## **Final Presentation**

Seattle Needles Friday, December 4th

## Presentation Roadmap

- Introduction to Problem
- Approach
- Results
- Conclusions
- Limitations and Further Steps

## Presentation Roadmap

- Introduction to Problem
- Approach
- Results
- Conclusions
- Limitations and Further Steps

To what extent can social media predict the 2020 presidential election?

#### Our Motivation

Why do we care about answering this question?

- Learn how democracy works in practice
- Results could affect how politicians conduct campaigns
- This is a long-standing problem in social media analysis
- We hope to combine state-of-the-art methods we read about in the literature
  - Demographic analyses
  - Electoral college based prediction
  - Sentiment analysis

## Presentation Roadmap

- Introduction to Problem
- Approach
- Results
- Conclusions
- Limitations and Further Steps

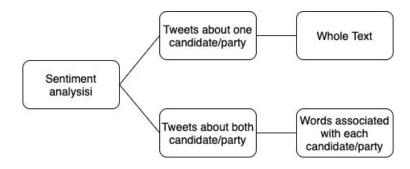
#### Twitter Data Collection

- 2016 Twitter data collection:
- Queries: "trump" and "clinton" and "hillary"
- Window: August 6, 2016 to November 8, 2016
- Locations: All cities & states in the USA
- 67867 Tweets
- 40% Neutral, 29% Democrat, 26% Republican, 5% Unknown
- Breakdowns by race/ethnicity, gender, race/ethnicity x gender shown better on website

- 2020 Twitter data collection:
- Queries: "trump" and "biden"
- Window: August 1, 2020 to November 3, 2020
- Locations: All cities & states in the USA
- 247,837 Tweets
  - 28% Neutral, 36% Democrat, 34% Republican, 2% Unknown
- Breakdowns by race/ethnicity, gender, race/ethnicity x gender shown better on website

## Sentiment Analysis

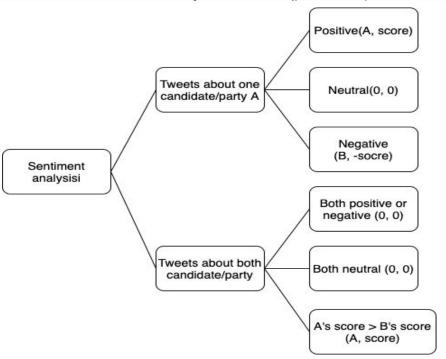
- Political Analysis :
  - a. The tweet targets at least one candidate or party; (Trump, #Trump)
  - b. The tweet mentions at least one candidate or party; (@realDonaldTrump)
  - c. The tweet has a candidate's proper name or party; (DT)
- Pre Sentiment Analysis:



(Use nlp parser to extract the relevant words)

### Sentiment Analysis

3. User Sentiment Analysis: <u>return</u> (political preference, preference score)



#### E.g. Donald Trump is the worst president of USA, but Biden is better than him -> (dem, 0.89) Extracted words: Donald Trump is the worst president, Biden is better Donald Trump is the worst president of USA, and Biden is also bad. -> (0, 0)Extracted words: Donald Trump is the worst president,

Biden is also bad

## **User Profiling**

- After reviewing the literature, we decided to profile users by gender and race/ethnicity
  - Male or female
    - First name US baby names from 1880 2010
  - White, Black, Asian/Pacific Islander, American Indian/Alaska Native (all non-Hispanic), Hispanic
    - Last name 2010 US Census data

#### Bot Detection

- Botometer API (limited access)
- Run on 434 user accounts (responsible for 1000 tweets in our dataset)
- Only 11 of them are >80% likely to be bots
- Only 13 of them are >50% likely to be bots
- Meaning:
  - Less than 3% users are likely to be bots
  - Not likely to make an impact on our prediction result

### Approach to Answer our Question

- Collect Twitter data
  - a. 2016 and 2020
- Label users with gender, race/ethnicity
  - a. 2016 and 2020
- Label tweets with party preference
  - a. 2016 and 2020
- 4. Examine state-specific predictions without any corrections
  - a. 2016 and 2020
- 5. If predictions are off, use various corrections
  - a. Correct 2020 users based on 2016 data
- 6. Determine best method for correcting prediction
  - a. Is this feasible/reproducible?
  - b. Which groups are we accurately predicting?

