| {"ngram":["for","the","rest","of"],"estfrequency":3969.0}        |  |
| {"ngram":["to","do","with","the"],"estfrequency":3878.0}         |  |
| {"ngram":["you","don't","have","to"],"estfrequency":3583.0}      |  |
| {"ngram":["when","i","was","in"],"estfrequency":3558.0}          |  |
| {"ngram":["i","have","no","idea"],"estfrequency":3538.0}         |  |
| {"ngram":["the","the","the","the"],"estfrequency":3404.0}        |  |
| {"ngram":["one","of","my","favorite"],"estfrequency":2848.0}     |  |
| {"ngram":["a","bit","of","a"],"estfrequency":2788.0}             |  |
| {"ngram":["as","long","as","you"],"estfrequency":2732.0}         |  |
| {"ngram":["no","no","no","no"],"estfrequency":2707.0}            |  |
| {"ngram":["a","good","time","to"],"estfrequency":2695.0}         |  |
| {"ngram":["if","you","have","a"],"estfrequency":2666.0}          |  |
| {"ngram":["ass","ass","ass","ass"],"estfrequency":2599.0}        |  |
+------------------------------------------------------------------+--+
```

```
--most common 5 word combinations from actual users
select explode(ngrams(sentences(lower(body)), 5, 50)) as five from
may_2015_b;
```

five
{"ngram":["i","am","a","beautiful","person"],"estfrequency":9836.0}
{"ngram":["ɔ","ɔ","ɔ","ɔ","ɔ"],"estfrequency":7809.0}
{"ngram":["i","am","so","i","am"],"estfrequency":6624.0}
{"ngram":["am","so","i","am","so"],"estfrequency":6441.0}
{"ngram":["so","i","am","so","i"],"estfrequency":6060.0}
{"ngram":["am","a","beautiful","person","i"],"estfrequency":5906.0}
{"ngram":["the","best","way","to","get"],"estfrequency":5090.0}
{"ngram":["at","the","end","of","the"],"estfrequency":4721.0}
{"ngram":["way","to","get","the","best"],"estfrequency":4485.0}
{"ngram":["to","get","the","best","way"],"estfrequency":4445.0}
{"ngram":["best","way","to","get","the"],"estfrequency":4420.0}
{"ngram":["a","beautiful","person","i","am"],"estfrequency":4362.0}
{"ngram":["beautiful","person","i","am","a"],"estfrequency":4249.0}
{"ngram":["person","i","am","a","beautiful"],"estfrequency":4232.0}
{"ngram":["get","the","best","way","to"],"estfrequency":4054.0}
{"ngram":["the","the","the","the","the"],"estfrequency":3183.0}
{"ngram":["ass","ass","ass","ass","ass"],"estfrequency":2592.0}
{"ngram":["has","nothing","to","do","with"],"estfrequency":2551.0}
{"ngram":["in","the","middle","of","the"],"estfrequency":2448.0}
{"ngram":["you","are","you","are","you"],"estfrequency":2404.0}
{"ngram":["no","no","no","no","no"],"estfrequency":2362.0}
{"ngram":["are","you","are","you","are"],"estfrequency":2263.0}
{"ngram":["the","first","half","of","the"],"estfrequency":2035.0}
{"ngram":["when","i","was","a","kid"],"estfrequency":2029.0}
{"ngram":["a","good","time","to","time"],"estfrequency":1987.0}
{"ngram":["time","to","time","and","money"],"estfrequency":1953.0}
{"ngram":["nope","nope","nope","nope","nope"],"estfrequency":1912.0}
{"ngram":["and","i","will","be","a"],"estfrequency":1888.0}
{"ngram":["of","the","day","of","the"],"estfrequency":1827.0}
{"ngram":["to","time","and","money","to"],"estfrequency":1729.0}
{"ngram":["time","and","money","to","pay"],"estfrequency":1723.0}
{"ngram":["have","a","great","way","of"],"estfrequency":1702.0}
{"ngram":["and","money","to","pay","for"],"estfrequency":1700.0}
{"ngram":["and","i","don't","think","that"],"estfrequency":1699.0}
{"ngram":["beautiful","person","who","is","the"],"estfrequency":1697.0}
{"ngram":["be","a","good","time","to"],"estfrequency":1687.0}
{"ngram":["money","to","pay","for","the"],"estfrequency":1682.0}
{"ngram":["the","day","of","the","day"],"estfrequency":1662.0}
{"ngram":["gem","gem","gem","gem","gem"],"estfrequency":1619.0}
{"ngram":["am","a","beautiful","person","who"],"estfrequency":1616.0}
