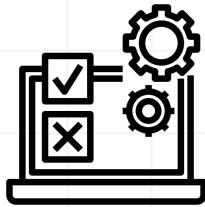


EXAMINING SHSAT SENTIMENTS

HELEN WANG

TABLE OF CONTENTS



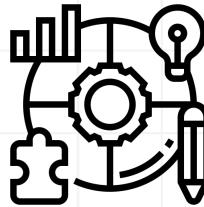
01

INTRODUCTION



02

DATA
COLLECTION



03

METHODS



04

RESULTS



05

NEXT
STEPS

01

INTRODUCTION

MOTIVATION & PROBLEM STATEMENT



SHSAT Predicts Whether Students Will Succeed in School, Study Finds

POLITICS & POLICY

August 29th, 2024

End Apartheid in Admissions to NYC's Elite High Schools

Don't Scrap the Kids Ace It

Commentary on New York's access over preparation.

Angry parents rage over SHSAT specialized

SPECIALIZED HIGH SCHOOL ADMISSIONS TEST(SHSAT)

As admissions controversy roils, data shows New York City's specialized high schools continue to accept few Black and Hispanic students

Specialized high school attack again

cial politics of entrance to high schools

The entrance exam for its elite schools has sparked a backlash to the backlash.

by Alvin Chang
Jun 14, 2018, 9:10 AM EDT



Fair and objective or useless and biased? A Chalkbeat guide to the case for and against New York City's specialized high school test

NEWS-AN

Candidates split on future of NYC's gifted and talented program

by Ariama C. Long
October 16, 2025



TIMELINE

1971
Hecht-Calandra
Act

2018
de Blasio's
proposal

2025
Mayoral
Elections

2013
Racial
discrepancies
discovered

2020
Discovery
Expansion

MAIN QUESTIONS

01

How have semantic associations with the SHSAT fluctuated over time?

02

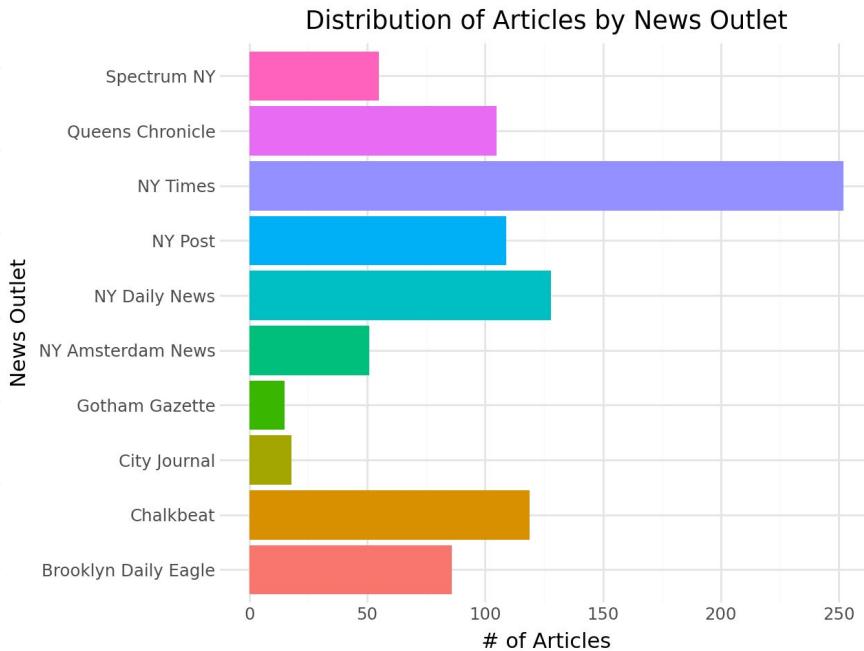
How have stances towards the SHSAT changed over time?

