Using Twitter for Nursing Research: A Tweet Analysis on Climate Change, Global Warming, Heat Related Illness, and Health

Daniel Smith

5/1/2019

Abstract

Twitter is a social media platform that allows users to create messages of 240 characters or less. These short messages are known as tweets and provide rich text data for analysis. The purpose of this project was to analyze tweets that mentioned the phrases 'climate change & health', 'global warming & health', and 'heat illness & health'. For each set of tweets term frequencies, word networks, and sentiment analysis were performed. For the 'climate change & health' tweets latent sentiment analysis was performed. This project provided evidence that Twitter is a viable tool for nursing research and can be used in various ways. However, using the set of analysis packages used in this project, Twitter could be especially useful in creating conceptual frameworks for studies and helping guide nurse researchers' thinking about the various concepts and ideas that appear in social media as relating to their topic of interest.

Introduction

Climate change is a topic of discussion in a plethora of disciplines and as environmental health researchers and nurses, we often see firsthand the effects that climate change has on individuals’ lives and health. The pathways in which climate change impact human health are complex and some examples include undernutrition, increased temperatures leading to increases in vector-borne disease and increases in heat-related illness, and increases in air pollution causing more cardiovascular and respiratory disease (Heaviside et al., 2016; Orru, Ebi, & Forsberg, 2017; Watts et al., 2017). However, as a term, ‘climate change’ encompasses the multiple ways in which the Earth’s climate is changing. Climate change can involve increasing instances of extreme events (i.e. hurricanes and polar vortices), increasing average global temperatures, decreasing ice caps, increasing sea levels, increasing levels or carbon dioxide in the atmosphere, increasing droughts, and changes in the amount of precipitation a region of the Earth receives (Shaftel, 2019). Given the many different parts that make up climate change, it is evident that multiple different conversations are happening within the climate change sphere.

Global warming is one piece of climate change that is particularly worrisome to the health of different populations around the globe. Heat waves from global warming will impact people with chronic disease. For example, people with hypertension, heart disease, diabetes, or asthma are at a greater risk for suffering complications related to their disease due to excess temperatures (Rossati, 2017). One study in Korea found that heat waves were associated with a 14% increase in out-of-hospital-cardiac arrest (Kang et al., 2016). However, the effects of increased heat are not just limited to the general population. Outdoor occupational groups, such as construction workers, miners, and agricultural workers, are particularly vulnerable to the impacts of increasing temperatures (Fleischer et al., 2013; Hunt, Parker, & Stewart, 2014; Levi, Kjellstrom, & Baldasseroni, 2018). As global temperatures continue to increase, heat-related illness is expected to become more common in outdoor workers. Studies have shown that heat-strain and dehydration cause agricultural workers to experience acute kidney injury during shifts (Mix et al., 2018; Moyce, Joseph, Tancredi, Mitchell, & Schenker, 2016; Moyce et al., 2017). Annually, farmworkers within the United States also experience heat-related deaths 20x the rate of the US civilian-workforce (Jackson & Rosenberg, 2010).

In order to better understand the nature of the conversations surrounding climate change and health on social media, this study will utilize Twitter (a social media site that allows users to tweet messages of 240 characters or less) to characterize and analyze the differences between tweets at various levels of the climate change conversation. Twitter has proved useful in many health-related studies through content analysis, surveillance, study recruitment, intervention delivery, and network analysis (Sinnenberg et al., 2017). This analysis will focus on content analysis of tweets through sentiment analysis (a metric for evaluating the opinions expressed within text as either positive or negative) and word networks within tweets. A comparison of word frequencies between the three groups will also be conducted. Keywords for identifying tweets from each group to be included in the analysis are ‘climate change + health’, ‘global warming + health’, and ‘heat illness + health’.

Methods

Data for this project was acquired from http://twitter.com and were acquired through Daniel Smith’s API account. All tweets were acquired on March 20, 2019, using the functions of the twitteR package. Key words/phrases were used to query tweets about climate change and health across multiple levels of the climate change spectrum. Starting with the idea of ‘climate change’, the keywords used to query tweets were ‘climate change + health’. Using the idea of ‘global warming’, keywords used in the query were ‘global warming + health’. Finally, to query tweets about heat illness, the keywords utilized in the search were ‘heat illness + health’. Using n=10000 and waiting for 10 minutes between requests form the API to avoid being rate limited were requested for each of the search terms. Tweets were also limited to the English Language.

