Data Viz Final Project Paper

Julian McClellan

Due 6/4/17

Introduction

The popular social media site, Reddit, is composed of a variety of sub-communities, or "subreddits". One such subreddit, r/ChangeMyView (CMV):

> . . . is a subreddit dedicated to the civil discourse of opinions, and is built around the idea that in order to resolve our differences, we must first understand them. We believe that productive conversation requires respect and openness, and that certitude is the enemy of understanding.

Essentially, participants in CMV participate in good-faith debate. Users post submissions, containing their views and supporting reasoning, and other users then attempt to change these views and the original poster (OP) of the submission often engages with responding users. If a user successfully changes the OP's view, then the OP can award a "delta" to the comment that changed his or her opinion. These deltas are tracked so that anyone can see if a view has been changed, or if it hasn't changed.

Unfortunately, the topics of the opinions, which can be visually skimmed from the subreddit itself, are largely unsearchable, thanks to Reddit's infamously inadequate search functions. Additionally, while the history of every author can be browsed via their Reddit user page, there is no such tool to quickly summarize the content of their submissions, or trends in submission activity and score. All of the current methods for exploring CMV submissions and their authors were created for participation, not exploration. This Flexdashboard/Shiny application is a tool built to allow users to explore long term delta trends in opinion (in)stability, with fine grained filtering procedures, to easily find information about individual posts and their authors by specific topic, and finally, to view the trends in an individual author's submission activity, score, subreddit participation.

Methods

Introduction and Glossary

Firstly, in order for the viewer to more easily interpret and utilize the application's visualizations, a quick explanation of what CMV is, what data is collected, and what a few key terms are, is presented. In this way, the following visualizations have the potential to be more truthful, functional, insightful, and enlightening. A link to the subreddit is also provided to allow the viewer to become more familiar with the CMV if they choose.

CMV Submissions Delta Trend

Beginning with a macro perspective, this first tab presents an interactive smoothed timeline (a frequency polygon) of stable opinions and changed opinions, a word cloud of the most frequently used words in the content of submissions (more frequently used words in larger font), and hidden behind a tabset, a searchable table of submissions. The word cloud provides a single image look at the submission content, while the searchable table offers a more detailed look into *all* of the text in individual submissions. The left sidebar offers 6 slider options to filter the submissions included in the timeline, word cloud, and table. The title on the timeline shows exactly how many submissions meet the user's filtering criteria. By default, all the submissions in 2016 are shown, and the user has the option to filter by the date of the submissions, adjust the number of days used to smooth the timeline, and can filter by specific attributes of the submissions. The user can filter by the range of the submission's sentiment, total comments, direct replies to the submission, and responses from the original poster (OP).

Search CMV Submissions by Title

In this tab, users are able to view individual CMV submissions by searching the title for keywords. In the upper section, viewers can see the username of the author, the sentiment on a scale from -1 to 1 (red to green) of the submission, the number of total comments, direct replies, and author comments of the submission (which they were able to filter by in the previous section), the number of submission prior to the CMV submission the author has made, the average and total scores of those submissions, and finally a measure of subreddit participation inequality based of the Gini Index, to give a quick numerical summary of the diversity of their subreddit participation.

In the lower section the viewers are initially presented with the content of the submission (in plain text form), but they also have access to a word cloud of the author's previous submission content, a similar function to the word cloud shown in the previous tab and another word cloud that contains the subreddits the author has posted most in (with more favored subreddits displayed in larger font). Similar to the first tab, there is also a searchable table of submissions for the author in question, in case the viewer wants a closer look at submission content.

CMV Author Submissions

Lastly, in this tab, the user can view information concerning a specific author's *entire* submission history. Using the left column, the viewer can search for a specific author, and then select what time of timeline to display, a frequency polygon timeline of that author's activity (with smoothing control available), or a timeline of that author's submission scores over time. In the upper section, the name of the author is displayed as well as the number of submission that author has ever made to CMV. In the timeline below, no matter what version of the timeline is shown, vertical colored lines indicate where a CMV submission took place (every author has at least one).