02

DATA COLLECTION

API, WEB-SCRAPING, AND MORE

DATA SOURCES

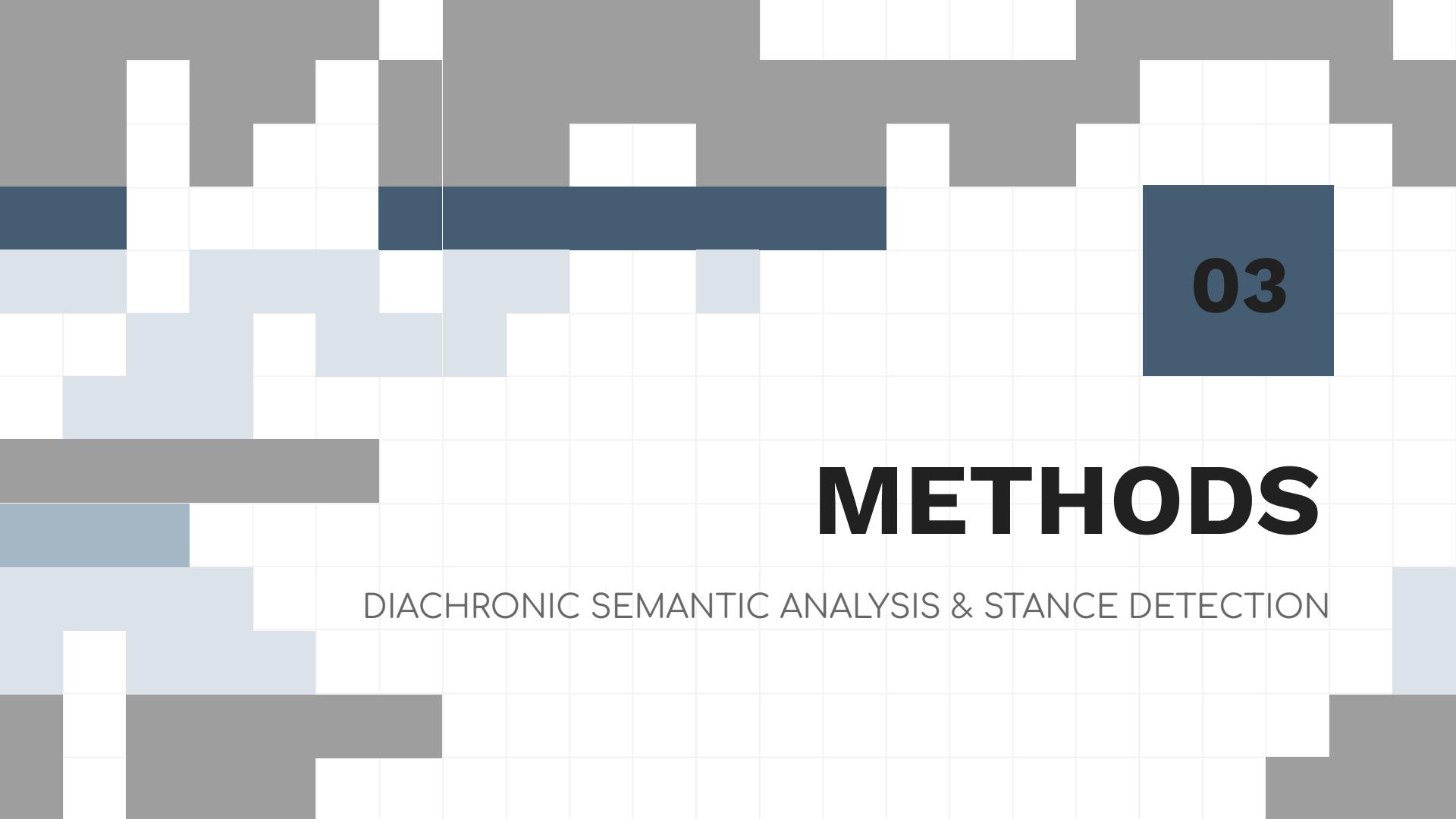


- Focused on gathering news articles from NYC based news outlets — ultimately 10
- Used the New York Times API , Google CSE API, and build web scraping pipelines for each news outlet
- Pulled headlines and abstracts/snippets
- Extracted all articles as of Nov 18, 2025



