Visualizing Common Words and Sentiment Trends Related to Burning Man from 2009-2015

Jennifer Lu
Stanford University
jenylu@stanford.edu

Christina Kao
Stanford University
chris18@stanford.edu

ABSTRACT

In this paper, we describe the motivation for and design of an interactive visualization on text-based Burning Man datasets and Twitter feeds, to help both outsiders and experienced Burners learn more about the cultural trends of the festival over the years. We look into comparing types of words based off of word count, themes, entities, queries, and categorization which is detailed further under Methods. User are able to filter and view words based off of these sections as well as by the type of dataset and by the year. In addition we present sentiment coloring of words to better visualize positive or negative emotions for certain topics. Finally, to view context of a word under Twitter, users can hover over a bolded and italicized word in order to read the tweet that it came from. At the end of this paper, we discuss the user studies and feedback we received as well as future directions for this project.

INTRODUCTION

What is Burning Man?

Burning Man is an annual music and art gathering that takes place in the Black Rock Desert in Nevada. The event name comes from the symbolic, ritual burning of a large wooden effigy called the Man. Each year a different theme is proposed for the festival which art installations base their work off of as well as camp themes, and the design of the Man.

Problem and Motivation

Currently, existing visualizations of Burning Man are mainly static, and focus less on the underlying cultural trends and the context of what is happening in the real world. We hope to expand on Burning Man visualizations by performing NLP analysis on datasets from different sources allowing us to examine (1) the correlation among Burning Man theme, event, art, camp datasets and Twitter feeds over the years (2) current event context (3) cultural trends (4) demographic trends and (5) sentiment. This would be interesting to see from a historical point of view or to allow people who are new to Burning Man to learn more about it.

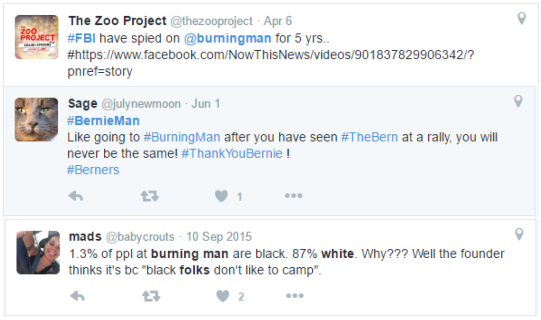


Figure 1: Burning Man tweets related to current events, politics, demographics

While sifting through Twitter feeds related to Burning Man, we were able to see current event references with #OccupyBurners and #BernieMan. There was one year in 2015 where it was released that Burning Man had been under FBI surveillance since 2010 for anti-terrorist security and the hashtag #fbi was most popular then. There have also been critiques on the demographic trends and lack of black attendees. Cultural trends include things such as mediation, yoga, and self exploration. Other negative sentiment was related to being unable to attend Burning Man and #fomo or fear of missing out. Just from looking at short snippet words from Twitter, we were able to see a lot of the past events that have happened during Burning Man, and also the festival's relation to outside current events.

RELATED WORK

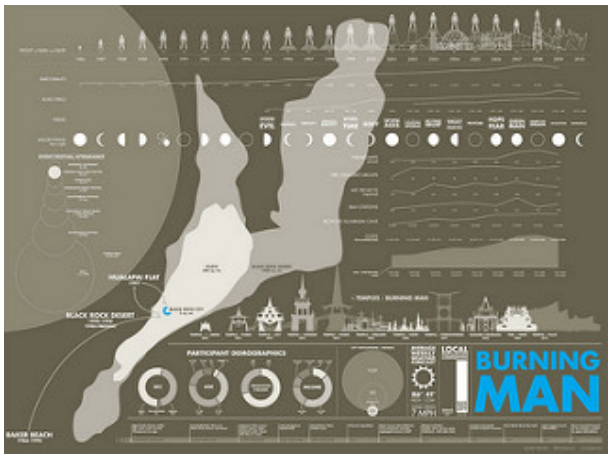
Burning Man Visualizations

Whether static or interactive, most of the existing Burning Man visualizations display fact-based information and do not incorporate deeper analysis on the content or outside data related to the event. Our project is similar to existing visualizations and projects in that we also utilize the official Burning Man events and arts datasets and compare trends across the years. Yet in addition, we also compare the Burning Man data with outside sources such as Twitter and perform NLP analysis in order to extract the underlying theme and cultural trends.

