

# A Content Analysis of Reddit User's Response Towards Students for Fair Admissions v. Harvard

Term paper for SURV/SURVMETH 727

Chloe Chen & Kangrui Liu

2023-12-11

# Contents

<b>1</b>	<b>Abstract</b>	<b>3</b>
<b>2</b>	<b>Introduction</b>	<b>4</b>
<b>3</b>	<b>Methodology</b>	<b>5</b>
3.1	Data source . . . . .	5
3.2	Analysis techniques . . . . .	5
<b>4</b>	<b>Results</b>	<b>8</b>
4.1	Analysis . . . . .	8
<b>5</b>	<b>Conclusion and Discussion</b>	<b>14</b>
	<b>References</b>	<b>15</b>

# 1 Abstract

This study investigates public sentiment and discourse on Reddit regarding the landmark Supreme Court decision in *Students for Fair Admissions v. Harvard*, which ruled race-based affirmative action in college admissions unconstitutional. Our research aims to understand the impact of this decision on public opinion, particularly among Reddit users. The study focuses on three key questions: (1) Did the Supreme Court’s decision incite increased discussion about race-based affirmative action among Reddit users? (2) How did public sentiment change over this period, particularly before and after the Supreme Court’s decision? (3) What specific topics dominated these discussions? To answer these questions, we collected data from Reddit comments, employed the Latent Dirichlet Allocation (LDA) model for content filtering and sentiment analysis, supplemented by time series analysis to track sentiment changes. We also analyzed word frequencies to understand the focus of discussions. Our findings indicate a noticeable increase in discussions about affirmative action on Reddit following the Supreme Court decision. The overall sentiment trended negatively over time, suggesting growing public disapproval of affirmative action policies on the platform. Word frequency calculations reveal “Asian” as a predominant focus in the debate, with differing emotional tones in discussions about different racial groups.

**GitHub Link:** [https://github.com/krliu67/Final-Project\\_727.git/](https://github.com/krliu67/Final-Project_727.git/)

## 2 Introduction

*Students for Fair Admissions v. Harvard* is a landmark decision of the Supreme Court of the United States, in which the court determined that race-based affirmative action in college admissions violates the Fourteenth Amendment’s Equal Protection Clause on 29th June, 2023 (Totenberg 2023). This ruling barred all public and private colleges from considering race in admissions decisions, significantly changing the landscape of affirmative action in higher education, reflecting the deep divide and ongoing controversy over these policies. The case began in 2013 when Students for Fair Admissions (SFFA) filed a lawsuit against Harvard University. SFFA claimed that Harvard’s admissions practices violated Title VI of the Civil Rights Act of 1964 by discriminating against Asian Americans. In 2017, under the Trump administration, the Department of Justice opened an investigation into the allegations against Harvard’s policies. The case reached the Supreme Court, and on June 29, 2023, the Supreme Court issued a decision by a vote of 6-2 that reversed the lower court’s ruling and held that affirmative action in college admissions is unconstitutional (Chamberlain 2023).

The debated focus in this case is “affirmative action”, which has long been a controversial and often poorly understood policy (Crosby, Iyer, and Sincharoen 2006). Affirmative action refers to policies that support members of disadvantaged groups that have previously suffered discrimination in areas such as education and employment. In the United States, affirmative action typically involves measures to improve the opportunities for historically excluded groups, particularly racial minorities, which can include setting goals for inclusion, outreach efforts, and in some cases, considering race or gender as one factor in university admissions or hiring processes (Rubinfeld 1997). The concept is controversial for several reasons. Firstly, critics argue that it can undermine the principles of meritocracy and fairness, as it may give less qualified candidates an advantage over more qualified candidates due to their race or gender. It can also lead to reverse discrimination, where members of a majority group feel they are being unfairly discriminated against in favor of minority groups. Besides, there is a concern that affirmative action can stigmatize beneficiaries, making it seem that they cannot succeed without special treatment, which can be detrimental to their self-esteem and societal views of their capabilities (Skrentny 2018). Further, the debate often centers around whether the focus should be on equality of opportunity (not discriminating and allowing everyone an equal chance) or equality of outcome (actively working to ensure that groups that have been historically disadvantaged are represented proportionally) (Fryer Jr and Loury 2005).