849

headlines and abstracts/snippets after preprocessing



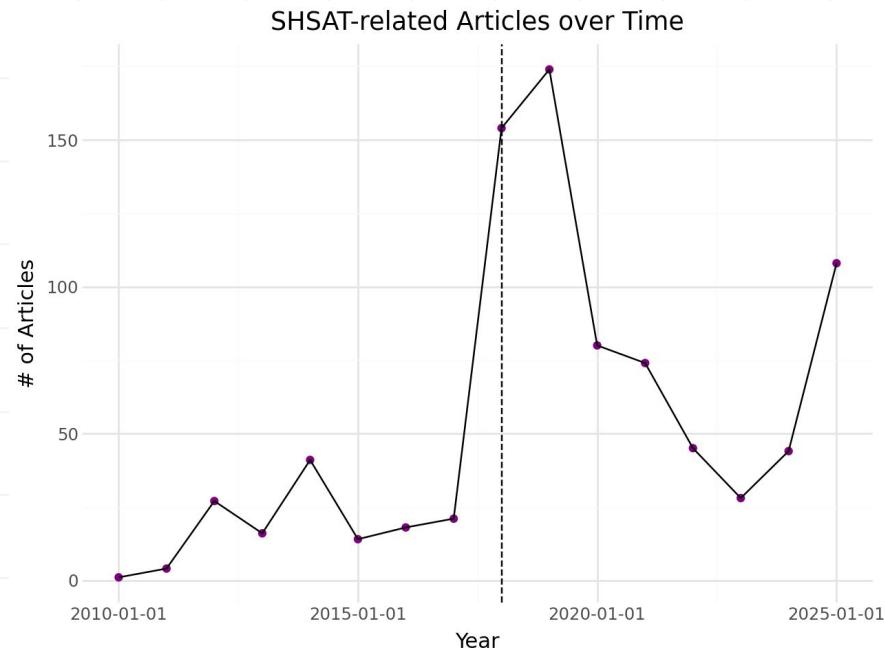
03

METHODS

DIACHRONIC SEMANTIC ANALYSIS & STANCE DETECTION

METHODOLOGICAL CHOICES

- Focused on comparing 2 time periods, pre- and post- 2018
 - Pre: 142
 - Post: 707
- Bootstrap Resampling to account for small corpus
- Analysis
 - Change in Individual words across time
 - Change in cosine similarity of words across time

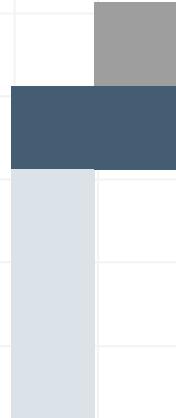


DIACHRONIC SEMANTIC ANALYSIS



ALIGNED TIME SERIES MODEL

Trains word embeddings on each time slice and aligns them post-hoc to assess semantic change



CHRONOLOGICALLY TRAINED MODEL

Initializes each time slice model using the weights of the previous $t - 1$ model, with the initial model being trained on the full corpus