In the lower section, similarly to the previous two tabs, a word cloud of submission content words is shown by default, with a word cloud to show favored subreddits and a searchable table.

Discussion

Truthful

The delta-trend tab provides an extraordinary amount of control over the submissions that the user desires to see, and the CMV submissions and CMV author submissions tabs offer a very detailed level of information concerning the individual submission, and an individual author, respectively. However, in the delta-trend and author tab, a filtering control has the potential to obfuscate the truth: the smoothing parameter. The smoothing parameter can be utilized at the viewer's discretion, with a smoothing of 1 day serving as the most fine-grained option that summarizes the least amount of information, but a timeline with 1 day smoothing is rather difficult to interpret, which can obfuscate easy viewing of information, and thus, up to 30 (365) days of smoothing is allowed for the delta-trend (author) tab, but even this

option can gloss over trends occurring within 30 (365) days. Used correctly, with a balance between too little and too much smoothing, the smoothing control more clearly communicates the truth of the opinion (in)stability trend. The default smoothing value of 7 (30) days, is an attempt to strike a balance between information overload and an overly parsimonious timeline.

The word cloud operates in a similar manner to the smoothing parameter, striking a balance between too much and too little information. While the user cannot adjust the cloud's settings, they can utilize the filters, or select the CMV submission or author, to adjust what submissions have their content displayed through the word cloud, in this way, the user can have a swift, one image summary of the content of various slices of submissions, and the subreddit's favored by an author. If the viewer desires more information/truth, then the searchable tables allow for in-depth exploration.

Functional

The application offers targeted functionality to view delta trends, individual CMV submissions and their authors, and the entire submission history of CMV submission author's in 2016. The delta-trends tab offers a fine-tuned level of filtering control, as well as dual levels of depth for analyzing submission content. The CMV submissions tab quickly summarizes important information about the submission, and the author, and displays the whole content of the submission, so that the viewer can have the whole of the opinion and it's reasoning available. If the viewer is then curious about the author's history before that submission, they can utilize the tabbed word cloud of their content and subreddits, as well as a more detailed searchable table. If the viewer is interested in the *all* of the submissions of that author, before and after, then they can utilize the author tab for a closer look.

Beautiful

The application makes use of timelines, intuitively displayed (and sometimes) colored text boxes, and if needed detailed tables. Smoothing parameters often offer adjustments to the "beauty" of timelines. Each tab is by default configured to show the most aesthetically pleasing and swift summary of the information, with the option, through tables, and adjustment, for the viewer to investigate further.

Insightful

Each tab in the application overs insights into overall trends, submissions and their authors, or authors individually. There are a plenty of features germane to each submission, but in the delta-trend tab, the viewer is at first presented with no restrictions based on these features. The viewer can then create their own filtering procedures, and have the opportunity for spontaneous insights by looking at the timelines at word clouds to see how the opinion (in)stability trend and content of the submissions are for their given slice of data. Additionally, knowledge building insights can come from one of the many data tables in all of the tabs. The other two tabs are conducive to spontaneous insights, as they also feature easy to read and clearly presented information with word clouds, timelines, and large textual boxes to summarize submission and author characteristics.

Enlightening

The application enables the exploration of a forum where good faith debate takes place, in regards to the forum's topics of debate and participants in those debates. In a polarized political climate, lending the spotlight to a niche of the internet where many types of views are engaged in all sorts to form a high quality discussion is a good demonstration of the alternatives to entrenched modes of thinking. At the same time, the application shows that not every opinion changes, and that time, sentiment, and author interaction affect the stability of an opinion. The application also sheds light on who puts their opinions out there for debate, showing what other communities they participate and their activity. Persuasion is a ubiquitous process, and exposing the unique landscape of r/ChangeMyView highlights this process in the internet medium.