



#2tweetornot2tweet



The Question:

What do the Senators from the 115th Congress (Jan 2017 - Jan 2019) think on twitter?



Data Gathering

Had to find determine if Senators had Twitter account.

If multiple Twitter accounts, which account was their active account (campaign accounts, office manager accounts, old accounts, etc.).

Used [Tweepy](#) (Python Twitter API Wrapper) to request tweets from each Senator.

In the end 299,200 tweets were collected.



Rules for the AI



“Negative” Tweets

Tweets attacking Politicians: asking rhetorical questions, attacking someone’s platform.

Tweets attacking Business Entities: talking about taxes, regulations, bailouts, etc.

Tweets attacking Opposing Ideas: demonizing political ideologies, threatening outcomes of political goals, etc.



“Neutral” Tweets

Tweets with Links: links to other websites or another person’s statement.

Tweets with Factual/Announcement Statements: announcing times of events or poll numbers; correcting a statement or stating law with no additional statement.

Fundraising Tweets: non-politically fueled fundraising tweets (not included if they included political agenda), or short statements that have little to no context.

Neutral Tweets: Positive of one side and negative to the other side.



“Positive” Tweets

Celebratory Tweets: discussing a passed law, team win, etc.

Supportive Tweets: discussing or stating support of individuals, teams, or entities.

Thanking Tweets: non-sarcastic thank you's.



Model Data Sets

Training Set: 1,000 Senator Tweets

467 positive tweets, **334** neutral tweets, and **199** negative tweets.

Test Set: 298,200 Senator Tweets



Model #1

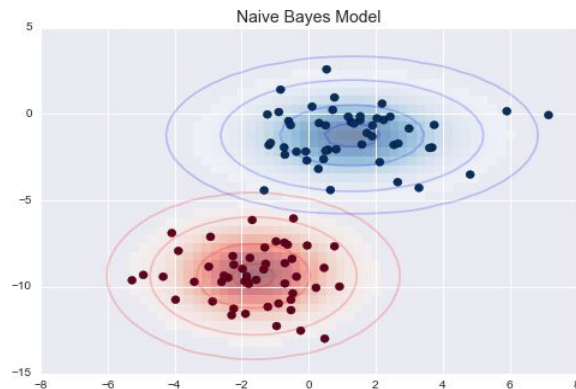
Naive Bayes: Bernoulli NB



```
from sklearn.naive_bayes import BernoulliNB
```

Each “feature” (text in Tweet) is independent of one another, in order to predict the category (negative, neutral, positive) of a given sample (tweet).

Naive Bernoulli Bayes ML model was about 60% accurate.





Model #2

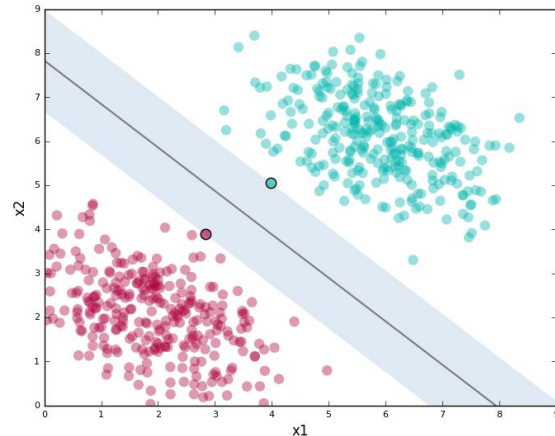
LinearSVC (Support Vector Classifier)



```
from sklearn.svm import LinearSVC
```

Classifies by using a “best-fit” method on a hyperplane to divide/categorize your data(negative, neutral, positive); providing features to the hyperplane will produce “predicted” class.

LinearSVC ML model was about 50% accurate.





Model #3

LinearSVC & Party Affiliation



```
from google_search import democrat, republican
```

Used LinearSVC model, but categorizing by political party (Democrat, Republican)

Gathered each Senator's birth year, state, and political party.

LinearSVC ML model was 80% accurate in predicting a tweet's party affiliation (Independent not included).



