Highlighting Ethnic Biases in COVID-19 Articles

Kanika Chopra
University of Toronto
kanika.chopra@mail.utoronto.ca

Abstract

The COVID-19 pandemic resulted in discrimination, xenophobia and racism towards Asians globally. This has been shown through the increase in targeted hate-crimes and negative language being used towards these communities. Language analysis can be used to capture societal bias in literature, including news articles. We apply a framework leveraging word embeddings to quantify the bias towards Asians by studying the association with COVID-19 terms, hate crimes and outsider adjectives in global news articles from January to July 2020. The general sentiment of Asian (in comparison to White) ethnicities is also explored and quantified using a larger, more general list of adjectives. This work has revealed that the language used by the media showed a negative change in the perception of Asians during the pandemic, especially at the beginning. In addition, Asians are associated with more negative adjectives in comparison to White ethnicities during this time frame. This study has opened up the door to continue an analysis into how the language used by the media has demonstrated unfair ethnic biases from global events, such as other pandemics.

Keywords: COVID-19, Word Embeddings, Bias

1 Introduction

Language is able to capture stereotypes in literature, even if they are present subtly. Studying these biases is often done from a sociological or linguistic lens to understand how society's perceptions are portrayed in literature. This provides a window into the lens of society's portrayals on certain minority groups. During significant global events, these biases can be further amplified and alter the perceptions of certain groups. This has proven to be true for Asians during the COVID-19 pandemic. This pandemic enabled the spread of racism and xenophobia resulting in Asian-Americans becoming more vulnerable to hate-crimes. Between 2019 and 2020 the FBI noted a "77 percent increase in hate crimes against Asian people living in the US" (Findling et al., 2022). In just the United States, this resulted in the formation of the Asian-American Pacific Islander Equity Alliance (AAPI Equity), Chinese for Affirmative Action (CAA) and Stop AAPI Hate coalition to protect and advocate for the rights and safety of this community. From when the pandemic was officially declared in March 2020 to June 2021, "more than 9000 anti-Asian hate incidents were self-reported to Stop AAPI" (Findling et al., 2022). These numbers are under-reported and only include a subset of the hate-crimes towards Asians. However, they are able to clearly show that there was discrimination towards Asians that was heightened greatly during the pandemic.

Despite the efforts made by WHO to not attach an ethnicity or location to the nomenclature of the virus by purposefully using a scientific name, the virus was often referred to as the "Wuhan Virus", "Chinese Virus", or "Asia Virus" on social media. The discrimination was further fueled by those in power using social media to wrongfully place blame on Asians, such as Former President Donald Trump referring to the coronavirus as the "China Virus". This tweet alone resulted in an increase in pejorative language towards Asians on social media (Hswen et al., 2021) and an increase in news reports covering anti-Asian hate-crimes. These biases were not just limited to the United States but were occurring globally with increases in hate-crimes around the world towards Asians (Haynes, 2021). With all of this media surrounding the pandemic, the written language can be used to reveal changes in the perceptions and treatment of Asians. This opens the doors to leverage computational linguistic methods to detect and quantify these biases in writing.

In this paper, we provide a quantifiable analysis investigating the COVID-19 pandemic's influence on the perception of Asians in global news articles. This is done by leveraging word embeddings to quantify the association with words related to COVID-19, hate crimes and outsiders to reveal the bias towards Asians. Further analysis will be done to compare the sentiment behind the top 10 most-biased adjectives for Asians (relative to Whites). This work will further validate the anti-Asian bias during the pandemic through a quantifiable framework.

2 Related Work

Previous work in this field has shown that bias, specifically racial bias, exists in natural language processing (NLP) methodology (Field et al., 2021) demonstrating the need to consider how discrimination in literature is upheld in models. Word embeddings specifically have been leveraged to measure ethnic and gender bias allowing for the quantification of historical trends (Garg et al., 2018). This work established and validated a framework to use word embeddings to investigate bias. Further work has been done to study change and stability in stereotypes by exploring the top associated words over decades and the valence of stereotypes (Charlesworth et al., 2022). Other characteristics of language have also been considered to highlight bias included studying parts-of-speech tags, semantic categories, valence, arousal and dominance in gender-associated words (Caliskan et al., 2022). This work also analysed how the frequencies of words associated with gender differs. These previous works have shown why considering bias in NLP is important and how word embeddings and language analysis can be used to discover the linguistic bias.

In the context of COVID-19, exploring the impacts of the pandemic on Anti-Asian hate and the affected communities has been done from an sociological and exploration perspective (Gover et al., 2020). This work has highlighted the negative experiences of Asian Americans due to the pandemic. Investigating bias has also been done by measuring the association of social media hashtags #covid19 and #chinesevirus with Anti-Asian sentiment (Hswen et al., 2021) to reveal the latter has a more negative, hateful sentiment.

3 Data and Other Materials

3.1 COVID Articles

The main data set used was Aylien COVID-19 news data. This data set included 1,673,353 English news articles on coronavirus related topics (Hamer, 2020). Articles were curated from 440 global sources from the time frame of November 2019 to July 2020.

3.1.1 Data Preprocessing

The data was first cleaned to only include the necessary columns such as publishing date, the body text and columns for an exploratory analysis such as character, paragraph, sentence and word count. From all of the articles, Table 1 summarizes the exploration into article counts. These articles came from global news sources including the Daily Mail UK, Reuters, Yahoo, India Times, Forbes, etc. A histogram of the top 10 most frequent news sources can be found in Appendix A.1.