Cleaning

Before analysis of tweets can occur, S. Yang & Zhang (2018) suggests that tweets should be converted to all lower case, have links, punctuation and stop words removed. Stop words are words such as ‘a’ and ‘to’ that are not meaningful in the textual analysis. S. Yang & Zhang (2018) also suggests removing usernames from tweets, However, given the topic and that certain politicians and other individuals are often known to tweet about “controversial” topics such as climate change and global warming, usernames were only removed during sentiment analysis. In order to clean the tweets, the package tidytext was used and tweets were cleaned/pre-processed individually for each analysis step. For each step, a list of “pure” tweets was used (i.e. retweets that were extracted through the API were removed leaving only original tweets).

For **term frequency** analysis, tweets were converted to lower case, punctuation removed, stop words removed, URLs removed, and unique IDs were assigned to each word in the tweets. When counting frequencies of words, it is important that all words are converted to lower case as “low” and “Low” would be counted as two separate words (S. Yang & Zhang, 2018). For **word network** analysis, tweets were converted to lower case, punctuation removed, stop words removed, URLs removed, unique IDs were assigned to each word, and bigrams (a pair of consecutive written units) had to be identified and counted. The cleaning for **sentiment**

analysis of the tweets varied when compared to the other two analyses above. The cleaning/preprocessing of tweets included maintaining the upper- and lower-case structure of the tweets, not removing punctuation nor stop words, but URLs, usernames, and hashtags were removed. This was done to preserve the original writing of the text that is important for sentiment analysis (S. Yang & Zhang, 2018). Finally, for the "climate change & health" tweets, **latent sentiment analysis** was performed. To do so, tweets were cleaned by removing all URLS and stop words. All words were also lemmatized.

Packages

Term Frequency, Word Networks, and Sentiment Analysis.

For each category of tweets, the top 15 appearing words were analyzed using tidytext, word networks were created using igraph and ggraph, and sentiment analysis was conducted using syuzhet. The function get_nrc_sentiment extracts sentiment using the NRC Word-Emotion Association Lexicon. This package assigns each word to 1 of 8 emotions (i.e. anger, fear, anticipation, trust, surprise, sadness, joy and disgust) and 1 of 2 sentiments (i.e. positive or negative). Tweets that are considered positive have a sentiment score of >0, neutral tweets have a sentiment score equal to 0, and negative tweets have a sentiment score of <0.

Latent Sentiment Analysis

For the 'climate change & health tweets', latent sentiment analysis was performed using the LSAfun and tm packages.

Results

In table 1, the number of tweets pulled for each set of keywords are presented. The keywords ‘climate change & health’ returned the most tweets with the requested number of 10,000 being returned. However, in each category a significant portion of the total number of retweets returned were retweets.

Table 1: Number of Tweets Pulled from Twitter Stratified by Search Terms.

	Climate Change & Health	*Global Warming & Health*	*Heat Illness & Health*
Number of Tweets	10000	715	5
Number of Tweets without Retweets	2512	301	5

Term Frequencies

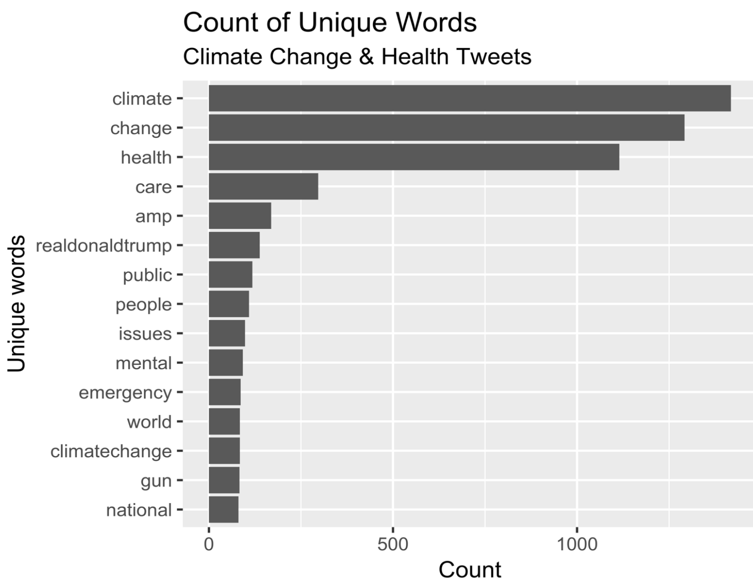

Figure 1: Term Frequencies of Climate Change & Health Tweets.