Therefore, studying public opinion towards *Students for Fair Admissions v. Harvard*, particularly the Supreme Court’s decision, is of paramount importance. It offers a comprehensive perspective on the attitudes of various stakeholders towards the policy. This analysis can hopefully reflect the role that the policy plays in contemporary American society, and yield insights for policy implications.

Specifically, we aim to answer: (1) Did the Supreme Court’s decision trigger increased discussions among Reddit users about race-based affirmative action in college admissions? (2) How did people’s sentimental tendencies change during that period (especially, would this trend shift before and after the announcement of the Supreme Court’s decision)? (3) What topics are of interest to these discussions? This research hopes to get insight into the above questions through the lens of social media with the aid of sentiment analysis. In the following section, we will specify the data sources and data gathering process.

## 3 Methodology

### 3.1 Data source

We collected Reddit comments posted before and after the Supreme Court decision leak (5/15/2015 - 11/30/2023), focusing on those containing keywords relevant to this topic, such as “affirmative action” “fair admission” “SFFA” and “admission discrimination”. Our search will be specifically targeted at certain subreddits. This approach is due to the difficulty in accurately identifying our intended content through a broad, site-wide search. To address potential representative biases that might arise from limiting our search to specific subreddits, we have intentionally chosen a diverse range of subreddits, including those named after various races and representing both major political parties, ensuring considerable heterogeneity in our data.

### 3.2 Analysis techniques

To access the public opinion, we resorted to Reddit, which is one of the most popular websites in the United States over the years. At the time of writing, Reddit is the 3rd most popular website nationally. The prevalence of Reddit suggests a wealth of data and the potential for capturing a broad range of views. Additionally, the platform’s anonymity encourages users to share their true thoughts without concern for social approval. Furthermore, Reddit is more conducive to research than other well-known social networks such as Facebook and Twitter, due to its provision of readily accessible data at no cost.

We collected relevant data using RedditExtractoR API on R. Briefly, we use keywords “affirmative action” “fair admission” “SFFA” and “admission discrimination”. Our search is specifically targeted at certain subreddits as listed in the graph that follows. This approach is due to the difficulty in accurately identifying our intended content through a broad, site-wide search. To address potential representative biases that might arise from limiting our search to specific subreddits, we intentionally chose a diverse range of subreddits, including those named after various races and representing both major political parties, ensuring considerable heterogeneity in our data.

We employ the Latent Dirichlet Allocation (LDA) model to filter posts unrelated to our investigation after initial data scraping. After this, we conducted sentiment analysis coupled with time series analysis to investigate changes in sentiment over time. We also organized the most commonly used words with wordcloud to depict word frequencies associated with different sentiments. Finally, we include a total of 14785 comments in our analysis. Our working flow is shown in the 1.

The LDA model successfully identified 30 prominent topics. These topics were characterized by distinct sets of keywords above. Below are examples of topics listed by LDA model that were strongly associated with our goal.

Table 1: LDA-Filtered Topics

topic_number	topic
Topic 2	supreme court kavanaugh confirmation justice brett trump senate amp vote
Topic 3	affirmative action hear realize affirmativeaction chance supreme michigan caucasians bet
Topic 4	asian harvard americans american amp admissions many years lower policies

---

topic_number	topic
Topic 5	capitol us trump supporters storm protesters building police protrump lawmakers
Topic 6	court supreme ban travel trump decision ruling upholds us trumps
Topic 8	asian asianamericans asianamerican used college racial law americans use hope
Topic 12	people minority two law marginalized race say goal mean population
Topic 13	believe today civil marriage justice don now equal women change
Topic 14	court whether opinion federal decision circuit gt case issue can
Topic 16	action affirmative admissions college gt thomas race strikes clarence harvard
Topic 19	case university court day father work vance nassar child action
Topic 20	black white racism racist women woman minorities poc ever just
Topic 22	school schools can research based education admission part students put
Topic 23	students university admissions higher colleges education said race school student
Topic 24	people lgbtq discrimination sexual title support act gender community federal