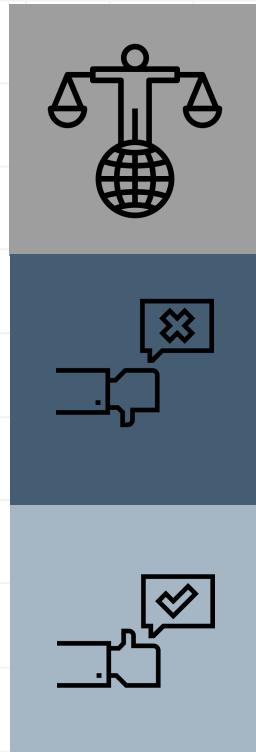


STANCE DETECTION



HUGGING FACE

rwillh11/mdeberta_NLI
_stance_NoContext



NEUTRAL

NEGATIVE

POSITIVE



SHSAT

EXAM

TEST

SPECIALIZED
HIGH SCHOOL

SPECIALIZED
HIGH SCHOOLS





04

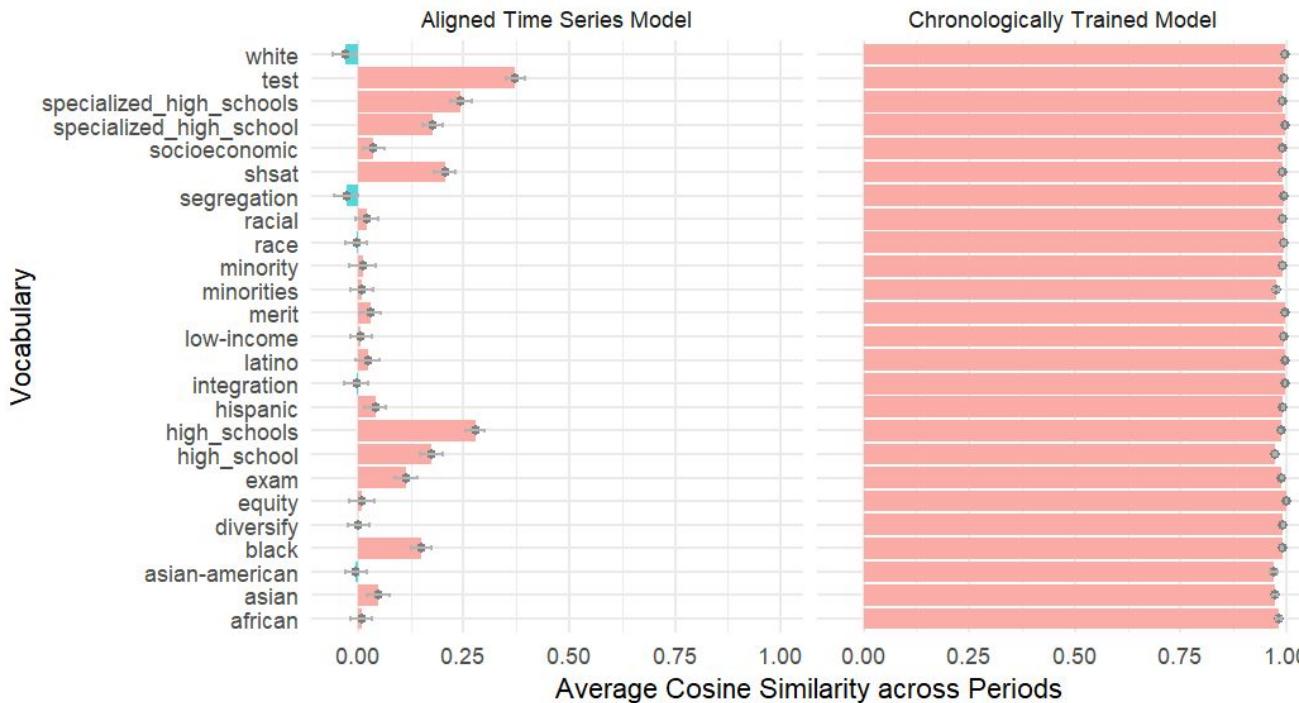
RESULTS

CHANGE OVER TIME

COMPARING MODELS

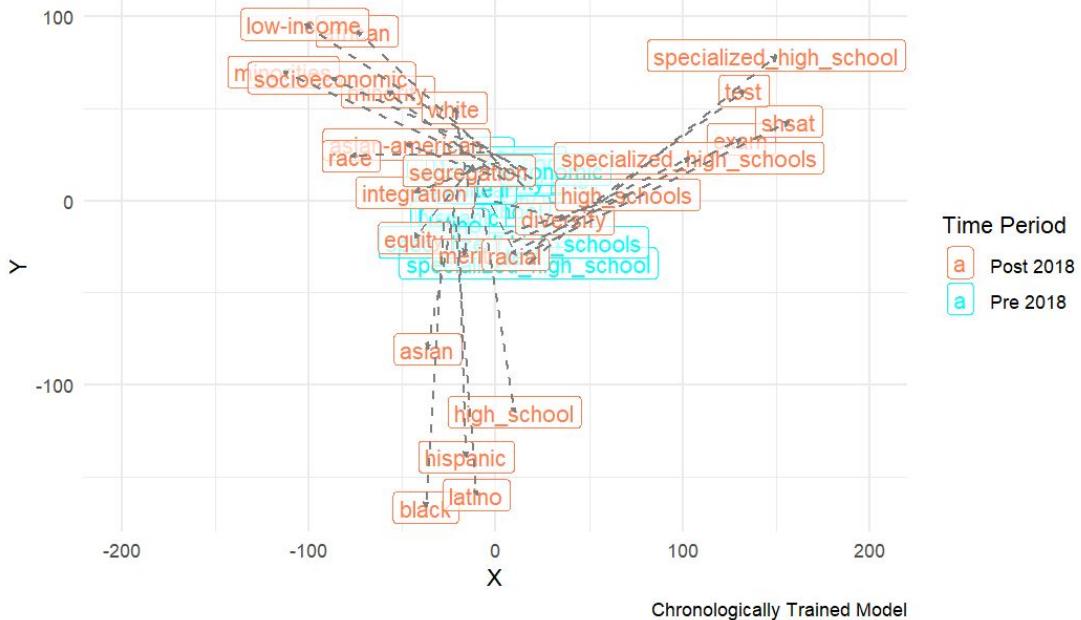
How Similar are Words Pre- and Post- 2018?

