CUNY School of Professional Studies | CUNY

M.S. in Data Science Capstone Presentations

Fall 2022

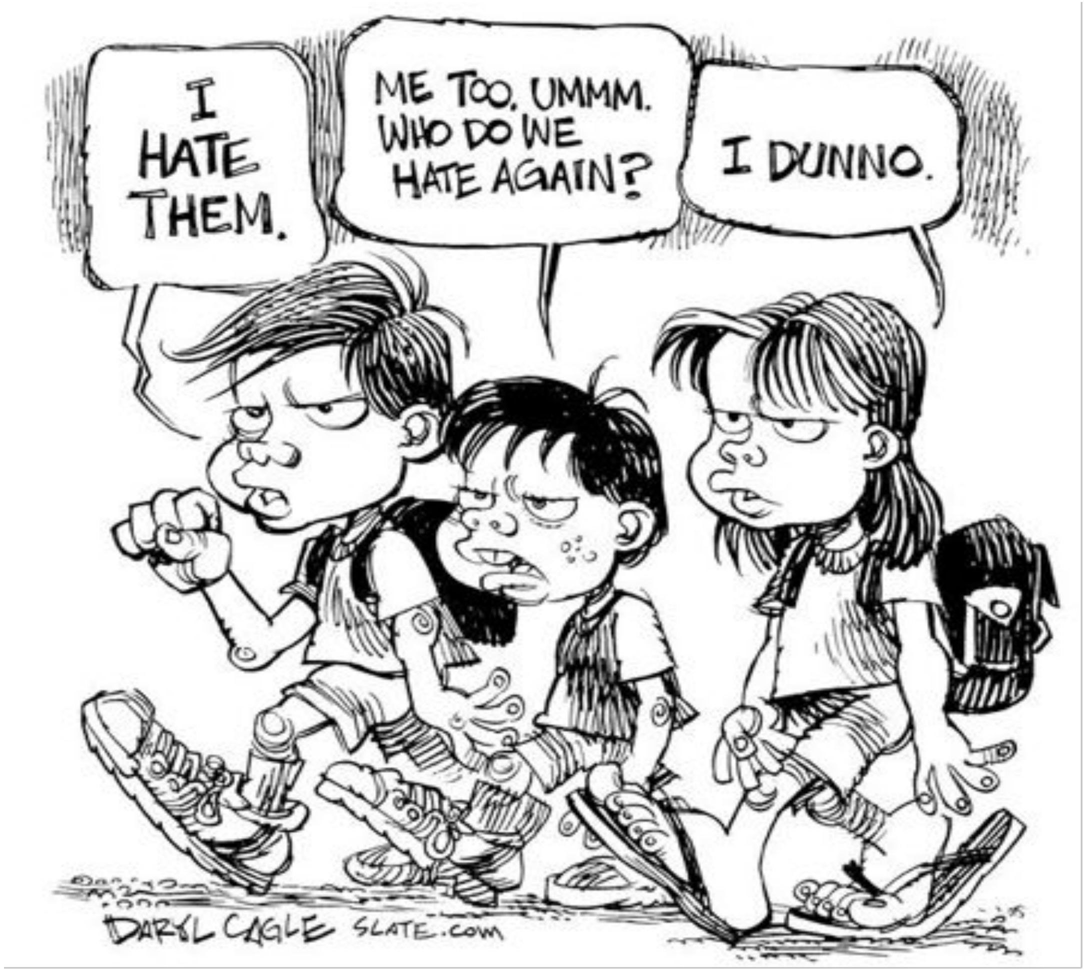
When: Wednesday, December 7, 2022 @9pm

Online Only: Join Us @ Zoom Meeting, Meeting ID: 885 6361 6684, Passcode: 829861

Presenters: Sie Siong Wong, Mario Pena, Joseph Shi

Paper: @Github Repo

Association of Hateful Tweets & Hate Crime in New York City



https://www.learningforjustice.org

Introduction

- Social media may have become a platform to spread hate
- Concerns of increase in crimes encouraged by such messages
- Finding alternative solutions that avoid censorship
- There may be a correlation between offensive tweets and crimes in a given community
- Our focus is on classifying and analyzing hateful tweets against hate crime data in the five boroughs.

