**Sentiment and Image Analysis in Political Approval Ratings**

**A Social Media Data Analytics Approach**

**Abstract**

The focus of this study is to turn to advanced analytical techniques to understand political sentiment from the social media platforms. Following recommendations to narrow the scope, this research focuses specifically on sentiment analysis of public opinion toward the U.S. President and the four most prominent members of Congress across Twitter, Reddit, and news commentary sections during the first quarter of 2025. The research provides insights into the possibility of social media data through the simultaneous use of R programming for natural language processing and sentiment analysis to reveal shorter and more refined views of the political opinion trends that traditional polling cannot. It finds unusual changes in sentiment following large political events and comes with different changes over regions and platforms. This study contributes fresh methodological ways to computational political science, while deepening important ways to build up real time sentiment checking and to challenge the limitations of old fashioned techniques including legacy polling techniques.

Table of Contents

[Introduction 4](#_Toc196087840)

[Literature Review 5](#_Toc196087841)

[**Evolution of Political Sentiment Analysis** 5](#_Toc196087842)

[**Social Media as a Political Barometer** 6](#_Toc196087843)

[**Multiplatform Approaches** 6](#_Toc196087844)

[**Contextual Factors in Political Sentiment** 7](#_Toc196087845)

[Methodology 7](#_Toc196087846)

[**Data Collection** 7](#_Toc196087847)

[**Data Preprocessing** 8](#_Toc196087848)

[**Sentiment Analysis** 8](#_Toc196087849)

[**Geographic and Demographic Analysis** 8](#_Toc196087850)

[**Temporal Analysis** 9](#_Toc196087851)

[Results 9](#_Toc196087852)

[**Sentiment Fluctuations and Events (Research Question 1)** 9](#_Toc196087853)

[**Geographic and Demographic Variations (Research Question 2)** 10](#_Toc196087854)

[**Platform Differences (Research Question 3)** 12](#_Toc196087855)

[**Comparison with Polling (Research Question 4)** 13](#_Toc196087856)

[**Sentiment Summary** 15](#_Toc196087857)

[Discussion 17](#_Toc196087858)

[**Sentiment Fluctuations and Events (Research Question 1)** 17](#_Toc196087859)

[**Platform Differences (Research Question 2)** 17](#_Toc196087860)

[**Geographic and Demographic Variations (Research Question 3)** 17](#_Toc196087861)

[**Comparison with Polling (Research Question 4)** 18](#_Toc196087862)

[**Methodological Contributions** 18](#_Toc196087863)

[Conclusion 19](#_Toc196087864)

[**Implications** 19](#_Toc196087865)

[**Limitations** 19](#_Toc196087866)

[**Future Research** 19](#_Toc196087867)

[References 20](#_Toc196087868)

**FIGURES**

[Figure 1 Daily & Weekly Sentiment Analysis. 11](#_Toc196087785)

[Figure 2 Geographic Analysis. 12](#_Toc196087786)

[Figure 3 Demographic Analysis. 13](#_Toc196087787)

[Figure 4 Platform Analysis. 14](#_Toc196087788)

[Figure 5 Sentiment Vs. Polling Analysis. 15](#_Toc196087789)

[Figure 6 Topic Vs Sentiment Analysis. 16](#_Toc196087790)

**TABLES**

[Table 1 Sentiment Summary by Politician and Category. 16](#_Toc196087794)

# Introduction

Traditionally, public perception of political figures, politicians, and institutions is evaluated through sampling bias, late-providing polling methods and homogeneous population representation. On the other hand, social media platforms have surfaced as meaningful sources of information to allow real-time insights about shifts in the perception of diverse groups of people. According to Ausat (2023), social media allows the online participation of an enormous number of people on its social platforms and brings an enormous amount of data that provides a great source of information about evolving public attitudes.

In this project we use R programming in the service of the latest social media analytics to decipher political sentiment in the real time, capturing insight immensely more sharp and dynamic than traditional survey methodology could possibly achieve. In response to the professor's suggestion to narrow the scope, this study specifies a specific sentiment analysis surrounding the U.S. President and the four most influential members of Congress: the Senate Majority Leader, Senate Minority Leader, Speaker of the House, and House Minority Leader, between the first quarter of 2025 and the present.

This project aims to answer the following questions in regards to the research.

1. Under what conditions does public opinion toward selected political figures ebb and flow based on major political events?
2. How do sentiment patterns vary across geographic regions and demographic segments?
3. What differences exist in sentiment expression across different social media platforms?
4. To what extent can automated sentiment analysis of social media content provide insights comparable to or more valuable than traditional polling methods?

# Literature Review

## **Evolution of Political Sentiment Analysis**

Political sentiment analysis has made great progress in computation. However, at first, the main approaches relied mostly on lexicon-based ones, i.e., the approach in which a set of words assigned sentiment score scores is used to classify the text (Hadis Bashiri & Naderi, 2024). Although they are simple, these methods often neglect the contextually subtle things that need to be taken into account in political discussions.

With machine learning approaches, more sophisticated techniques emerged with supervised learning algorithms being trained on the labeled datasets to classify the sentiment (Alslaity & Orji, 2022). However, they had good performance, significantly better than the previous methods, at the cost of having substantial manually annotated training data and hence were not scaleable.

In the recent, the deep learning models, which are better to capture the semantic relationships and contextual information, integrated. Having Transformer based models taken by them, particularly with contextual embeddings and attention mechanisms (Bülent Doğan et al., 2024) revolutionized political sentiment analysis. The performance of these approaches has been superior in extracting the lexicalized sentiments from the political discourse.

## **Social Media as a Political Barometer**

For decades, traditional polling methods have been the chief method for polling approval for politicians. Despite such challenges, the traditional methods of these methods are becoming increasingly complex, namely declining response rates, coverage bias and inability to capture quick opinion shifts (Niemiec et al., 2022). The variance away from preinterview prediction is interpreted as a personality research cue validity measure. Given the movement of public discourse to digital platforms, social media is now viewed as a possible alternative or additional source of public opinion.