The Burning Man Official site has a showcase of open source projects, mainly mobile applications that map out events and locations at Black Rock City, and allows filtering of events, camps and art listings. Flint Hahn Hahn [2010] created a Burning Man poster that shows how the event changed over

the years, on information such as the height of the man, number of participants, ticket price, official theme, moon phase on burn night, the structure of the Temple etc. Another static visualization by Unknown [2015] shows the amount of human activity at each location in Black Rock City during the event in 2015. The activity data is tracked with human.co. Finally an interactive timeline Ranger et al. on the official Burning Man site illustrates how the event changed over the years for the mans structure and number of participants. Clicking on each year shows more detailed information regarding Burning Man that year, such as themes, photos, interviews, art works etc.

Figure 2: Burning Man Graphic by Flint Hahn



Music Festival Visualizations

As we look at visualizations for music festivals in general, we found more variety, a lot of which are designed with specific purposes in mind. The Spotify visualization McDonald helps to connect people, the Belgium mobile system is designed for safety by Barco [2016], and the Uber Fan by Schaffer [2013] shows demographic profiles to let marketers, companies or organizations better understand their audiences. Our project is similar as we focus on solving a specific issues or achieving a particular goal as well - namely to surface the underlying cultural trends of Burning Man events over the years, and how outside events influence the festival and vice versa. However, the exact problem we aim to solve is different from the prior works that we have found. In addition, these three examples also gave us inspiration for the different aspects of our dataset that we could look into if time permits, such as how we could connect people with similar interests at the event, or give outsiders better insight into the demographics of Burning Man audiences.

Spotify Visualizes the Entire Planet as a Never-ending Music Festival: an interactive map that shows when two people are listening to the exact same song at the exact same moment all over the world. The International Music Festival, Belgium - a mobile, networked visualization system that help keep festival-goers safe - describes a large scale visualization system deployment for monitoring the music festival, generating real-time views in order to keep participants safe. Finally the Uber Fan is a static visualization showing the technology, entertainment and social spendings and preferences of a typical Uber Fan.

Social Media Visualizations

We found several visualizations related to social media and more specifically Twitter. The Nokia Internet Pulse Kaye et al. [2012] for instance visualizes recent Twitter discussion around a particular topic. There were four main ways that this was displayed: time series of stacked tag clouds, words sized proportionally to frequency, words colored according to the emotional content of the tweet, and clicking on words to show a list of tweets that contained these words. Similarly we wanted to visualize popular and relevant word clouds for every year at Burning Man and to be able to look closer at a specific word.

Another paper showed traffic information based on tweets and presented it in map view for mobile applications Endarnoto et al. [2011]. The traffic information included time, origin, destination, and conditions. This inspired us to think about whether or not we could use our twitter data for a specific purpose such as creating a visualizing of which locations at Burning Man are most popular.

A third paper Smith et al. [2014] discussed different ways to map Twitter topic crowds such as with a polarized crowd, tight crowd, brand clusters, community clusters, broadcast network, and support network. This would show the relationships that communities had over different topics on Twitter; for instance, where a polarized crowd would be between two opposing sides. We thought about how we could display patterns in Twitter conversations about Burning Man and its relation to outside conversation. These papers offer lots of inspiration for displaying words, relationships among crowds on Twitter, and use of real-time updates. One thing that we wanted to focus on however, was the inclusion of multiple types of datasets while these papers only focus on one dataset. In our case, we would include data on themes, events, art, and camps provided by the official Burning Man site and also look at the relationship of Twitter conversations directly related to Burning Man and events happening outside.

DATA

Burning Man has a public dataset on their site from the years of 2009-2015 that includes information on their past themes, art, camps, and events in csv form. For the art dataset, each row includes information on distance, time address, description, artist, url, id, year, and name. The camp dataset includes information on the url, camp id, year, name, and description. The events dataset includes the camp host name, description, title, url, year, time, id, printed description, event type, and location. The themes dataset was one column description.

Our second main dataset was tweets from 2009-2015 during the Burning Man festival periods with the term ”burningman” included, scraped using Jefferson Henrique’s github project GetOldTweets-java’s command-line queries cited under Henrique. An example query for tweets from the year

2015 is as follows, "java -jar got.jar querysearch=burningman since=2015-08-30 until=2015-09-07".