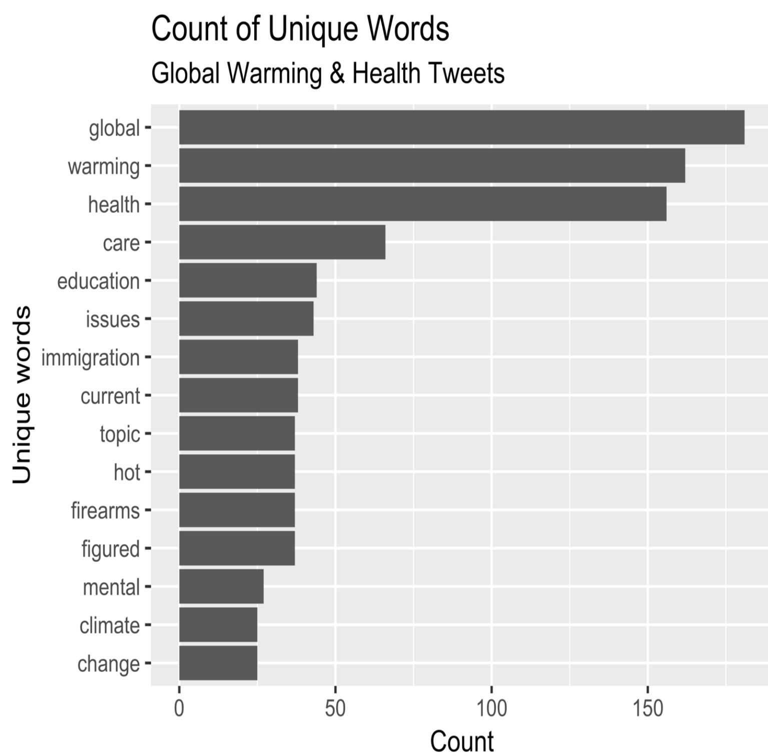

Figure 2: Term Frequencies of Global Warming & Health Tweets.

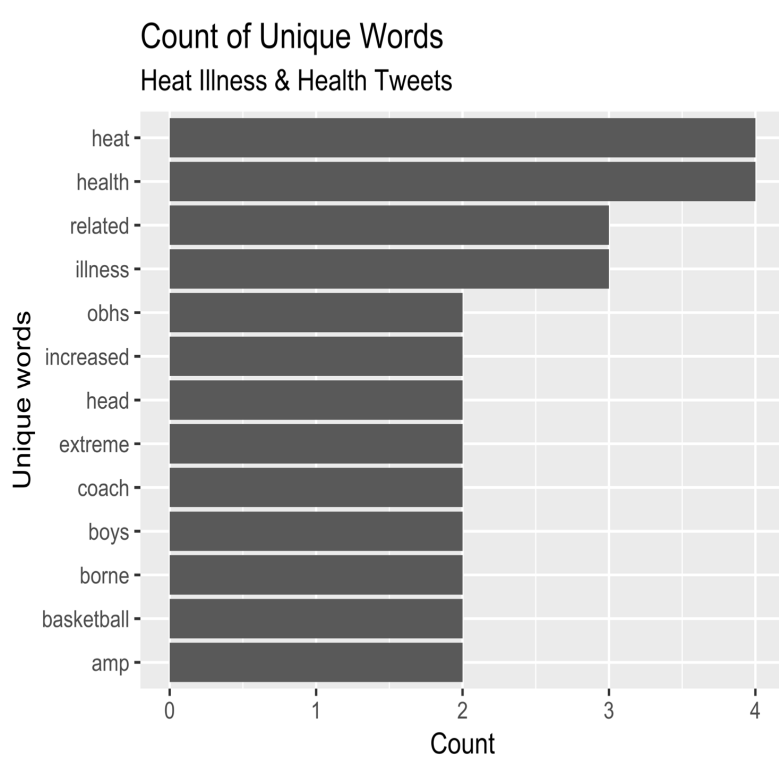

Figure 3: Term Frequencies of Heat Illness & Health Tweets.

The top 15 words in the tweets of “climate change & health” and “global warming & health” are show below. Due to the fewer number of tweets (and subsequent number of words) collected about “heat illness & health”, the top 13 words are shown displaying on those words that had frequencies of 2 or higher. In Figure 1, the top 15 words from the ‘Climate Change & Health’ are presented. The top word in this set of tweets was ‘climate’ with 1418 observations. In Figure 2, the top 15 words from the ‘Global Warming & Health’ tweets are presented. The top word in this set of tweets was ‘global’ with 181 observations. In Figure 3, the top 13 words from the ‘Heat Illness & Health’ tweets are presented. The top words in this set of tweets were ‘health, heat’ with 4 observations each.

Word Networks.

