Project Resilience

Investigating Resilience Using Twitter

Aysha Emmerson

December 13, 2019

Abstract

"Project Resilience" is the launchpad into a larger exploration of the themes and attitudes that surround the construct of resilience, conducting a survey of how "resilience" enters into public discourse—specifically, via Twitter. By downloading tweets mentioning the word resilience that were posted over a nine day period, the project constructs a shiny app displaying a time series plot, two sentiment analysis histograms, and a word cloud plot, in addition to creating the opportunity for users to explore individual tweets by keywords, in order to elucidate the different contexts in which the construct of "resilience" is salient. Overall, the project shows that term resilience is applied to an array of circumstances but is especially associated with mental well-being, In addition, it tends to be affiliated with other positive sentiments.

Introduction

Resilience, defined by researchers at Columbia's Department of Human Development, "refers to a dynamic process encompassing positive adaptation within the context of significant adversity" (Luthar et al). It is a construct that seeps into both academic and popular spheres, offering a paradigm to investigate the manifold factors that inform the experience of adversity and trauma, as well as what conditions need to be in place, on individual and communal levels, so that humans can endure and flourish in spite of these challenges. This project aims to conduct preliminary research on the concept of resilience and the different ways the public employs the term to disclose the important themes and sentiments orbiting it. Specifically, the project begins with the question, "How is the term "resilience" used by the general public and what social valences does this word possess?" From here, it aims to investigate how resilience enters into public discourse through Twitter, a platform that acts as a barometer for popular sentiments and understandings, reflecting and forecasting the changing attitudes of different groups.

Hypothesis

There will be an increase in tweets mentioning resilience over November 11th, which is Veteran's Day. Resilience will most often be associated mostly with positive sentiments and it will be used in many different contexts, particularly those pertaining to health and the environment. Moreover, the six different users will shed light onto the many different contexts that resilience appears in.

Methods

Data Collection

The project uses Twitter's API to download three kinds of tweets:1) tweets that mention the word "resilience"; 2) tweets that use #resilience; 3) tweets from six users (all non-profit organizations) with a twitter name that includes the word resilience. After filling out an online API request at httpss://developer.twitter.com/, I ascertained access keys and consumer codes, which authorized me to download this data. In addition, I installed the rtweet and twitteR packages. All of the tweets I downloaded were posted between 2019-11-06 and 2019-11-15 as the stream_tweets function only allows tweets within a 6-9 day span to be scraped. I thought that this time frame would prove especially interesting, as it includes Veterans Day. When downloading the tweets that include the word "resilience," I included retweets as they seem indicative of what ideas of resilience resonate with the public. I set retryonratelimit as TRUE to automate the process of conducting big searches as well as set number of tweets to download as 100,000 to gather as wide and varied a sample as possible. Conversely, I set retweets as false for the other two downloads as I only wanted to see the tweets original to every user.

The six users were selected for their diversity, each one aspiring to engender resilience in a different context (social, environmental, economic, etc). The maximum number of tweets that can be downloaded using user_tweets is 3200; however, all of the users fell short of this number anyways. After downloading users' tweets in this way, I immediately converted the information into a dataframe and combined all of their tweets into a single table, which only included the variables of interest.

Initial Data Sorting/Cleaning

All of the raw and cleaned tweet data was stored in directories, allowing it to be imported into my shiny app. I performed an initial cleaning of the data for #resilience and resilience tweets to keep the most important values of the dataset (screen_name, time created, number of retweets, number of favorites) and exclude tweets that were not in english. I also chose to exclude all retweets in this clean. Furthermore, I created separate data frames with only the top fifty most popular tweets for these two datasets, as measured by the number of times the tweet was favorited. In addition, to construct my time series plot later on, I created a new dataframe using the word resilience tweets but with fewer variables, as the original data frame was too large to push to GitHub. Lastly, I was interested in gaining a sense of the six users' presence on Twitter, so I used the mutate function to create a series of new variables for some basic twitter metrics including mean length of tweets, number of tweets, and mean number of favorites and retweets for twitter users.