---

## Methods

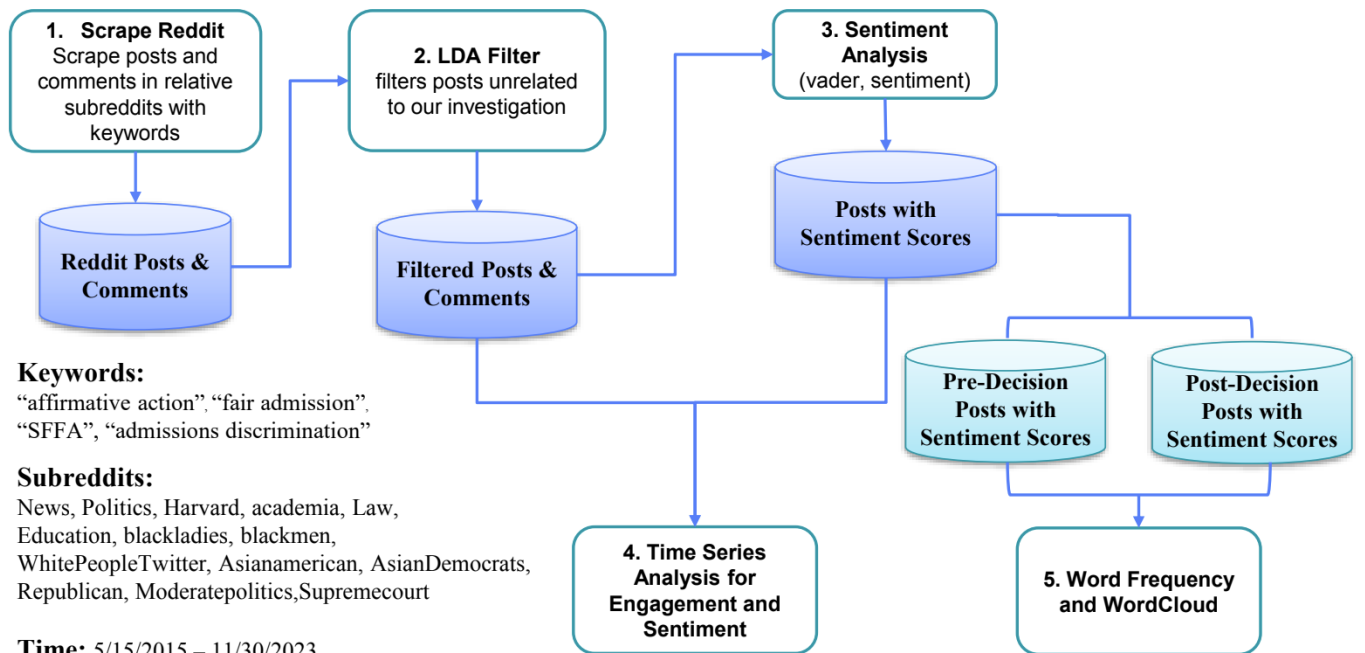


Figure 1: Flowchart

## 4 Results

### 4.1 Analysis

#### 4.1.1 Reddit Engagement Trend Over Time

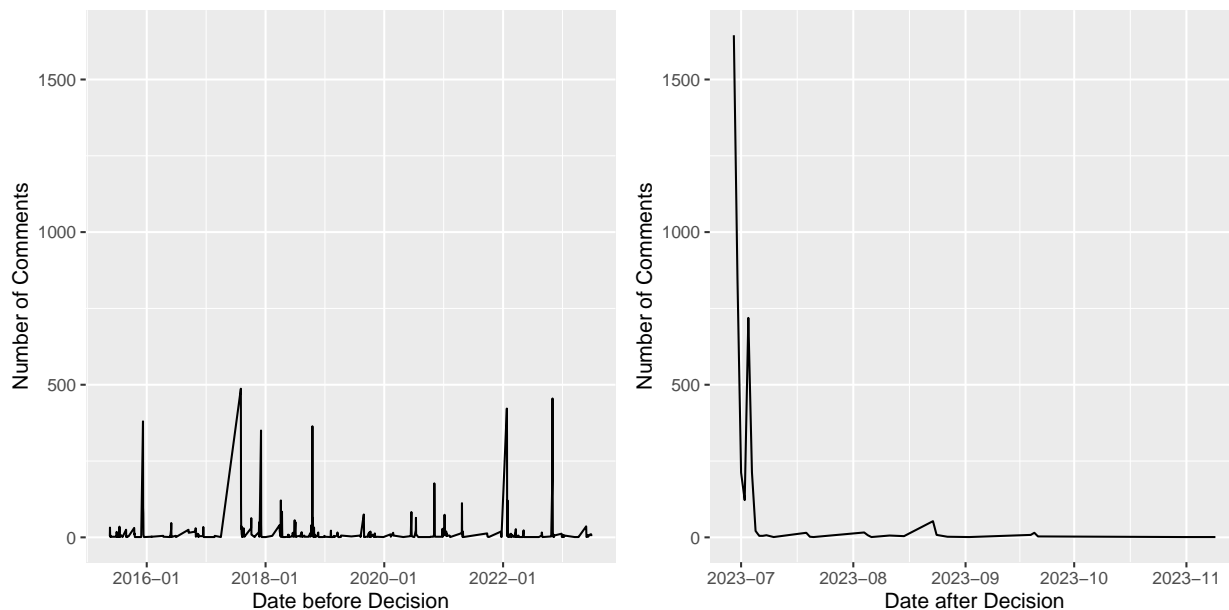


Figure 2: Number of Comments

We first calculated the number of posts before and after the decision of the Supreme Court, we can see a steep increase in the engagement in Reddit about relative topics after the decision date as expected. This feature shows in 2.

#### 4.1.2 Sentiment Trend Over Time

**4.1.2.1 Vader** We analyzed the sentiment trend by calculating vader sentiment score. we smoothed the day-by-day jitters by plotting the rolling means across a period of 15 days for the data before the decision in 3, and a period of 3 days for the data after the decision to make the plot more readable. The plot shows that people’s sentiment towards this incident fluctuates a lot before and after the decision.

In order to have a overview of the general trend statistically, we employed the Mann-Kendall trend test to examine whether there is a monotonic trend in the sentiment change over time. As is shown in the results, the Mann-Kendall trend test indicates a downward trend in the overall sentiment. This suggests that, over the entire period from 2015 to 2023, public sentiment on Reddit regarding affirmative action became more negative. Below is our Mann-Kendall trend test result.