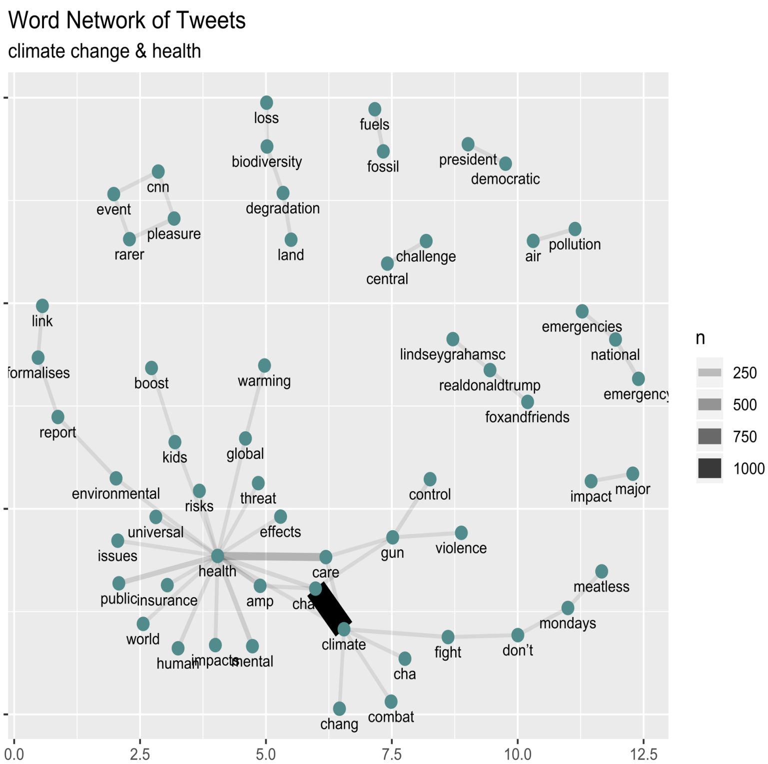

Figure 4: Work Network Analysis of Climate Change & Health Tweets.

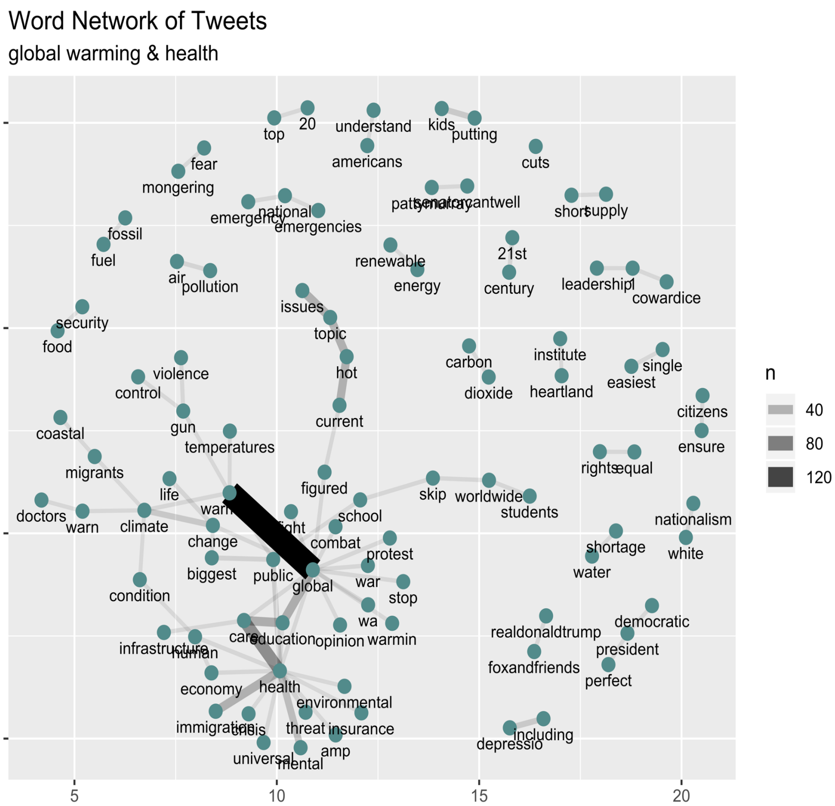

Figure 5: Word Network Analysis of Global Warming & Health Tweets.