METHODS

We had two main types of analysis for our data which were word count and sentiment.

Word Count

For word count, we eliminated common words like 'the', 'and', 'burning man', 'black rock city'. The rest of the words excluded can be found in the appendix section. Then for each source and year, the top 20 most frequent words were outputted in order.

Sentiment Analysis

For our sentiment analysis, we used a Sentiment Analysis, Categorization and Named Entity Extraction Tool called Semantria from Lexalytics. For each of our datasets, we filtered out either the description or name (if description was not available) to run through the tool. This gave us back sentiment analysis for four different extraction techniques, - themes, entities, queries, and categories.

Themes are lexically important noun phrases. They represent the buzz words of the documents and give a feel of what people are saying. Entities are proper nouns such as person, place, or product. Categorization labels large bodies of text by popular topics taken from the top 500
Wikipedia articles transferred into about 120 categories like 'Math', 'Automobiles', 'Boats' etc. An example would be with the sentence "I wonder who will win in the California mid-term congressional elections?" categorized as the topic "Politics". Finally queries are simple search categories or boolean phrases.

Each one of these sections lists the top words found and the count for negative, neutral, and positive sentiment as well as a word cloud. The word cloud has each word colored by its sentiment from a gradient scale of red to green where red is negative and green is positive and size based off of the frequency of the word.

For each word, we calculated a sentiment score based on the number of occurrences of the negative, neutral, and positive words. This was calculated by (negative * -1 + neutral * 0 + positive * 1) / (total count)). We correspond this score with the word's sentiment color from a gradient of red to green where red is negative and green is positive. Thus, -1 is most red to 0 which is yellow to 1 which is most green.

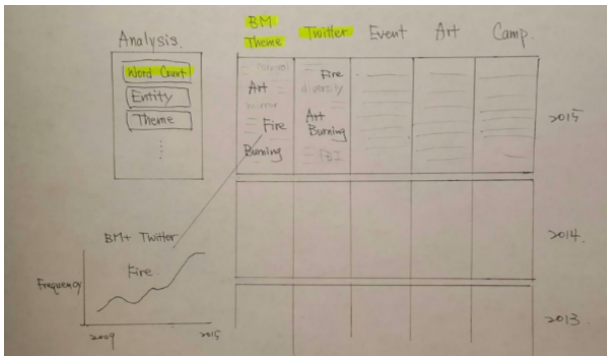
PROCESS

With the goal of analyzing the content and cultural trends over the years based on type of dataset and top entity words or themes, we came up with two initial visualization ideas.

Design Idea - 1 : Grid Layout

The first idea (Figure 3) lays out the words in grids - each grid corresponds to a unique dataset / year / NLP technique combination. Choosing multiple datasets highlights their intersection, and clicking on a word shows how its trend changed over the years.

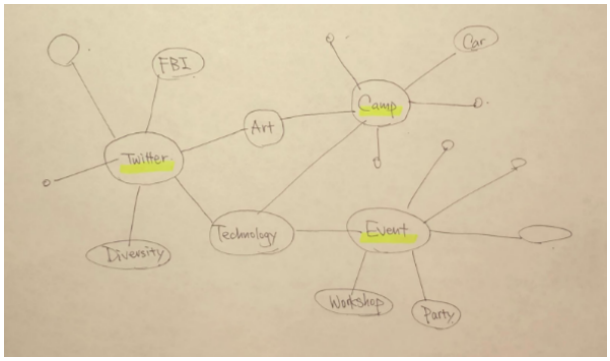
Figure 3: Grid Layout Visualization



Design Idea - 2 : Network / Node Layout

The second idea (Figure 4) shows each dataset type and word as a node, and each dataset node is connected to its top words. If there are word overlaps among datasets, only one word node would be shown, and the datasets that contain the word would have a link to it. There would be one network / node graph for each unique year and NLP Technique combination.

Figure 4: Network / Node Layout Visualization



After receiving feedback from the teaching staff and classmates, we decided to adopt the first design idea, as it's more compact and easier to compare the trends over the years. We decided not to incorporate the functionality that highlights intersecting words on additional dataset selection, as having no overlapping words would gray out all of the words in the selected datasets, and might confuse the user. In addition, instead of having a pop-up chart that shows an individual word's trend, when the user clicks on a word, we highlight the word across all selected datasets and years.

