

Reddit Threads Monitoring

This Project will help investigator explore what's going on Reddit more easily and find out the attitude of people towards threads.

Author: Yi Qin, He Huang

Project page (on Github): <https://github.com/qyyMriel/Reddit-Monitor>

Working demo: <http://redditmonitor17.herokuapp.com/>

Problem Statement

One sentence summary: our project aims to help investigators explore what's going on Reddit, what top threads are and the attitude of people on these threads and what time people contribute to these threads.

Exploring threads activity over time and people's attitude on those threads is difficult for investigators if they look at the Reddit page directly. The raw thread page has so much information that investigator can't extract valuable information they needed, especially the activity of thread over time and the attitudes of people towards the thread.

Thus we want to help investigators filter and visualize the useful thread data so that they can understand the activity of thread and people's attribute on threads.

Visualization Questions

Given the nature of our data, we assume few things about investigator: the investigator has the intention to explore more about the dataset and perhaps want to pinpoint a certain entity for investigation.

According to our data from Reddit api, Reddit threads/posts could be used for this entity.

Questions about the threads:

- What top 30 threads are, what are their title keywords and Reddit score ?
- What subreddit do they belong to ?
- What attitudes do people have on these threads, positive or negative ?
- What time slot does certain thread attract users' focus?

Data Attributes

Our original data is from Reddit API and because we use Node.js to build backend framework, we use snoowrap package to extract thread data. We only extract some useful attributes through the API and transfer these attributes into new form so that we can use them at front-end.

The derived attributes are as below:

Title keyword: We used “keyword-extractor” to extract keyword from title, the reason we do this is that original title sometimes is too long and can't be understood by the investigator quickly.. Extract only meaningful keywords from title can provide user concise and direct information about thread title.

Comment_distrubution: we want to analyze the comment time distribution of thread, so we first extract time of each comment and do summation of each time range, finally we build an array to record these summations, index of array represents time range.

Sentiment: we analyzed the top 30 comments in each thread. Then calculate a value that can represent the overall sentiment of these comments in each thread . In this way, user can know the attitude of people towards this thread.

Here is form that contains all the attributes:

Attribute Name	Attribute Type	Attribute Meaning	Value Range/Categories	Derived or not
threadId	categorical	the id of a particular forum	All threads	Not
title	unstructured text	Thread title	All threads	Not
titleKeyword	Array of string	Thread title keyword	Depends	Yes
urls	categorical	Thread content url	All threads	Not
score	Quantitative	Score of thread		Not
subReddit	categorical	subReddit thread belongs to	All threads	Not
value	Numerical		50	Yes
comment_dis	Numerical Array	The distribution of comment across 24 hour, each slot represents 1 hour range.	Depends.	Yes
sentiment	Numerical	Sentiment from thread comments	(-1,1)	Yes
radius	unstructured text	Control	50	No
max	Numerical	The maximum number of comment_dis	Depends	yes

Data(JSON) exported to front-end is like below:

Related Works

We take some ideas from following projects

Previous works has been done on other kinds of events monitoring or recording.

Digital Transformations Event Tree Map

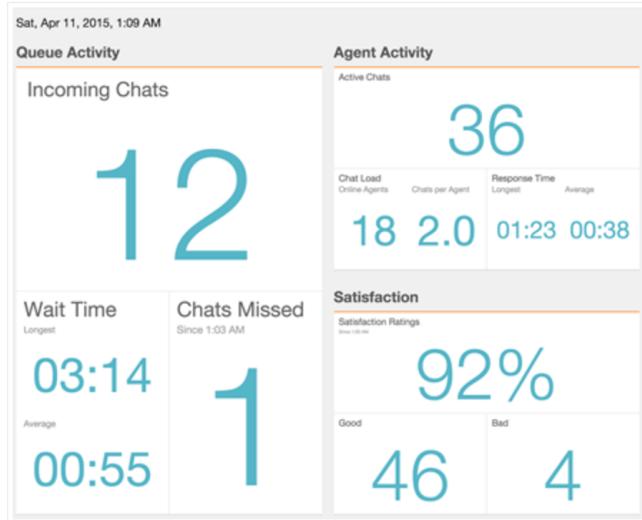
By [Wendy Matcham](#)



- Each section represents groups of challenges which the attendees thought were similar.
- Hover Over or Tap any challenge to see the longer description submitted by the attendees
- Click or Tap a colour name to see a video describing the area challenges
- **Relation:** This project groupby the similar answers together and uses different colors to identify different groups, that means we could group our thread object data by subReddit section and each section has unique color hue as identifier. Another hint we get from this treemap is that they just show short description of contents, in our project, we only show the title keyword, not title itself.