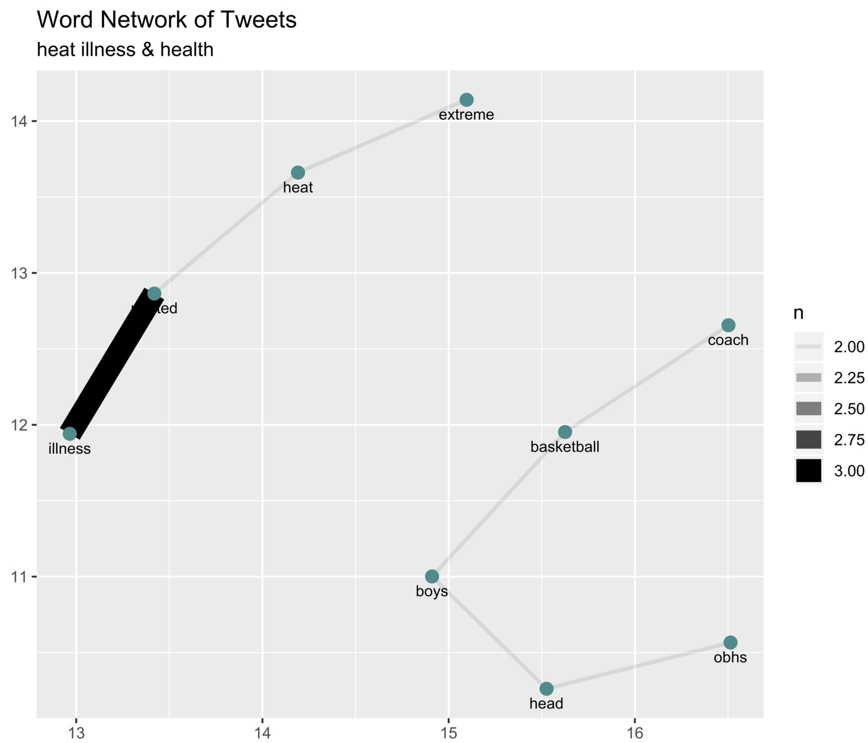

Figure 6: Word Network Analysis of Heat Illness & Health Tweets.

In Figure 4, the word network created using bi-grams of the 'Climate Change & Health' tweets is presented. The most frequent bi-gram was 'climate-change' with 1249 observations. In Figure 5, the word network created using bi-grams of the 'Global Warming & Health' tweets is presented. The most frequent bi-gram was 'global-warming' with 158 observations. In Figure 6, the word network created using bi grams of the 'Heat Illness & Health' tweets is presented. The most frequent bi-gram was 'related-illness' with 3 observations.

Sentiment Analysis

In figure 7, a pie chart of the sentiment analysis results for the 'climate change and health' tweets is presented. For this set of tweets 30% were classified as positive, 52% as neutral, and

18% as negative. In figure 8, a pie chart of the sentiment analysis results for the 'global warming and health' tweets is presented. For this set of tweets 38% were classified as positive, 41% as neutral, and 21% as negative. In figure 9, a pie chart of the sentiment analysis results for the 'heat illness and health' tweets is presented. For this set of tweets no tweets (0%) were classified as positive, 40% as neutral, and 60% as negative.