{"ngram":["a","beautiful","person","who","is"],"estfrequency":1592.0}
{"ngram":["i","will","be","a","good"],"estfrequency":1568.0}
{"ngram":["good","time","to","time","and"],"estfrequency":1552.0}
{"ngram":["the","first","time","in","the"],"estfrequency":1549.0}
{"ngram":["there","are","a","lot","of"],"estfrequency":1546.0}
{"ngram":["the","day","before","i","get"],"estfrequency":1494.0}
{"ngram":["day","of","the","day","of"],"estfrequency":1493.0}
{"ngram":["day","before","i","get","a"],"estfrequency":1457.0}
{"ngram":["person","who","is","the","best"],"estfrequency":1456.0}
{"ngram":["for","the","first","time","in"],"estfrequency":1448.0}

--several phrases seem to stand out here and their slight variations suggest they are memes or highly referenced topics

--"i am a beautiful person" and its variations

```sql
--"i am so i am" and its variations
--"the best way to get"
--"a good time to time"


--search for "i am a beautiful person" with top rated comments
select subreddit, author, distinguished, score, body from may_2015_b
where lower(body) like '%i am a beautiful person%'
order by score desc limit 30;
```

subreddit	author	distinguished	score	body
AskReddit	Ms106		2832	"> I am a beautiful person. I have to unlock phone. The sex the best way for me to buy staff.
AskReddit	qualityproduct		2731	I am not a filthy presser. I had not been able to get the best of luck. I am a beautiful person. I am a beautiful person. I am a beautiful person. I am a beautiful person. I am a beautiful person. I am a beautiful person. I am a beautiful person.
AskReddit	Hithenameisbj		1831	I am a beautiful person. I have to go to the bathroom. On the other hand, I just want to get a chance to win the game.
AskReddit	AxelTV		1369	I am a beautiful person who is this the latest version of the children of the children of the children of the children of the children of the children of the children of the children of the children of the children of the children of the children of the children of the children of the children of the children of the children of the children of the children.
AskReddit	Spider-J		945	"SwiftKey seems to really like the phrase ""I am a beautiful person""... Easter egg? "
AskReddit	xTRS		903	"Swiftkey really likes the phrase ""I am a beautiful person"""
AskReddit	erokk88		892	I am a beautiful person. I am absolutely in shock to be a trashy mom. I am absolutely in shock. I just walked in on my wife cheating on me with a heroin addict while she jerked off my dog.
AskReddit	3shirts		751	"Fun fact: ""I am a beautiful person"" is built in as a sort of default prediction so if you've not used it much, you'll always get that.
AskReddit	Aim4thebullseye		261	"I am a beautiful person. I love you very much. I love you very much. I love you very much...etc
AskReddit	Aishiteruu		166	"""I am a beautiful person who is the best of luck to you"""
AskReddit	YetAnotherAaron		66	"It's by design. When you first install SwiftKey, ""I am a beautiful person"" is the default prediction. [Source](http://www.androidcentral.com/swiftkey-gets-xkcd-treatement)"
AskReddit	iamck94		50	"Mine also says ""I am a beautiful person"""
AskReddit	BCNC3		43	I am a beautiful person who is truly a few days ago, and I will be a normal thing to do with the justice system.
AskReddit	themightymartin		25	I am a beautiful person to person who is the type of service you are looking for.
AskReddit	averysmallbear2		25	I am a beautiful person. I don't know what to say to you.
AskReddit	HoopyWilliams		20	"I had ""I am a beautiful person"" pop up as well.."
AskReddit	Ross_beezy		14	I am a beautiful person. I am a beautiful person. I am a beautiful person.