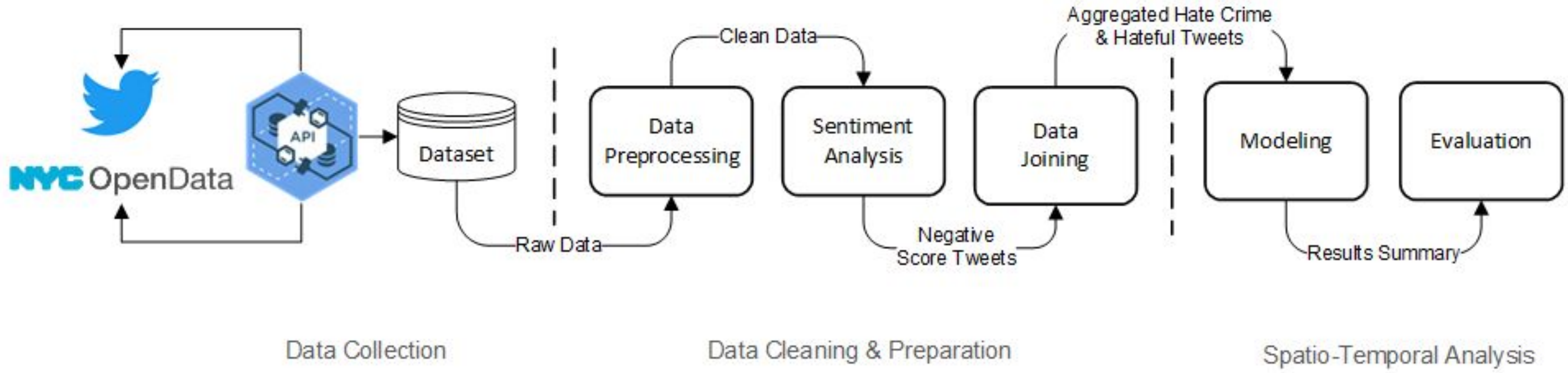
Previous Findings

- Different types of hate speech target specific groups and individuals (Miró-Llinares & Rodriguez-Sala, 2016)
- ML models are used to classify and annotate hateful, and non-hateful tweets (*Williams et al., 2019*)
- Modeling techniques have acquired mixed performances, with deep learning approaches obtaining higher results (Lee et al., 2022)
- Hate speech detection remains a challenge for AI (Matsaki L., 2018)
- Research has also found mixed results in the association of hate speech and hate crimes using spatio-temporal analysis (Curiel et al., 2020)
- Other factors, such as weather, can also be considered to predict crime, thus enhancing hate speech and hate crime modeling (Chen et al., 2015)

Challenges

- Data Collection:
  - Twitter academic research developer account
  - Potential hateful tweets collection for each borough
- Spatio-temporal analysis
  - INLA package installation
  - Bayesian modeling
  - Univariate and bivariate choropleth map

Architecture of Proposed Methodology



Scope and Limitations

- Focus on hateful tweets from NYC 5 boroughs
- Exclude video contents
- Socio-economic factors are not considered
- Only 2.5 % of collected tweets have geographic coordinates
- No guarantee that every negative attitude tweets is 100% hateful tweets

Spatio-Temporal Bayesian Modeling

$$Y_{ij} \sim Binomial(n_{ij}, p_{ij})$$

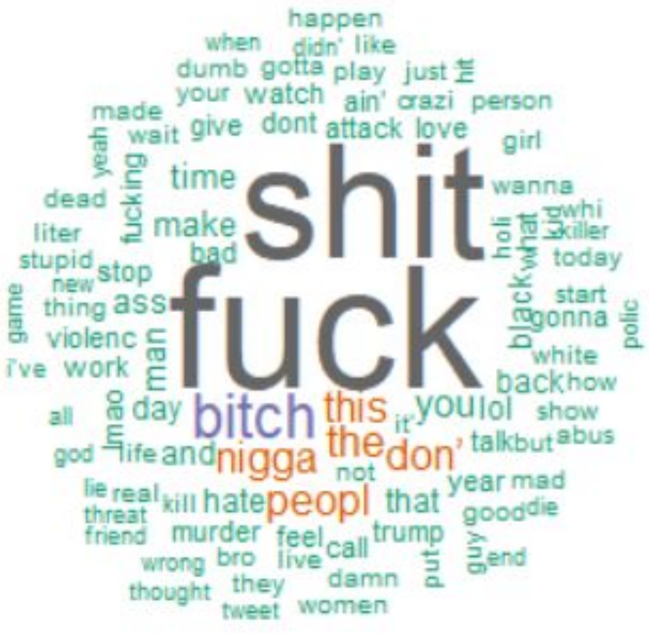
$$logit(p_{ij}) = \alpha + \beta X_{ij} + u_i + s_i + \gamma t_j + \sigma_i t_j$$