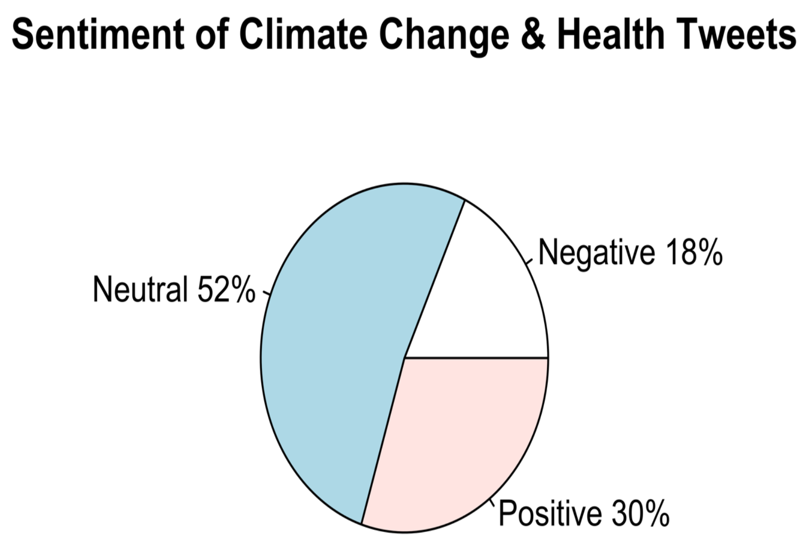

Figure 7: Pie Chart of Sentiment Analysis of Climate Change & Health Tweets

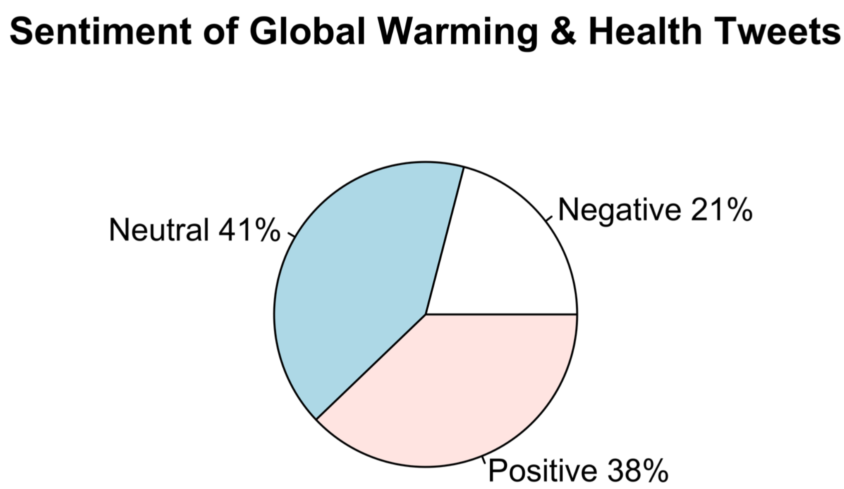

Figure 8: Pie Chart of Sentiment Analysis of Global Warming & Health Tweets

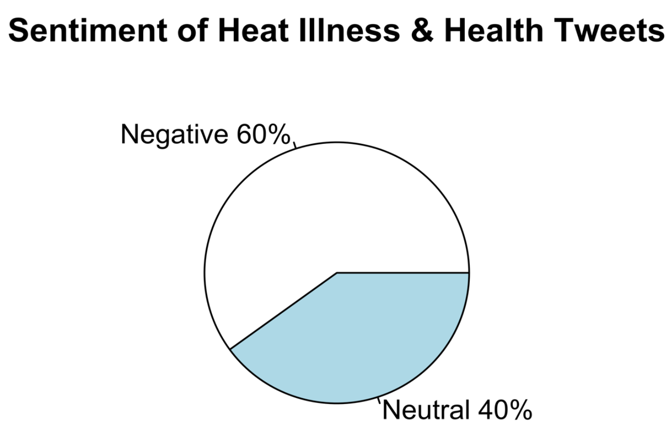

Figure 9: Pie Chart of Sentiment Analysis of Heat Illness & Health Tweets

Latent Sentiment Analysis

In figure 10, The results of latent sentiment analysis of the climate change and health tweets is presented.

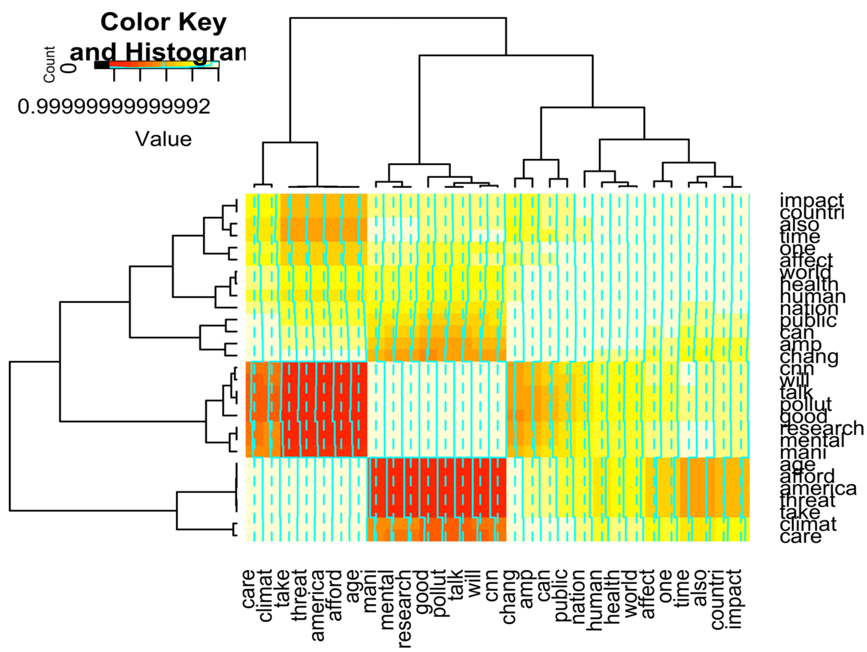

Figure 10: Heat Map of Climate Change and Health Tweets

Discussion

When pulling tweets from Twitter's standard API, common terms and phrases should be used. In this analysis, the most tweets originated using the phrase 'climate change and health' and the least tweets originated from 'heat illness and health'. These results indicates that more Twitter users are discussing the interaction of climate change and health and not the interaction between heat-related illness and health. One possible explanation is that 'climate change' has become a buzzword in contemporary discourse and is a 'hot topic' for researchers across multiple disciplines(Barnes et al., 2013; Uguru, Baiyeri, & Aba, 2011). Therefore, a topic such as heat-related-illness, which is used as an indicator of climate change (*Climate change indicators in the united states, 2016*, 2016), that does not carry the same 'buzzword status' is likely to be less discussed than topics that are commonly considered as buzzwords. Analysis of the 'heat illness & health' tweets is limited across this project due to there being only being 5 tweets pulled form twitter. However, if access to the premium API were had, a Twitter exploration of lesser discussed topics such as 'heat illness' or 'chronic kidney disease of unknown etiology' could potentially be completed.