Comparing Different Diachronic Models

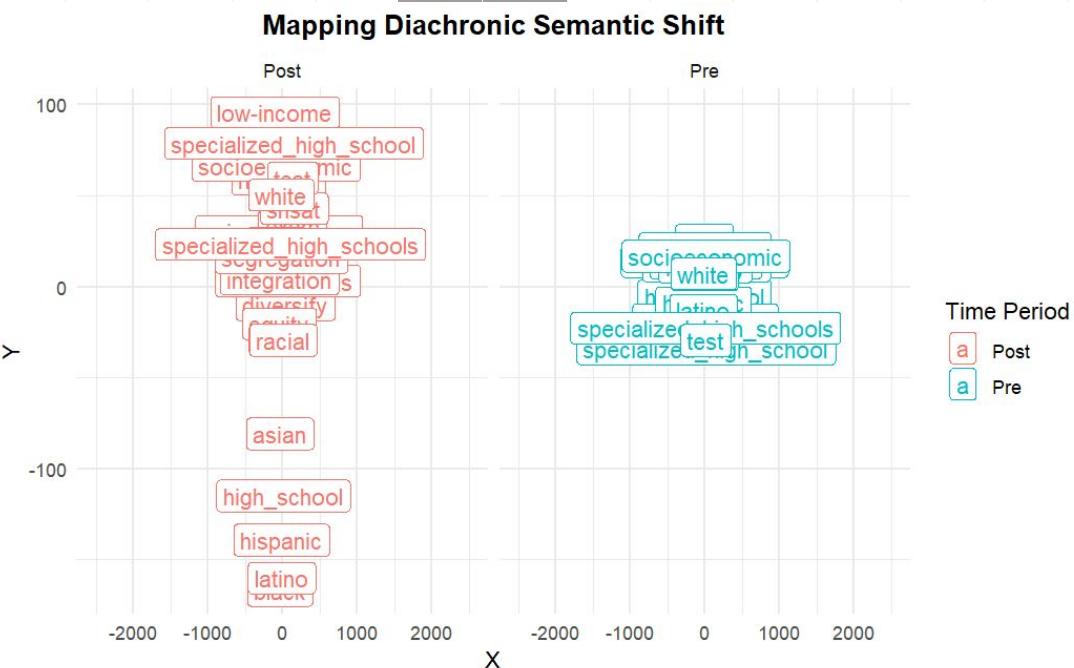


TEMPORAL SEMANTIC CHANGE

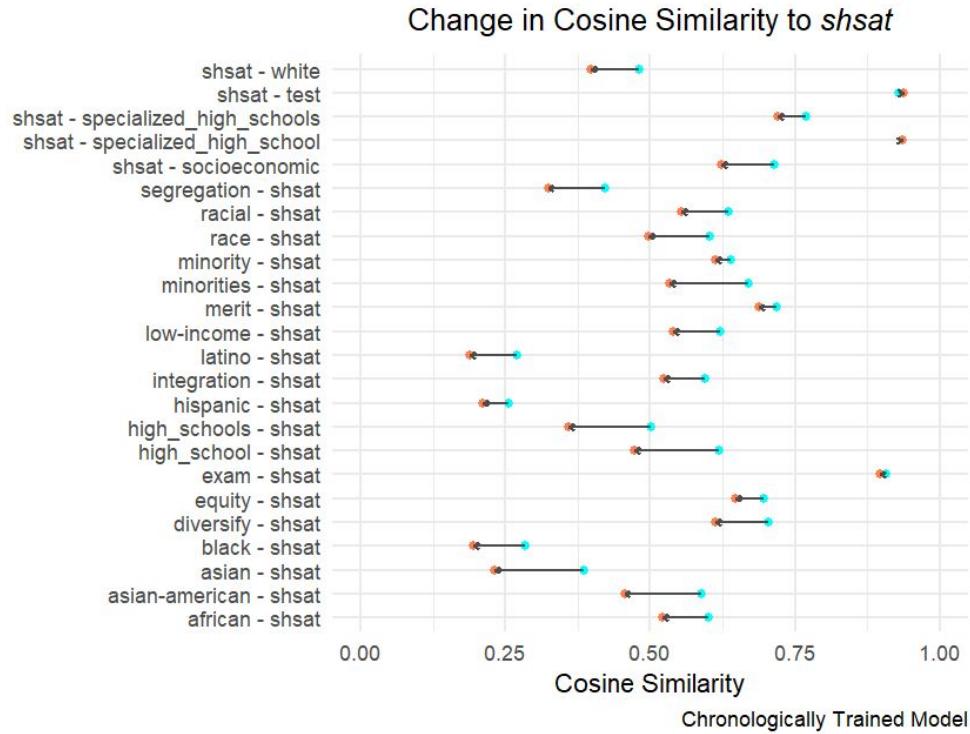
Mapping Diachronic Semantic Shift



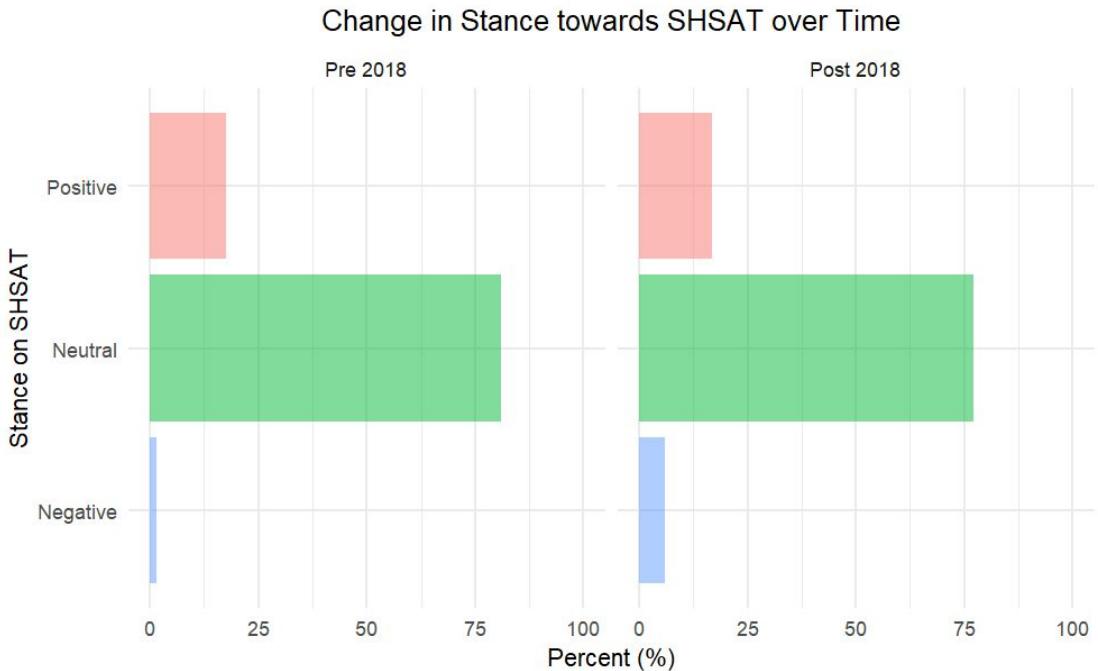
TEMPORAL SEMANTIC CHANGE



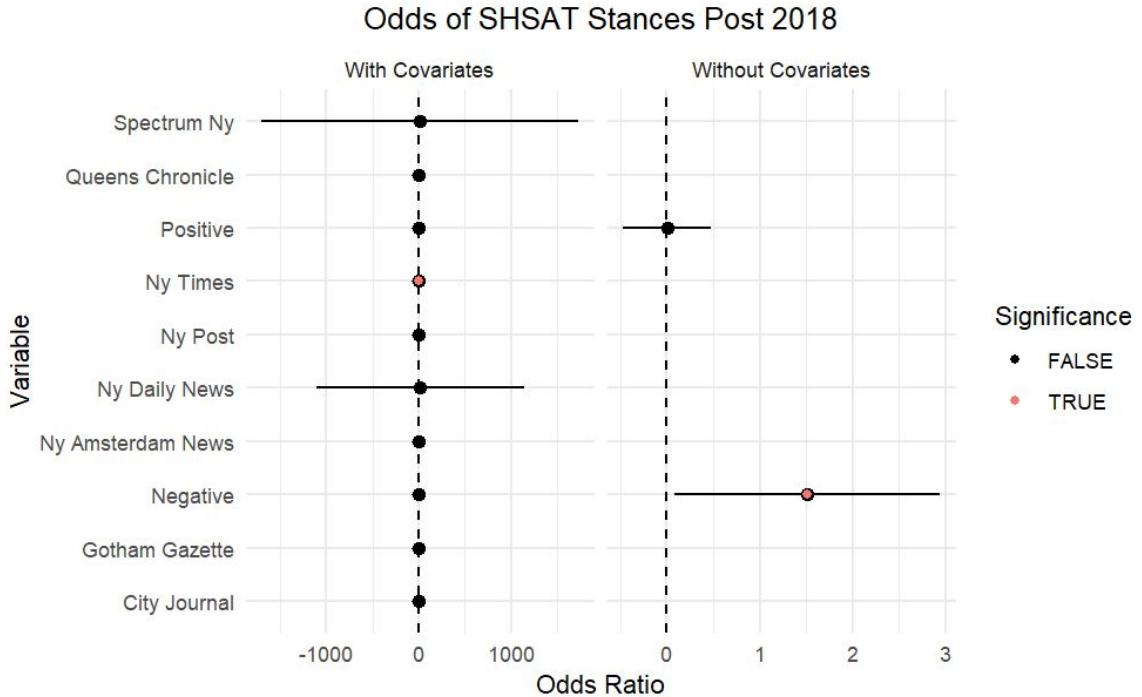
TEMPORAL SEMANTIC CHANGE



STANCE DETECTION



STANCE DETECTION



VALIDATION

Human Coders



Fleiss Kappa:
0.465 😐



Note: Validation was Conducted with a Random Sample of 200 Documents and 3 Individual Coders

Accuracy: 0.64 😢

