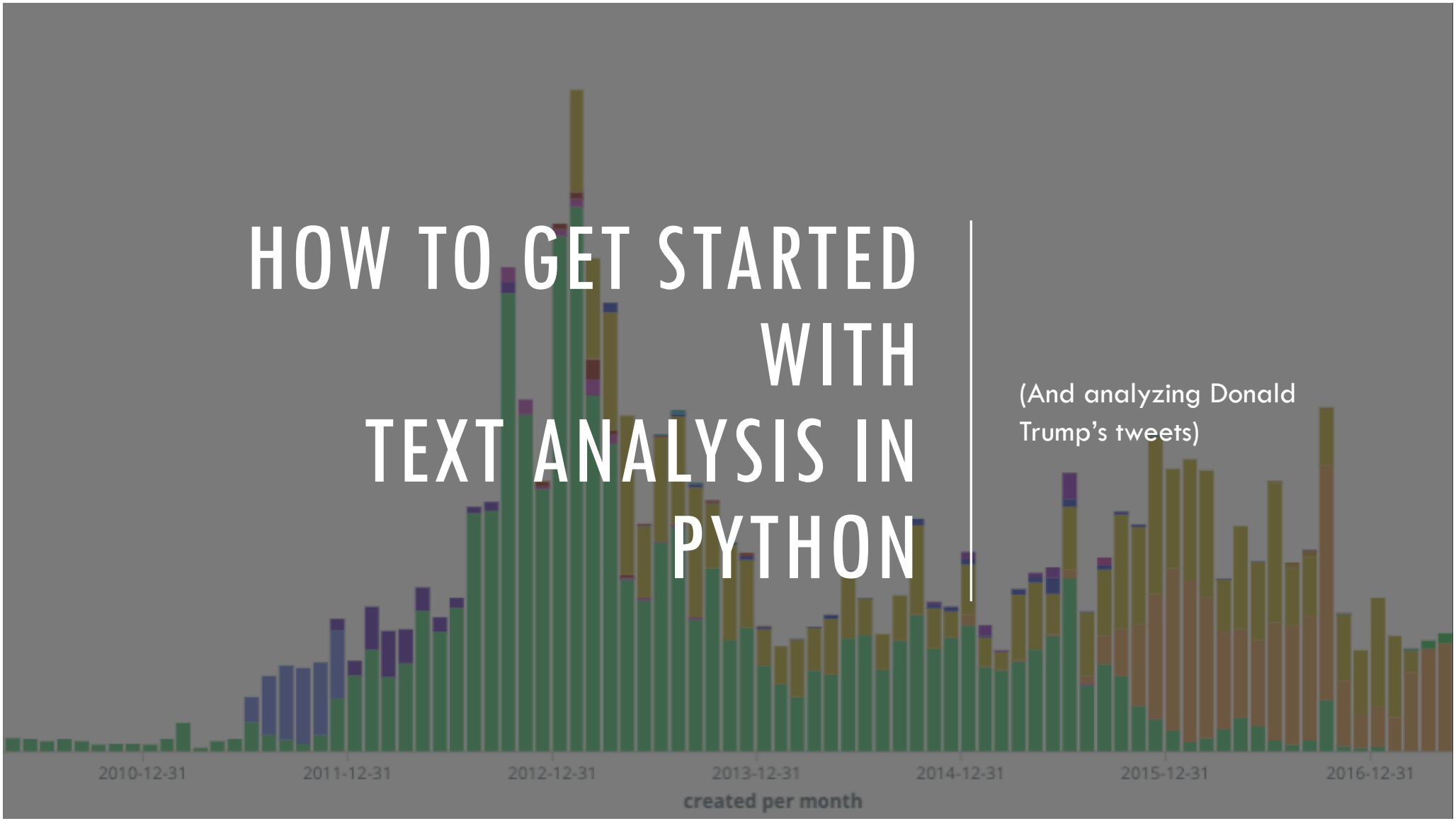
HOW TO GET STARTED WITH TEXT ANALYSIS IN PYTHON

(And analyzing Donald Trump's tweets)



Please keep in mind:	Options are my own
	I'm not an expert on anything
	It's a Python conference