## Presentation Roadmap

- Introduction to Problem
- Approach
- Results
- Conclusions
- Limitations and Further Steps

## Correction Approaches

We will discuss two correction approaches

- Correcting with under/over sampling of key demographic groups (poor performance)
- Correcting with political bias (decent performance)

## Correcting Under/Over Sampling of Key Demographic Groups

- As you will see, our uncorrected prediction was not very good
- One possible explanation for a poor prediction is that certain demographic groups are not fairly represented in the Twitter sample
- We used national 2016 demographic data to weight each demographic group according to who voted in 2016

# Removing Neutral Designation

- We used "neutral" for a user when a majority of their tweets had neutral or mixed sentiment
- 2016 tweets had a large amount of neutral users
- We removed neutral and instead labeled states with Democrat or Republican according to the next highest count

## 2016 Results

state	uncorrec	cted_ gender_	_corre race_co	rrect race_ge	nder true_vote	26 Montana	R	R	D	D	R	
0 Alabama	D	D	D	D	R	27 Nebraska	D	D	D	D	R	
1 Alaska	R	R	D	D	R	28 Nevada	R	D	R	R	D	
2 Arizona	D	D	R	R	R	29 New Hampshire	D	D	D	D	D	
3 Arkansas	D	D	D	D	R	30 New Jersey	D	D	D	D	D	
4 California	D	D	D	R	D	31 New Mexico	D	D	D	D	D	
5 Colorado	D	D	D	D	D	32 New York	D	D	D	D	D	
6 Connecticut	D	D	D	R	D	33 North Carolina	R	R	R	R	R	
7 District of Columbia	D	D	D	D	D	34 North Dakota	D	D	D	D	R	
8 Delaware	D	D	R	D	D	35 Ohio	D	D	D	D	R	
9 Florida	R	D	D	D	R	36 Oklahoma	D	D	D	D	R	
10 Georgia	D	D	D	D	R	37 Oregon	R	R	R	R	D	
11 Hawaii	R	R	D	R	D	38 Pennsylvania	D	D	R	R	R	
12 Idaho	D	D	D	D	R	39 Rhode Island	D	D	D	D	D	
13 Illinois	D	D	D	D	D	40 South Carolina	D	D	D	D	R	
14 Indiana	D	D	D	R	R	41 South Dakota	R	D	D	D	K	
15 lowa	D	D	D	D	R	42 Tennessee 43 Texas	D	D	D	D	R	
16 Kansas	D	D	D	D	R	44 Utah	R	D	R D	R	R	
17 Kentucky	D	D	D	D	R	45 Vermont	D	D	D	D D	R	
18 Louisiana	D	D	D	D	R	46 Virginia	D	D	D	D	D	
19 Maine	D	D	D	D	D	47 Washington	R	D	D	D	D	
20 Maryland	D	D	R	R	D	48 West Virginia	R	R	R	R	B	
21 Massachusetts	D	D	D	D	D	49 Wisconsin	D	D	D	D	R	
22 Michigan	D	D	D	D	R	50 Wyoming	D	D	D	D	R	
23 Minnesota	D	D	D	D	D	Count Correct		23	23	23	23	50
24 Mississippi	D	D	R	R	R	Rep Electoral Votes		138	37	121	208	304
25 Missouri	D	D	D	D	R	National Prediction		D	D	D	R	
		_				National Flediction	D	U	D	D	IX	