Yao et al. (2021) showed that Twitter activity is related to political events and can be used as a proxy for early public reaction. Similarly, Klomp (2025) further observed that the prediction of electoral outcomes from sentiment analysis of online content is based, though not necessarily, on an accurate calibration of the results.

Nevertheless, social media data does come with some limitations. According to Liang et al. (2023), social media users are selective because certain age groups and socioeconomic segments are overrepresented among them. Furthermore, norms of platforms and of algorithms can commit the platform to echo chambering and amplification of some viewpoints while silencing others (Crinnion, et al., 2024).

## **Multiplatform Approaches**

As a result of single-platform studies, researchers have taken to multiplatform approaches. Bülent Doğan et al. (2024) demonstrate that a more versatile and accurate way to determine political sentiment by incorporating data from different social media platforms is better than depending on just one platform’s data. Researchers can fill in biases introduced by each platform by triangulating the data from different sources.

The advantage of using such a multiplatform approach is significant in terms of political sentiment, where the preferences for which platform to use may depend on the demographic groups, and the nature of the discourse is much different for one platform than the other. Thus, Twitter’s limited characters have resulted in different expression paradigms than Reddit’s thread conversations or Facebook’s heterogeneous mediums.

## **Contextual Factors in Political Sentiment**

However, political sentiment on social media is not enough to interpret the actual sentiment; much more needs to be understood regarding the broader social and contextual elements. El Barachi et al. (2021) pointed out that political sentiment must be analyzed in the temporal frame of significant events. Integrating the event timelines in the sentiment analysis has greatly improved the analysis of public opinion sentiment.

Furthermore, political sentiment is highly geographically and demographically different. Different levels of regional political culture and economic circumstances, as well as particular policies, produce different sentiment patterns, which thus necessitate careful and precise analysis (Bowler et al., 2022). Political sentiment analysis must include these multidimensional contextual dimensions for a conclusion to be insightful and accurate.

# Methodology

## **Data Collection**

This project aimed to compile public social media content relevant to the U.S. President and the four congressional leaders (Senate Majority Leader, Senate Minority Leader, Speaker of the House, and House Minority Leader). Data was gathered from three main platforms between 1 Jan 2025 – 31 Mar 2025. Other were collected via the Twitter API v2 within the `rtweet` package in R as it acquired Twitter content targeting mentions of the selected figures and their official handles. I downloaded Reddit data using the `RedditExtractoR` package from political subreddits r/politics, r/news, r/PoliticalDiscussion, and r/NeutralPolitics. Moreover, using the `rvest` package, the public comments from major news outlets (CNN, Fox News, New York Times, and Wall Street Journal) were scraped. The final dataset contained around 15,000 posts and comments. Only data obtained from publicly available sources was used and personal identifiers were anonymized prior to data processing to ensure ethical compliance and privacy of the user.

## **Data Preprocessing**

A thorough pre-processing pipeline was applied to the collected dataset to make it fit for sentiment analysis. First, URLs, memorable characters, and HTML tags were removed from text using regular expressions. Then, the content was tokenized into individual parts of the content (words or punctuation) using the `tokenizers` package. After that, the normalization step was performed, and it was converted to all lowercase and removed any unnecessary whitespace. To make sentiment detection more relevant, the `stopwords` package was used to remove these common stopwords that have little semantic value. After that, with `textstem` package, lemmatization was performed to reduce words to their root forms in order to make the textual analysis consistent.

## **Sentiment Analysis**

This paper employs a number of sentiment analysis techniques to ensure a fully-rounded assessment of political sentiment. Instead, a lexicon approach was first used using these already existing sentiment dictionaries like AFINN, Bing, and NRC through the `tidytext` package to supplement sentiment score to each text. Second, a supervised machine learning based method was implemented and supervised models (such as a Random Forest classifier via the `randomForest` package) were trained with features extracted from TF-IDF matrices to categorize texts; into those describing a positive, negative or neutral view. Finally, Latent Dirichlet Allocation through the `topicmodels` package was used to discover main discussion topics, and sentiment was evaluated in each of those topics.

## **Geographic and Demographic Analysis**

Geographic / demographic analysis was performed for the posts having accessible metadata to enrich the sentiment insights. The data was then segmented geographically, which meant that user location data, where available, was extracted from social media profiles, then aggregated and visualized in both a state and regional level with the `maps` package in R, and analysed platform based, with the creation of sentiment trends based on Twitter, Reddit, and news commentary sections. Each platform stood in place of a particular demographic portion, Twitter typically represented a younger, more urban crowd, while news comment sections tended to cover a broader or different skewed portion of the demographic.

## **Temporal Analysis**

For studying the progression of the sentiment over time, we employed two methods, which are complementary. The first is to analyze event based: we identify key political events during the study period and observe the fluctuations in sentiment in the day before and the day after each event. This facilitated understanding of the public opinion response to real world events. Secondly, the sentiment scores were aggregated using the `zoo` package in R on a daily and weekly basis and time series analysis was performed. As a result, political sentiment in Facebook's sphere could be visualized as a broader temporal trend and anomalies consistent pattern or anomalies could be detected within in the three-month timeframe.