DISCLAIMERS

OVERVIEW

01

Text Analysis with TextBlob

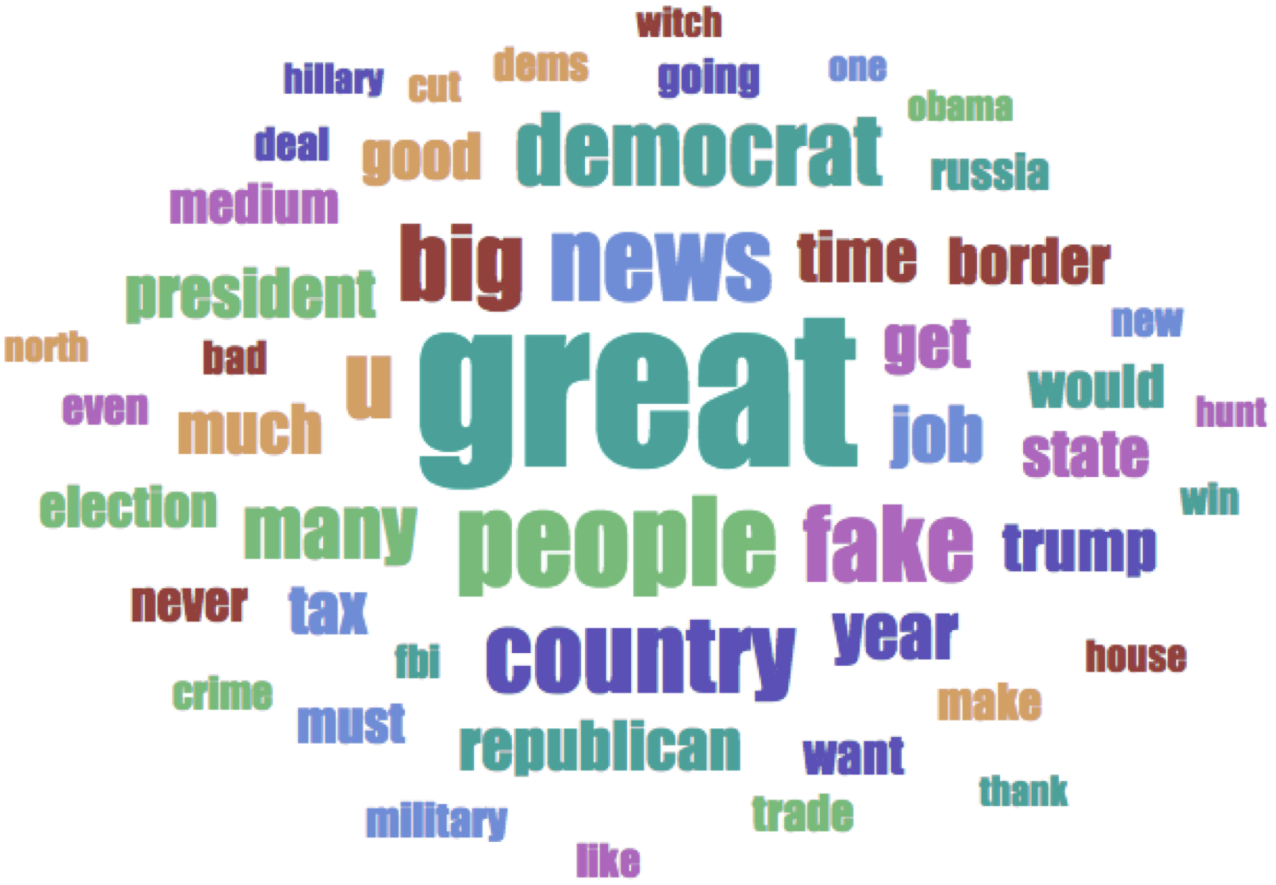
02

Text Classification Basics

03

Machine Learning with Naïve Bayes

INTRO TO TEXTBLOB