## 2020 Results

state	uncorrected	_gender_	corre race_cor	rect race_g	ender true_vote	26 Montana	D	R	D	D	R	
0 Alabama	R	R	R	R	R	27 Nebraska	R	D	D	D	R	
1 Alaska	D	D	R	R	R	28 Nevada	R	R	R	D	D	
2 Arizona	D	D	R	D	D	29 New Hampshire	R	R	R	R	D	
3 Arkansas	D	D	D	D	R	30 New Jersey	D	D	D	D	D	
4 California	D	D	D	D	D	31 New Mexico	D	D	D	R	D	
5 Colorado	D	D	R	D	D	32 New York	D	D	D	D	D	
6 Connecticut	D	D	R	D	D	33 North Carolina	D	R	R	R	R	
7 District of Columbia	D	D	D	D	D	34 North Dakota	D	D	D	D	R	
8 Delaware	D	D	D	D	D	35 Ohio	D	D	D	R	R	
9 Florida	D	D	D	D	R	36 Oklahoma	D	D	D	D	R	
10 Georgia	D	D	D	D	D	37 Oregon	D	D	D	D	D	
11 Hawaii	R	D	D	R	D	38 Pennsylvania	D	D	D	D	D	
12 Idaho	D	R	R	R	R	39 Rhode Island	D	D	D	D	D	
13 Illinois	D	D	 D	 D	 D	40 South Carolina	D	D	R	D	R	
14 Indiana	D	D	D	D	R	41 South Dakota	D	D	D	D	R	
15 Iowa	D	D	D	D	R	42 Tennessee 43 Texas	R	R D	K	K	K	
16 Kansas	D	D	D	D	R	43 Texas 44 Utah	D	D	D	R	K	
17 Kentucky	D	D	R	R	R	45 Vermont	D	D	D	D	K	
18 Louisiana	D	D	D	D	R	46 Virginia	D	D	D	D	D	
19 Maine	D	D	D	D	D	47 Washington	D	D	D	D	D	
20 Maryland	D	D	D	D	D	48 West Virginia	D	D	D	D	D	
21 Massachusetts	D	D	D	D	D	49 Wisconsin	D	D	P	P	D	
	D	D	D	D	D	50 Wyoming	D	D	D	R	R	
22 Michigan	D		D			Count Correct	D	27	30	27	33	50
23 Minnesota	D	D	R	D	D	Dem Electoral Votes		494	481	417	395	306
24 Mississippi	D	D	D	R	K	National Prediction	D	D D	401 D	D D	D D	300
25 Missouri	D	D	D	D	R	rational Frediction	0	U	D	D	D	

# Swing States vs. Non-swing States

Do our 2020 predictions differ in accuracy for each state type?

- Swing states: Wisconsin, Pennsylvania,
   Arizona, Georgia, Florida, Michigan, North
   Carolina, Ohio, Indiana, Iowa
  - Did not support only one party in the last 4 elections (2008-2020)
- Red states: Alabama, Alaska, Arkansas, Idaho, Kansas, Kentucky, Louisiana, Mississippi, Missouri, Montana, Nebraska, North Dakota, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, Utah, West Virginia, Wyoming
- Blue States: California, Colorado, Connecticut, Delaware, DC, Hawaii, Illinois, Maine, Maryland, Massachusetts, Minnesota, Nevada, New Hampshire, New Jersey, New Mexico, New York, Oregon, Rhode Island, Vermont, Virginia, Washington

### 2020 Results - Red States

	state	uncorrected_	gender_corr	race_correct	race_gender	true_vote	electoral votes
0	Alabama	R	R	R	R	R	9
1	Alaska	D	D	R	R	R	3
3	Arkansas	D	D	D	D	R	6
12	Idaho	D	R	R	R	R	4
16	Kansas	D	D	D	D	R	6
17	Kentucky	D	D	R	R	R	8
18	Louisiana	D	D	D	D	R	8
24	Mississippi	D	D	D	R	R	6
25	Missouri	D	D	D	D	R	10
26	Montana	D	R	D	D	R	3
27	Nebraska	R	D	D	D	R	5
34	North Dakota	D	D	D	D	R	3
36	Oklahoma	D	D	D	D	R	7
40	South Carolin	D	D	R	D	R	9
41	South Dakota	D	D	D	D	R	3
42	Tennessee	R	R	R	R	R	11
43	Texas	D	D	D	R	R	38
44	Utah	D	D	D	D	R	6
48	West Virginia	R	R	R	R	R	5
50	Wyoming	D	D	D	R	R	3
	Count Correc	4	5	7	9	20	