Democrat/Republican

	precision	recall	f1-score	support
Democrat	0.81	0.81	0.81	36633
Independent	0.37	0.68	0.48	877
Republican	0.82	0.81	0.81	37290
micro avg	0.80	0.80	0.80	74800
macro avg	0.67	0.76	0.70	74800
weighted avg	0.81	0.80	0.81	74800



Model #4

LinearSVC & Generation



Generation: 71% accuracy

	precision	recall	f1-score	support
Baby Boomer	0.82	0.80	0.81	51252
Gen X	0.50	0.53	0.52	11945
Silent	0.48	0.50	0.49	11603
micro avg	0.71	0.71	0.71	74800
macro avg	0.60	0.61	0.61	74800
weighted avg	0.72	0.71	0.71	74800



Model #5

LinearSVC & State



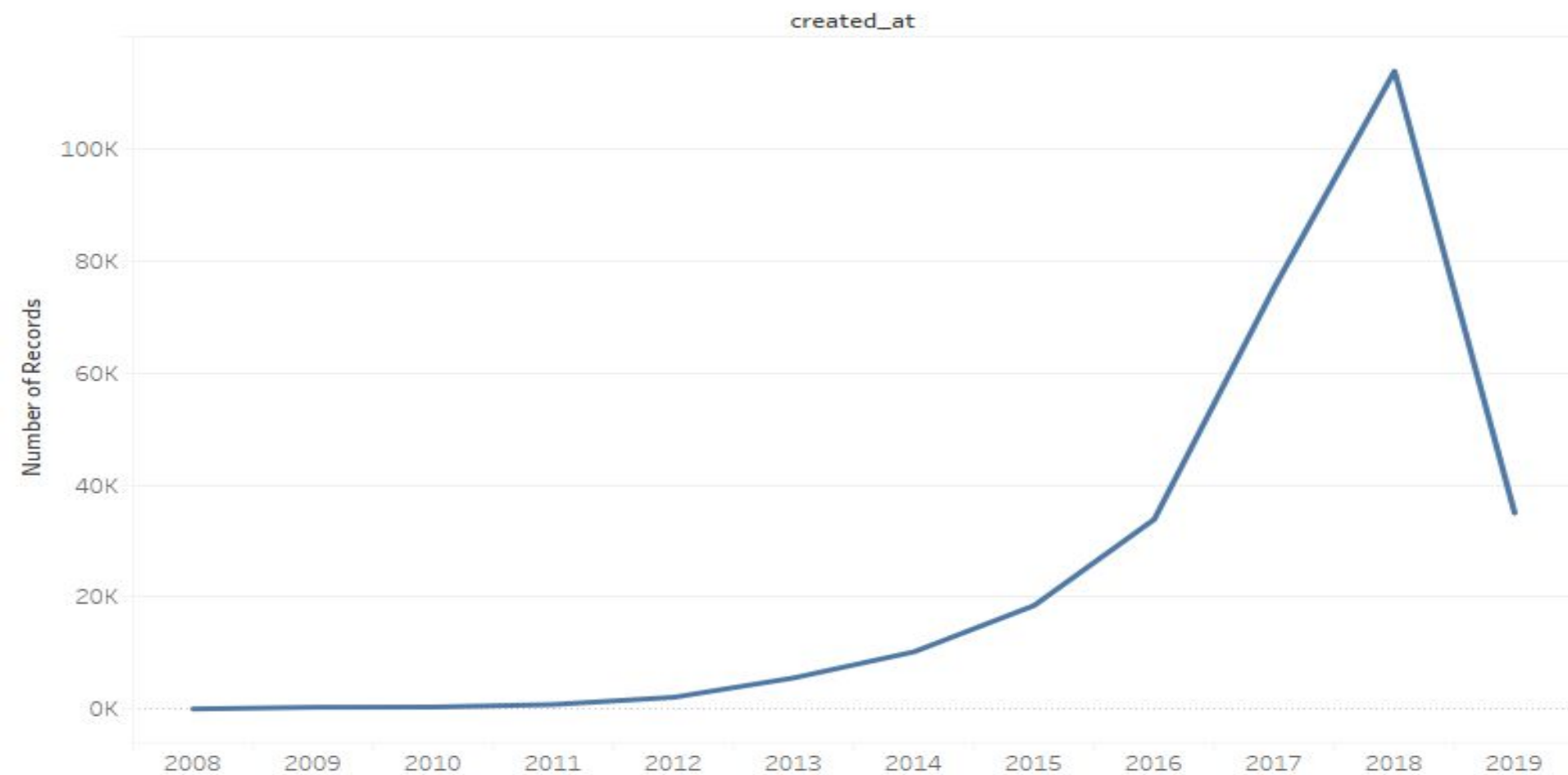
State: 52% Accuracy

Overall, it predicted correctly about 52% of the time, unless the senator was from Indiana, Iowa, or North Dakota (~70%).