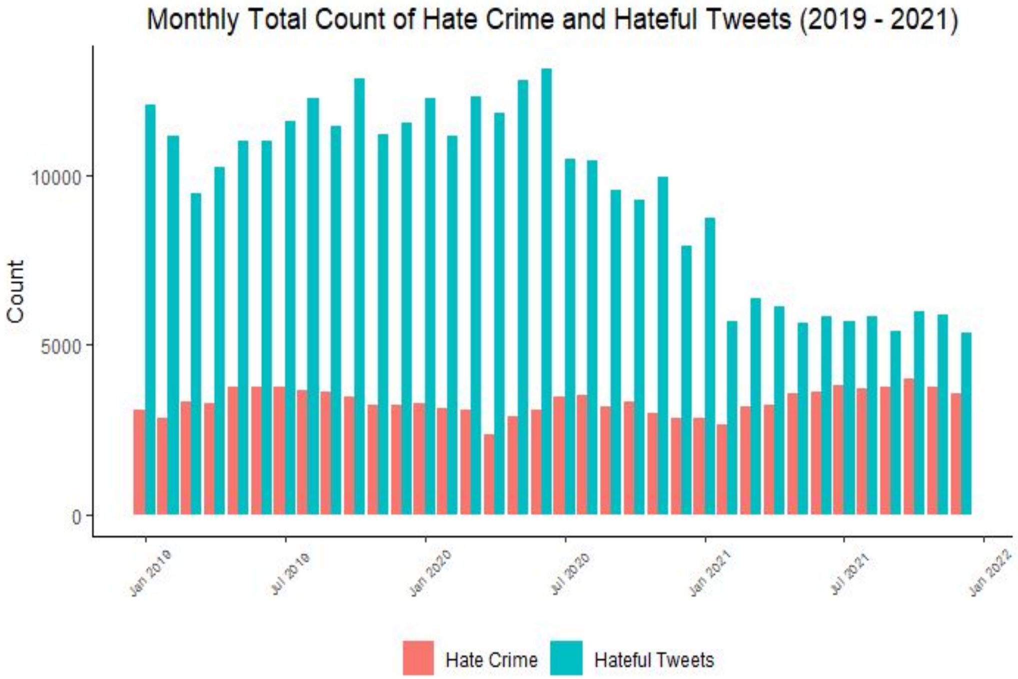
$$Y_{ij} \sim Poisson(E_{ij}\theta_{ij})$$

$$log(\theta_{ij}) = \alpha + \beta X_{ij} + u_i + s_i + \gamma t_j + \sigma_i t_j$$

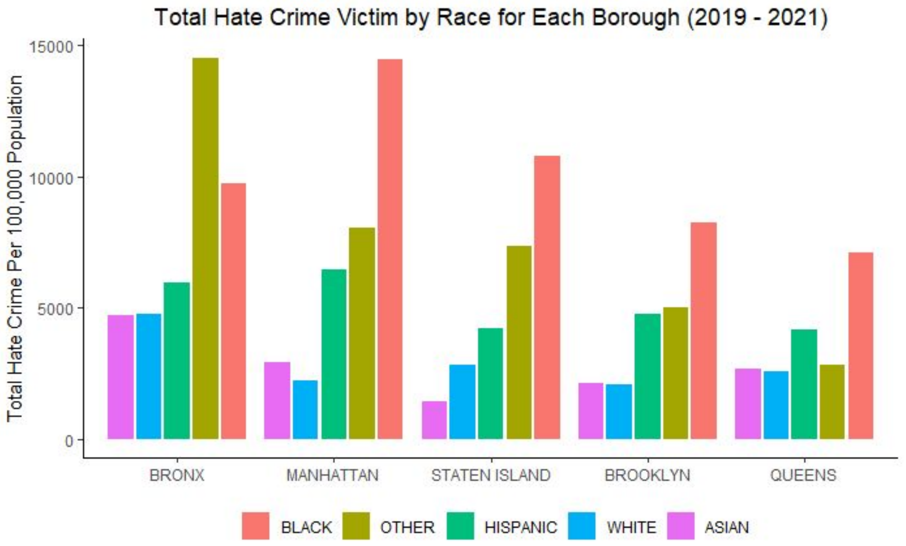
(Moraga, 2019; Hu et al., 2019)