Monitoring real-time chat metrics

By [Nora Mullen](#)



Description: an overview of key chat metrics, including queue size, customer wait times, and chat satisfaction, on a single screen.

- Real-time dashboard for recording important chat metrics.
- Three main blocks, queue activity, agent activity and satisfaction
- Each main block is divided by specific chat metrics and they are shown real-time.
- **Relation:** In our project, we groupby data according to their subreddit. In our original thought, when we step into thread level, we still use this kind of block to show different statistics of thread.

Keyhole

One of top 25 social media monitoring tools

Real-time Tracker: **Soccer**

DEC 21 - DEC 22



Top Posts

RT / Likes	Klout	Recent
 ren @milinghao Dec 2		1,604 ↗
allee looks like a proud mother cheering for her sons after winning a soccer match https://t.co/WfssnsWV5j		
 Soccer @TrueSoccerLife Dec 21		1,346 ↗
Bob Marley: Soccer is a whole skill to itself. A whole world. A whole universe to itself. Soccer is freedom. https://t.co/qFVgePCTT1		
 Football Funnys @FootballFunnys 7:57 am		994 ↗
 Soccer is a whole skill to itself. A whole world. A whole universe to itself. Soccer is freedom. - Bob Marley. https://t.co/ZbAdicPou		
 Morgan Magazine @supermorgy 4:37 am		977 ↗
Passenger PLANE with SOCCER TEAM CRASHES in COLOMBIA https://t.co, #Colombia #plane_crash #Soccer #chancenope #Colombia https://t.co/		

Related Topics

Most Influential

A horizontal scrollable bar displaying a variety of Twitter user profiles and logos, likely representing different brands or entities. The visible items include:

- Engagement
- Klout
- Frequency

The scrollable area contains numerous small profile pictures and logos, such as:

- TRUE CRIME LIFE
- THE WALKING DEAD
- USA
- FOX NEWS
- SPIDER KING
- TWITTER
- LEAGUE OF LEGENDS
- YOUTUBE

Recent Users



KEYHOLE

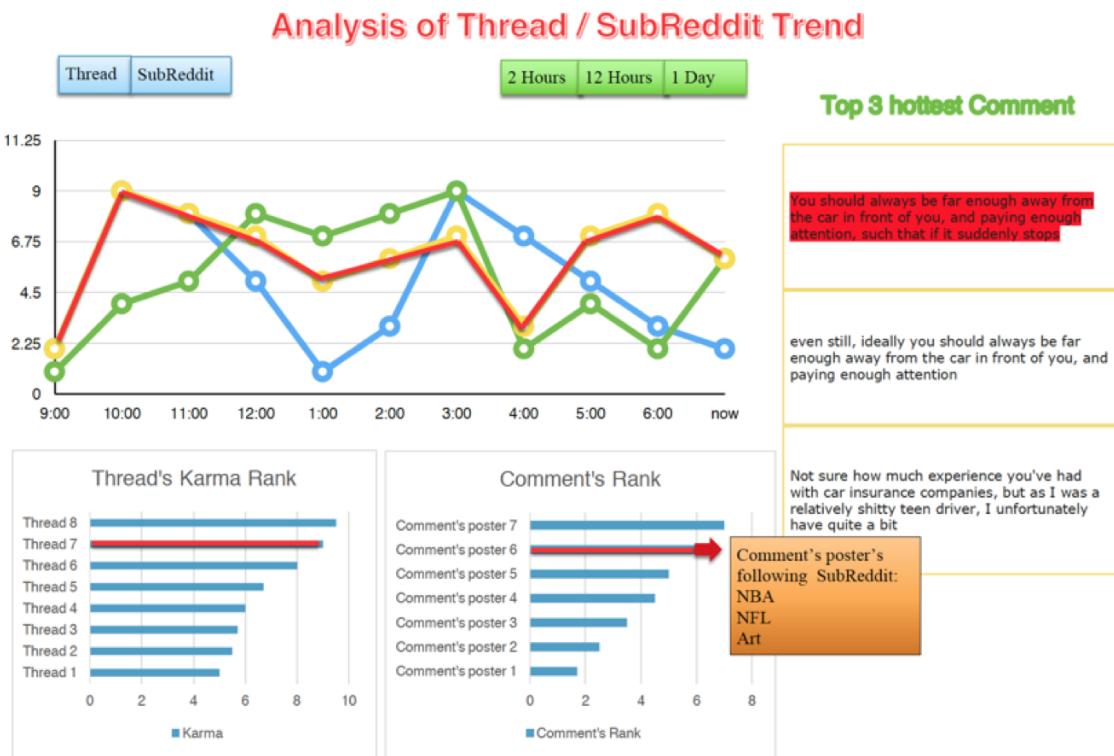
Description: Real-time media monitoring visualization and analysis