Data Visualizations

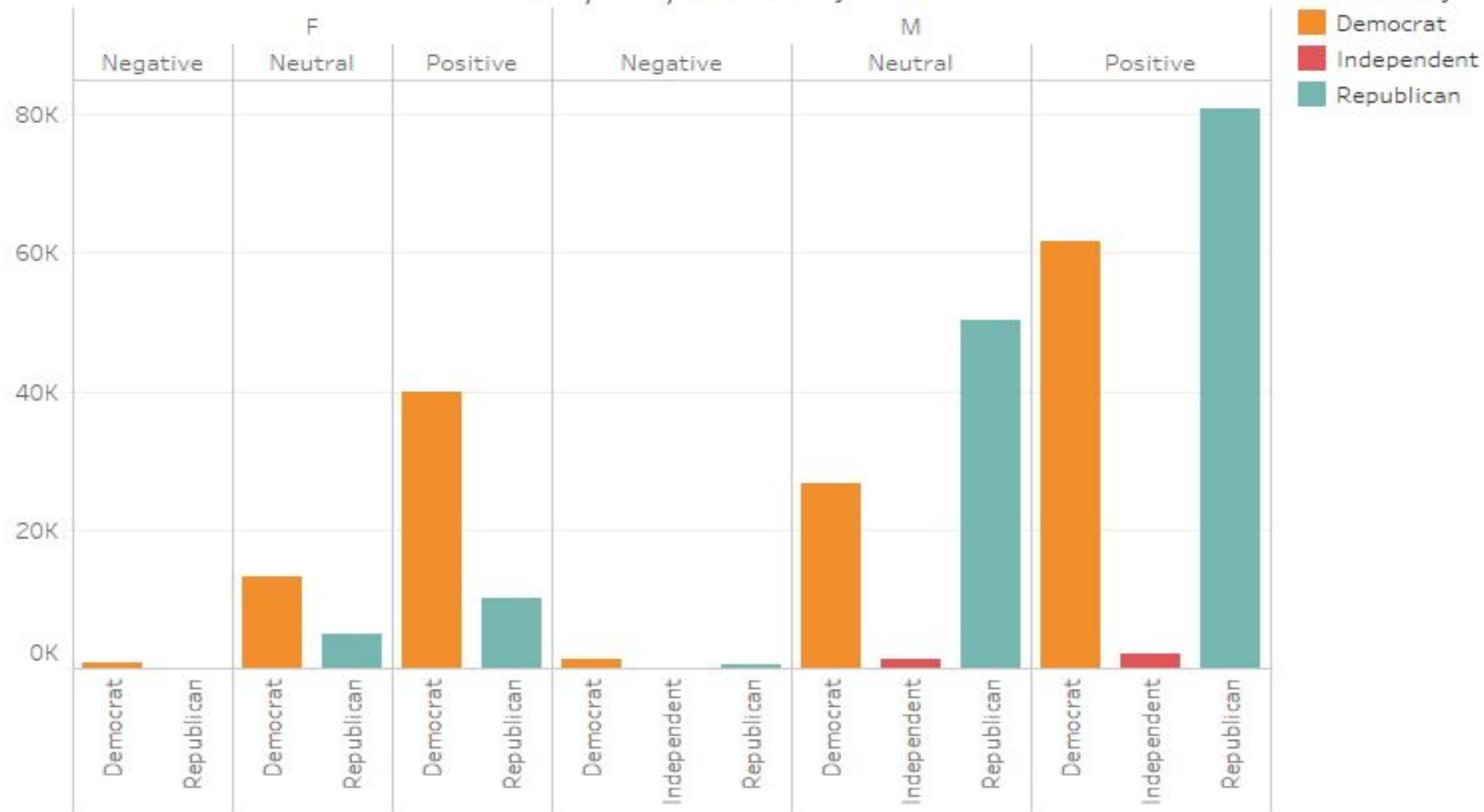
Number of Tweets of all Senators by Year