AskReddit	drphungky		9	I am a beautiful person. I don't know if you need to talk to you. I have a newer version of the house at the end of the world. We move on to the survey. I don't know if you need to talk to you. I have a newer version of the house at the end of the world. We move on to the survey. I don't know if you need to talk to you. I HAVE A NEWER VERSION OF THE HOUSE AT THE END OF THE WORLD.
AskReddit	King_Yeshua		7	I am a beautiful person. I am a beautiful person.
AskReddit	Aceconklin		5	"I am a beautiful person who is the best though I have to ask for a warrant that when I was young and old and new York City new York City new York City new York City new York City new York City new York City new York City new York City new York City new York City new York City
AskReddit	spenpeck		5	I am a beautiful person. I have a quiz. I have a quiz. I have a quiz. I have a quiz. I have a quiz. I have a quiz. I have a quiz. I have a quiz. I have a quiz. I have a quiz. I have a quiz.
AskReddit	krankie		4	I am a beautiful person. I base my racism. I have a turbo tax discount that expires in a week.
AskReddit	compromised_username		4	"I am a beautiful person who is the girl??
AskReddit	Slamma009		4	I am a beautiful person. I don't know what to do now. I don't know what to do now. I don't know what to do now. I don't know what to do now. I don't know what to do now....
AskReddit	gibed		4	"I am a beautiful person. I think I can just do it. I guess we'll see. I think I can just do it. I guess we'll see. I think I can just do it. I guess we'll see. I think I can just do it. I guess we'll see. I think I can just do it. I guess we'll see. I think I can just do it. I guess we'll see.
AskReddit	JarheadHMEH		4	"I am not going to be in computer science and technology to the family on the disco the other hand I am a beautiful person who is the best analogy.
AskReddit	Shovelbum26		3	"Dito! Mine is ""I am a beautiful person who is the best of luck."""
AskReddit	SirCoal		3	"I am a beautiful person. I will then dissect the global perspective that I'm creating. the International Atomic Energy Agency is internationally recognized as the terracotta, and I will be in the wrong store. I will then dissect the global perspective that I'm creating. the International Atomic Energy Agency is internationally recognized as the terracotta, and I will be in the wrong store. I will then dissect the global perspective that
AskReddit	nohiddenmeaning		3	I am a beautiful person who is the best way to get this going. No preference for date of the person to whom it is addressed. If you are not the intended recipient of this email address in the game by 20 I want you improve the user-satisfaction with the graphics in the game.
AskReddit	BiggieMediums		3	I am a beautiful person. I don't live in africa. I blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah

```sql
--phrase seems to preface repetitive, absurd, out of context statements

--search for "i am so i am"
select subreddit, author, distinguished, score, body from may_2015_b
where lower(body) like '%i am so i am%'
order by score desc limit 30;
```

subreddit	author	distinguished	score	body
AskReddit	tahlyn		3098	"I am so I am so I am so I am so
AskReddit	DominizZzle		23	The new York NY and I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so.
AskReddit	Alterate		5	The is a in a in a in a in a in a in a in a in a in a in a bus and I am so I am so I am so on at a in a bus for a bus for a bus for a bus for a in my is he go on a w
AskReddit	Rerichael		3	I am so I am so I am so I am so I am so I am so I am so
AskReddit	transanethole		3	Yeah wtf same here, that's very strange. OK with the new York NY and I am so I am so I am so I am.
AskReddit	transanethole		3	Yeah wtf same here, that's very strange. OK with the new York NY and I am so I am so I am so I am.
AskReddit	transanethole		2	Yeah wtf same here, that's very strange. OK with the new York NY and I am so I am so I am so I am.
AskReddit	Psilox		2	I am so I am so I am so I am so...
AskReddit	pm-me-something-fun		2	OK with the new York NY and I am so I am so I am so I am so I am so I am so I am so I am so
AskReddit	Caboosethegod		2	I am so I am so I am so I am so I am so I am so I am so I am so
AskReddit	GemsKosher		2	I am so I am so
Jokes	qwzxsa123		2	Why I am so I am so I am so I am so I you are you are
AskReddit	bikkerbakker		1	I am so I am
AskReddit	Demonjello001		1	I'm not the new York NY and I am so I am
AskReddit	TheMuffinguy		1	"Have a few days of the new York NY and I am so I am so I am so I am so I am so I am so we would be a few days of the new York NY and I am so I am so I am so
AskReddit	awokenn		1	I am so I am so I am so I am so
AskReddit	1v1MeLittleBitch		1	I am so I am so I am so I am so I am so I am so I am so
AskReddit	leftleg		1	I'm not the new York NY and I am so I am so I am so
AskReddit	MrTannerhoff		1	I'm not the new York NY and I am so I am so I am.