- This dashboard tool shows many information about certain topic

- It shows timeline, top posts, related topics, keywords and user information
- **Relation:** our project shows the comment distribution across time and we also extract keywords from title.

Design Iterations

Initial Mockup:



Our initial mockup wants to show the thread activity, its relative rank and most popular comment as well as hottest comment, the top line graph shows the timeline of thread across time. The left graph on the bottom shows thread ranks according to their score. The right graph on the bottom shows comment rank according to comment karma score, its vertical axis shows the comment poster name. The right graph shows the specific hottest comment of thread. If we click a comment, the thread containing this comment will be highlighted, in addition, comment content will also be highlighted.

Our initial mockup is denied by professor, so we change our idea about project in the progress report.

Progress Report:

What's going on in Reddit now?

This bowl season, the biggest game is also the most entertaining

Jaguars, Broncos combined for 875 punting yards

Indianapolis Colts vs. New York Jets: LIVE score updates, stats, fan chat (12/5/16), NFL Week 13

HEISMAN 2016: The spotlight isn't always easy for Jackson

The Latest: Incoming NC gov. talks unity after concession

As Cowboys are first to clinch a berth, NFL playoff picture begins to take shape

AP Poll: Defending champion Villanova is the new No. 1

Jeff Brohm moving from Western Kentucky to Purdue

Purdy: The Melanon signing shows Giants are same—but it's

Rays Could be without Team for

Bulls suspend Rajon Rondo one game for conduct

Cavaliers' Smith exits game with knee injury

Amazon's new grocery store is missing one of the most important things about its business

House GOP Signals Break With Trump Over Tariff Threat

Supreme Court Brexit case: 'No need' for MPs to get final say

Apple's Critique Of Driverless Car Policy Doesn't Clarify Its Own Plans

Mexico Oil Auction Succeeds In Drawing World's Biggest Drillers

Italy Vote Increases Pressure on Banks

Arts and Humans Argue a Merger's Benefits as Arbitrator Trial Begins

Bank strike lift Dow to a new record, and tech shares climb

Castro, The Mob And Capitalism's Full Circle

Review: Oculus' New Controllers Fix the Rift's Biggest Shortcoming

The Last Guardian review: can heart make up for technical problems?

For Europe's Unity, 2017 Will Be a Year of Reckoning

Eying climate role, Ivanka Trump to meet with A.G. Gatz

Carney Lays Out Vision to Revive Benefits of Globalization

Bank strike lift Dow to a new record, and tech shares climb

Georgia father sentenced to life in prison for the hot car death of his 22-month-old son

Review: Oculus' New Controllers Fix the Rift's Biggest Shortcoming

The Last Guardian review: can heart make up for technical problems?

Asia stocks bounce as risk appetite returns after Italy vote

Ford Officially Eyes Rate Hike See Path Fed to Fiscal Policies

Trump's Health Care Plan: What It Would Mean for You

Global central bankers to meet in December

French left stakes its hopes on Manuel Valls in presidential race

Review: Oculus' New Controllers Fix the Rift's Biggest Shortcoming

The Last Guardian review: can heart make up for technical problems?