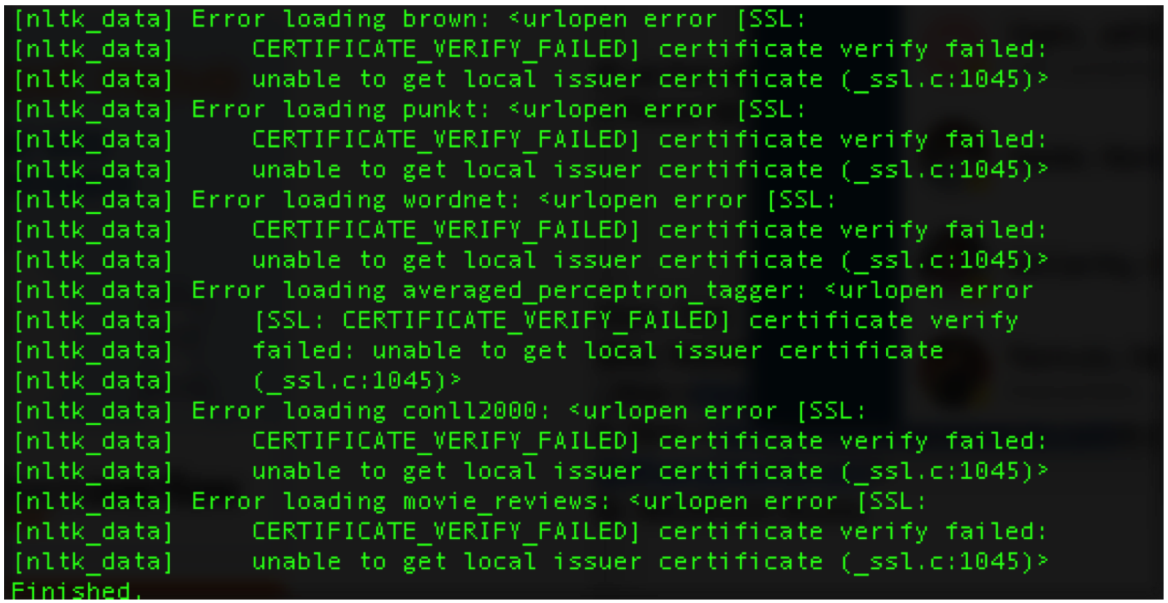


TEXTBLOB BASICS

Term Counts	Parts of Speech	Tokenization
Sentiment (Subjectivity / Polarity)	Language Detection / Translation	Spelling