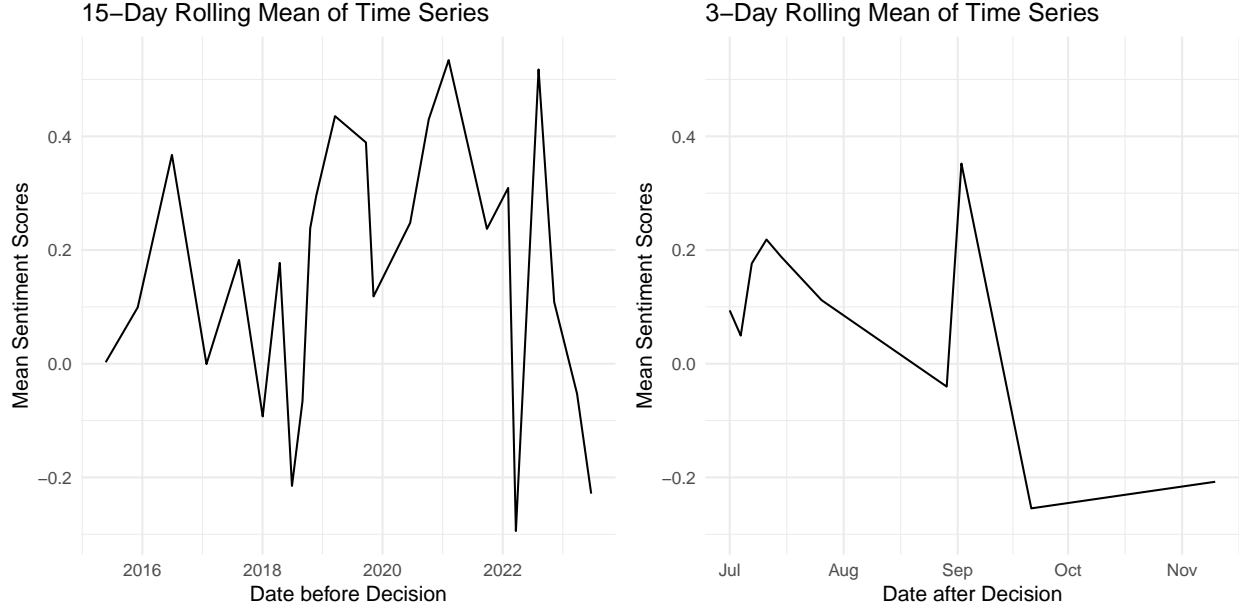


Figure 3: Rolling Mean of Time Series

Table 2: Mann-Kendall Trend Test by Vader

	vader__tau	vader__p
data_overall	-0.0623	0.0490
data_before	-0.0381	0.2861
data_after	-0.1170	0.3550

**4.1.2.2 loughran** In order to get a more nuanced view of the public sentiment change, we further used the Loughran-McDonald Sentiment Word Lists. The Loughran-McDonald Sentiment Word Lists are often used in finance and accounting research to assess the tone of financial documents. Unlike typical sentiment analysis models that classify words into positive, negative, and neutral categories, the Loughran-McDonald lists were specifically designed to account for the context in which words are used in financial reports. The result shows in 4.