Further Analysis and Visualizations

I constructed a series of visualizations to analyse and display key takeaways from the data, all of which served to help answer my guiding research question. I created three explore pages—for the top 50 "word resilience" tweets, the top 50 #resilience tweets, and all of the six users' tweets"—to allow users to explore the tweets themselves, according to keywords. However, perhaps most notably, I created a time series plot, two sentiment analysis histograms, and a word cloud, which draw more explicit conclusions.

The time series plot display date on the x-axis and the number of tweets posted that mention the word resilience on the y-axis. I was able to create this graphic fairly easily, using rtweet's ts_plot function (Appendix 1). The sentiment analysis histograms aimed to provide general information about the tone used in tweets containing the word resilience. The sentiment analysis was conducted on tweets including the designated key word "resilience" and I chose to use the bing and nrc lexicons. After locating this keyword and assigning it a relevant keyword value, the text was broken up to determine the relevant sentiments for all words in the text. Next, I grouped by each keyword value tally and the total of different sentiments, so that the resulting findings are the sentiment for the word resilience in each tweet, given the words surrounding it. The bing analysis provides information about whether the tweets express, in general, positive or negative sentiments by assigning a binary value based on the aggregate rating of each word in the tweet. Instead of just assigning binary values, the nrc analysis assigns a more specific emotions, including anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. I then used ggplot to graph the results (Appendix 2).

The "Word Cloud" tab included a word cloud of tweets with #resilience. Word clouds are an effective way to present information about the frequency of words in a visually appealing, clear way. The size of the word in the cloud depends on how often it is employed. I used the "wordcloud," "SnowballC," and "tm" packages to construct the cloud. First, the data was cleaned and text extracted as a vector, which included removing common stop words in order to produce meaningful results. The text was also cleaned by converting all words to lowercase and removing punctuation or unnecessary symbols. In addition, I removed the word resilience because it was too large. From here, I made a "document-term" matrix that creates a dataframe containing each word in the first column and its frequency in the second. This data was then stored before being made into a word cloud, using the wordcloud2 function.

Discussion of Results

The time series plot shows a spike in tweets mentioning the word resilience on November 11th, rising from approximately 8000 to 16000 tweets. This is consistent with my hypothesis, which anticipated that the concept of resilience would be employed most frequently on Veterans Day. The nrc sentiment analysis showed that for all of the tweets mentioning the word resilience, the term was most commonly associated with a 'positive sentiment'—accounting for 41584 tweets. This was followed by trust (21963), anticipation (15527), joy (14265), and negative (13105). The bing sentiment analysis showed a similar trend, with 29788 positive tweets and 13270 negative ones.

Why this range in sentiments? When looking at the individual tweets in the dataset, the word resilience most often arises as either a plea/call to action "x happened therefore we require greater resilience" or as a statement of affirmation/admiration "y is so resilient," which provides ample opportunity for it to be surrounded by words with a wide array of connotations. The word itself is not necessarily seen as a bad thing but it often arises out of unfavourable circumstances.