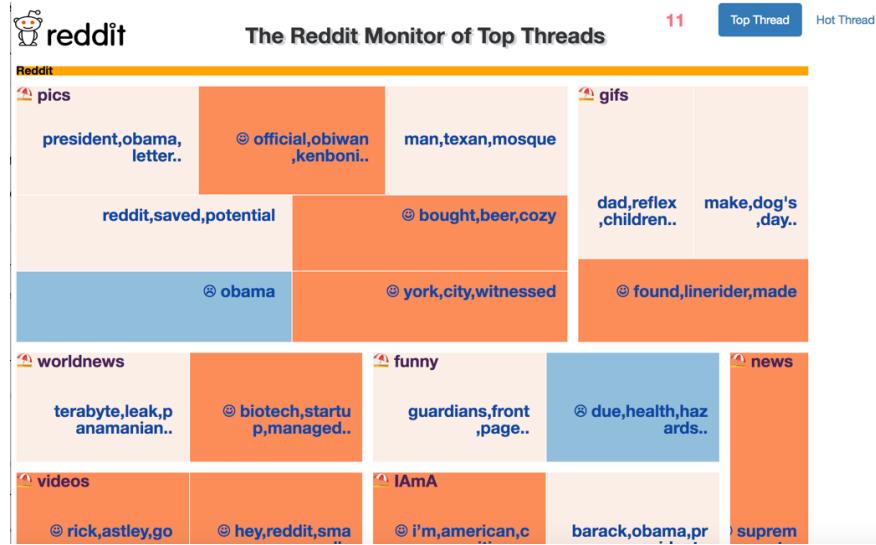
This time, we changed our project visualization structure into treemap, each small block represents a thread and shows the thread title contents. If threads belong to same subreddit, then we groupby them together, different color hue represents different subreddit. Here, the color saturation in subreddit represents the attitude of people towards thread, the color is more darker, corresponding attitude is more negative. As for interaction, if we click one thread, a small window will show us the details of thread, like url, contents and keyword of the thread.

Final Visualization

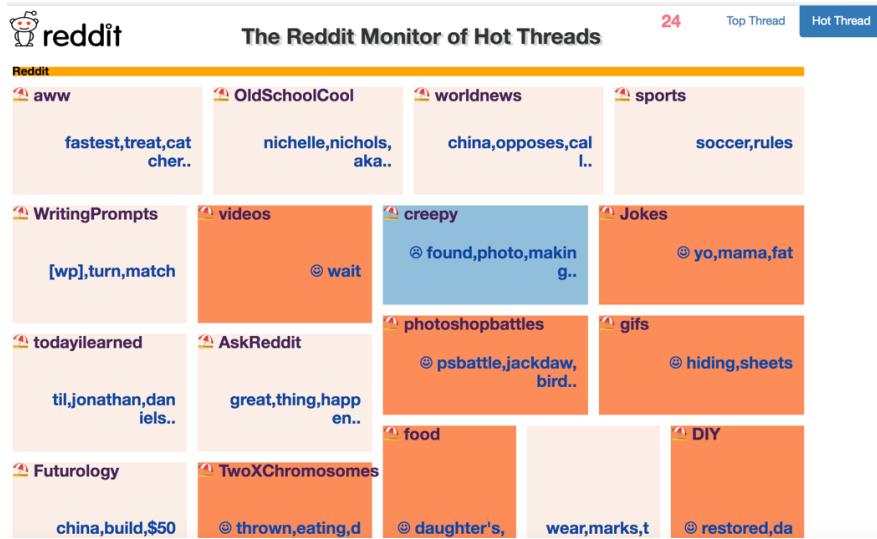
The Treemap in our final visualization has three levels. Besides the treemap in each level, each level contains a control panel and the directory. In the control panel, there is count down timer, it will shows how many seconds left to refresh the data(We refresh the data every 60 seconds). And there is a pill that we can choose the top threads to visualize or hot threads to visualize. The Directory can show the user which level you are at, and if you click the directory, you can return to the previous level.