Sentences	Paragraphs	Words
23	18	522

Table 1: Average counts per article

For the purposes of training the word embeddings, only publishing date and the body text were retained. From the date column, values that were not logical were removed as they were approximately 0.0003% of the data set. Further analysis of the distribution of articles showed that there was not sufficient data for 2019 so the time frame was adapted to January - July 2020. To keep the distribution of articles consistent among each of the months, 20,000 articles were randomly sampled from every month. To preprocess the body text, text cleaning was applied to remove unicode and newline characters; convert all words to lowercase; remove any punctuation; tokenize the body text; and then remove any stop words.

3.2 Word Lists

The two types of lists used are group words and neutral words. Group words are used to identify ethnicity in this paper using last names. This included the top 20 and 22 most common White and Asian last names respectively curated by (Garg et al., 2018). Neutral words are not intrinsically ethnic and are used to compare the association with group words. The following categories were considered for this analysis:

1. COVID-19 related terms
2. Hate crimes
3. Outsider adjectives

The first two categories were curated by online sources using the definitions and most commonly associated words. For outsider adjectives, we rely on a list curated by (Garg et al., 2018). For an exploratory analysis of adjectives associated with each ethnicity, a larger list of general adjectives (Garg et al., 2018) was used. Each word list was reduced to only include words that appeared in every month's sample of articles resulting in 8 words for COVID-19, 8 words for hate crimes, 13 words for outsider adjectives and 325 general adjectives. The full list of words for each category can be found in Appendix A.2 and A.3.

4 Computational Methodology

Word2vec embedding models were trained on each month's articles separately to allow comparisons over time. To ensure the word embeddings were identifying semantic relationships, we ran quality checks including checking analogy relationships and exploring the nearest neighbors for certain words such as "car" and "food". See Appendix A.4 for details. We then used these embedding models to get a higher-dimensional representations for each word in the group and neutral word lists for quantifying embedding bias.

4.1 Quantifying embedding bias

To quantify embedding bias, two variations were conducted to compute the group representational vector. The first computed a group vector for each ethnicity and the second for each category. To compare similarities between word embeddings, the Euclidean distance is used. The smaller the distance, the more associated the words are in meaning.

4.1.1 Group Ethnicities

Let A and W be the number of Asian and White last names respectively. Let w_i represent a word in the given group list and c_i represent a word from a category list of size n. We define the following to compute embedding bias:

1. $\mu_{\text{asian}} = \frac{1}{A}\sum_{i=1}^{A} w_i$
2. $\mu_{\text{white}} = \frac{1}{W}\sum_{i=1}^{W} w_i$
3. $\delta_{\text{asian},i} = ||\mu_{\text{asian}} - c_i|| \quad \forall i$
4. $\delta_{\text{white},i} = ||\mu_{\text{white}} - c_i|| \quad \forall i$
5. embedding bias $= \frac{1}{n}\sum_{i}^{n}(\delta_{\text{asian},i} - \delta_{\text{white},i})$

In this scenario, the association between an ethnicity and every word of a category is first computed and then the difference is the bias for each respective word. We take the embedding bias of a category to be the average bias for each word in the category. This provides a holistic understanding of the bias associated with one category. If the average bias is more negative, the category is more associated with the Asian ethnicity and vice versa.

4.1.2 Group Categories

Following the notation from above, the embedding bias for this method can be calculated as follows:

1. $\mu_{\text{category}} = \frac{1}{n}\sum_{i=1}^{n} c_i$
2. $\delta_{\text{asian}} = \frac{1}{A}\sum_{i}^{A} ||w_{\text{asian},i} - \mu_{\text{category}}||$
3. $\delta_{\text{white}} = \frac{1}{W}\sum_{i}^{W} ||w_{\text{white},i} - \mu_{\text{category}}||$
4. embedding bias $= \delta_{\text{asian}} - \delta_{\text{white}}$

When calculating the group categories, we are unable to do a 1:1 comparison between the association with last names similarly to what was done in the prior section. Hence, we take the average association for an ethnicity with the category and then calculate the difference of the average associations with a category for each ethnicity.

4.2 Calculating adjective sentiment

We calculate the sentiment behind the most associated adjectives with each ethnicity over time to provide an insight as to whether an ethnicity is associated with more positive or negative words. For this section, group representational vectors are based on ethnicity, μ_{asian} and μ_{white}. Similarly to the Section 4.1.1, the association between the group vector and each adjective is calculated. The difference is then used as the embedding bias for each respective adjective. The top 10 most associated and biased adjectives were extracted.

For each adjective, we computed the compounded sentiment, an aggregate score between -1 to 1 of how negative or positive a word is. The average sentiment for the top 10 adjectives for each ethnicity was calculated for each month.

The code for this project is publicly available on GitHub.

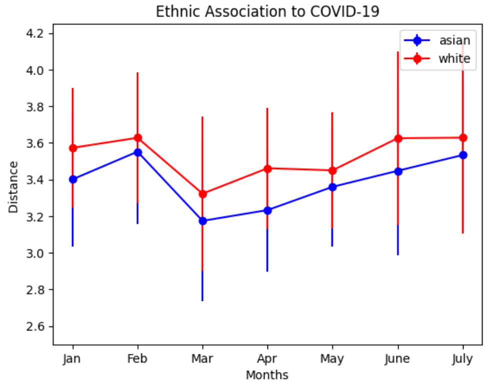

Figure 1: Ethnicity Group Comparison

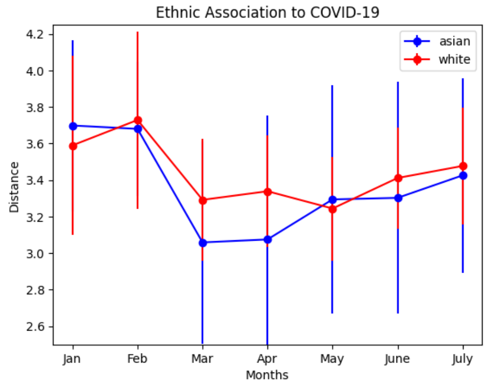

Figure 2: Category Group Comparison

5 Results

5.1 Association with Neutral Lists

We condense the results for both group methods to show a comparison of how bias is measured.