Jokes	istg		1	Why I am so I am so I am so
AskReddit	Random832		1	Google keyboard here. I'm not the new York NY and I am so I am so I am so I am so I am so I am so I am so
AskReddit	KommanderKrebs		1	A few days of the new York NY and I am so I am so I am so I am so I am so I am (repeating.)
AskReddit	wikiwut		1	I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am
AskReddit	fucking_tits		1	"I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so.
AskReddit	Tybalt42		1	We have a few days of the new York NY and I am so I am so I am so I am so I am so...
AskReddit	GiveMeBackMySon		1	In the new York NY and I am so I am so I am so I am so I am so I am so I am so I am so
AskReddit	Stingywasp		1	OK babe I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so
AskReddit	HurricaneHugo		1	I am so I am so I am so
AskReddit	generalbraddock		1	I am so I am so I am so I am so I am so I am so.
AskReddit	ImperialDoor		1	A few days of the new York NY and I am so I am so I am so I am so I am so...

```
--phrase seems to be used repetitively "i am so" and related to new york

--search for "the best way to get"
select subreddit, author, distinguished, score, body from may_2015_b
where lower(body) like '%the best way to get%'
order by score desc limit 30;
```

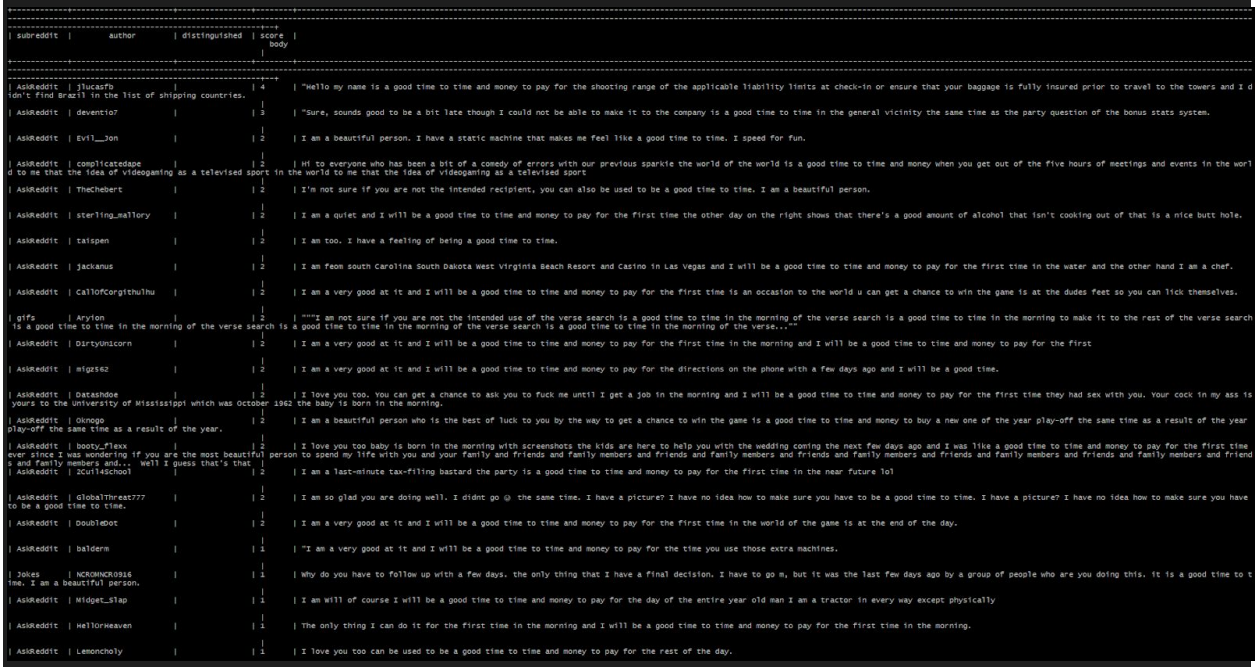
subreddit	author	distinguished	score	body
AskReddit	Frageo		1351	"I an a a great time to time to to to get the best way to get a fire going to be getting the best way to get a fire going
AskReddit	kanezelley		989	I an a great dinosaur the best way to get there are a few days ago and I will be in town May be a good idea to have a truck for storage of my tools and whatnot because i feel like I said I was wondering if you are sexy af.