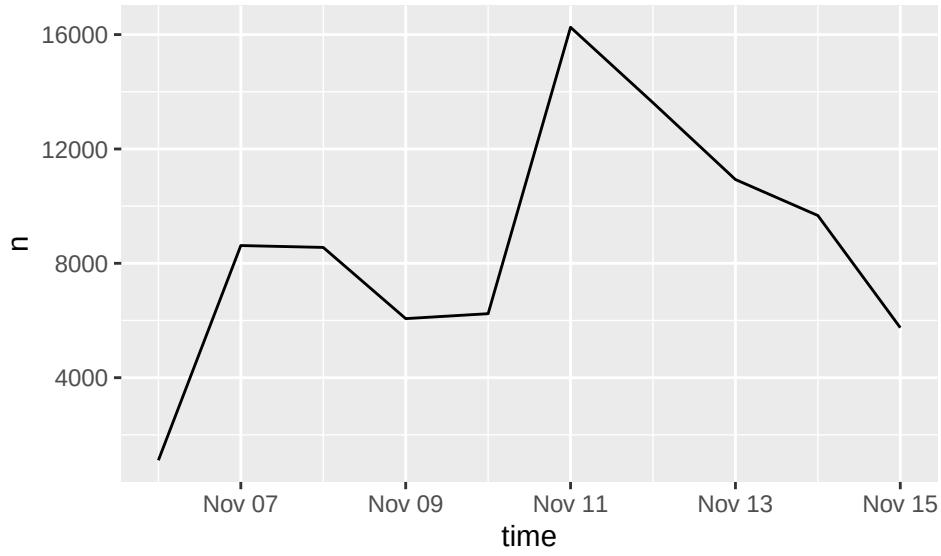
Finally, the word cloud plot disclosed important information about the concepts and words most often affiliated with resilience. From the relative size of the words, we can see that "can" (480), "mentalhealth" (290), "today" (234), "wellbeing" (197), "build" (163), "leadership" (160), "psychology" (153), "self-improvement" (149), "climate" (134), "support" (127), "community" (121), and "stress" (101), are among the most common words, which aligns with the initial hypothesis that the concepts most commonly associated with resilience would be cross-disciplinary; however, it exhibited a greater emphasis on mental wellbeing than I had expected.

Conclusion

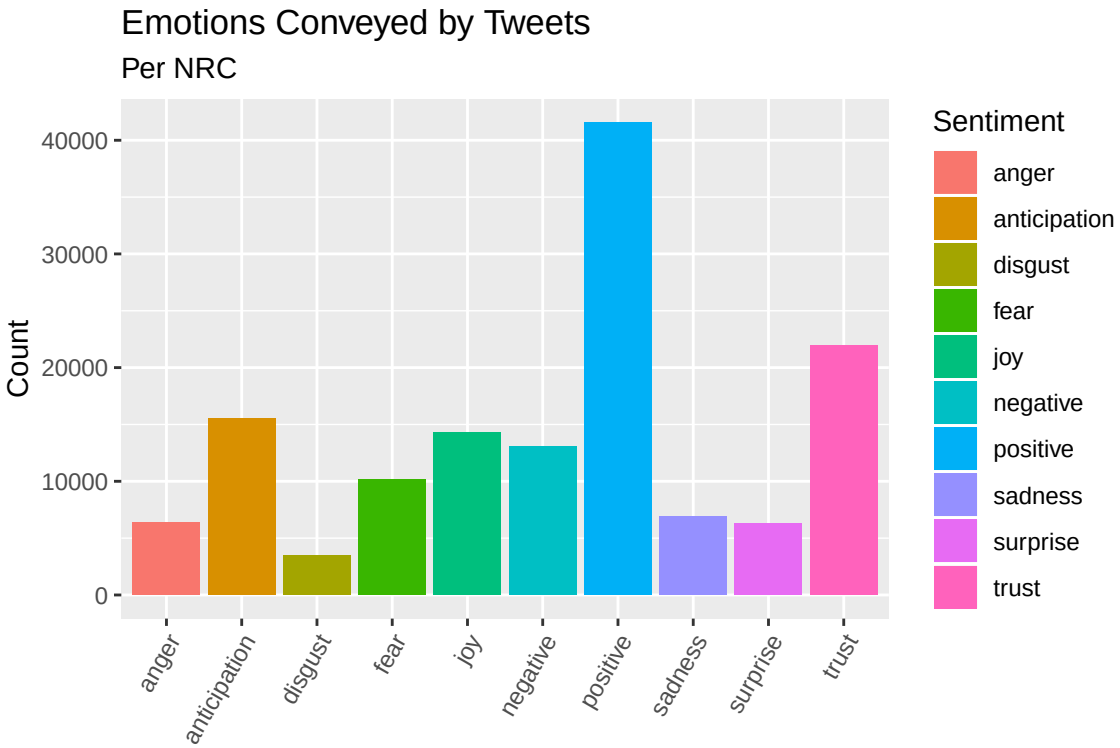
By downloading tweets that were posted between November 6th and November 15th 2019, using the Twitter API, this project aimed to investigate this research question and make a case for the salience of this construct to the public sphere and the importance of understanding the ways in which it is represented. My goal was to make my code as readable and replicable as possible, so that I can use it again in the future to perform similar analyses. I hope to use all of this code as I continue on in my studies, to create similar apps for different periods of time. All in all, it served its purpose, providing a portrait of the ways resilience is used in the public sphere; which, is information that I hope to use to guide deeper research into this construct.