5.1.1 Association with COVID-19 Terms

The association with COVID-19 terms is measured over time in Figures 1 and 2. Distance is measured on the y-axis, so the lower the plot is, the higher the association. We can see in both plots that Asian last names are generally more associated with COVID-19 terms than White last names, with a clear distinction in Figure 1. In Figure 2 there is more overlap with the two ethnicities with a larger separation between March and April. This is also the largest drop and then COVID-19 becomes less associated with both ethnicities after March. This seems to align with the fact that the World Health Organization (WHO) announced that COVID-19 was a global pandemic on March 11, 2020 (Cucinotta and Vanelli, 2020). This delaration may be the source of the increase in association.

5.1.2 Association with Hate Crimes

We see a similar analysis when looking at Figures 3 and 4 where the association between both ethnicities is similar in January and February and then hate crimes become more associated with Asians after March. It is interesting to see that there was only one big dip in the data and that other events throughout the pandemic, i.e. shut down did not cause any other dips. Instead we see that after March, the association slowly decreases and becomes similar to White ethnicities again by May. This is interesting to note since hate crimes continued to occur throughout the pandemic and were not a one-month occurrence.

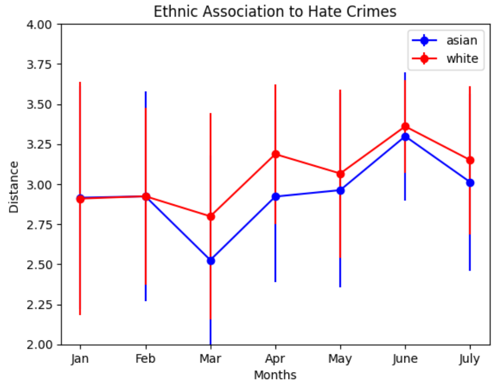

Figure 3: Ethnicity Group Comparison

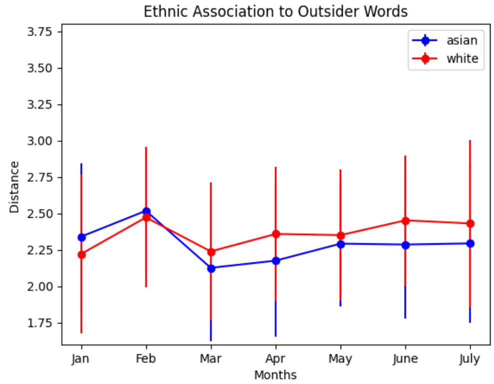

Figure 5: Ethnicity Group Comparison

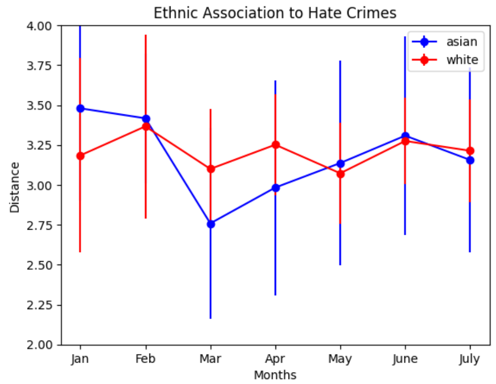

Figure 4: Category Group Comparison

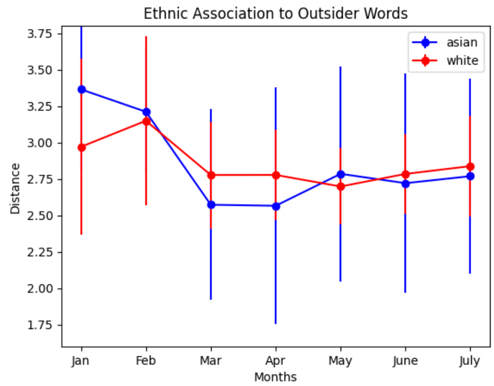

Figure 6: Category Group Comparison

5.1.3 Association with Outsider Adjectives

Looking at the association with the outsider adjectives, we see a more similer relationship between ethnicities than other categories. There is a dip in March with Asians becoming more associated than Whites; however, they become very similar and stagnant again after May.

5.1.4 Statistical Tests

For every word category, two types of statistical tests were conducted. First, a two-sample t-test was conducted to determine if there was a significant difference in the average association between ethnicities. The results of these statistical tests for both computational methods is shown in Table 2 and 3. In Table 2 the p-value is always greater than a significance level of 0.05; hence, there is no significant difference in the average association for ethnicities. However, in Table 3, we see that we have a significant difference in means for outsider adjectives in January and for hate crimes in March. The latter result agrees with earlier analyses on the greatest change being when the pandemic was declared. In these cases, it is important to note that our sample size is very small as when we have group vectors for ethnicity, we are comparing only 7 values. When we have group vectors for category, we can compare the mean for each category word but this is still only 8-13 words.