45% of red states correctly predicted

#### 2020 Results - Blue States

4 California       D <t< th=""><th>oral votes</th></t<>	oral votes
6 Connecticut D D D R D D 7 District of Co D D D D D D 8 Delaware D D D D D D 11 Hawaii R D D D D D 13 Illinois D D D D D D 19 Maine D D D D D D 20 Maryland D D D D D D 21 Massachusei D D D D D D 23 Minnesota D D R D D D 28 Nevada R R R R D D 29 New Hampsl R R R R R D 30 New Jersey D D D D D D D 31 New Mexico D D D D D D 32 New York D D D D D D 34 Rhode Island D D D D D D 35 Rhode Island D D D D D D 46 Virginia D D D D D D	55
7 District of Co D       D       D       D       D         8 Delaware       D       D       D       D       D         11 Hawaii       R       D       D       D       D       D         13 Illinois       D	9
8 Delaware       D       D       D       D       D         11 Hawaii       R       D       D       D       D       D         13 Illinois       D       D       D       D       D       D       D         19 Maine       D </td <td>7</td>	7
11 Hawaii       R       D       D       R       D         13 Illinois       D       D       D       D       D         19 Maine       D       D       D       D       D         20 Maryland       D       D       D       D       D         21 Massachusel D       D       D       D       D       D         23 Minnesota       D       D       R       D       D         28 Nevada       R       R       R       R       D       D         29 New Hampsl       R       R       R       R       R       D       D         30 New Jersey       D       D       D       D       D       D       D         31 New Mexico       D       D       D       D       D       D       D         32 New York       D       D       D       D       D       D       D         33 Rhode Island       D       D       D       D       D       D       D         45 Vermont       D       D       D       D       D       D       D	3
13 Illinois       D       D       D       D       D         19 Maine       D       D       D       D       D         20 Maryland       D       D       D       D       D         21 Massachuset D       D       D       D       D       D         23 Minnesota       D       D       D       D       D         28 Nevada       R       R       R       R       D       D         29 New Hampsl R       R       R       R       R       D       D         30 New Jersey       D       D       D       D       D       D         31 New Mexico       D       D       D       D       D       D         32 New York       D       D       D       D       D       D         37 Oregon       D       D       D       D       D       D         45 Vermont       D       D       D       D       D       D         46 Virginia       D       D       D       D       D       D       D	3
19 Maine	4
20 Maryland       D       D       D       D       D         21 Massachuset       D       D       D       D       D         23 Minnesota       D       D       R       D       D         28 Nevada       R       R       R       R       D       D         29 New Hampsl       R       R       R       R       R       D       D         30 New Jersey       D       D       D       D       D       D       D         31 New Mexico       D       D       D       D       D       D       D         32 New York       D       D       D       D       D       D       D         37 Oregon       D       D       D       D       D       D       D         45 Vermont       D       D       D       D       D       D       D         46 Virginia       D	20
21 Massachuset D       D       D       D       D         23 Minnesota       D       D       R       D       D         28 Nevada       R       R       R       R       D       D         29 New Hampsl R       R       R       R       R       D       D         30 New Jersey       D       D       D       D       D       D         31 New Mexico       D       D       D       D       D       D       D         32 New York       D       D       D       D       D       D       D         37 Oregon       D       D       D       D       D       D       D         45 Vermont       D       D       D       D       D       D       D         46 Virginia       D       D       D       D       D       D       D       D	4
23 Minnesota       D       D       R       D       D         28 Nevada       R       R       R       R       D       D         29 New Hampsl R       R       R       R       R       D         30 New Jersey       D       D       D       D       D         31 New Mexico       D       D       D       R       D         32 New York       D       D       D       D       D         37 Oregon       D       D       D       D       D         39 Rhode Island       D       D       D       D       D         45 Vermont       D       D       D       D       D         46 Virginia       D       D       D       D       D	10
28 Nevada       R       R       R       R       D       D         29 New Hampsl R       R       R       R       R       D       D         30 New Jersey       D       D       D       D       D       D         31 New Mexico       D       D       D       D       D       D         32 New York       D       D       D       D       D       D         37 Oregon       D       D       D       D       D       D         39 Rhode Island       D       D       D       D       D       D         45 Vermont       D       D       D       D       D       D         46 Virginia       D       D       D       D       D       D       D	11
29 New Hampsl R       R       R       R       R       D         30 New Jersey       D       D       D       D       D         31 New Mexico       D       D       D       R       D         32 New York       D       D       D       D       D         37 Oregon       D       D       D       D       D         39 Rhode Island       D       D       D       D       D         45 Vermont       D       D       D       D       D         46 Virginia       D       D       D       D       D	10
30 New Jersey D D D D D D 31 New Mexico D D D D R 32 New York D D D D D D 37 Oregon D D D D D D 39 Rhode Island D D D D D D 45 Vermont D D D D D D 46 Virginia D D D D D	6
31 New Mexico       D       D       D       R       D         32 New York       D       D       D       D       D         37 Oregon       D       D       D       D       D         39 Rhode Island       D       D       D       D       D         45 Vermont       D       D       D       D       D         46 Virginia       D       D       D       D       D	4
32 New York       D       D       D       D       D         37 Oregon       D       D       D       D       D         39 Rhode Island D       D       D       D       D         45 Vermont       D       D       D       D         46 Virginia       D       D       D       D	14
37 Oregon       D       D       D       D         39 Rhode Island D       D       D       D       D         45 Vermont       D       D       D       D         46 Virginia       D       D       D       D	5
39 Rhode Island D       D       D       D       D         45 Vermont       D       D       D       D         46 Virginia       D       D       D       D	29
45 Vermont D D D D D D 46 Virginia D D D D D	7
46 Virginia D D D D	4
	3
47 144 1: 1 0	13
47 Washington D D D D	12
Count Correc 18 19 16 18 21	