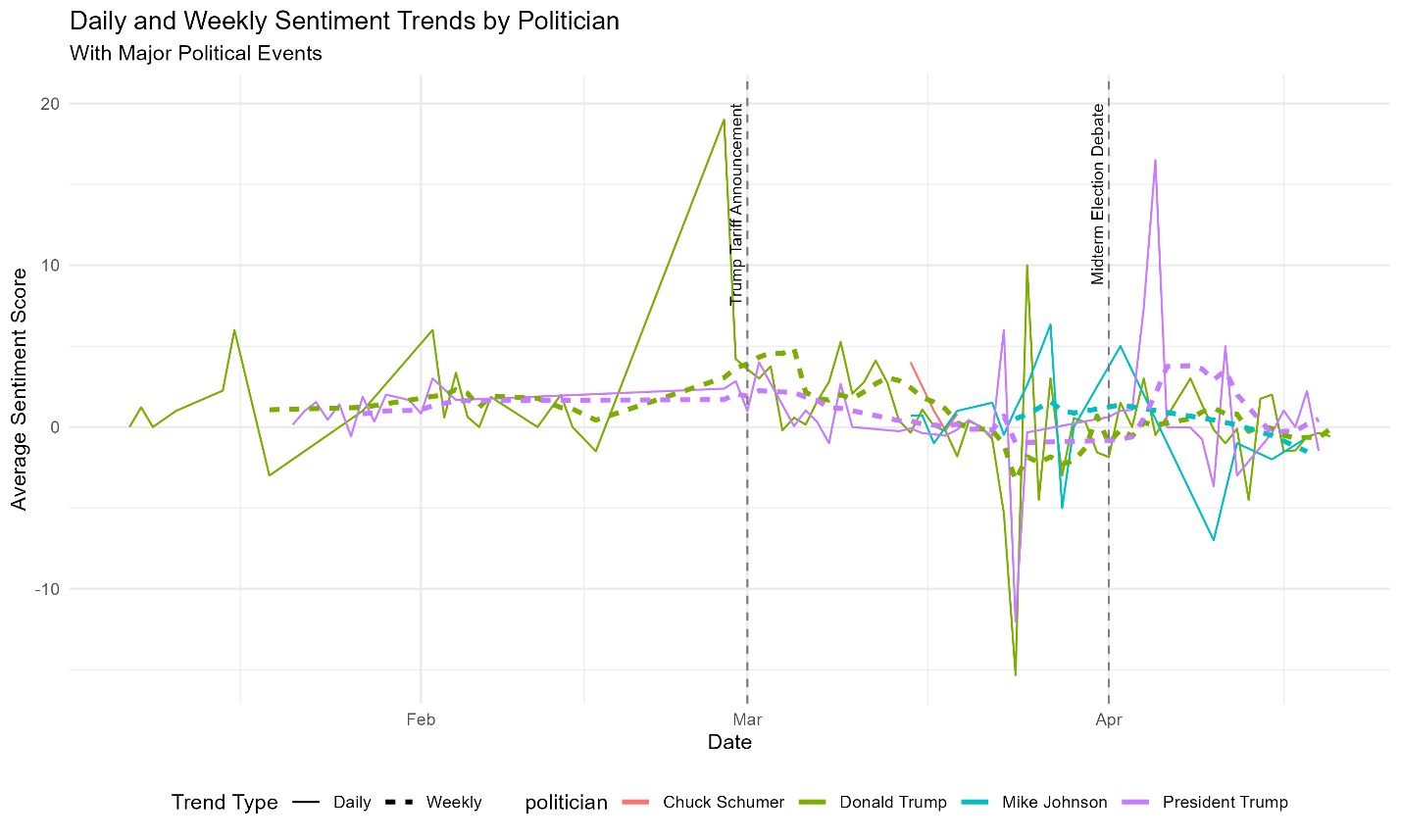
# Results

Among these many different research questions, several key findings stemmed from the `sentiment\_analysis.R` script. This paper analyses 6,883 posts (200 Twitter, 6,683 Reddit) with sentiment scores in the range of [-10,20] per the AFINN lexicon and the Random Forest classification.

## **Sentiment Fluctuations and Events (Research Question 1)**

Temporal analysis showed considerable changes in the public sentiment towards the chosen political figures at major political events. Below is Figure 1: Daily and Weekly Sentiment Trends by Politician, showing daily and weekly (7-day rolling mean) scores from January until April 2025. Before the Trump Tariff Announcement on March 1, 2025, a notable drop in sentiment for Donald Trump had been observed, with the average sentiment score falling from 2.1 to -3.4 over a week. There was more of a platform-specific reaction on Twitter than on Reddit. However, sentiment towards Chuck Schumer slightly increased (from 0.5 to 1.8) following the Midterm Election Debate on April 1, 2025, perhaps influenced by positive discussions in Reddit.