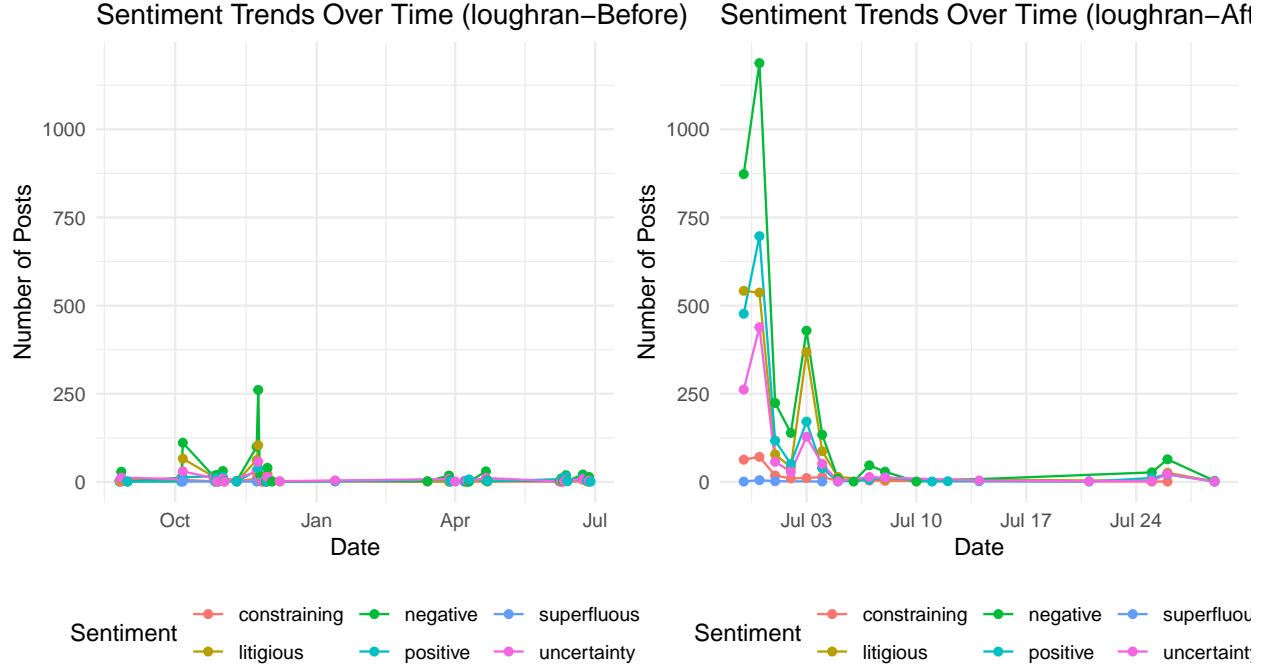


Figure 4: Loughran-McDonald Sentiment Word Lists

The result correspond the the former one, as the plot shows a sharp increase in negative sentiment after the decision, and the overall count of positive comments is relatively low. Among the sub-category, we can see a relatively considerable percentage of *litigious* comments all across the time period, this reflects the controversial nature of the topic.

**4.1.2.3 NRC** The NRC Emotion Lexicon, also known as EmoLex, provides another way of looking into the sentiment categories. It is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive) developed by Saif Mohammad and Peter Turney. The lexicon categorizes each word into these emotions and sentiments based on human annotations.

Observing 5 and 6, Both positive and negative comments came to a peak after the leak of the decision. This corresponds with the increased engagement revealed in our first step of analysis.

Looking at the trends of comments with positive sentiment, “trust” is the most noticeable component all across the period, followed by “anticipation”. The condition is more complicated for comments with negative sentiment, “anger” “sadness” and “fear” seem to overlap with each other during this period.