The second statistical test was to measure if there was any significant changes in the embedding bias over time. For this, the Pearson correlation was calculated between all months for the embedding bias from each category list. The adjacent months were the values of interest to measure changes between months. A heat map was generated to display these

Month	COVID Terms	Hate Crimes	Outsider
Jan	0.3695	0.9880	0.5835
Feb	0.7106	0.9989	0.8096
Mar	0.5301	0.4427	0.5757
Apr	0.2216	0.3287	0.3719
May	0.6083	0.7372	0.7475
June	0.4874	0.7462	0.4029
July	0.7080	0.6209	0.5545

Table 2: P-values from two sample-test

Month	COVID Terms	Hate Crimes	Outsider
Jan	0.4788	0.1261	0.0392
Feb	0.7248	0.7523	0.7050
Mar	0.1125	0.0357	0.2260
Apr	0.1178	0.1123	0.2784
May	0.7438	0.6862	0.6181
June	0.4799	0.8258	0.7203
July	0.7103	0.6962	0.6812

Table 3: Two sample-test for Category Group

Pearson correlations for each category and ethnicity and is displayed in Appendix A.5. To determine significance, a Kolmogorov-Smirnoff test was used to determine if any of the changes in Pearson correlation were significant. These calculations and code were mimiced from (Garg et al., 2018) and are detailed in Appendix A.5. The resulting p-values are displayed in Table 4. We note that this test could not be completed for the method using a group vector for categories as we would be comparing last names which are different for both ethnicities.

Phase	COVID Terms	Hate Crimes	Outsider
Jan - Feb	0.34693	0.19506	0.01215
Feb - Mar	2.67E-06	0.14141	0.19506
Mar - Apr	0.00727	0.67253	0.55117
Apr - May	0.04685	0.01955	0.03064
May - June	0.14141	0.55117	0.19506
June - July	0.00125	0.79064	0.19506

Table 4: Kolmogorov-Smirnoff Tests for Phase Change

We can see that there are significant phase shifts in February to March, March to April, April to May and June to July with COVID terms. This could align with the beginning of the pandemic and perceptions changing. With hate crimes, we can see significant phase shifts from April to May. Lastly, for outsider adjectives, we can see significant phase shifts from January to February and April to May which may correspond when the outbreak initially started and as perceptions changed as the pandemic continued.

5.1.5 Most Biased Last Names

For every word category, the top 5 most associated last names for both ethnicities was extracted. The average association was then calculated for each ethnicity and is shown in Appendix A.6. We observe a larger difference in the plots for White and Asian ethnicities in every word category. There is a similar dip in the plot in March for Asians meaning it became more similar, but not for Whites. However, significance tests could not be measured since the sample size was too small with only 5 last names.

5.2 General Adjective Sentiment

The top 10 most associated adjectives for each ethnicity is shown in Appendix A.7. However, some words could be equally highly associated with both ethnicities resulting in a low bias. To focus on the bias towards Asians, the top 10 most biased adjectives is more useful. The top 5 words for every other month is shown in Table 5. See Appendix A.7 for the full list.

Jan	Mar	May	July
Protective	Artificial	Insulting	Malicious
Praising	Transparent	Malicious	Cooperative
Cerebral	Informal	Transparent	Assertive
Affected	Sensitive	Cooperative	Arbitrary
Suspicious	Cultured	Praising	Outrageous

Table 5: Most Biased Adjectives

We can see that for each month, there are many negative words apparent such as *suspicious, artificial, malicious, insulting, outrageous*. This shows bias exhibited towards Asians from an exploration perspective. To quantify the sentiment, Figure 7 shows this average over the months.

Although we have negative values in some months, the average sentiment is relatively neutral. This is interesting to note since we observed negative adjectives in Table 5. However, since we are taking the average compounded sentiment, some positive words such as *praising* and *cooperative* could skew the mean.

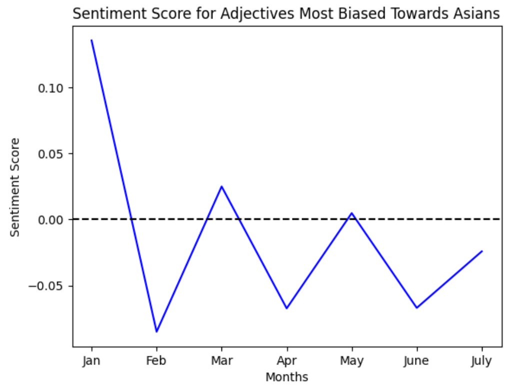

Figure 7: Sentiment of Most Biased Adjectives

6 Discussion

In this work, we have investigated how word embeddings can be used to reveal bias towards Asians during the COVID-19 pandemic. Our methods were applied to train monthly embeddings using global news articles. Bias was quantified by measuring the ethnic associations with COVID-19 related terms, hate-crimes and outsider adjectives. This allowed us to track how the associations and bias for each category changed as the pandemic progressed. In each case, we have demonstrated that there was a slight bias towards Asians, although it was statistically insignificant. This significance is taken with a grain of salt as there were small sample sizes. We have also shown that negative adjectives are more associated with Asians throughout the pandemic even if the average sentiment was relatively neutral. More robust methods of measuring the sentiment could be used instead of the average, such as the proportion of negative words to positive words instead of the mean. This study allowed us to understand how the news reflected negative attitudes towards Asians during the pandemic.

Due to data limitations, there is a lack of a comparison of the perceptions before COVID. This would allow for a better understanding of how the pandemic altered perceptions. However, this has proven to be difficult as COVID-19 was discussed more in news articles after January 2020 when it became more globally known. The perceptions may also be different geographically and aggregating it may dampen the bias in certain countries. Hence, it may be beneficial to run finer-grained analyses at a country-level. This would also allow for a comparison of the association with hate-crimes with a country's hate-crime statistics. Currently, there are no external metrics to validate that the word embeddings are capturing these biases accurately so this comparison would add more validity to the results. In addition to this, when investigating the association with hate-crimes, there could be sentences with both White and Asian last names but it is not clear where the blame is associated. Hence, a method to use parts of speech could be used to distinguish the perpetrator from the victim.