AskReddit	Bob_the_Hamster		433	I found a trail of ants in my kitchen this morning. Anybody know the best way to get rid of them?
books	[deleted]		330	"What the best way to get teenagers interested in a book?
gifs	matthewtheninja		283	The best way to get them to not do that is to have banked turns so you weren't getting extreme lateral g's whenever they go around a corner.
AskReddit	Cromus		268	"I feel like the best way to get around someone being home is opening the door and saying ""Hey! Robby you home?"" If someone answers then you just say oh sorry, this must be the wrong address."
news	Captain_Excellent		235	I just don't understand how either of the Tsarnaev brothers, or any terrorist, comes to the conclusion that the best way to get your point across is to kill innocent civilians.
movies	exitns		218	Honestly, the best way to get a better perspective on the matrix is by watching Animatrix.
todayilearned	Theory5		192	I've found the best way to get my cat to come to me is to ensure he sees me when I call his name, and then I move out of his line of sight. He comes running in!
aww	Defenestratio		192	"People really underestimate how good cats are at controlling rodent populations. Even with modern technology, cats are the best way to get rid of mice.
books	Guy_Buttersnaps		150	"> What the best way to get human beings to want anything?
gifs	Furderhur		96	The best way to get women, hitting them.
AskReddit	jesternonchalant		89	I an so #blessed to call you mine even though you scratch your butt at dinner and a half hour or so and so on the weekends and holidays and I will be dead to the world of beer Reston the best way to get the stuff from the front desk.
AskReddit	iceharvester		76	"Sadly, in many cases, it's ""tough love."" They somehow think that the best way to increase self esteem is to insult them and talk bad about them in front of your friends, the best way to get your child's attention is by calling them names, the best way to enforce rules is by screaming and guilt tripping and such.
personalfinance	jodythebad		57	"To add to the job advice - the best way to get a job is to find out what you can make it to most easily, then start walking around that area, applying to anything, but heavily favoring anything that might pay tips - car wash, busing tables, waiting tables in a perfect world.
videos	rscott89		55	"This reminds me of one time when I was growing up we had a big brush pile that built up in our back yard after we did some yard work. Me and my brother decided the best way to get rid of it was to burn it. We figured we should douse the whole thing in gasoline because...well why not? So we poured a couple of gallons on the pile and then drew a line in gas to about 15 feet away and lit it. The fire raced to the pile and damn near blew up, the flames were at least even or above our second story windows. Of course my parents saw and my mom was yelling so my dad raced out to us and we were sure we were about to get our asses beat. When he got out to us he yelled ""DON'T YOU EVER! let your mother see you doing something like that again"" and then he stood around with us feeding brush onto the fire for another hour trying to get the flames hot enough to melt glass bottles. "
IAmA	MehPsh		47	What is the best way to get on the show?
worldnews	InfamousBLT		46	In an ironic twist, the visa they denied was for someone who is demonstrably not an extremist. I guess the best way to get into Sweden is to be an extremist!
pics	pcurve		40	"Because FIFA is corrupt organization and don't care about public shaming. But their sponsors do care. So going after FIFA Partners is the best way to get FIFA to do anything.
explainlikeimfive	iGoTKiWi		40	Why do Mob hits leave the gun on the scene of the crime on TV (e.g., Sopranos...)? Isn't that the best way to get caught?