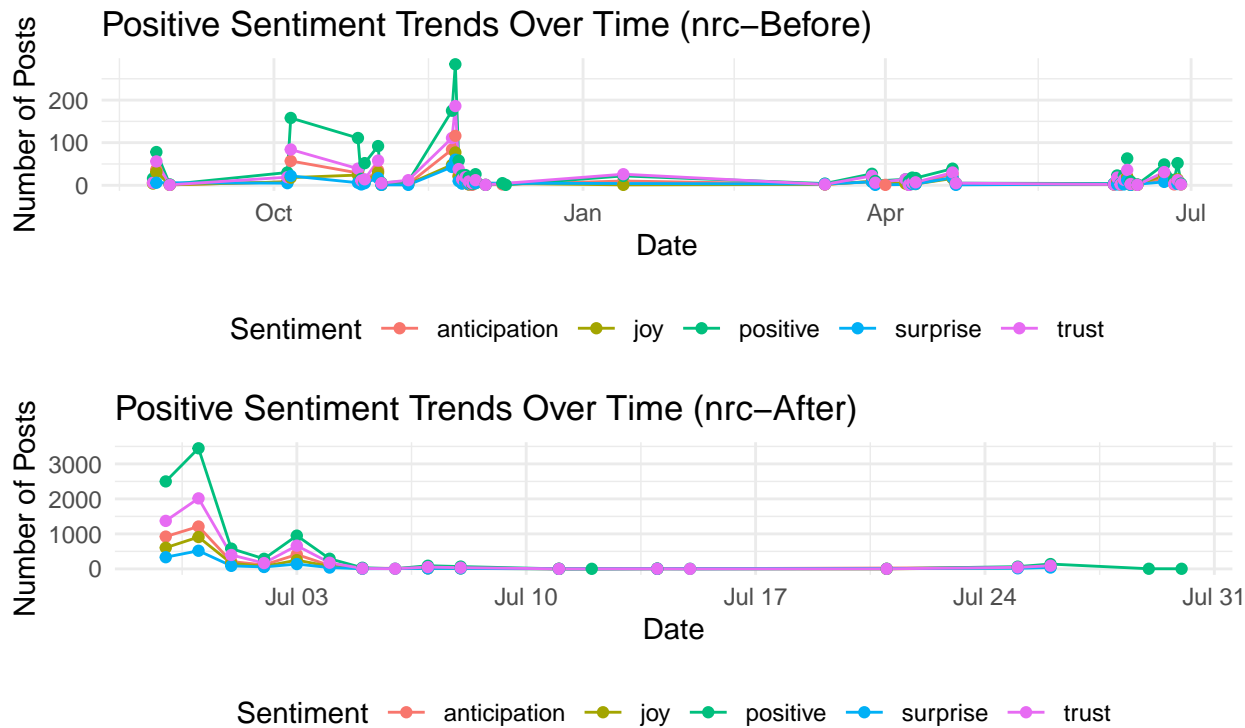


Figure 5: NRC Emotion Lexicon

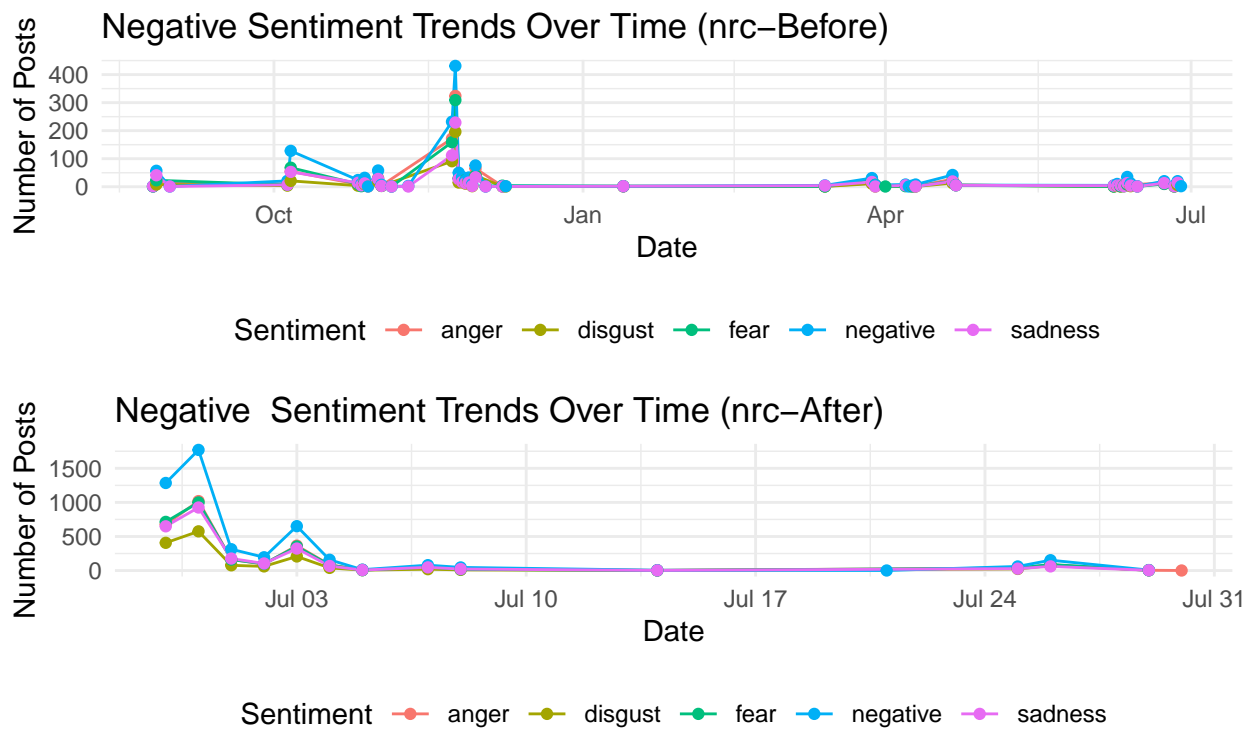


Figure 6: NRC Emotion Lexicon

### 4.1.3 Word Frequency by Sentiment

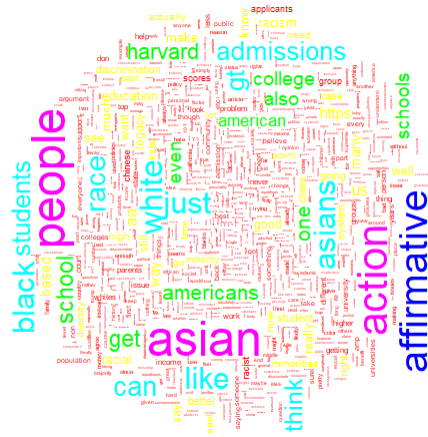


Figure 7: WordCloud for all comments



Figure 8: WordCloud for positive comments

We then look into the most commonly used words in different sentiments, which are 7, 8 and 9. The calculation of word frequency indicates the debate overall is focused on “Asian”. There are nuanced differences between negative comments and positive posts, as “black” is a considerable component of comments with negative emotions. We further got a view of which words we might see more of if related words such as “asian”, “american”, and “affirmative”, “action” are removed, since it seems these are being picked up the most due to the nature of the topic of discussion. We can see that among the comments related to negative feelings, racism is the dominant accusation.