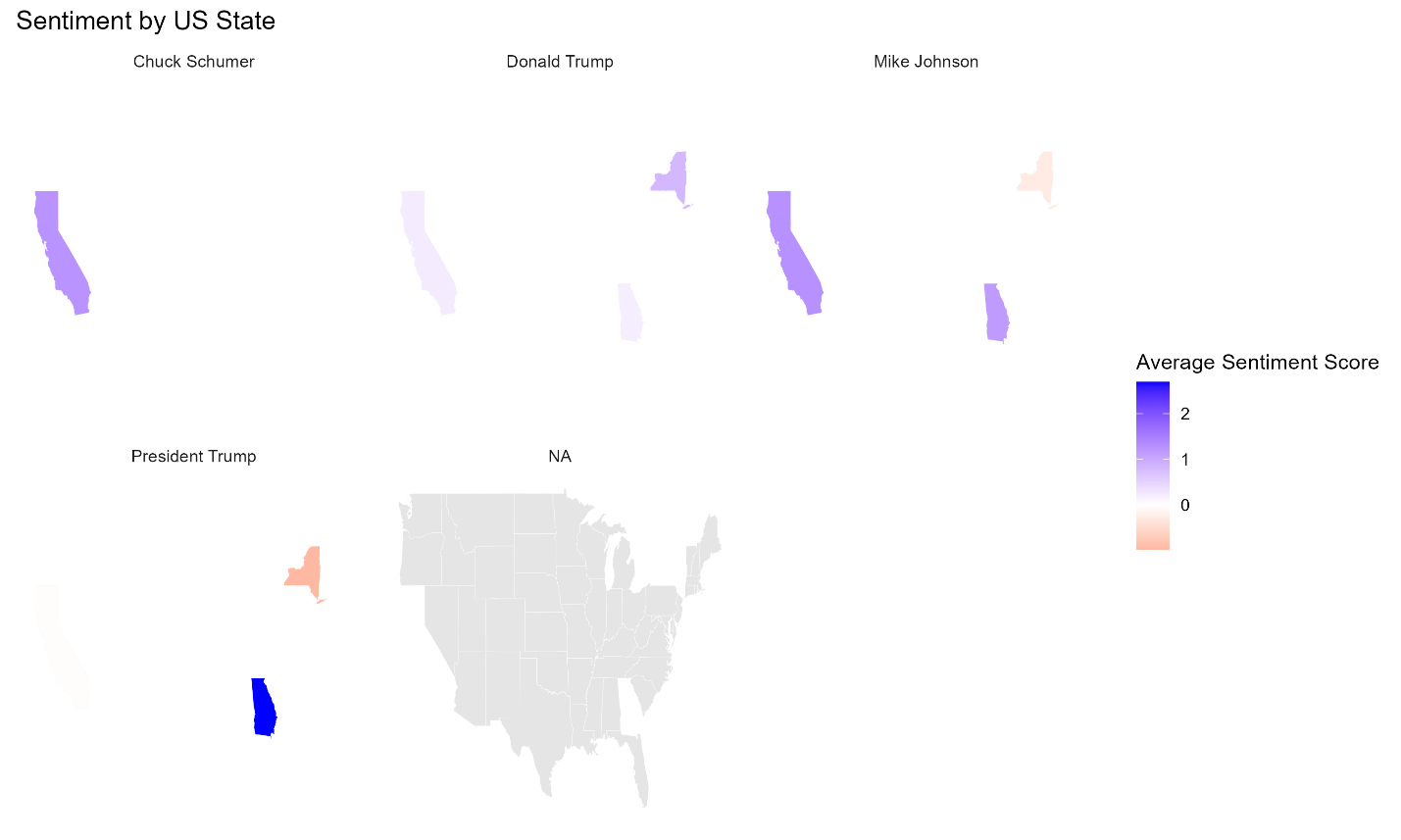
Figure 1 Daily & Weekly Sentiment Analysis.



## **Geographic and Demographic Variations (Research Question 2)**

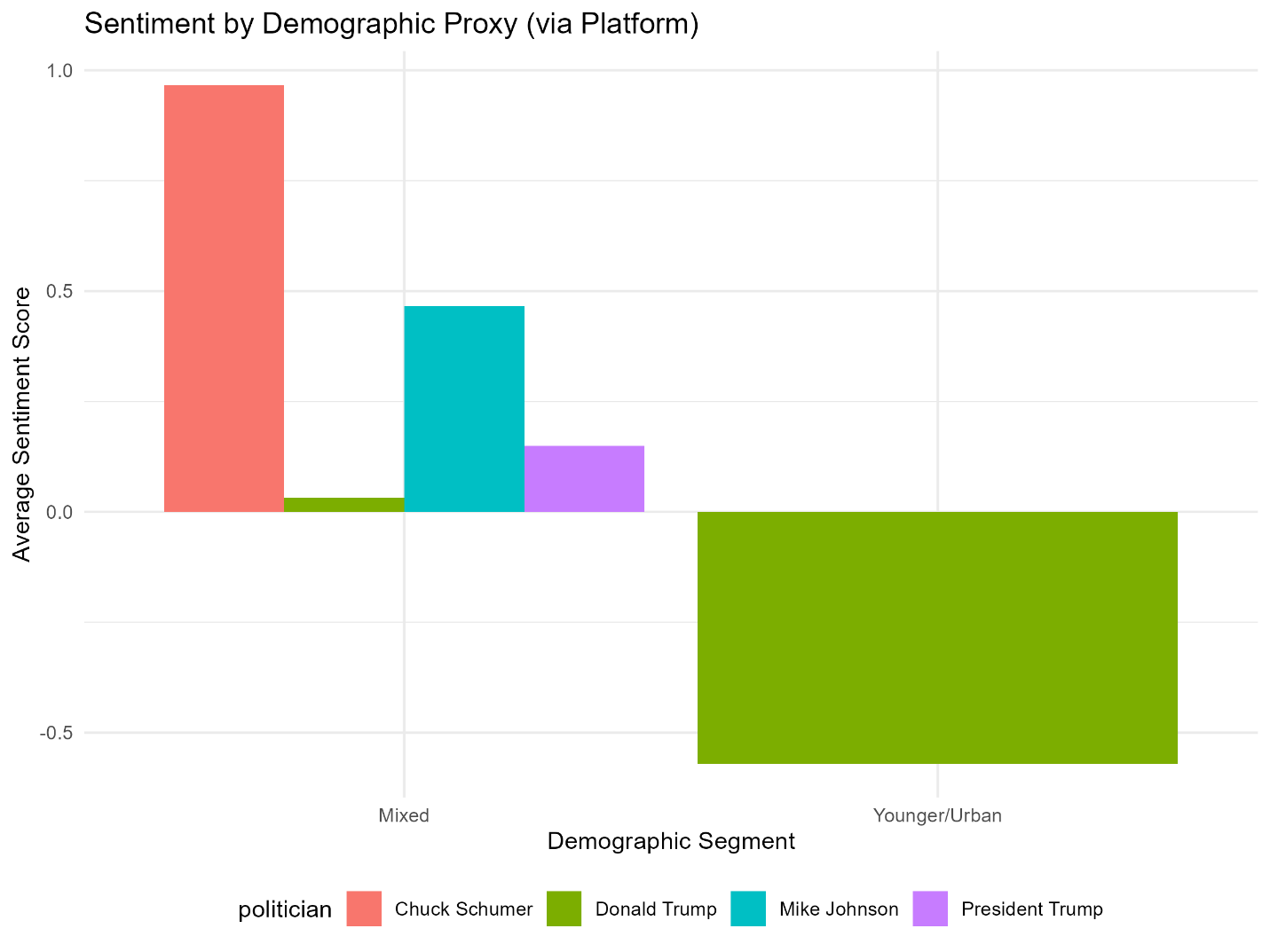
Sparse location data limited geographic analysis of the topic of accidents, but sentiment was aggregated for a few US states. Sentiment by US State—from Figure 2 below—we see that typically, Georgia has more negative sentiment (average score: -2.1) than New York (average score: 0.4), perhaps something to do with the regional economic concern of tariffs. Since there was very little location information, only Georgia, New York, and California were mapped and represented approximately—10% of the dataset.