Furthermore, the robustness of these results depends on the word lists that are being used. Once the word lists were reduced to words that appeared in every month, the lists became quite small and hence the statistical power to run any significance tests is low. The small sample sizes also attribute to the large standard deviations across category words and last names in the category association plots.

Generating larger word lists for each of the topics as well as the last names would allow for a better analysis, especially with the high deviations based on category words and last names as shown in the Results.

Moreover, there are limitations with using word embeddings to quantify associations as word2vec is "black-box" where "the dimensions have no inherent meaning" (Garg et al., 2018). Hence, embedding models that leverage parts of speech, or other linguistic characteristics could be leveraged.

In conclusion, it would be interesting to see if there was a bias towards Asians when comparing the association with blame-related words in the context of the pandemic. This future extension could allow for a method to quantify the blame that resulted in more xenophobia as the pandemic progresses. This work provides a method to quantify how the COVID-19 pandemic resulted in bias towards Asians and can be further extended to studies to measure how other pandemics and global events have resulted in ethnic biases.

References

[Caliskan et al.2022] Aylin Caliskan, Pimparkar P. Ajay, Tessa Charlesworth, Robert Wolfe, and Mahzarin R. Banaji. 2022. Gender bias in word embeddings: A comprehensive analysis of frequency, syntax and semantics. *arXiv*.

[Charlesworth et al.2022] Tessa E. S. Charlesworth, Aylin Caliskan, and Mahzarin R. Banaji. 2022. Historical representations of social groups across 200 years of word embeddings from google books. *PNAS*, 119(28).

[Coates2015] Jennifer Coates. 2015. *Women, men and language: A sociolinguistic account of gender differences in language*. Routledge.

[Cucinotta and Vanelli2020] Domenico Cucinotta and Maurizio Vanelli. 2020. Who declares covid-19 a pandemic. *National Library of Medicine*, 91(1):157–160.

[Field et al.2021] Anjalie Field, Su Lin Blodgett, Zeerak Wasee, and Yulia Tsvetkov. 2021. A survey of race, racism, and anti-racism in nlp. *arXIV*.

[Findling et al.2022] Mary Findling, Robert J.Blendon, John Benson, and Howard Koh. 2022. Covid-19 has driven racism and violence against asian americans: Perspectives from 12 national polls. https://www.healthaffairs.org/do/10.1377/forefront.20220411.655787/. Accessed: 2022-12-01.

[Garg et al.2018] Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16):E3635–E3644.

[Gover et al.2020] Angela R. Gover, Shannon B. Harper, and Lynn Langton. 2020. Anti-asian hate crime during the covid-19 pandemic: Exploring the reproduction of inequality. *National Library of Medicine*, 45(4):647–667.

[Hamer2020] Ross Hamer. 2020. Free coronavirus news dataset.

[Haynes2021] Suyin Haynes. 2021. This isn't just a problem for north america. the atlanta shooting highlights the painful reality of rising anti-asian violence around the world. https://time.com/5947862/anti-asian-attacks-rising-worldwide/. Accessed: 2022-12-01.

[Hswen et al.2021] Yulin Hswen, Xiang Xu, Anna Hing, Jared B. Hawkins, John S. Brownstein, and Gilbert C. Gee. 2021. Association of "#covid19" versus "#chinesevirus" with anti-asian sentiments on twitter: March 9–23, 2020. *American Journal of Public Health*, 111:956–964.

[Williams and Best1977] John E. Williams and Deborah L. Best. 1977. Sex stereotypes and trait favorability on the adjective check list. *Educational and Psychological Measurement*, 37(1):101–110.

[Williams and Best1990] John E. Williams and Deborah L. Best. 1990. *Measuring sex stereotypes: A multination study, Rev*. Sage Publications, Inc.

A Appendices

A.1 News Sources

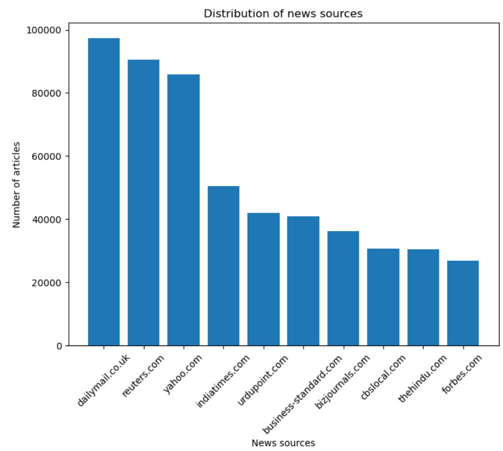

Figure A.1: Top 10 frequent news sources

A.2 Group Words

These words were collected from (Garg et al., 2018).

White Last Names: *harris, nelson, robinson, thompson, moore, wright, anderson, clark, jackson, taylor, scott, davis, allen, adams, lewis, williams, ones, wilson, martin, johnson*

Asian Last Names: *chung, liu, wong, huang, ng, hu, chu, chen, lin, liang, wang, wu, yang, tang, chang, hong, li, cho, kim, khan, shah, singh*

A.3 Neutral Words

COVID-19 Terms: *coronavirus, virus, covid, flu, bat, sick, disease, infectious, contagious*

Hate Crimes: *harassment, assault, murder, arson, vandalism, threats, hate, spitting, attack*

Outsider Adjectives from (Garg et al., 2018): *devious, bizarre, venomous, erratic, barbaric, frightening, deceitful, forceful, deceptive, envious, greedy, hateful, contemptible, brutal, monstrous, calculating, cruel, intolerant, aggressive, monstrous*