1. The first level shows all the general observations of each thread. Different subreddits are divided by larger white gaps and subreddit names are in black font color which are shown on the top left of frame. Each small box represent a thread which are divided by smaller gaps. The blue name shows the title keyword of thread. At this level, the size of small box doesn't represent anything, we make them in same size because in our view, each thread is equally

important. If their sizes are decided by other thing like thread score or something else, bigger box has more chance to be clicked than smaller box. The emoji at the front of thread **title keyword** and the **background** of each small box represents the attitudes of people towards thread. Smiling **emoji** and orange background means positive, sadness emoji and dark blue **color** means negative, and all the others illustrates the sentiment is mild.



The Top thread monitor first level interface



The Hot thread monitor first level interface

2. The graph below shows the second level, when you click any thread on first level, the visualization will be transferred into graph below with slow transitioning effect. The second level is based on subreddit and subreddit information is shown in the directory, for example, below graph represents the thread in subreddit *pics*. Its attributes **title keyword**, **background color** and the **emoji** still represent the same thing in the first level, besides these attributes, we add **hottest comment** in each thread at this level, so the user can see the comments details for the corresponding thread.

The Reddit Monitor of Top Threads

Reddit / [pics](#)

reddit,saved,potential	bernie	obama
But what actually happened with those two girls?	Its been quite a while since ive been rolled so well.	You know, as a conservative libertarian I swore I'd h
divorced,lost,job	president,obama,letter	man,texan,mosque
Stupid question: How are people making the captions f	I'm surprised he could do that back flip with balls t	Excited to see McDonald's start using this on all of

The Top thread monitor second level interface

The Reddit Monitor of Hot Threads

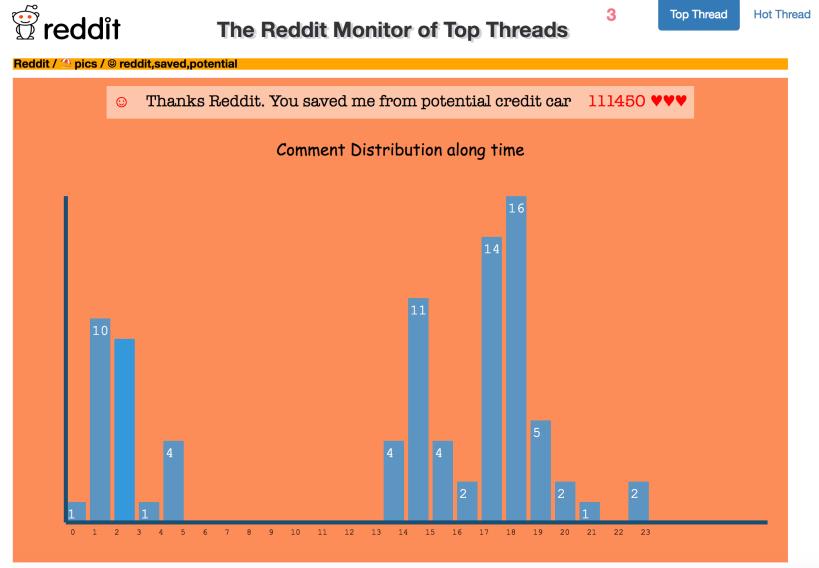
Reddit / [worldnews](#)

china,opposes,call
2016: when China became the reasonable voice in

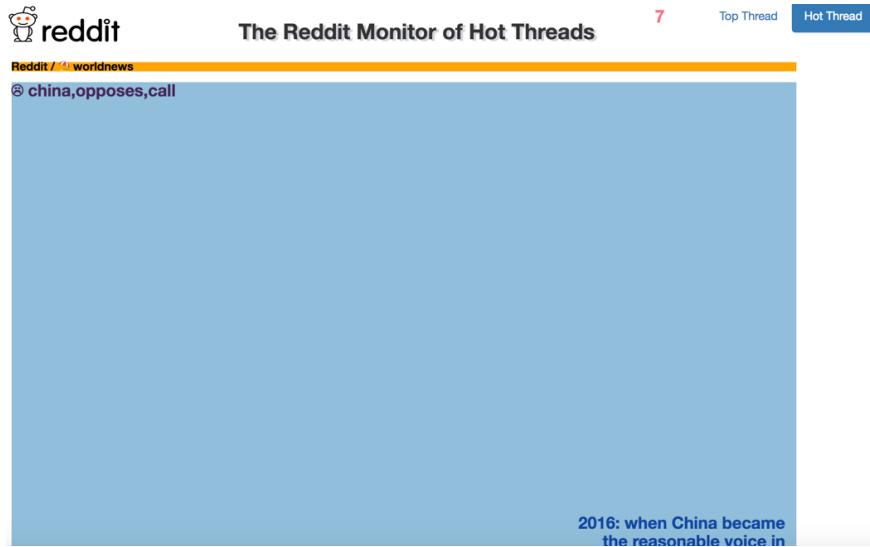
The Hot thread monitor second level interface