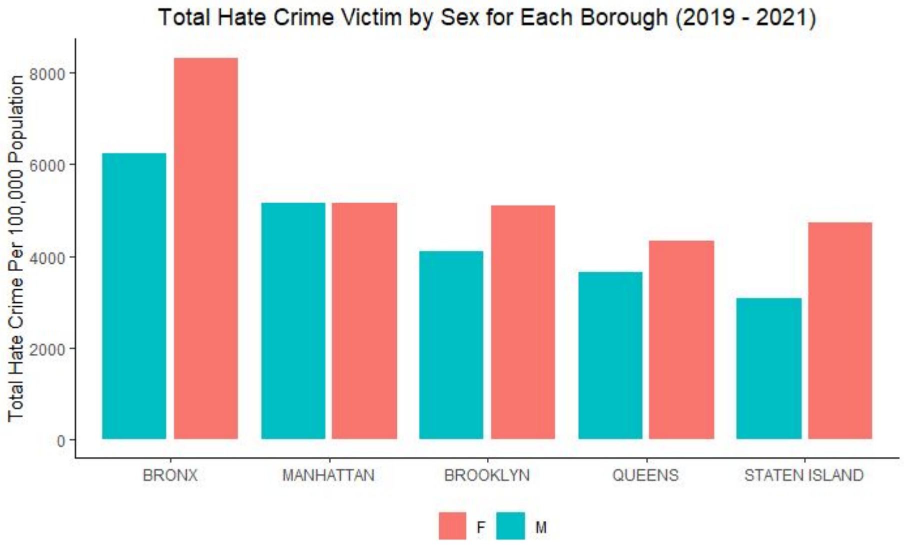
Results of Sentiment Analysis



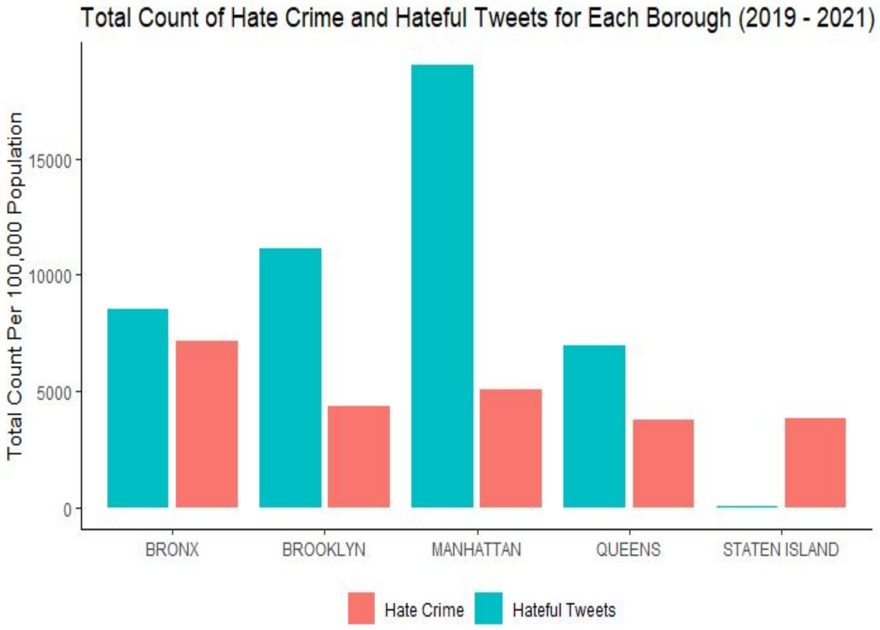

Monthly Total Count of Hate Crime and Hateful Tweets (2019 - 2021)