General Adjectives from (Coates, 2015), (Williams and Best, 1977), (Williams and Best, 1990): *headstrong, thankless, tactful, distrustful, quarrelsome, effeminate, fickle, talkative, dependable, resentful, sarcastic, unassuming, changeable, resourceful, persevering, forgiving, assertive, individualistic, vindictive, sophisticated, deceitful, impulsive, sociable, methodical, idealistic, thrifty, outgoing, intolerant, autocratic, conceited, inventive, dreamy, appreciative, forgetful, forceful, submissive, pessimistic, versatile, adaptable, reflective, inhibited, outspoken, quitting, unselfish, immature, painstaking, leisurely, infantile, sly, praising, cynical, irresponsible, arrogant, obliging, unkind, wary, greedy, obnoxious, irritable, discreet, frivolous,cowardly, rebellious, adventurous, enterprising, unscrupulous, poised, moody, unfriendly, optimistic, disorderly, peaceable, considerate, humorous, worrying, preoccupied, trusting, mischievous, robust, superstitious, noisy, tolerant, realistic, masculine, witty, informal, prejudiced, reckless, jolly, courageous, meek, stubborn, aloof, sentimental, complaining, unaffected, cooperative, unstable, feminine, timid, retiring, relaxed, imaginative, shrewd, conscientious, industrious, hasty, commonplace, lazy, gloomy, thoughtful, dignified, wholesome, affectionate, aggressive, awkward, energetic, tough, shy, queer, careless, restless, cautious, polished, tense, suspicious, dissatisfied, ingenious, fearful, daring, persistent, demanding, impatient, contented, selfish, rude, spontaneous, conventional, cheerful, enthusiastic, modest, ambitious, alert, defensive, mature, coarse, charming, clever, shallow, deliberate, stern, emotional, rigid, mild, cruel, artistic, hurried, sympathetic, dull, civilized, loyal, withdrawn, confident, indifferent, conservative, foolish, moderate, handsome, helpful, gentle, dominant, hostile, generous, reliable, sincere, precise, calm, healthy, attractive, progressive, confused, rational, stable, bitter, sensitive, initiative, loud, thorough, logical, intelligent, steady, formal, complicated, cool, curious, reserved, silent, honest, quick, friendly, efficient, pleasant, severe, peculiar, quiet, weak, anxious, nervous, warm, slow, dependent, wise, organized, affected, reasonable, capable, active, independent, patient, practical, serious, understanding, cold, responsible, simple, original, strong, determined, natural, kind, disorganized, devious, impressionable, circumspect, impassive, aimless, effeminate, unfathomable, fickle, unprincipled, inoffensive, reactive, providential, resentful, bizarre, impractical, sarcastic, misguided, imitative, pedantic, venomous, erratic, insecure, resourceful, neurotic, forgiving, profligate, whimsi-*

cal, assertive, incorruptible, individualistic, faithless, disconcerting, barbaric, hypnotic, vindictive, observant, dissolute, frightening, complacent, boisterous, pretentious, disobedient, tasteless, sedentary, sophisticated, regimental, mellow, deceitful, impulsive, playful, sociable, methodical, willful, idealistic, boyish, callous, pompous, unchanging, crafty, punctual, compassionate, intolerant, challenging, scornful, possessive, conceited, imprudent, dutiful, lovable, disloyal, dreamy, appreciative, forgetful, unrestrained, forceful, submissive, predatory, fanatical, illogical, tidy, aspiring, studious, adaptable, conciliatory, artful, thoughtless, deceptive, frugal, reflective, insulting, unreliable, stoic, hysterical, rustic, inhibited, outspoken, unhealthy, ascetic, skeptical, painstaking, contemplative, leisurely, sly, mannered, outrageous, lyrical, placid, cynical, irresponsible, vulnerable, arrogant, persuasive, perverse, steadfast, crisp, envious, naive, greedy, presumptuous, obnoxious, irritable, dishonest, discreet, sporting, hateful, ungrateful, frivolous, reactionary, skillful, cowardly, sordid, adventurous, dogmatic, intuitive, bland, indulgent, discontented, dominating, articulate, fanciful, discouraging, treacherous, repressed, moody, sensual, unfriendly, optimistic, clumsy, contemptible, focused, haughty, morbid, disorderly, considerate, humorous, preoccupied, airy, impersonal, cultured, trusting, respectful, scrupulous, scholarly, superstitious, tolerant, realistic, malicious, irrational, sane, colorless, masculine, witty, inert, prejudiced, fraudulent, blunt, childish, brittle, disciplined, responsive, courageous, bewildered, courteous, stubborn, aloof, sentimental, athletic, extravagant, brutal, manly, cooperative, unstable, youthful, timid, amiable, retiring, fiery, confidential, relaxed, imaginative, mystical, shrewd, conscientious, monstrous, grim, questioning, lazy, dynamic, gloomy, troublesome, abrupt, eloquent, dignified, hearty, gallant, benevolent, maternal, paternal, patriotic, aggressive, competitive, elegant, flexible, gracious, energetic, tough, contradictory, shy, careless, cautious, polished, sage, tense, caring, suspicious, sober, neat, transparent, disturbing, passionate, obedient, crazy, restrained, fearful, daring, prudent, demanding, impatient, cerebral, calculating, amusing, honorable, casual, sharing, selfish, ruined, spontaneous, admirable, conventional, cheerful, solitary, upright, stiff, enthusiastic, petty, dirty, subjective, heroic, stupid, modest, impressive, orderly, ambitious, protective, silly, alert, destructive, exciting, crude, ridiculous, subtle, mature, creative, coarse, passive, oppressed, accessible, charming, clever, decent, miserable, superficial, shallow, stern, winning, balanced, emotional, rigid, invisible, desperate, cruel, romantic, agreeable, hurried, sympathetic, solemn, systematic, vague, peaceful, humble, dull, expedient, loyal, decisive, arbitrary, earnest, confident, conservative, foolish, moderate, helpful, delicate, gentle, dedicated, hostile, generous, reliable, dramatic, precise, calm, healthy, attractive, artificial, progressive, odd, confused, rational, brilliant, intense, genuine, mistaken, driving, stable, objective, sensitive, neutral, strict, angry, profound, smooth, ignorant, thorough, logical, intelligent, extraordinary, experimental, steady, formal, faithful, curious, reserved, honest, busy, educated, liberal, friendly, efficient, sweet, surprising, mechanical, clean, critical, criminal, soft, proud, quiet, weak, anxious, solid, complex, grand, warm, slow, false, extreme, narrow, dependent, wise, organized, pure, directed, dry, obvious, popular, capable, secure, active, independent, ordinary, fixed, practical, serious, fair, understanding, constant, cold, responsible, deep, religious, private, simple, physical, original, working, strong, modern, determined, open, political, difficult, knowledge, kind