3. The graph below shows the third level, when you click any thread on second level, the visualization will be transferred into graph below with slow transitioning effect. The **title keyword and subreddit information** will be displayed in the **directory**. The **background color** represents the attitudes of people towards this thread, this color inherits from second level. The text within white frame is the **actual title** of thread, this title could be clicked which render a page showing thread contents. Moreover, the number following the title represents the **thread score**. The **bar chart** below shows the distribution of comments across time. Because Reddit API rule, we couldn't extract all

comments. And we assume those comments could represent whole comments distribution across time.



The Top thread monitor third level interface



The Hot thread monitor third level interface

Findings

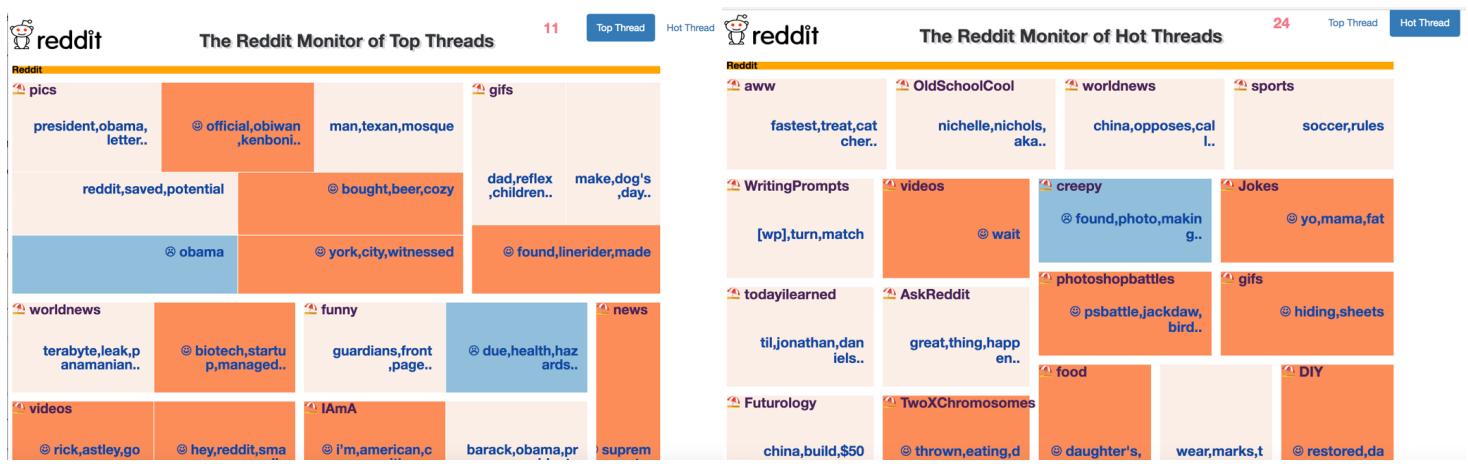
Our project implements real-time Reddit Monitoring, the investigator could know what's happening on Reddit, including current top threads and current hot threads, so our project don't have much data to do analysis but we can still find some interesting patterns from project. Here, top threads are ranked by total scores of threads since they are posted. Threads become hot threads because they are discussed and viewed most in most recent time.

1. Top threads have less subreddits than hot threads.

After clarifying these two concepts, we can find that some interesting phenomenon exists from below two graphs. As we can see, in left graph, there exists only 7 subreddits and *pics* subreddit contains five threads. In right graph, there exists 25 subreddits, each subreddit just hold only one thread. Not like left top threads graph, right hot threads graph changes frequently and each subreddit only holds one thread at most time.