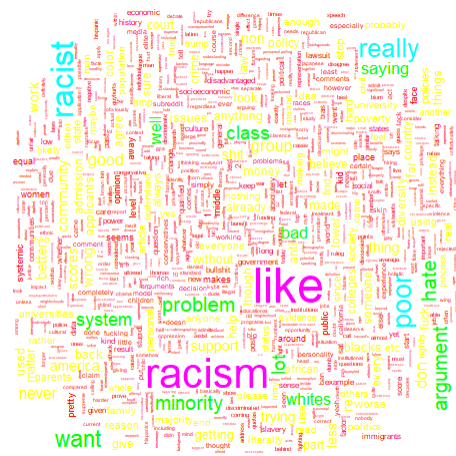


Figure 9: WordCloud for negative comments

## 5 Conclusion and Discussion

In conclusion,

1. The decision of the Supreme Court might be the trigger of public engagement on Reddit about this topic.
2. Regarding the trend of attitude over time, we observed volatility in public sentiment on Reddit towards affirmative action during this period. This volatility might reflect the public’s reaction to various events related to affirmative action during this period. This phenomenon also corresponds with existing studies, which revealed perceptions of affirmative action fluctuate substantially and in an irregular manner over time[Ewoh and Elliott (1997)](Schuman et al. 1997). The variability in support for affirmative action may be anticipated, considering the ambiguity and breadth of the concept, as well as the prevalent misunderstandings regarding its definition(Crosby and Cordova 1996).
3. But the downward trend in the overall sentiment is also noteworthy, suggesting public sentiment on Reddit regarding affirmative action became more negative.
4. The calculation of word frequency indicates the debate overall is focused on “Asian”. There are nuanced differences between negative comments and positive posts, as “black” is a considerable component of comments with negative emotions. After removing words used for searching, we can see racism is the dominant accusation among the comments related to negative feelings.

There are many limitations of the current study. As mentioned above, we do not know how representative the subreddit sample was of the whole U.S. population. Therefore, we could not blindly generalize the findings to the whole United States. Also, the LDA model hasn’t been trained to the utmost, so the current posts may still include some noises.

## References

- Chamberlain, S. 2023. "Supreme Court Outlaws Affirmative Action in College Admissions in Landmark Decision." *The New York Post*.
- Crosby, F. J. and D. Cordova. 1996. "Words Worth of Wisdom: Toward an Understanding of Affirmative Action." *J. Soc. Issues* 52:33–49.
- Crosby, F. J., A. Iyer, and S. Sincharoen. 2006. "Understanding Affirmative Action." *Annu. Rev. Psychol.* 57:585–611.
- Ewold, A. I. and E. Elliott. 1997. "End of an Era? Affirmative Action and Reaction in the 1990s." *Rev. Public Pers. Admin.* 38–51.
- Fryer Jr, R. G. and G. C. Loury. 2005. "Affirmative Action and Its Mythology." *Journal of Economic Perspectives* 19(3):147–62.
- Rubinfeld, J. 1997. "Affirmative Action." *Yale LJ* 107:427.
- Schuman, H., C. Steeh, L. Bobo, and M. Krysan. 1997. *Racial Attitudes in America: Trends and Interpretations*. Revised. Cambridge, MA: Harvard Univ. Press.
- Skrentny, J. D. 2018. *The Ironies of Affirmative Action: Politics, Culture, and Justice in America*. University of Chicago Press.
- Totenberg, N. 2023. "Supreme Court Guts Affirmative Action, Effectively Ending Race-Conscious Admissions." *NPR*.