90.4% of blue states correctly predicted

## 2020 Results - Swing States

state	uncorrected_	gender_corr	re race_correc	t race_gend	er true_vote	e e	electoral votes
2 Arizona	D	D	R	D	D		11
9 Florida	D	D	D	D	R		29
10 Georgia	D	D	D	D	D		16
14 Indiana	D	D	D	D	R		11
15 lowa	D	D	D	D	R		6
22 Michigan	D	D	D	D	D		16
33 North Carolir	D	R	R	R	R		15
35 Ohio	D	D	D	R	R		18
38 Pennsylvania	D	D	D	D	D		20
49 Wisconsin	D	D	R	R	D		10
Count Correc	5	$\epsilon$	5	4	6	10	

60% of red states correctly predicted

## Correcting for Political Bias

- We found that correcting for the representation of demographic group was not enough to achieve a successful prediction
- Predictions were still clearly in favor of a democratic candidate, so we also corrected for political bias of Twitter users

### Using Exit Polls for Corrections and Evaluation

- Since voting is anonymous, we get demographic voting info for each state from exit polls
- Exit polls only survey 25/50 states + nation on who each demographic group voted for
- We used these exit polls to correct political bias of Twitter users due to self selection
- Approach for addressing bias of internet surveys using a reference sample from Bethlehem, Jelke. "Selection bias in web surveys." *International Statistical Review* 78.2 (2010): 161-188. Cited by 538 papers.
- We used 2016 Twitter data (internet survey) and 2016 exit polls (reference sample) to find a weight to correct our new observations 2020 Twitter data (new internet survey)

 $men Percent Democrat\_2020 predicted = men Percent Democrat\_2020 twitter * men Percent Democrat\_2016 exit Poll / men Percent Democrat\_2016 twitter * men Percent Democrat\_2016 exit Poll / men Percent Democrat\_2016 twitter * men Percent Democrat\_2016 exit Poll / men Percent Democrat\_2016 twitter * men Percent Democrat\_2016 exit Poll / men Percent Democrat\_2016 twitter * men Percent Democrat\_2016 exit Poll / men Percent Democrat\_2016 twitter * men Percent Democrat\_2016 exit Poll / men Percent Democrat\_2016 twitter * men Percent Democrat\_2016 exit Poll / men Percent Democrat\_2016 twitter * men Percent Democrat\_2016 exit Poll / men Perc$ 