2. *Pics* and *gifs* are always the the most popular subReddits in Top Thread

We find that *pics* and *gifs* are always the most popular subreddit, there are always at least 3 top threads ranked in the top 30. We can find that people more like the picture-type threads than other types like videos, text etc. Picture is a very direct and easiest way for user to understand what's happening now or express something.

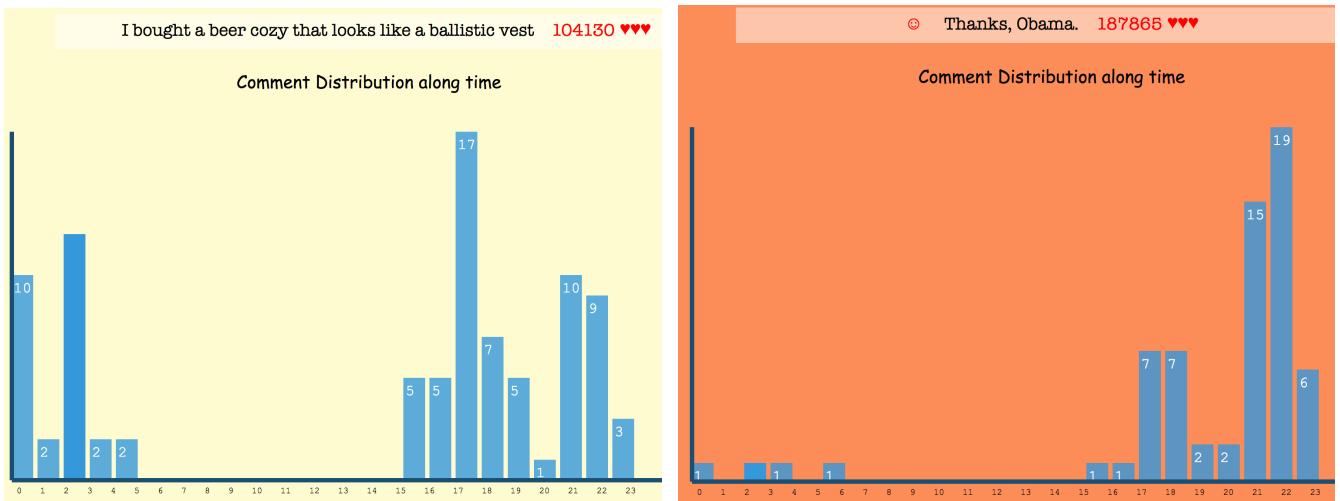


Top thread

Hot Thread

3. Users are always active at some time slot.

The most active time slots of all threads are from 17:00 to 23:00, even though it has already been proved by tons of surveys from social scientists, we also find the same result. We visualized the comments distribution of top threads in detail, as you can see the graph below, Reddit users' most active time is in the night, maybe the night time is the causal time after work and study.



4. Attitudes of users tend to be neutral

As to most trending threads, people will take a neutral attitude about them, in another word, the amount of positive sides or negative sides are equal, every user will express his real feeling or attitude about threads, they may debate with each other, the final result of people's attitude is neutral. Not too many threads have obvious positive or negative attitude unless they are some hot-button issues.

The screenshot shows a grid of 16 subreddits from the 'Hot Thread' section of The Reddit Monitor of Hot Threads. Each subreddit has a small icon, a name, and a brief description.

- aww**: fastest,treat,cat cher..
- OldSchoolCool**: nichelle,nichols, aka..
- worldnews**: china,opposes,cal l..
- sports**: soccer,rules
- WritingPrompts**: [wp],turn,match
- videos**: @ wait
- creepy**: @ found,photo,makin g..
- Jokes**: @ yo,mama,fat
- todayilearned**: til,jonathan,dan iels..
- AskReddit**: great,thing,happ en..
- photoshopbattles**: @ psbattle,jackdaw, bird..
- gifs**: @ hiding,sheets
- Futurology**: china,build,\$50
- TwoXChromosomes**: @ thrown,eating,d
- food**: @ daughter's,
- DIY**: @ restored,da

Limitations and Future Works

The major limitations of our project for now is that we don't have enough data analysis. If we had more time, we'd like to add user data to find some connections between user activity and threads. For example we could show which user contribute most to this thread, what their comments are, how often their activity in this thread or what time they posted their comments.