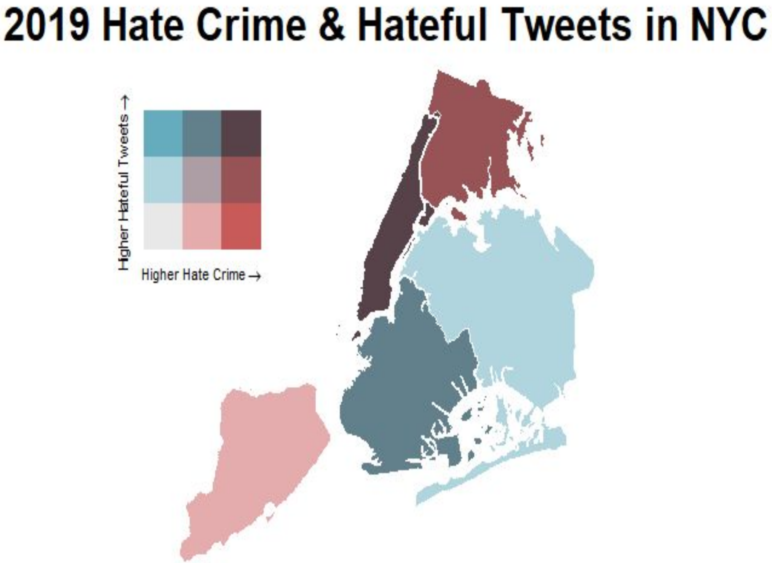
Results of Minority Group Analysis


Total Hate Crime Victim by Race for Each Borough (2019 - 2021)


Total Hate Crime Victim by Sex for Each Borough (2019 - 2021)

Results of Hate Crime vs Hateful Tweets





Results of Models Performance

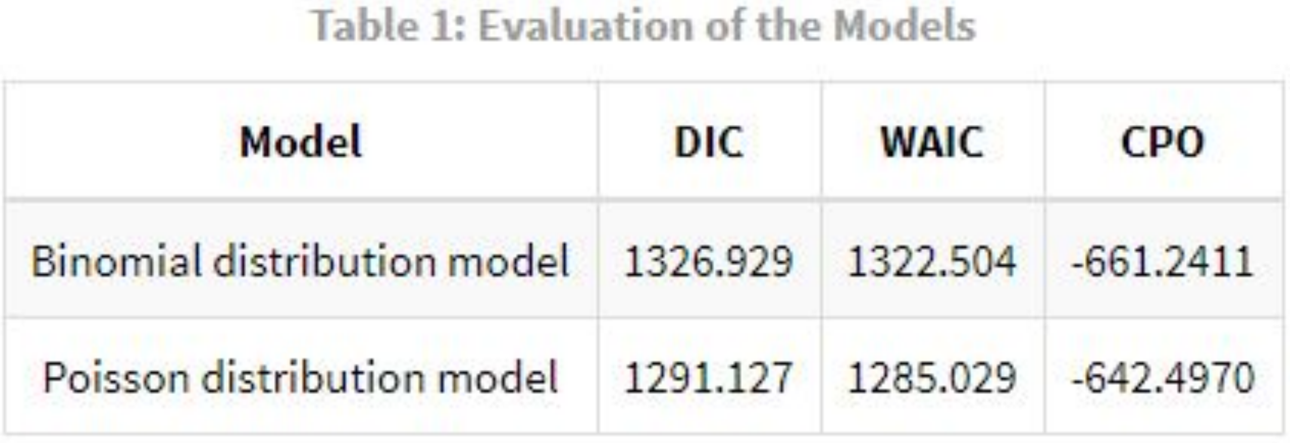


Table 1: Evaluation of the Models

Model	DIC	WAIC	CPO
Binomial distribution model	1326.929	1322.504	-661.2411
Poisson distribution model	1291.127	1285.029	-642.4970