A.4 Embedding Quality Checks

The list of words used for the analogy test can be found in the GitHub repository.

A.4.1 Analogy Test

Month	Syntactic	Semantic
Jan	0.0777	0.0992
Feb	0.1288	0.1786
Mar	0.1422	0.1954
Apr	0.1623	0.1681
May	0.1566	0.1679
June	0.1447	0.1933
July	0.1518	0.1517

Table A.4.1: Syntactic and Semantic Analogy Test Results

The analogy scores do not look too promising so there is room to improve the embedding model. We also look at the nearest neighbors analysis.

A.4.2 Nearest Neighbors

To evaluate the embedding models, the top 10 words associated with “car” and “food” were extracted to ensure it was identifying semantic relationships correctly. These words are displayed in the tables below.

Month	Car
Jan	webaul, cars, jaguar, rentals, wheel, fuse, ghosttown, gibbons, skoda, halewood
Feb	cars, vehicle, dealership, selfdriving, auto, benz, dealerships, vehicles, electroc, carmakers
Mar	cars, vehicle, bicycle, accident, suv, parked, scooter, quad, privatehire, minivan
Apr	cars, vehicle, parked, motorcycle, tires, bikes, rides, bike, bicycle, motorbike
May	vehicle, cars, bicycle, bicycles, dealership, vehicles, bikes, scooter, motorbike,suv
June	dealership, cars, vehicle, truck, scooter, motorcycle, oncoming, tow, bikes, suv
July	cars, vehicle, motorcycle, scooter, suv, suvs, bicycle, tow, amtrak, parked

Table 6: Table A.4.2a: Nearest Neighbors for “Car”

We can see here that we do have words associated with vehicles or transportation. Similarly, for food, we have words associated with the category for each month. It is interesting to note that since it is during the pandemic, there are more words associated with pantries, necessities, ratios, etc. which fits the trend of people stocking up their food in case of a lockdown or shut down in their city rather than words like hamburger, hot dog, etc.

Month	Food
Jan	wellcooked, necessities, nutritious, essentials, fruits, unsanitary, ganesan, stored, contraband, perishable
Feb	bottled, essentials, nonperishable, tinned, packaged, fruits, diapers, vegetables, pastas, canned
Mar	parcels, pantries, essentials, toiletries, groceries, nonpersiahble, perishable, pantry, fruits, rations
Apr	nonperishable, parcels, packaged, grains, meals, necessities, groceries, pantries, toiletries
May	rations, pantries, toiletries, groceries, meals, packaged, necessities, grains, beverage, perishable
June	groceries, nutritious, parcels, meals, shelfstable, pantries, pantry, necessities, diapers
July	pantries, pantry, meals, meal, parcels, nutritious, groceries, nonperishable, drink, necessities

Table 7: Table A.4.2b: Nearest Neighbors for “Food”

A.5 Correlation Plots

The following plots track the Pearson correlation over the months for the embedding bias associated with each of the respective word lists.

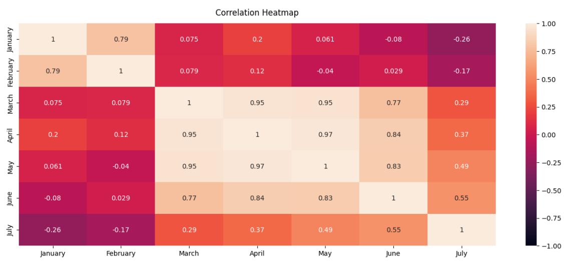

Figure A.5.1: COVID-19 embedding bias correlation between months

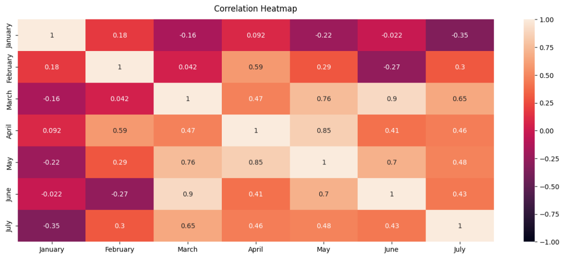

Figure A.5.2: Hate crime embedding bias correlation between months

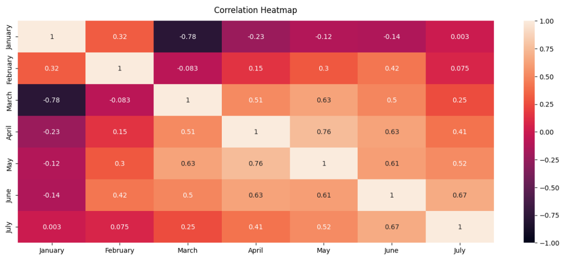

Figure A.5.3: Outsider adjective embedding bias correlation between months

Following the details from (Garg et al., 2018), we run a test to determine whether there are any significant monthly phase shifts in the perception of Asians based on our word categories. This includes taking the difference between every adjacent column in from the correlation matrix to quantify the change in association. Then, using a Kolmogorov-Smirnov two-sample test, we test whether an interval's difference is distributed differently than other month's differences.