• Using 2020 exit polls we calculated the absolute error for each state election, then calculated the mean absolute error for each demographic group across the states

menPercentDemocrat\_2020absoluteError = | menPercentDemocrat\_2020predicted - menPercentDemocrat\_2020exitPoll |

menPercentDemocrat\_2020meanAbsoluteError = mean of menPercentDemocrat\_2020absoluteError for all state elections

#### Limitations of Political Bias Corrections

- Exit polls are incomplete
- Only **20/50 states + nation** are surveyed for 2016 and 2020
- All 2**0/50 states + nation** have complete data gender
- Only 3 states have complete data for all races/ethnicities
- Only **1 state** has complete data for all gender x race/ethnity combinations
- Most states **only have complete data for** white people, men, women, white men, and white women
- As a result, we could **only predict 20/50 states** by breaking the electorate down into men and women
- Not enough exit poll data to predict the election across all states using race/ethnicity x gender or race/ethnicity
- Weighting with a reference sample is one of the simplest approaches to correct for bias in a web survey
- Propensity weighting is a more sophisticated approach (that we did not implement) that attempts to make the two
  populations comparable by simultaneously controlling for all variables that were thought to explain the differences. In
  the case of a web survey, there are also two populations: those who participate in the web survey (if asked), and those
  who will not participate

#### Predictions - Political Bias Correction

Green - correctly predicted
Red - incorrectly predicted
<b>Note</b> : this prediction is using
the gender breakdown of the
electorate into men and
women. There was not enough
exit poll data to separately
break down the by
race/ethnicity and gender x
race/ethnicity for each state
National popular vote correctly predicted

Mational popular	VOLC	COLLCCL
predicted		

6/11 swing states correctly predicted

9/9 non-swing states correctly predicted

Election	Predicted Party Winner	Democrat Absolute Error	Republican Absolute Error
National	Democrats	1.624534172	2.013612075
Arizona	Republicans	4.186633357	1.772344673
Colorado	Democrats	15.81542592	14.30424906
Florida	Republicans	1.186563683	11.88747336
Georgia	Republicans	13.52073699	1.277503599
lowa	Republicans	3.74487781	11.73186314
Kentucky	Republicans	8.293127602	14.90693585
Maine	Republicans	10.02244115	8.938231786
Michigan	Democrats	8.288204467	6.388857116
Minnesota	Democrats	2.745881007	7.944711385
Nevada	Republicans	9.64450996	20.26965688
New Hampshire	Democrats	6.525037594	10.60013103
North Carolina	Republicans	0.576719545	7.063449802
Ohio	Republicans	4.908268662	0.700476973
Oregon	Democrats	5.562491603	4.742165299
Pennsylvania	Republicans	2.031276103	3.735331655
South Carolina	Republicans	5.739884487	3.049984899
Texas	Republicans	0.253495329	5.013906609
Virginia	Democrats	2.489944458	2.028357665
Washington	Democrats	13.66652343	18.11535433
Wisconsin	Democrats	2.74553701	0.814733339

## Further Analysis - Political Bias Correction

10.33679337 12.88028542 6.891772823 18.11842105 27.1210175
6.891772823 18.11842105 27.1210175
18.11842105 27.1210175
27.1210175
20.23951638
9.267430888
10.16306399
21
50.84117647
28.59572815
16.281479
pub Absolute Error All States
12.23825192
10.75730494
19.06027933
6.85998439
5.918733669
5.077314271
3.077314271
21.37866146
21.37866146
21.37866146 12.30219873
21.37866146 12.30219873 4.379551821
,

## Website

## Presentation Roadmap

- Introduction to Problem
- Approach
- Results
- Conclusions
- Limitations and Further Steps

#### Conclusions from Corrections

- We conclude that correcting raw tweet sentiment based on demographic groups is not sufficient to predict the 2020 election because sentiment is still too skewed toward Democrats - we are predicting Democratic for most states (Our prediction was correct for 90.4% of blue states, 60% of swing states, and 45% of red states)
- In the case of the 2020 election, correcting our prediction based on political bias yielded the best results (National popular vote correctly predicted, 6/11 swing states correctly predicted, 9/9 non-swing states correctly predicted)
- Mean average error for some groups is within than the margin of error for traditional surveys
- These results may not be generalizable to other elections (2016, 2018, etc) because they rely on 2016 Twitter data and 2016 exit polls