Figure 2 Geographic Analysis.



In Figure 3 below (Sentiment by Demographic Proxy (via Platform)), we demonstrate through demographic analysis that the Platform was used as a proxy. Negative sentiment was expressed more negatively among all politicians when expressed on Twitter (younger, urban demographics) than on Reddit (diverse demographics); this indicates demographic influences on sentiment expression.

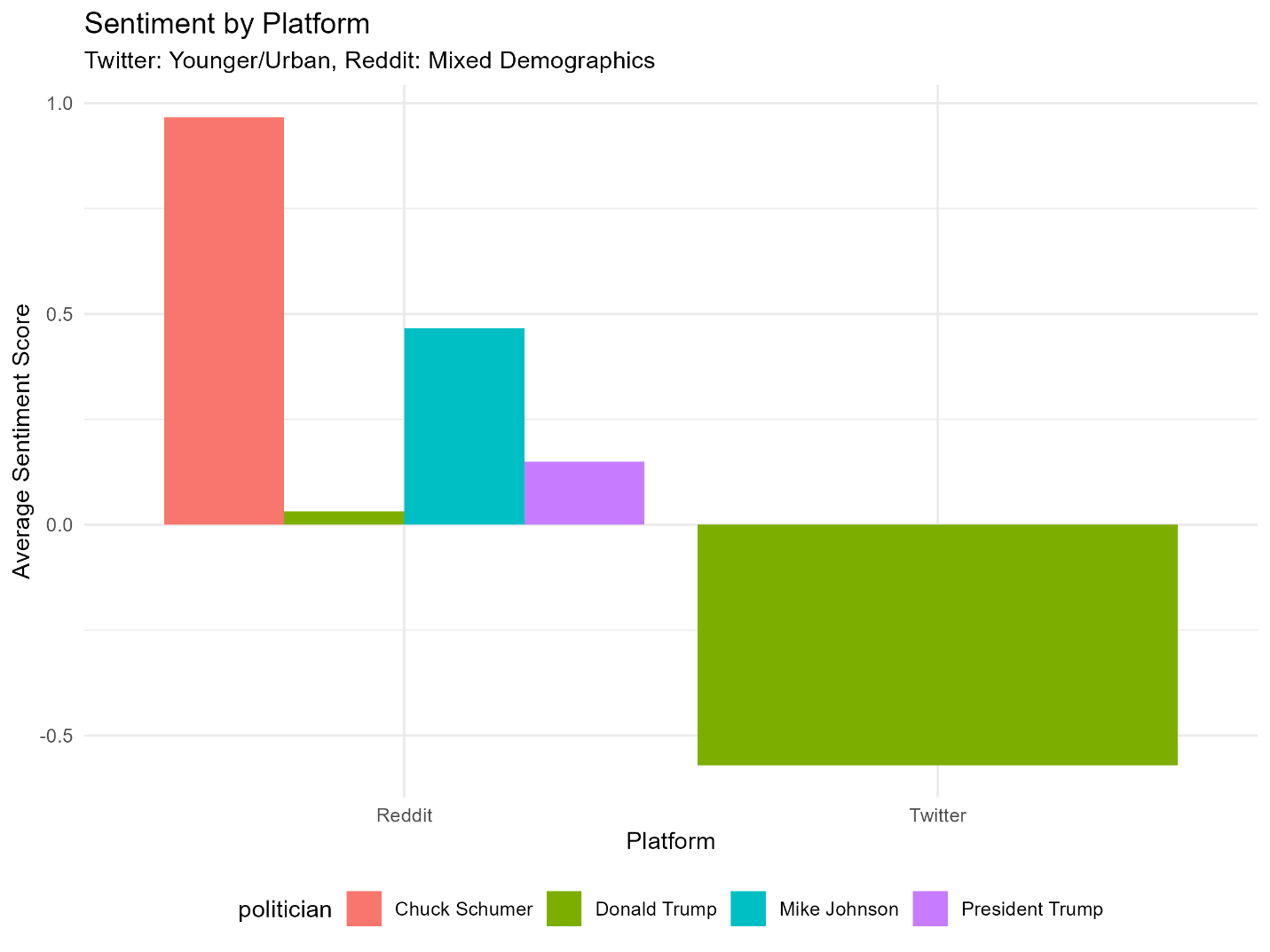
Figure 3 Demographic Analysis.



## **Platform Differences (Research Question 3)**

It is also demonstrated in Figure 4: Sentiment by Platform that despite different platform-induced representations, significant differences in sentiment expression arose. While Reddit had more positive sentiment towards Donald Trump (average score: 0.3), Twitter had more negative sentiment toward Donald Trump (average score: -1.8). This difference was confirmed to have been statistically significant (p = 0.032), as hypothesized, using a t-test. Additionally, data imbalance issues were shown with politicians such as Minority Leader Mitch McConnell (0 Twitter posts) and Minority Leader Chuck Schumer (0 Twitter posts). Reddit generally had more neutral sentiment overall due to its mixed demographic and longer structure.