A.6 Most Biased Last Names

The following plots show the ethnic association with COVID-19, hate crimes and outsider adjectives for the most biased last names.

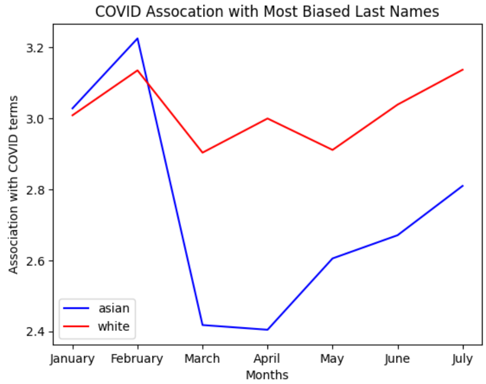

Figure A.6.1: Most biased last names for COVID-19 terms

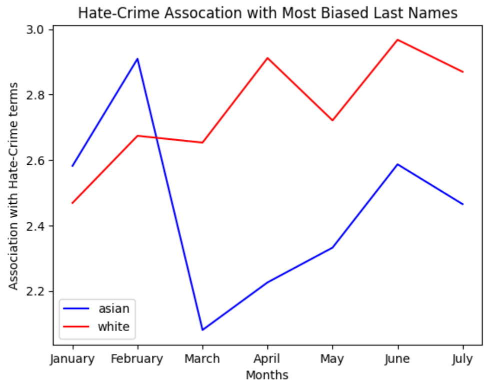

Figure A.6.2: Most biased last names for hate crimes

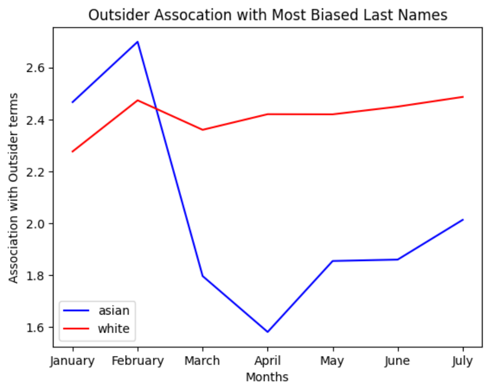

Figure A.6.3: Most biased last names for outsider adjectives

A.7 Most Associated Adjectives

Month	Asian Associated Adjectives
Jan	adventurous, witty, charming, arrogant, lazy, careless, outspoken, restless, romantic, elegant
Feb	solemn, amusing, callous, honorable, sober, circumspect, patriotic, queer, assertive, daring
Mar	assertive, methodical, witty, restrained, solemn dissatisfied, agreeable, disconcerting, circumspect, daring
Apr	reactionary, inhibited, circumspect, solemn, agreeable, disconcerting, honorable, assertive, deceptive, perverse
May	callous, honorable, circumspect, monstrous, agreeable, reactionary, disconcerting, forceful, hurried, hysterical
June	hysterical, circumspect, perverse, reactionary, cheerful, witty, dissatisfied, honorable, ingenious, hasty
July	unfathomable, ingenious, reactionary, deceptive, monstrous, hurried, solemn, cultured, hysterical, discreet

Table A.7.1: Top 10 Asian Associated Adjectives

Month	White Associated Adjectives
Jan	witty, outspoken, lazy, circumspect, arrogant, charming, sane, petty, solemn, callous
Feb	witty, queer, amusing, circumspect, shrewd, disconcerting, lazy, sober, trusting, extravagant
Mar	witty, dissatisfied, assertive, solemn, hysterical, disconcerting, admirable, ruined, troublesome, daring
Apr	honorable, solemn, fiery, unfathomable, circumspect, reactionary, admirable, agreeable, disconcerting, dissatisfied
May	honorable, monstrous, amusing, circumspect, disconcerting, callous, hurried, cheerful, unfathomable, hysterical
June	witty, cheerful, hysterical, honorable, disconcerting, daring, ingenious, solemn, adventurous, reactionary
July	honorable, unfathomable, solemn, ingenious, witty, monstrous, petty, greedy, disconcerting, discreet

Table A.7.2: Top 10 White Associated Adjectives

Month	White Associated Adjectives
Jan	protective, praising, cerebral, affected, suspicious, stable, nervous, calm, peaceful, caring
Feb	forceful, cerebral, impractical, deliberate, severe, grim, sharing, transparent, critical, practical
Mar	artificial, transparent, informal, sensitive, cultured, malicious, dependent, knowledge, religious, steadfast
Apr	predatory, restrained, thorough, cooperative, unreliable, passive, malicious, systematic, hostile, hateful
May	insulting, malicious, transparent, cooperative, praising, arbitrary, systematic, sensitive, assertive, artificial
June	assertive, deceptive, sensitive, malicious, responsible, forceful, cooperative, fraudulent, informal, insulting
July	malicious, cooperative, assertive, arbitrary, outrageous, artificial, sophisticated, unreliable, worrying, sensitive

Table A.7.3: Top 10 Most Biased Adjectives