Results of Fixed Effects Significant

Posterior Distribution

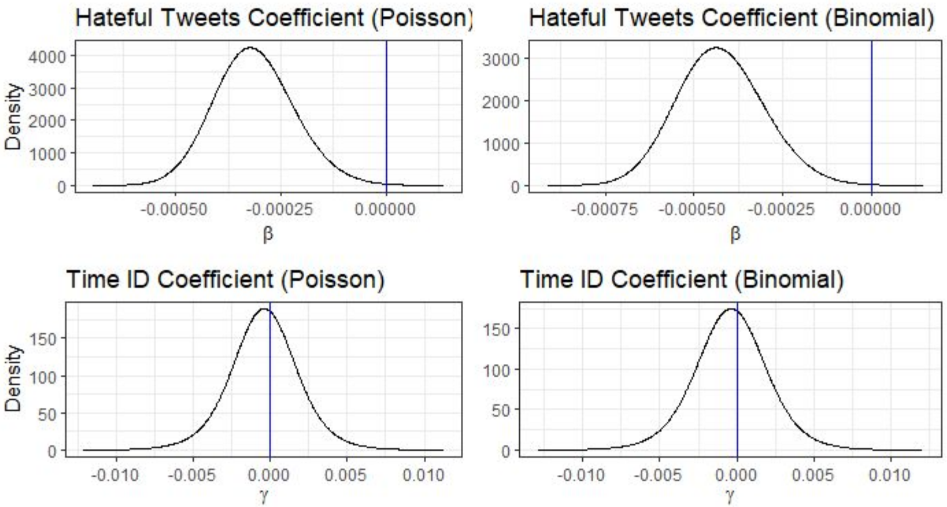
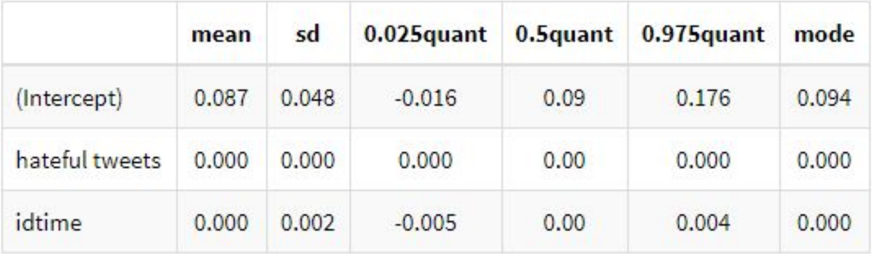


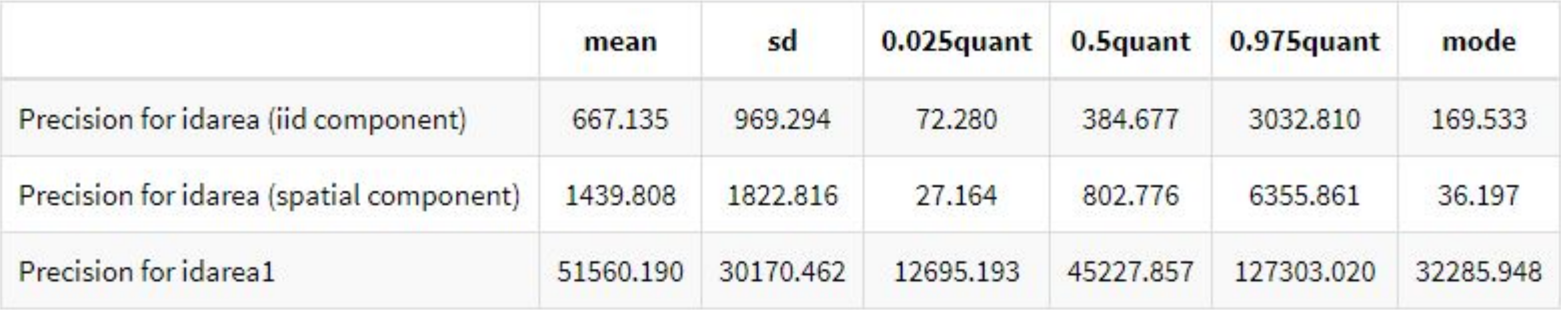
Table 2: Fixed Effects Coefficient at 95 % CI



	mean	sd	0.025quant	0.5quant	0.975quant	mode
(Intercept)	0.087	0.048	-0.016	0.09	0.176	0.094
hateful tweets	0.000	0.000	0.000	0.00	0.000	0.000
idtime	0.000	0.002	-0.005	0.00	0.004	0.000