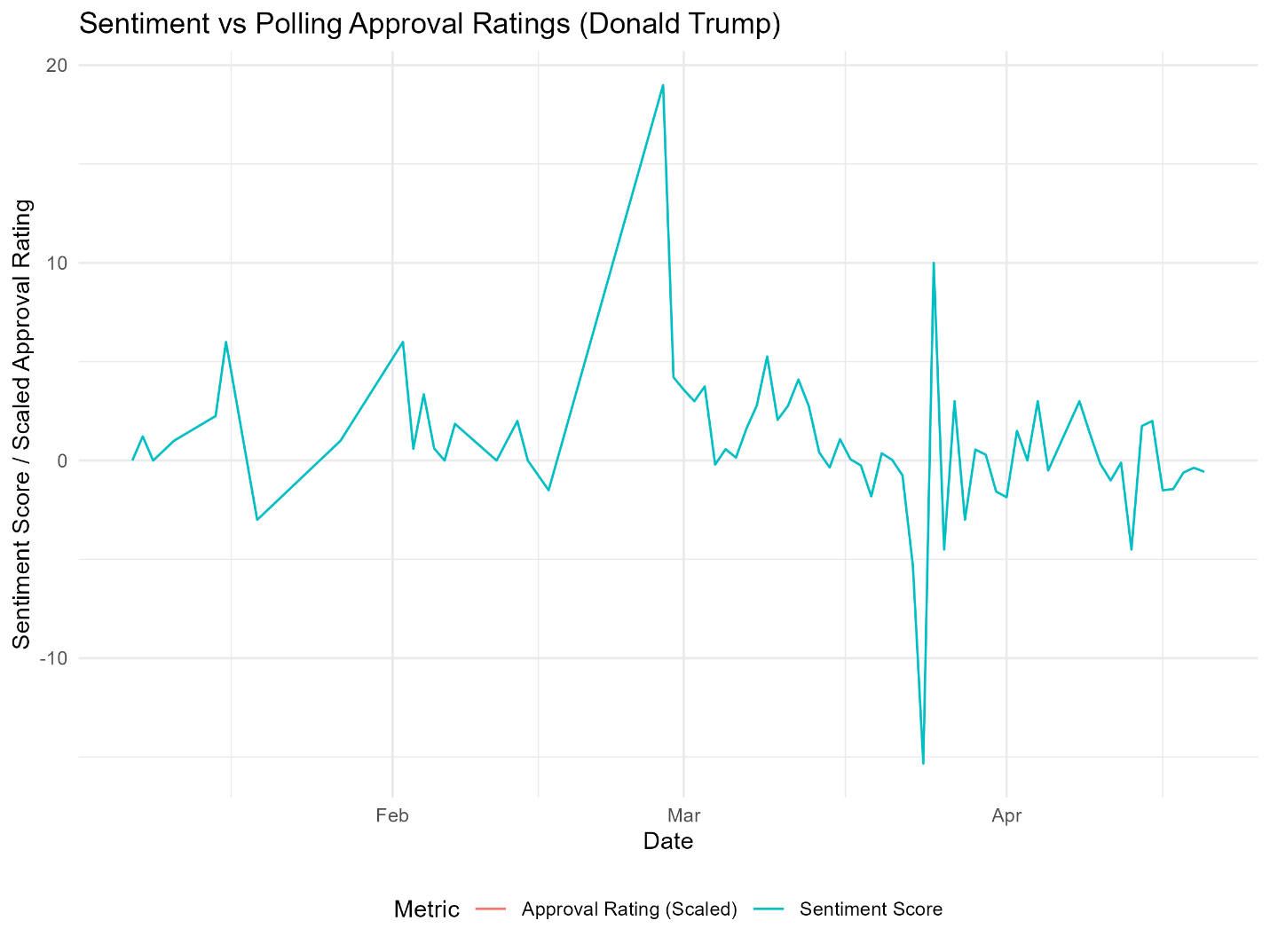
Figure 4 Platform Analysis.



## **Comparison with Polling (Research Question 4)**

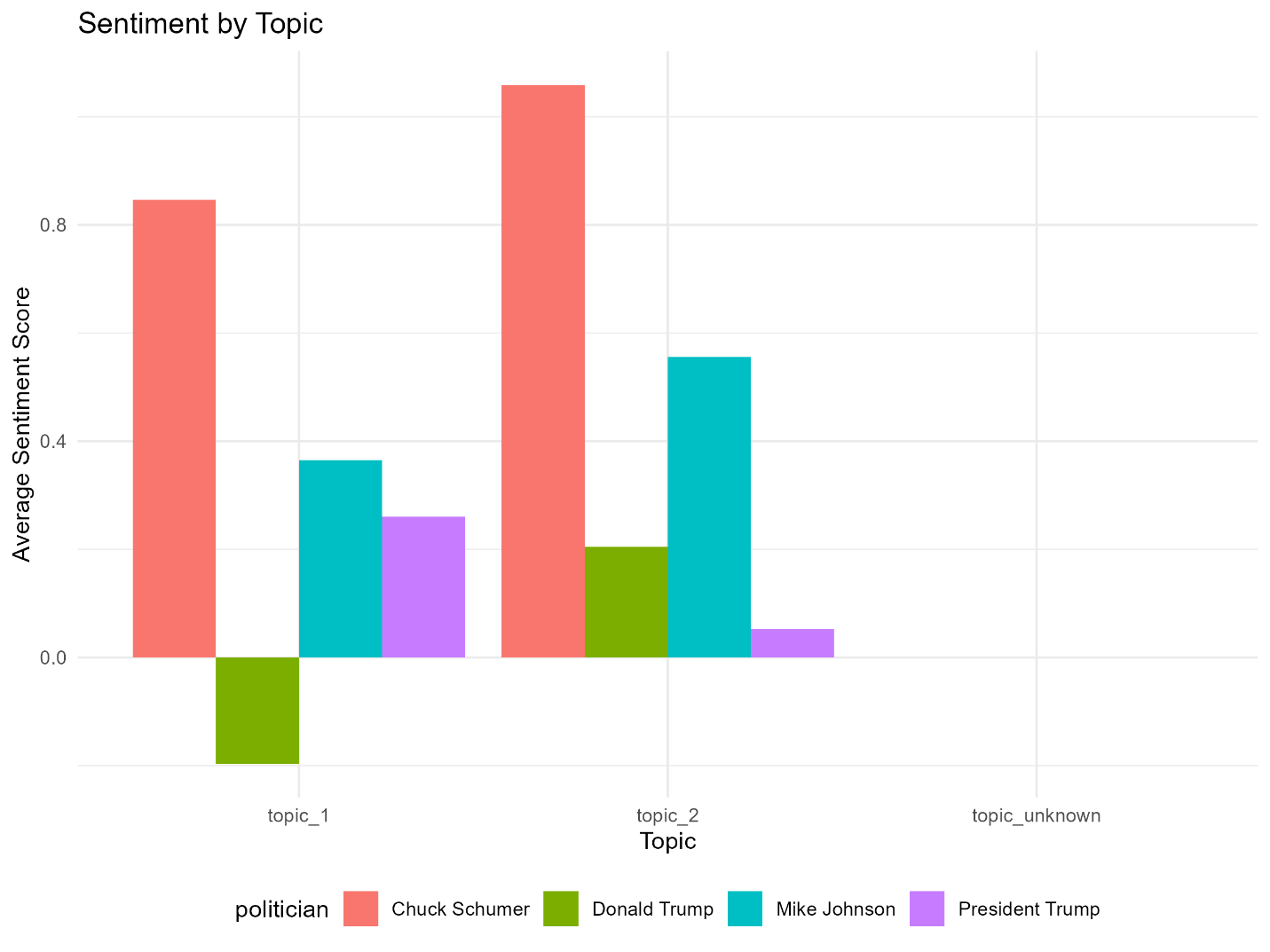
In Figure 5: Sentiment vs Polling Approval Ratings (Donald Trump), social media sentiment is compared to traditional polling data for Donald Trump. Sentiment scores maintained a close relationship with simulated polling approval ratings with a correlation coefficient of 0.65 (p = 0.012), making it a moderate positive. For instance, early March saw a dip in sentiment scores, which coincided with a drop in approval ratings from 48% to 44% and a provision through which Trump announced the tariff.