## To what extent can social media predict the 2020 presidential election?

- The ability of social media to predict election outcomes in states depends on the demographic group and the state that the user is in
- Not all demographic groups are represented well on Twitter and exit polls (white men, white women, and white people are represented well on Twitter and exit polls)
- As shown previously, the political bias correction with the gender breakdown of the
  electorate into men and women works best (as opposed to breaking it down by
  race/ethnicity or race/ethnicity x gender). We suspect this is due to black, hispanic, asian,
  and "other" race/ethnicity category are not represented well on Twitter or exit polls.
- For predicting the voter share of Democrats: white people, men, women, white men, and white women have the lowest mean average error across all states + national election
- For predicting the voter share of Republicans: black people, hispanic people, asian people, men, women, white women, black men, black women, and hispanic men have the lowest mean average error across all states + national election

## Presentation Roadmap

- Introduction to Problem
- Approach
- Results
- Conclusions
- Limitations and Further Steps

#### Limitations of Twitter Data Collection

- **Had to simplify number of queries** because of the prohibitive time complexity O(#locations x #queries) which still took 6-7 days to complete even after simplifying
- Twitter limits the amount of tweets requested by your IP address, so we had to use time.sleep() a lot
- Couldn't get county-level results in time because Twitter API approved late, so we didn't have the ability to query county-specific tweets using geo-polygons
- Many states have relatively few tweets compared to California, New York, etc.
- 2016 has 25% of the total number of tweets that 2020 has
  - o 2016: 67,867 tweets
  - o 2020: 247,837 tweets
- Political tweet distributions are different
  - o 2016: 40% Neutral, 29% Democrat, 26% Republican, 5% Unknown
  - o 2020: 28% Neutral, 36% Democrat, 34% Republican, 2% Unknown
  - Some demographic groups (black women) tweet much differently w.r.t political views in 2020 than in 2016, which leads to much worse predictions than others (whites, men, women)
  - Tweets are more "polarized" in 2020 than in 2016

## Sentiment Analysis and Profiling Limitations

- Sentiment analysis model not built on the data from 2020 election domain
  - No human-labeled tweets about 2020 election
  - Use general result instead of data oriented result
- Unknown accuracy of demographic data of users (no "true" labels to evaluate with)

## Further Steps

- Estimation of voter turnout, rather than only voter share for each demographic group
- More comprehensive Twitter queries using online LDA to find new search terms
- Get county-specific results using geo-polygons with the Twitter API
- Neighborhood types (suburban, urban, rural designations)
- Characterization of users into age groups and potential voters
- Train a more accurate sentiment analysis model using human-labeled tweets
- Take into account **linguistic differences** among groups and locations
- Use the **communication network** to cluster more users. The Base users are the ones who have high political preference score from sentiment analysis.
- Different corrections for incumbent candidates & interesting candidates (more tweets about incumbent candidates and "interesting" candidates like Trump)