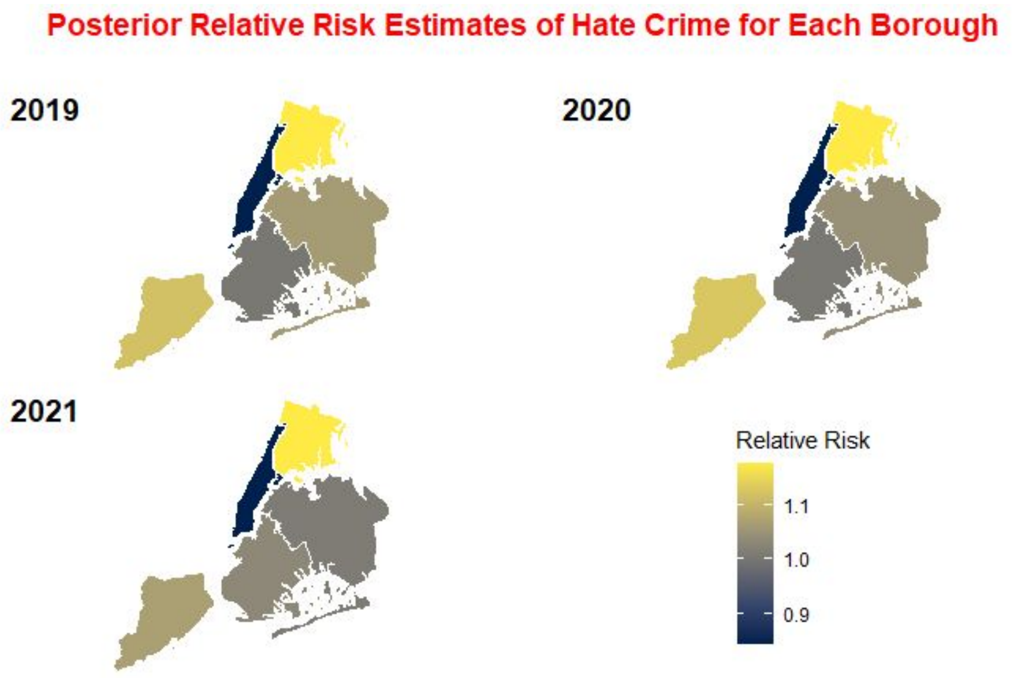
Results of Random Effects Significant

Table 3: Random Effects Coefficient at 95 % CI



	mean	sd	0.025quant	0.5quant	0.975quant	mode
Precision for idarea (iid component)	667.135	969.294	72.280	384.677	3032.810	169.533
Precision for idarea (spatial component)	1439.808	1822.816	27.164	802.776	6355.861	36.197
Precision for idarea1	51560.190	30170.462	12695.193	45227.857	127303.020	32285.948

Results of Relative Risk


Posterior Relative Risk Estimates of Hate Crime for Each Borough
2019
2020
2021
Relative Risk
1.1
1.0
0.9

Conclusion

- Hateful tweets covariate is not statistically significant correlated with hate crime.
- Black, other unidentified races and female are mostly the victim.

Future Research

- Expand the scope to the whole New York state and then do the analysis at the county level
- Incorporate topic modeling to further filter out non-hateful tweets
- Conduct the prediction of hate crime for locations

References:

- Moraga P. (2019, November 25). *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. Chapman & Hall/CRC Biostatistics Series. Retrieved from https://www.paulamoraga.com/book-geospatial/sec-arealdataexamplest.html#model-2.

- Hu et al. (2018, October 31). *Urban crime prediction based on spatio-temporal Bayesian model*. National Library of Medicine. Retrieved from https://ncbi.nlm.nih.gov/pmc/articles/PMC6209226/.

- Williams et al. (2019, July 23). *Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime*. The British Journal of Criminology. Retrieved from https://academic.oup.com/bjc/article/60/1/93/5537169.

- Matsaki L. (2018, September 26). *To Break a Hate-Speech Detection Algorithm, Try 'Love'*. Wired. Retrieved from https://www.wired.com/story/break-hate-speech-algorithm-try-love/.

References Cont..:

- Lee et al. (2022, January). *Racism Detection by Analyzing Differential Opinions Through Sentiment Analysis of Tweets Using Stacked Ensemble GCR-NN Model*. ResearchGate. Retrieved from https://www.researchgate.net/publication/357916429_Racism_Detection_by_Analyzing_Differential_Opinions_Through_Sentiment_Analysis_of_Tweets_Using_Stacked_Ensemble_GCR-NN_Model.

- Miró-Llinares F. & Rodriguez-Sala J.J. (2016, July). *Cyber Hate Speech on Twitter: Analyzing Disruptive Events from Social Media to Build a Violent Communication and Hate Speech taxonomy*. ResearchGate. Retrieved from https://www.researchgate.net/publication/308487177_Cyber_hate_speech_on_twitter_Analyzing_disruptive_events_from_social_media_to_build_a_violent_communication_and_hate_speech_taxonomy.

- Curiel et al. (2020, April 02). *Crime and its fear in social media*. Nature. Retrieved from https://www.nature.com/articles/s41599-020-0430-7#Sec8.

- Chen et al. (2015, June 8). *Crime prediction using Twitter sentiment and weather*. IEEE Xplore. Retrieved from https://ieeexplore.ieee.org/abstract/document/7117012/authors#authors.

Q
A