When considering the term frequency of tweets returned during the three different searches it is unsurprising that the keywords used for each search were the most frequent terms. This patter was found across all three tweet groups with the words 'climate', 'change', and 'health' appearing most frequently for the 'climate change and health' tweets; the words 'global', 'warming', and 'health' appearing most frequently for the 'global warming and health' tweets; and the words 'heat', 'related', 'illness', and 'health' appearing most frequently for the 'heat related illness and health' tweets. The main purpose of analyzing the top terms from the tweets is to verify that the correct tweets were pulled from twitter.

For the sentiment analysis, the results were surprising. The terms 'climate change', 'global warming', and 'heat related illness' often carry negative connotations. It has even been suggested in the popular media that when discussing the topics of climate change and global warming that these terms are not used in order to avoid the negative connotation and political factors associated with the terms (Aubuchon, 2018). The results of sentiment analysis for the 'climate change & health' tweets returned that 82% of tweets were considered either positive or neutral; the ‘global warming & health’ tweets were 79% either positive or neutral; and the ‘heat illness & health’ tweets were only 40% neutral and had no positive tweets. Given that ‘climate change’ and ‘global warming’ are both terms that often carry negative connotations, the findings of approximately 8 out of 10 tweets being considered as either positive or neutral is noteworthy. The tweets on ‘heat illness & health’ suffer from a low n and the findings that none of these tweets are considered positive should be considered with caution.

Using network analysis, multiple tangential topics from ‘climate change & health’ and ‘global warming & health’ appeared. For example, in the ‘climate change & health’ a network appears of ‘lindseygrahamsc-realdonaldtrump-foxandfriends’, which provide insight into the political nature of climate change. Other networks that appeared in the ‘climate change & health’ tweets included ‘loss-biodiversity-degradation-land’, ‘emergencies-national-emergency’ and ‘meatless-Mondays’. For the ‘global warming & health’ tweets topics that appeared included ‘white-nationalism’, ‘fear-mongering’, ‘renewable-energy’, ‘water-shortage’. and ‘realdonaldtrump-foxandfriends’. Each of these networks provides a distinct insight into the set of tweets and helps build a complete picture of the interactions of climate change and global warming with health. Additionally, using latent sentiment analysis, the words that are connected to health can also be further analyzed using heatmap. The heat map as the one produced in this analysis, provides the correlation between the nearest neighboring words to the word of interest (in this case, health) and adds additional quantitative evidence of the importance of the connections between words. For this study, only the ‘climate change & health tweets’ were analyzed using latent sentiment analysis. Figure 10 shows the 20 closest words to health in the tweet set. By focusing on the areas in red (or the highest correlations), the words research, good, pollut(ion), talk, will, cnn, and chang(e) were found to be the most important to health in the set of tweets about ‘climate change & health’.

Being able to visualize topics that are on the periphery of keywords (i.e. climate change and health; global warming and health) through network analysis as well as using a focused analysis of words correlated with the term ‘health’ through latent sentiment analysis has several implications for nursing research. Given that the world in which we live is becoming increasingly reliant on digital technology (Smith, 2017), nursing researchers can turn to twitter as a tool of theory development for studies. Now, this is not to say that a twitter analysis can completely replace the tried and true methods for developing middle-range theory. However, twitter can aid in the development of theories by highlighting important secondary topics that surround the primary topic of interest and providing a detailed view of words related to the primary topic of interest. By using two different approaches for tweet analysis (word network analysis and latent sentiment analysis), there are multiple lines of evidence for including a certain topic in a theory. For example, in the ‘climate change & health tweets’ the idea of ‘politics’ appears through the ‘lindseygrahamsc-realdonaldtrump-foxandfriends’ network in the word network and through ‘CNN’ appearing as one of the closest words to health in this set of tweets. While the politics behind climate change are fairly well establish (Bolin, 2007; Harrison & Sundstrom, 2007), if the collected tweets were from a developing area of nursing researcher this would give researchers evidence and support for potentially including new concepts/ideas in their theoretical frameworks.

In conclusion, using twitter as a tool for nursing research is both feasible and practical. As completed in this project, tweet analysis allows for the visualization of words and ideas that are being used as part of the discourse around a given topic. Nurses can utilize twitter in theory development or as a standalone project. However, as expressed in Sinnenberg et al. (2017), the use of twitter in health research is not limited to an analysis of concepts such as this project. Thus, nursing researchers should become familiar with twitter as an essential tool for conducting nursing research that can be adapted to a range of projects.

References

Aubuchon, A. R. (2018, June 27). Let’s stop using terms ’climate change’ and ’global warming’. Web Page. Retrieved from https://theglobepost.com/2018/06/27/climate-change-global-warming-pollution/

Barnes, J., Dove, M., Lahsen, M., Mathews, A., McElwee, P., McIntosh, R., … others. (2013). Contribution of anthropology to the study of climate change. *Nature Climate Change*, *3*(6), 541.