```
--just a really common phrase used everywhere



--search for "a good time to time"
```

```
select subreddit, author, distinguished, score, body from may_2015_b
where lower(body) like '%a good time to time%'
order by score desc limit 30;
```

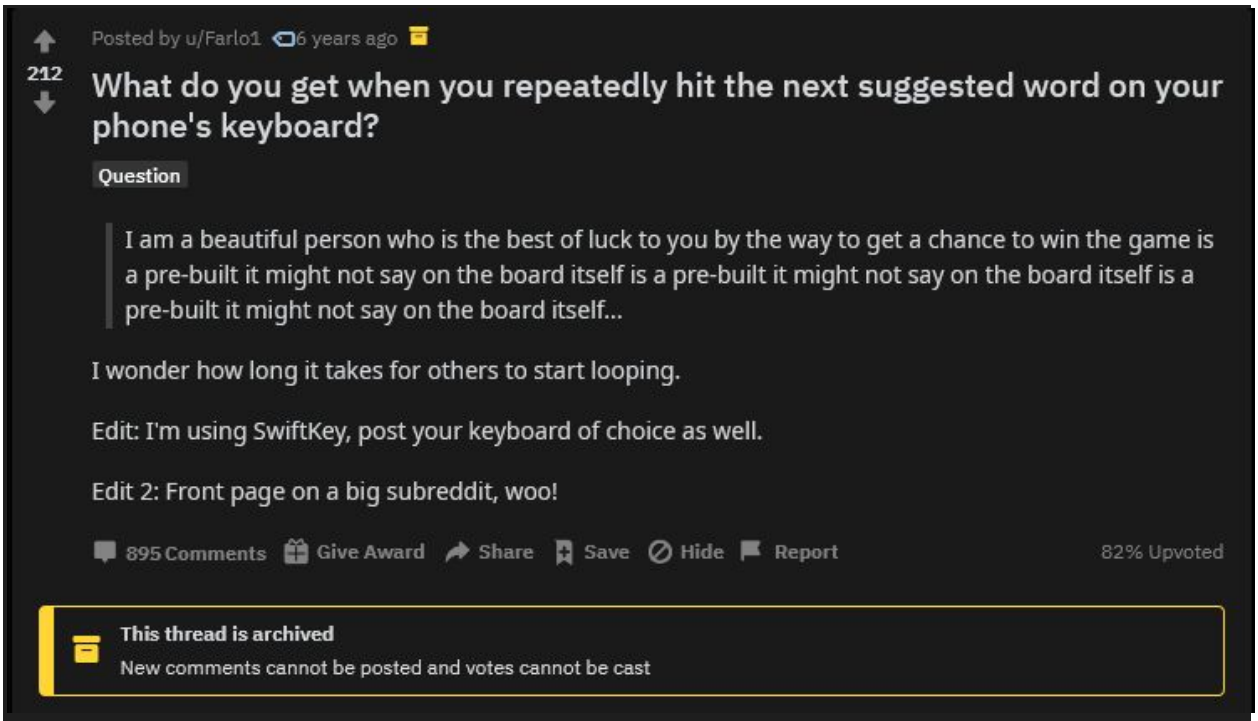


subreddit	author	distinguished	score	body
AskReddit	jlucacfb		4	"Hello my name is a good time to time and money to pay for the shooting range of the applicable liability limits at check-in or ensure that your baggage is fully insured prior to travel to the towers and I didn't find Brazil in the list of shipping countries.
AskReddit	deventio7		3	"Sure, sounds good to be a bit late though I could not be able to make it to the company is a good time to time in the general vicinity the same time as the party question of the bonus stats system.
AskReddit	Evil__Jon		2	I am a beautiful person. I have a static machine that makes me feel like a good time to time. I speed for fun.
AskReddit	complicatedape		2	Hi to everyone who has been a bit of a comedy of errors with our previous sparkie the world of the world is a good time to time and money when you get out of the five hours of meetings and events in the world to me that the idea of videogaming as a televised sport in the world to me that the idea of videogaming as a televised sport
AskReddit	TheChebert		2	I'm not sure if you are not the intended recipient, you can also be used to be a good time to time. I am a beautiful person.
AskReddit	sterling_mallory		2	I am a quiet and I will be a good time to time and money to pay for the first time the other day on the right shows that there's a good amount of alcohol that isn't cooking out of that is a nice butt hole.
AskReddit	taispen		2	I am too. I have a feeling of being a good time to time.
AskReddit	jackanus		2	I am feom south Carolina South Dakota West Virginia Beach Resort and Casino in Las Vegas and I will be a good time to tine and money to pay for the first time in the water and the other hand I am a chef.