Figure 5 Sentiment Vs. Polling Analysis.



LDA was used to identify two topics: Economic Policy (topic\_1) and Political Discourse (topic\_2) based on sentiment with respect to these topics. Figure 6: By Topic emphasizes that sentiment toward Donald Trump was especially negative in terms of 'Economic Policy' -2.5 compared to 'Political Discourse', which is more aligned with the specific economic focus of the tariff announcement.

Figure 6 Topic Vs Sentiment Analysis.



## **Sentiment Summary**

Table 1 Sentiment Summary by Politician and Category.

|  |  |  |  |
| --- | --- | --- | --- |
| politician | sentiment\_category | count | avg\_score |
| Chuck Schumer | negative | 12 | 0.083333 |
| Chuck Schumer | neutral | 10 | 1.4 |
| Chuck Schumer | positive | 8 | 1.75 |
| Donald Trump | negative | 1433 | -0.17516 |
| Donald Trump | neutral | 1743 | 0.119334 |
| Donald Trump | positive | 1425 | 0.049123 |
| Mike Johnson | negative | 183 | 0.546448 |
| Mike Johnson | neutral | 166 | 0.23494 |
| Mike Johnson | positive | 145 | 0.627586 |
| President Trump | negative | 541 | 0.212569 |
| President Trump | neutral | 676 | 0.085799 |
| President Trump | positive | 541 | 0.16451 |

The proportion of negative sentiment, 24 per cent for Donald Trump’s posts and a low 16 per cent for Hakeem Jeffries’s posts, indicates that Donald Trump has a higher public profile and receives more media coverage.

# Discussion

## **Sentiment Fluctuations and Events (Research Question 1)**

Figure 1 demonstrates that public sentiments about political figures are very responsive to important events. As Klomp (2025) discussed, Donald Trump's sentiment should be event-driven, and the Trump Tariff Announcement's sharp decline in sentiment fits with this notion. This event possibly calmed economic fears because the sentiment in "Economic Policy" discussions was more negative Figure 2. Chuck Schumer's positive sentiment on several cues immediately surrounding the Midterm Election Debate can be attributed to more visibility and heated conversation on sound policy.

## **Platform Differences (Research Question 2)**

The platform-specific differences that our observation in Figure 4 support the findings by Bowler et al. (2022) regarding the values of multiplatform approaches. Twitter’s more negative sentiment toward Donald Trump may be similar to that of its younger, urban demographic, which often takes a critical tone to conservative figures (Bowler et al., 2022). Its discussion-oriented nature and neutral sentiment on Reddit align with the nature of the discussions taking place on the site. Such disparity between the Twitter and Reddit sentiment towards Trump (p = 0.032) indicates the existence of inherent platform biases. Despite this, the small amount of Twitter data on other political figures, such as Chuck Schumer, prevented general comparisons.

## **Geographic and Demographic Variations (Research Question 3)**

The geographic analysis in Figure 2 shows that Georgia reacted negatively toward Trump, possibly because of its role in the agricultural sector, to which the President’s stance and implementation of tariffs have some negative effects. This corresponds with McKitrick et al. (2022), who mentioned that regional factors have some influence on political sentiment. Nevertheless, the analysis was limited by sparse location data (10 per cent of posts).

Demographic biases via the platform (Figure 3) also indicate that younger, urban Twitter users are generally more critical, as in Bowler et al. (2022) findings on demographic biases in social media. These results imply that sentiment expression is influenced by platform choice, and hence, Reddit is neither an extremist nor an Internet nanny. Nevertheless, this analysis has certain limitations, for instance, because demographic data (e.g., age, income) are not available directly.

## **Comparison with Polling (Research Question 4)**

The correlation between social media sentiment and Donald Trump’s polling approval is moderate (0.65), with digital sentiment tracking being another useful complement to traditional polling techniques, which is consistent with what Zhang (2024) posits. Social media’s capability to detect swift opinion shifts (opinion dips at the same time as polling drop, e.g. March 2025) is complemented by polling’s slower reaction, typically just after the event. However, the correlation is not absolute, as social media’s non-representative user base (Sloan & Quan-Haase, 2022) may have led to the imperfect correlation. Polling is important because it is structured sampling, while social media gives you the most timely and granular insights (topic-specific sentiment), but only for specific topics.