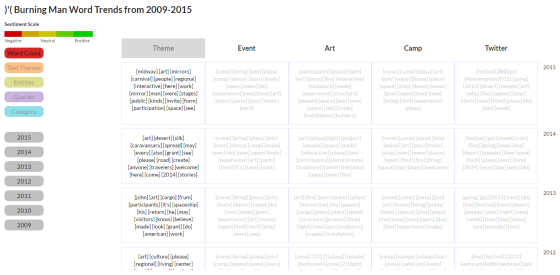
RESULTS

Our System and its Visualizations

Our system is a web-based application that uses D3.js for processing and dynamically updating data, and Semantic UI for styling and layout design. Figure 5 shows the layout of the system's interface. On the left, there are two sets of filters - the NLP techniques (as described in the previous section)

are in colored buttons, and the years in gray buttons. And on top of the filters is a color scale that shows the labeling of different sentiments. To the right is the interactive visualization, with all words grayed out initially as no datasets are selected by default. Each grid contains the top twenty words in descending order, sorted based on word counts, for the corresponding year, NLP technique and dataset.

Figure 5: System interface



There are three main interactive features that can be performed on the visualization:

(1) Filter results by NLP technique, year and dataset(s)

Figure 6 and Figure 7 show results from different NLP techniques. Figure 8 shows results from only two years, 2015 and 2009, and Figure 9 has the Theme, Event and Art datasets selected.

Figure 6: NLP Filter - Results by Word Count

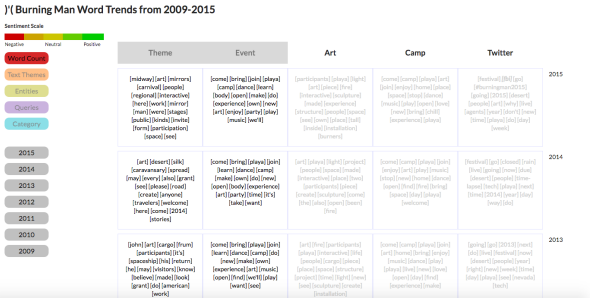
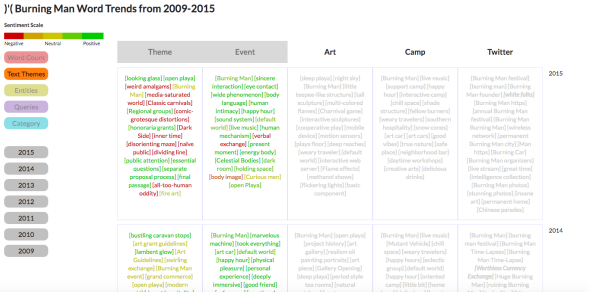


Figure 7: NLP Filter - Results by Theme



(2) Select and highlight individual words among datasets of interest

Selecting any non-grayed out word highlights the word amongst selected datasets, and grays out everything else, allowing the

Figure 8: Year Filter - Results from 2015 and 2009 for Word Count

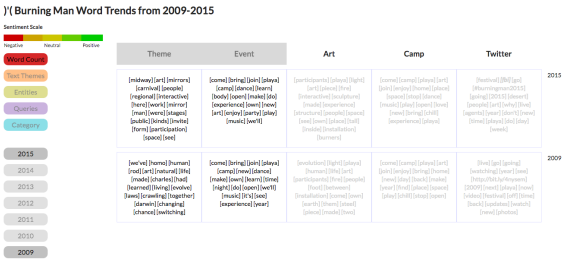
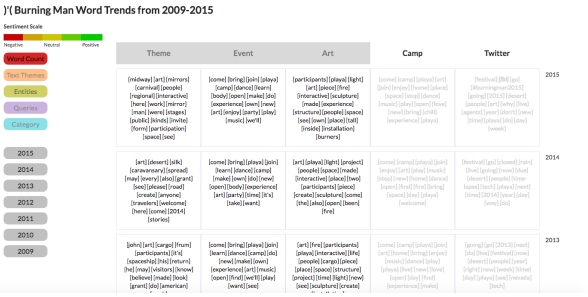
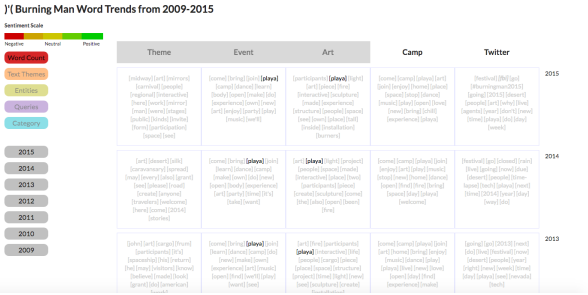


Figure 9: Dataset Filter - Results for Theme/Event/Art for Word Count



user to focus on specific phrases or themes of interest, and see how the trend(s) for the word changed over the years and how it compares across different datasets. Figure 10 highlights the word "playa" amongst selected datasets, indicating its importance in both events and arts, yet not a common term in the official Burning Man theme descriptions.