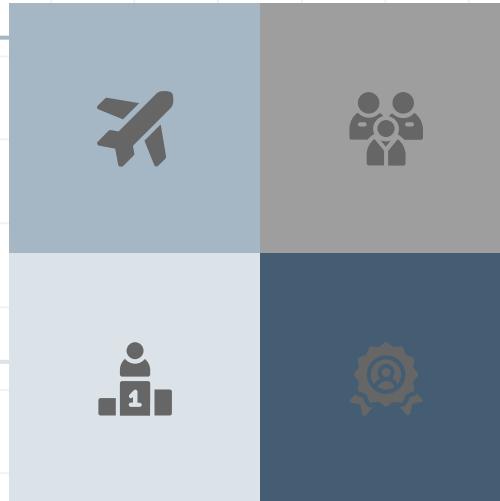
TAKEAWAY RESULTS

MODELS

Chronologically Trained Models appear to be more stable than Aligned Time Series Models

SEMANTIC

Words appear to become less associated over time — greater separation



STANCE

Slight increase in anti-SHSAT headlines but effect disappears if you control for source

ACCURACY

Performance of model is generally poor



05

NEXT STEPS

FUTURE DIRECTIONS, LIMITATIONS, AND LESSONS LEARNED

LIMITATIONS & THE FUTURE



SMALL CORPUS

Bootstrapping
stabilized results but....



SCRAPING

Greater range of
outlets, years, etc.

APPROACHES

Explored 2 modeling
approaches but...



ROBUSTNESS

Explore Temporal Word
Embeddings with a
Compass (TWEC)



LIMITATIONS & THE FUTURE



TRANSFORMER

Trained on political
corpus



COMPARISON

Other off-the-shelf
transformers and/or
dictionary methods

TARGET

May not have captured
all words + maximum
rather than averaging



SYSTEMIC

More systemic
approach to deciding
target words

THANKS!

ANY QUESTIONS?

SLIDEDECK FROM SLIDESGO