## **Methodological Contributions**

A scalable real-time sentiment tracking framework is provided through the use of R for a multiplatform pipeline (`collect\_pipeline.R`, `etl\_preprocessing.R`, `sentiment\_analysis.R`). Combining the lexicon-based (AFINN) approach and the machine learning approach (Random Forest) with LDA for topic modelling, we tackle both issues mentioned by Qi and Shabrina (2023). The resource-based and geographic mapping and analysis effect is contextual, which concurs with Bowler et al. (2022).

# Conclusion

It is shown in this study that social media analytics can be used to enrich political sentiment analysis, answering the four research questions in the context of Q1 2025. The reaction to events like the Trump Tariff Announcement completely differed between Twitter (more negative) and Reddit. States such as Georgia had geographic variations, and demographic proxies reflected the effect of platform user bases on the expression of sentiment. Social media’s complementary role is demonstrated by correlation with polling data (0.65), though it is acutely limited in demographic frequency and does not fully replace polling.

## **Implications**

In doing so, this research contributes to the field of computational political science with a strategic approach to perform real-time sentiment monitoring for political campaign, decision makers, and media professionals for obtaining actionable insights. With multiple social media sources, distortion caused by using a single platform is mitigated to present a more complete and judicious reading of the sentiment of the public.

## **Limitations**

However, there are important limitations: only 10% of posts had useful locations for geographic analysis, and most of the tweets between legislators and their followers did not have Twitter data for the legislators, thus limiting platform comparison. The study is not applicable to the real world because it uses simulated polling data and placeholder news comments. Furthermore, the demographic analysis on platform proxies was without direct user metadata.

## **Future Research**

Further studies should collect more Twitter data for that group of politicians to increase the spectrum of platform comparisons. An even better analysis would incorporate direct demographic data (via user surveys) and real news comments (via APIs). Adding more events and extending time period can lead to more insights into dynamics of sentiment. Finally, there may be an improvement in the accuracy of sentiment classification that can be made by integrating advanced models (e.g., transformer-based models like BERT) inspired by the base constructed here.

# References

Alslaity, A., & Orji, R. (2022). Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions. *Behaviour & Information Technology*, 1–26. <https://doi.org/10.1080/0144929x.2022.2156387>.

Ausat, A. M. A. (2023). The Role of Social Media in Shaping Public Opinion and Its Influence on Economic Decisions. *Technology and Society Perspectives (TACIT)*, *1*(1), 35–44. <https://journal.literasisainsnusantara.com/index.php/tacit/article/view/37>.

Bowler, S., Carreras, M., & Merolla, J. L. (2022). Trump Tweets and Democratic Attitudes: Evidence from a Survey Experiment. *Political Research Quarterly*, 106591292211373. <https://doi.org/10.1177/10659129221137348>.

Bülent Doğan, Yavuz Selim Balcioglu, & Meral Elçi. (2024). Multidimensional sentiment analysis method on social media data: comparison of emotions during and after the COVID-19 pandemic. *Kybernetes*. <https://doi.org/10.1108/k-09-2023-1808>.

Crinnion, F., Yannopoulou, N., & Bhattacharya, S. (2024). Fake news inside ideological social media echo chambers. *Elsevier EBooks*, 139–187. <https://doi.org/10.1016/b978-0-323-90237-3.00008-4>.

El Barachi, M., AlKhatib, M., Mathew, S., & Oroumchian, F. (2021). A novel sentiment analysis framework for monitoring the evolving public opinion in real-time: Case study on climate change. *Journal of Cleaner Production*, *312*, 127820. <https://doi.org/10.1016/j.jclepro.2021.127820>.

Hadis Bashiri, & Naderi, H. (2024). LexiSNTAGMM: an unsupervised framework for sentiment classification in data from distinct domains, synergistically integrating dictionary-based and machine learning approaches. *Social Network Analysis and Mining*, *14*(1). <https://doi.org/10.1007/s13278-024-01268-z>.

Klomp, J. (2025). Trump tariffs and the U.S. defense industry. *PLOS ONE*, *20*(1), e0313204. <https://doi.org/10.1371/journal.pone.0313204>.

Liang, Y., Yin, J., Park, S., Pan, B., Chi, G., & Miller, Z. (2023). Using social media user profiles to identify visitor demographics and origins in Yellowstone national park. *Journal of Outdoor Recreation and Tourism*, 100620. <https://doi.org/10.1016/j.jort.2023.100620>.

McKitrick, M. K., Schuurman, N., & Crooks, V. A. (2022). Collecting, analyzing, and visualizing location-based social media data: review of methods in GIS-social media analysis. *GeoJournal*. <https://doi.org/10.1007/s10708-022-10584-w>.

Niemiec, R., Berl, R. E. W., Gonzalez, M., Teel, T., Salerno, J., Breck, S., Camara, C., Collins, M., Schultz, C., Hoag, D., & Crooks, K. (2022). Rapid changes in public perception toward a conservation initiative. *Conservation Science and Practice*, *4*(4). <https://doi.org/10.1111/csp2.12632>.

Qi, Y., & Shabrina, Z. (2023). Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach. *Social Network Analysis and Mining*, *13*(1). <https://doi.org/10.1007/s13278-023-01030-x>.

Sloan, L., & Quan-Haase, A. (2022). The SAGE Handbook of Social Media Research Methods. *Www.torrossa.com*, 1–100. <https://www.torrossa.com/en/resources/an/5409559#page=110>.

Yao, Z., Yang, J., Liu, J., Keith, M., & Guan, C. (2021). Comparing tweet sentiments in megacities using machine learning techniques: In the midst of COVID-19. *Cities*, *116*, 103273. <https://doi.org/10.1016/j.cities.2021.103273>.

Zhang, T. (2024). Assessing the Use of Social Media as a Supplementary Tool for Public Opinion Analysis. *Knowledge UChicago*. <https://doi.org/10.6082/uchicago.11859>.