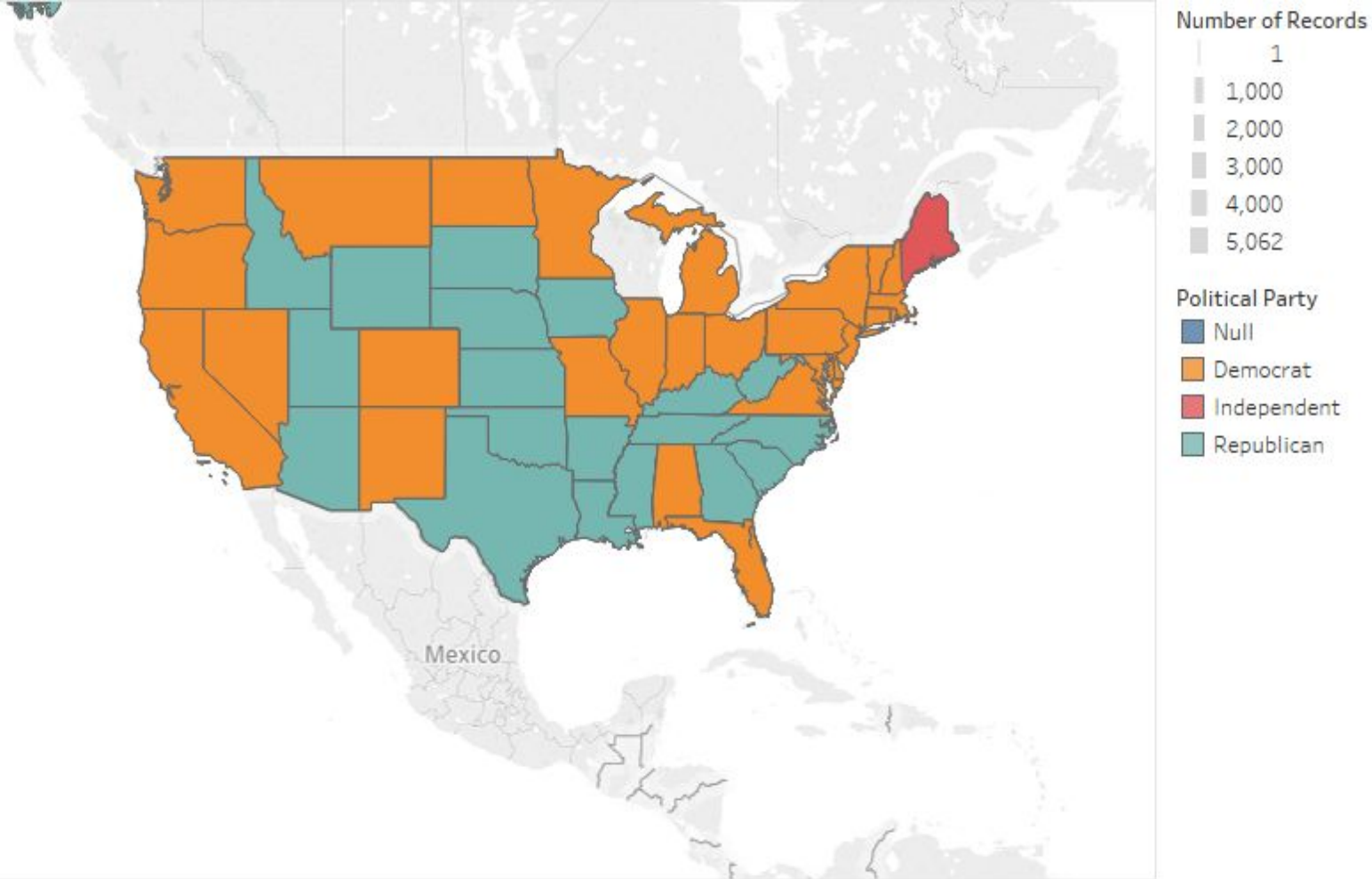
The trend of sum of Number of Records for created_at Year.