Figure 10: Highlighting Individual Words



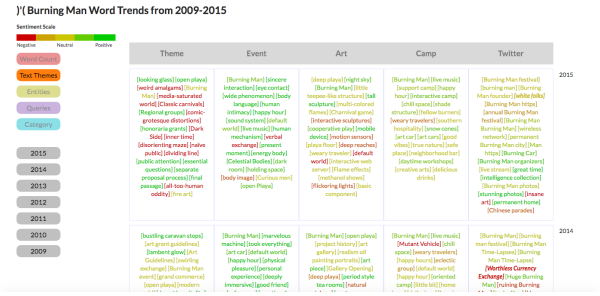
(3) Color words based on sentiment

Sentiment data is included for all NLP techniques except word count. As described in the Methods section, green indicates positive and red negative sentiments. Figure 11 shows the sentiment labels for the top words in Theme, Event and Art when analyzed by the algorithm for theme extraction.

(4) Show tweet context for words under Twitter

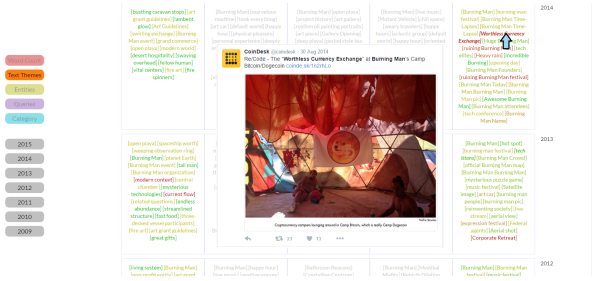
For a select few words under the Twitter column, users can hover over a bolded and italicized word to see the tweet that

Figure 11: Words Colored based on Sentiment



the word came from and read about the word's context as seen in Figure 12. We felt that this background added much more to the cultural trends that these words had and also understanding to what was happening at Burning Man in relation to the word. Because of time constraints and resources however, we currently only show this for a select few words under the technique, Text Themes. Trends that we were able to see with the added tweets were things like discovering the FBI had surveillience over Burning Man for the past 5 years. We can also see that Burning Man has become more and more tech saturated over the years as words like 'tech titan' and 'bitcoin' emerge. We will discuss more about the trends we discovered in the Discussion section.

Figure 12: Words Tweet Hover over Bold and Italicized Word



User Studies and Feedback

We conducted user studies with fellow students and colleagues and received feedback from them on the interaction. Currently it takes the site sometime to load all of the words when toggling between techniques which can lead to some confusion for users. One thing we can do to combat this is add in a loading icon. Currently the technique tags will all become a lighter shade instead of one being opaque if it is still loading.

Another type of feedback we saw was that users did not know how to immediately interact with the site. They did not always know that the column headers could be selected and would try to click the words first to find that nothing changed. We decided to have the Theme column already selected as the default state in order to solve this issue so that users would see that the columns can be selected and deselected. Some inquired as to why the datasets were not all selected by default. We felt that the large number of words and sentiment coloring would make the information appear too overwhelming and cluttered if they were all shown at once.

Our users also gave us suggestions or requested features on the design of the visualization to enable better comparison among the different results or for more targeted insights. Some examples include sizing words by frequency, allowing the user to enter specific queries to find trends, or when individual words are highlighted, take out the grayed out words completely to make it easier to compare across the years.

DISCUSSION

The main contribution of our project is successfully enabling users to, in a much more efficient manner, accomplish the goals we initially set out to achieve - namely to examine (1) the correlation among Burning Man theme, event, art, camp datasets and Twitter feeds over the years, (2) current event context, (3) cultural trends, (4) demographic trends, and (5) sentiment.