Appendix

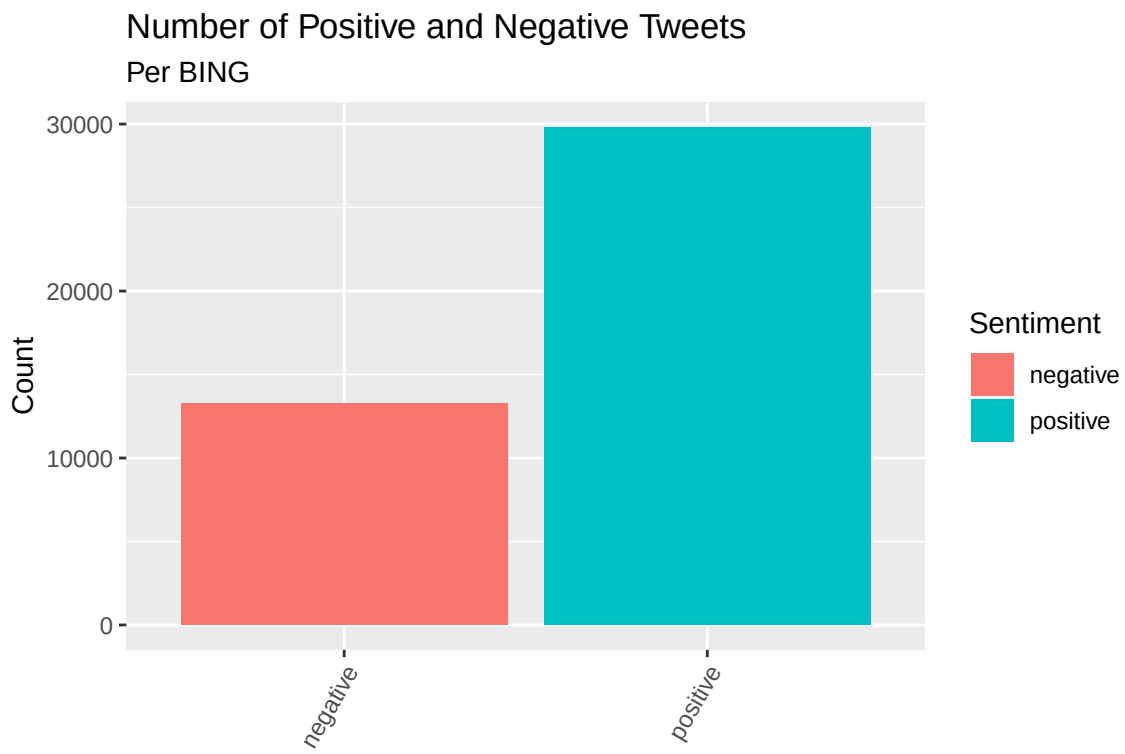
Appendix 1



Appendix 2



Appendix 3


Number of Positive and Negative Tweets
Per BING
Count
30000
20000
10000
0
negative
positive
Sentiment
negative
positive