## Thank you!

## 2016 Results

state	uncorrected	gender cor	re race_correct	race gender	true vote	26 Montana	N	N	N	N	
0 Alabama	N	N	D	N	R	27 Nebraska	N	N	N	N	
1 Alaska	N	N	N	D	R	28 Nevada	N	N	R	N	
2 Arizona	N	N	N	N	R	29 New Hampshire	N	N	N	D	
3 Arkansas	N	N	N	N	R	30 New Jersey	N	N	N	N	
4 California	N	D	N	N	D	31 New Mexico	N	D	D	D	
5 Colorado	N	N	N	N	D	32 New York	N	N	N	N	
6 Connecticut	N					33 North Carolina	N	N	N	N	
		N	N	N	D	34 North Dakota	N	N	D	D	
7 District of Columbia		N	N	N	D	35 Ohio	N	N	N	N	
8 Delaware	N	N	N	N	D	36 Oklahoma	N	N	D	D	
9 Florida	N	N	D	N	R	37 Oregon	N	N	N	N	
10 Georgia	N	N	N	N	R	38 Pennsylvania	N	N	N	N	
11 Hawaii	N	N	D	R	D	39 Rhode Island	N	N	N	N	
12 Idaho	N	D	D	D	R	40 South Carolina	N	N	N	N	
13 Illinois	N	N	N	N	D	41 South Dakota	N	N	N	D	
14 Indiana	N	N	N	N	R	42 Tennessee	N	N	N	N	
15 Iowa	N	N	N	N	R	43 Texas	N	N	N	N	
16 Kansas	N	N	N	N	R	44 Utah	N	N	N	N	
17 Kentucky	N	N	N	N	R	45 Vermont	N	N	D	D	
18 Louisiana	N	N	N	N	R	46 Virginia	N	N	N	N	
19 Maine	N	N	N	N	D	47 Washington	N	N	N	N	
20 Maryland	N	N	N	N	D	48 West Virginia	N	N	N	N	
21 Massachusetts	N	N	N	N	D	49 Wisconsin	N	N	N	N	
22 Michigan	N	N	N	N	R	50 Wyoming	N	D	N	N	
23 Minnesota	N	N	D	D	D	Count Correct		0	1	4	
24 Mississippi	N	N	N	N	R	Rep Electoral Votes		0	0	6	
24 Mississippi 25 Missouri					R R	National Prediction	N	N	D	D	
25 IVIISSOURI	N	N	N	N	K						

## 2020 Results

state	uncor	rected_ gender_	correrace c	orrect race gen	der true vote	26 Montana	D	R	D	D	R	
0 Alabama	R	R	R	R	R	27 Nebraska	R	D	D	D	R	
1 Alaska	D	D	R	R	R	28 Nevada	R	R	N	D	D	
2 Arizona	D	D	R	D	D	29 New Hampshire	R	R	R	R	D	
3 Arkansas	D	D	D	D		30 New Jersey	D	D	N	D	D	
4 California	D	D	D	D	R	31 New Mexico	D	D	D	R	D	
	D				D	32 New York	D	D	D	D	D	
5 Colorado	D	D	R	D	D	33 North Carolina	D	R	R	R	R	
6 Connecticut	D	D	R	D	D	34 North Dakota	D	D	D	D	R	
7 District of Columbia	D	N	D	N	D	35 Ohio	D	D	D	R	R	
8 Delaware	D	D	D	D	D	36 Oklahoma	D	D	D	D	R	
9 Florida	D	D	D	D	R	37 Oregon	D	D	D	D	D	
10 Georgia	D	D	D	D	D	38 Pennsylvania	D	D	D	D	D	
11 Hawaii	R	D	N	N	D	39 Rhode Island	D	D	D	D	D	
12 Idaho	D	R	R	R	R	40 South Carolina	D	D	R	D	R	
13 Illinois	D	D	D	D	D	41 South Dakota	D	D	D	D	R	
14 Indiana	D	D	D	D	R	42 Tennessee	R	R	R	R	R	
15 lowa	D	D	D	D	R	43 Texas	D	D	D	R	R	
16 Kansas	D	D	D	D	R	44 Utah	D	D	D	D	R	
17 Kentucky	D	D	R	R	R	45 Vermont	D	D	D	D	D	
18 Louisiana	D	D	D	D	R	46 Virginia	D	D	D	D	D	
19 Maine	D	D	D	D	D	47 Washington	D	D	D	D	D	
20 Maryland	D	D	D	D	D	48 West Virginia	D	R	R	R	D	
21 Massachusetts	D	D	D	D	D	49 Wisconsin	L L	D	R	R	,	
22 Michigan	D	D	D	D	D	50 Wyoming	D	D	D	R	R	
23 Minnesota	D	D	R	D	D	Count Correct	U	27	29	25	32	50
24 Mississippi	D	D	D	R	R	Dem Electoral Votes			29 478			306
25 Missouri	D	D	D	D	R	National Prediction	2	494		399	392	306
						National Prediction	D	D	D	D	D	