Gender / Sent / Political Party

Number of Records

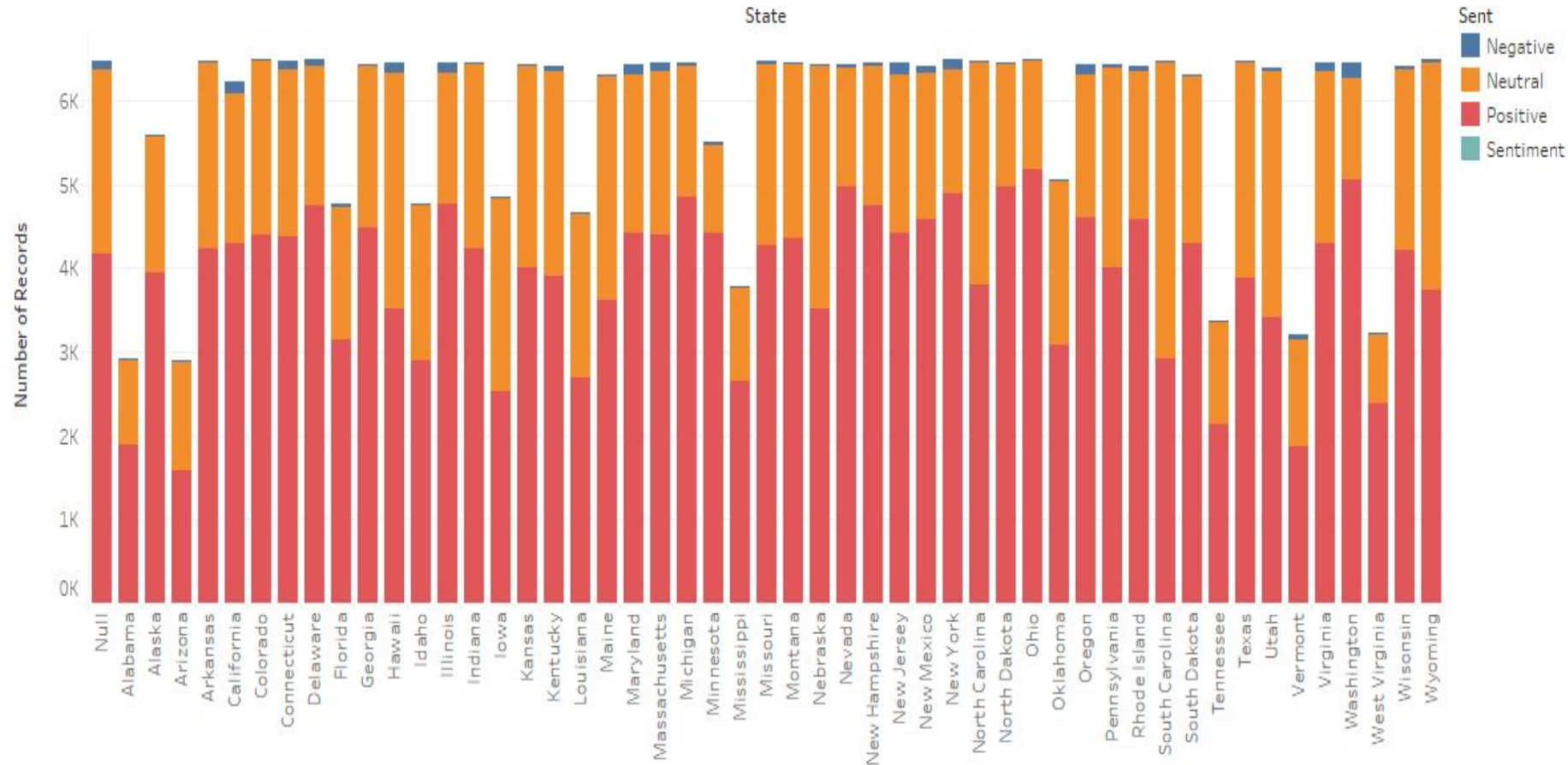


Sum of Number of Records for each Political Party broken down by Gender and Sent. Color shows details about Political Party. The view is filtered on Political Party, which has multiple members selected.