Whether it be people who did not even know about the festival or those who are experienced "Burners", most were able to observe the general change in trends, identify specific words or phrases of interest, or comment on the sentiment changes after interacting with the system in a minute or two. For instance, while looking through the Tweets over the years, we can see that Burning Man has become increasingly tech saturated as mentioned in the Results. Words like "open source cell phone service", "tech titans", and "bitcoin" have emerged and queries like "technology" have been prominent over the last 6 years. This has also translated into the latest demographic and diversity critiques that Burning Man has a lack of black attendees and is becoming saturated with white "techies".

Our current design requires some initial exploration by the user in order to extract insights. Most of the feedback we received has been on increasing the flexibility and diversity of user queries and more compact visual representations to better compare different results and learn about the underlying trends. Based on our development process and the feedback we received, we detail some of our ideas for further directions and research in the following section.

FUTURE WORK

We have five ideas for future work - four of which improve on our current system, and one generalizes it to able to carry out analysis on other types of text-based datasets. (1) Automate end-to-end workflow: currently we process the NLP results separately, and our implementation directly visualizes on the processed data. To increase the utility and scalability of the product, we could automate the end-to-end workflow - to allow the user to upload raw .csv or .json files, performs analysis, and generate interactive visualizations in real time. This would then set the foundation for enabling the system to handle more (2) diverse user queries, allowing the user to tailor the results to their particular interests or questions in mind. (3) Provide the user with more context: create functionalities that provide the user with more context of the dataset(s), for instance, total word count. (4) Improve

comparisons of words across years since users need to scroll through years with the current implementation. Taking out irrelevant words when selecting an individual word could be helpful or somehow incorporating a second graph that shows yearly trends. (5) Generalize to other text-based content: test our product on other types of text-based datasets, such as blogs, news articles, conference papers or books, and improve the system to allow it to carry out efficient and effective analysis and comparisons on a wider array of text-based datasets.

APPENDIX

List of common words removed from "word count"

[the, The, is, are, there, They, a, and, of, will, your, to, in, for, you, that, with, our, on, be, at, or, we, from, by, when, as, an, This, it, can, us, this, out, into, get, some, -, &, have, their, A, which, its, has, We, about, they, who, up, all, so, just, was, what, than, not, i, burning, man, burningman, each, like, no, while, other, but, more, one, if, through, how, only, where, none, :, ., i'm, my, black, rock, city, #burningman, @burningman, me, !, these, those, am, within, around, many, http://, https://, http://www., ?, (,), it's, —, via, ..., rt, https://www., 's, –]

References

Barco. Barco, Feb 2016. URL https://www.barco.com/en/references/2016-02-08---international-music-festival-belgium.aspx.

Sri Krisna Endarnoto, Sonny Pradipta, Anto Satriyo Nugroho, and James Purnama. Traffic condition information extraction & visualization from social media twitter for android mobile application. In *Electrical Engineering and Informatics (ICEEI), 2011 International Conference on*, pages 1–4. IEEE, 2011.

Flint Hahn. Burning man infographic - flint hahn, Nov 2010. URL http://fh.io/burning-man-infographic/.

Jefferson Henrique. A project written in java to get old tweets, it bypass some limitations of twitter official api. URL https://github.com/jefferson-henrique/getoldtweets-java.

Joseph'Jofish' Kaye, Anita Lillie, Deepak Jagdish, James Walkup, Rita Parada, and Koichi Mori. Nokia internet pulse: a long term deployment and iteration of a twitter visualization. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 829–844. ACM, 2012.

Kyle McDonald. Spotify visualizes the entire planet as a neverending music festival. URL http://www.fastcodesign.com/3034685/spotify-visualizes-the-entire-planet-as-a-neverending-music-festival.

Danger Ranger, Brian Doherty, and Kevin Evans. 1986. URL http://burningman.org/timeline/#!/2011.

Jared Schaffer. The uber fan - concerts and festivals, Apr 2013. URL http://visual.ly/uber-fan-concerts-festivals.

Marc A. Smith, Lee Rainie, Ben Shneiderman, and Itai Himelboim. Mapping twitter topic networks: From polarized crowds to community clusters, Feb 2014. URL http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/.

Unknown Unknown. Human activity at black rock city - burning man [oc], Sep 2015. URL https://www.reddit.com/r/dataisbeautiful/comments/3k4l09/human_activity_at_black_rock_city_burning_man_oc/.