Climate change indicators in the united states, 2016. (2016) (4th ed.). Retrieved from www.epa.gov/climate-indicators

Fleischer, N. L., Tiesman, H. M., Sumitani, J., Mize, T., Amarnath, K. K., Bayakly, A. R., & Murphy, M. W. (2013). Public health impact of heat-related illness among migrant farmworkers. *Am J Prev Med*, *44*(3), 199–206. Journal Article. doi:10.1016/j.amepre.2012.10.020

Harrison, K., & Sundstrom, L. M. (2007). The comparative politics of climate change. *Global Environmental Politics*, *7*(4), 1–18.

Heaviside, C., Tsangari, H., Paschalidou, A., Vardoulakis, S., Kassomenos, P., Georgiou, K. E., & Yamasaki, E. N. (2016). Heat-related mortality in cyprus for current and future climate scenarios. *Sci Total Environ*, *569-570*, 627–633. Journal Article. doi:10.1016/j.scitotenv.2016.06.138

Hunt, A. P., Parker, A. W., & Stewart, I. B. (2014). Heat strain and hydration status of surface mine blast crew workers. *J Occup Environ Med*, *56*(4), 409–14. Journal Article. doi:10.1097/JOM.0000000000000114

Jackson, L. L., & Rosenberg, H. R. (2010). Preventing heat-related illness among agricultural workers. *J Agromedicine*, *15*(3), 200–15. Journal Article. doi:10.1080/1059924X.2010.487021

Kang, S.-H., Oh, I.-Y., Heo, J., Lee, H., Kim, J., Lim, W.-H., … others. (2016). Heat, heat waves, and out-of-hospital cardiac arrest. *International Journal of Cardiology*, *221*, 232–237.

Levi, M., Kjellstrom, T., & Baldasseroni, A. (2018). Impact of climate change on occupational health and productivity: A systematic literature review focusing on workplace heat. *Med Lav*, *109*(3), 163–79. Journal Article. doi:10.23749/mdl.v109i3.6851

Mix, J., Elon, L., Mac, V., Flocks, J., Economos, E., Tovar-Aguilar, A. J., … McCauley, L. A. (2018). Hydration status, kidney function, and kidney injury in florida agricultural workers. *Journal of Occupational and Environmental Medicine*, *60*(5), e253–e260. Journal Article. doi:10.1097/JOM.0000000000001261

Moyce, S., Joseph, J., Tancredi, D., Mitchell, D., & Schenker, M. (2016). Cumulative incidence of acute kidney injury in california's agricultural workers. *J Occup Environ Med*, *58*(4), 391–7. Journal Article. doi:10.1097/JOM.0000000000000668

Moyce, S., Mitchell, D., Armitage, T., Tancredi, D., Joseph, J., & Schenker, M. (2017). Heat strain, volume depletion and kidney function in california agricultural workers. *Occup Environ Med*, *74*(6), 402–409. Journal Article. doi:10.1136/oemed-2016-103848

Orru, H., Ebi, K. L., & Forsberg, B. (2017). The interplay of climate change and air pollution on health. *Curr Environ Health Rep*, *4*(4), 504–513. Journal Article. doi:10.1007/s40572-017-0168-6

Rossati, A. (2017). Global warming and its health impact. *Int J Occup Environ Med*, *8*(1), 7–20. Journal Article. doi:10.15171/ijoem.2017.963

Shaftel, H. (2019, March 6). Web Page. Retrieved from https://climate.nasa.gov/evidence/

Sinnenberg, L., Buttenheim, A. M., Padrez, K., Mancheno, C., Ungar, L., & Merchant, R. M. (2017). Twitter as a tool for health research: A systematic review. *Am J Public Health*, *107*(1), e1–e8. Journal Article. doi:10.2105/AJPH.2016.303512

Smith, A. (2017). Record shares of americans now own smartphones, have home broadband. Journal Article. Retrieved from https://www.pewresearch.org/fact-tank/2017/01/12/evolution-of-technology/

Uguru, M., Baiyeri, K., & Aba, S. (2011). Indicators of climate change in the derived savannah niche of nsukka, south-eastern nigeria. *Agro-Science*, *10*(1).

Watts, N., Adger, W. N., Ayeb-Karlsson, S., Bai, Y., Byass, P., Campbell-Lendrum, D., … Costello, A. (2017). The lancet countdown: Tracking progress on health and climate change. *The Lancet*, *389*(10074), 1151–1164. Journal Article. doi:10.1016/s0140-6736(16)32124-9

Yang, S., & Zhang, H. (2018). Text mining of twitter data using a latent dirichlet allocation topic model and sentiment analysis. *OSI, Comp & Info Eng*, *12*(7), 525–529. Journal Article.