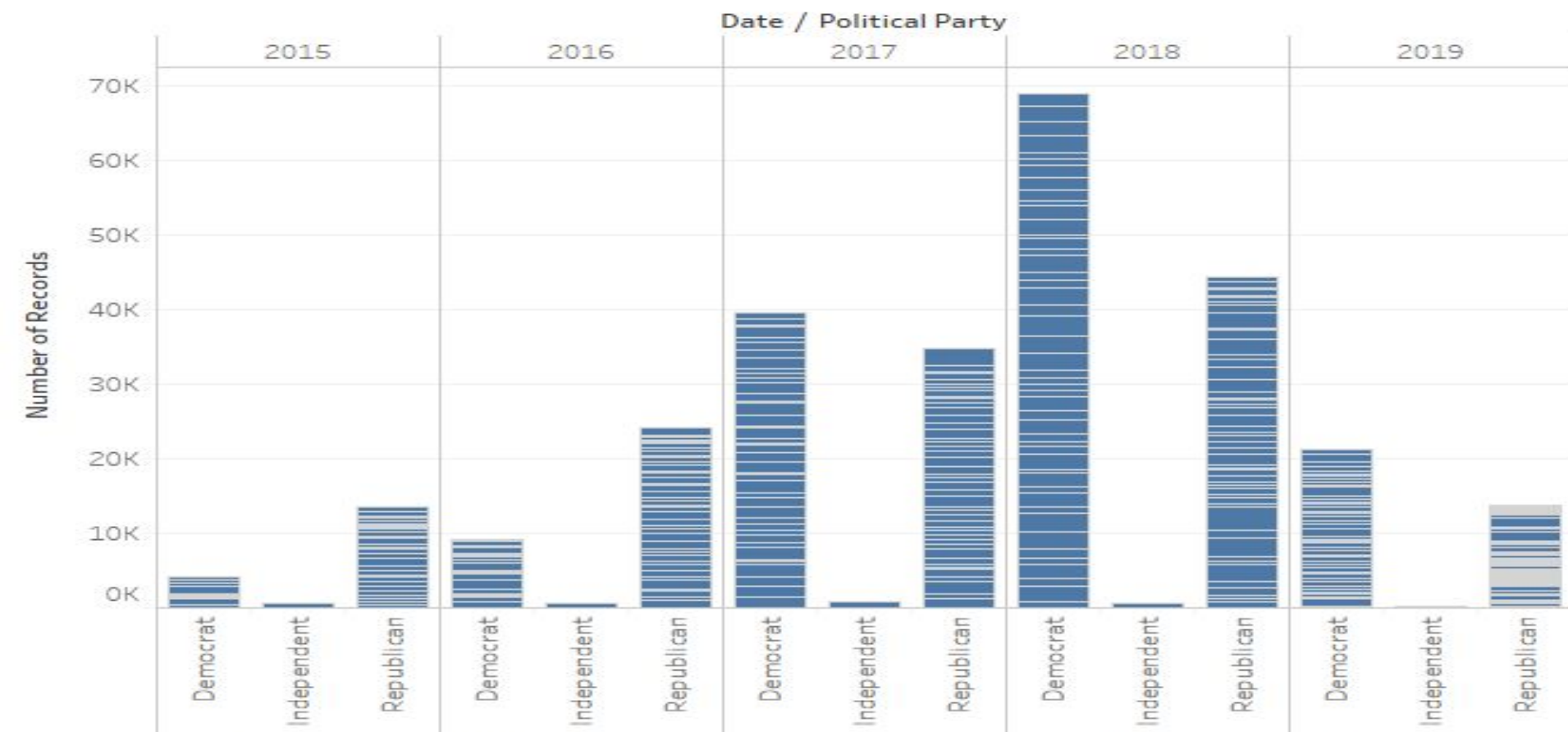


Map based on Longitude (generated) and Latitude (generated). Color shows details about Political Party. Size shows sum of Number of Records. Details are shown for State and Sent. The view is filtered on Sent, which keeps Negative, Neutral, Positive and Sentiment.

Positive, Neutral, Negative Senator Tweets by State



Tweets by Political Party



Sum of Number of Records for each Political Party broken down by Date Year. Details are shown for Senator. The view is filtered on Date Year and Political Party. The Date Year filter keeps 2015, 2016, 2017, 2018 and 2019. The Political Party filter keeps Democrat, Independent and Republican.

Sample Size Accuracy



Sum of Percent Correct for each Number Shuffled broken down by Target. Color shows sum of Percent Correct.