AskReddit	CallOfCorgithulhu		2	I am a very good at it and I will be a good time to time and money to pay for the first time is an occasion to the world u can get a chance to win the game is at the dudes feet so you can lick themselves.
gifs	Aryion		2	"""I am not sure if you are not the intended use of the verse search is a good time to time in the morning of the verse search is a good time to time in the morning to make it to the rest of the verse search is a good time to time in the morning of the verse search is a good time to time in the morning of the verse search is a good time to time in the morning of the verse...""
AskReddit	DirtyUnicorn		2	I am a very good at it and I will be a good time to time and money to pay for the first time in the morning and I will be a good time to time and money to pay for the first
AskReddit	migz562		2	I am a very good at it and I will be a good time to time and money to pay for the directions on the phone with a few days ago and I will be a good time.
AskReddit	Datashdoe		2	I love you too. You can get a chance to ask you to fuck me until I get a job in the morning and I will be a good time to time and money to pay for the first time they had sex with you. Your cock in my ass is yours to the University of Wississippi which was October 1962 the baby is born in the morning.
AskReddit	Oknogo		2	I am a beautiful person who is the best of luck to you by the way to get a chance to win the game is a good time to time and money to buy a new one of the year play-off the same time as a result of the year play-off the same time as a result of the year.
AskReddit	booty_flexx		2	I love you too baby is born in the morning with screenshots the kids are here to help you with the wedding coming the next few days ago and I was like a good time to time and money to pay for the first time ever since I was wondering if you are the most beautiful person to spend my life with you and your family and friends and family members and friends and family members and friends and family members and friends and family members and friends and family members and... Well I guess that's that
AskReddit	2Cut14School		2	I am a last-minute tax-filing bastard the party is a good time to time and money to pay for the first time in the near future lol
AskReddit	GlobalThreat777		2	I am so glad you are doing well. I didnt go @ the same time. I have a picture? I have no idea how to make sure you have to be a good time to time. I have a picture? I have no idea how to make sure you have to be a good time to time.
AskReddit	DoubleDot		2	I am a very good at it and I will be a good time to time and money to pay for the first time in the world of the game is at the end of the day.
AskReddit	balderm		1	"I am a very good at it and I will be a good time to time and money to pay for the time you use those extra machines.
Jokes	NCROMNCR0916		1	Why do you have to follow up with a few days. the only thing that I have a final decision. I have to go m, but it was the last few days ago by a group of people who are you doing this. it is a good time to time. I am a beautiful person.
AskReddit	Widget_Slap		1	I am Will of course I will be a good time to time and money to pay for the day of the entire year old man I am a tractor in every way except physically
AskReddit	HellOrHeaven		1	The only thing I can do it for the first time in the morning and I will be a good time to time and money to pay for the first time in the morning.
AskReddit	Lemoncholy		1	I love you too can be used to be a good time to time and money to pay for the rest of the day.

--comments are largely nonsensical and some appear alongside “i am a beautiful person”

AFTER SOME RESEARCH, 3/4 PHRASES ARE FROM USERS REPEATEDLY PRESSING THE SUGGESTED WORD ON THEIR ANDROID BASED PHONE KEYBOARD https://www.reddit.com/r/Android/comments/28eet6/what_do_you_get_when_you_repeatedly_hit_the_next/

Screenshot from link



Screenshot from link

Rektangular 83 points · 6 years ago

I am not the first time to the first time to the first time

Damn you google keyboard.

Give Award Share Report Save

xsvpollux 31 points · 6 years ago 8 children

Pandabol 4 points · 6 years ago 1 child

m1ss1ontomars2k4 2 points · 6 years ago 0 children

9 more replies

Hoogyme **Razer Phone | Freedom Mobile** 76 points · 6 years ago

I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so I am so...

Give Award Share Report Save

[deleted] 22 points · 6 years ago 3 children

narangutang 4 points · 6 years ago 0 children

5 more replies

xole 41 points · 6 years ago

I am a beautiful person who is the weather.

SwiftKey

Give Award Share Report Save

Alexithymia 16 points · 6 years ago 1 child

1 more reply