DEMO

DOWNLOAD CERTIFICATE PROBLEM



/Applications/Python 3.X/Install Certificates.command

MACHINE LEARNING

This Photo by Unknown Author is licensed under CC BY-NC-ND

MACHINE LEARNING

01

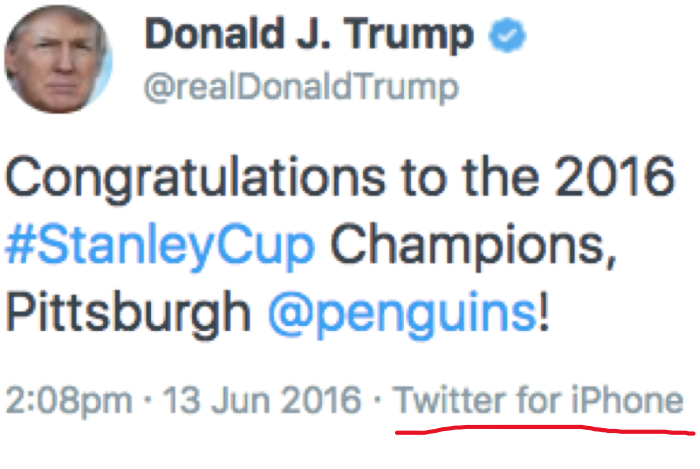
Problem

02

Text

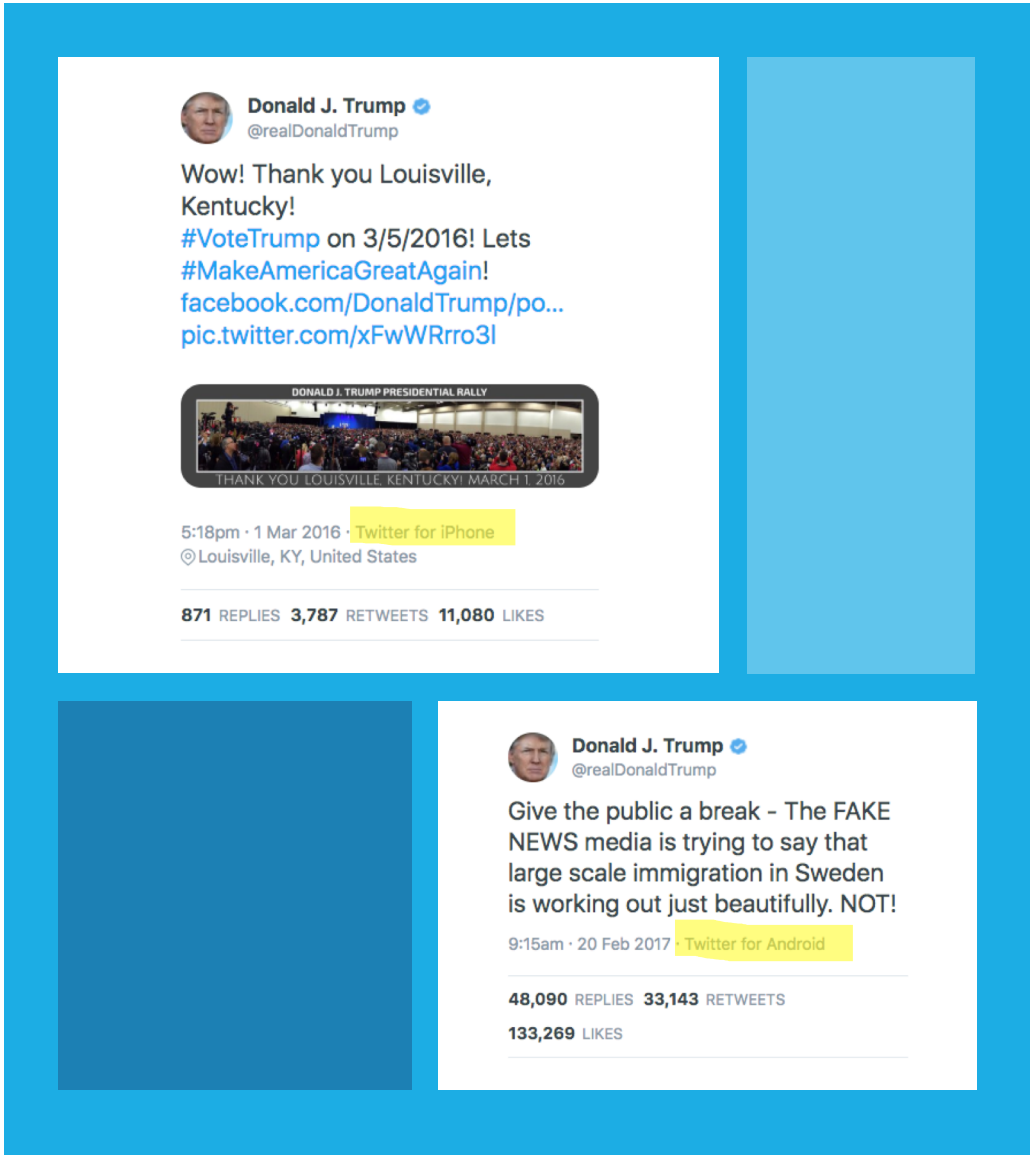
03

Tools



KNOW YOUR PROBLEM

KNOW YOUR PROBLEM



VARIANCE EXPLAINED

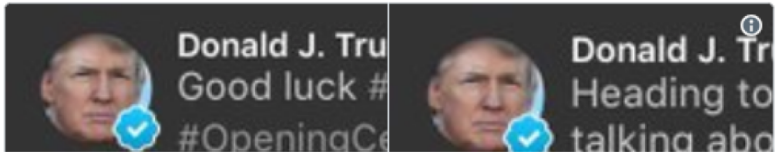
ABOUT ME POSTS LEARN R TEXT MINING IN R INTRODUCTION TO EMPIRICAL BAYES

David Robinson

Chief Data Scientist at DataCamp, works in R and Python.

Text analysis of Trump's tweets confirms he writes only the (angrier) Android half

I don't normally post about politics (I'm not particularly savvy about polling, which is where data science has had the largest impact on politics). But this weekend I saw a hypothesis about Donald Trump's twitter account that simply begged to be investigated with data:



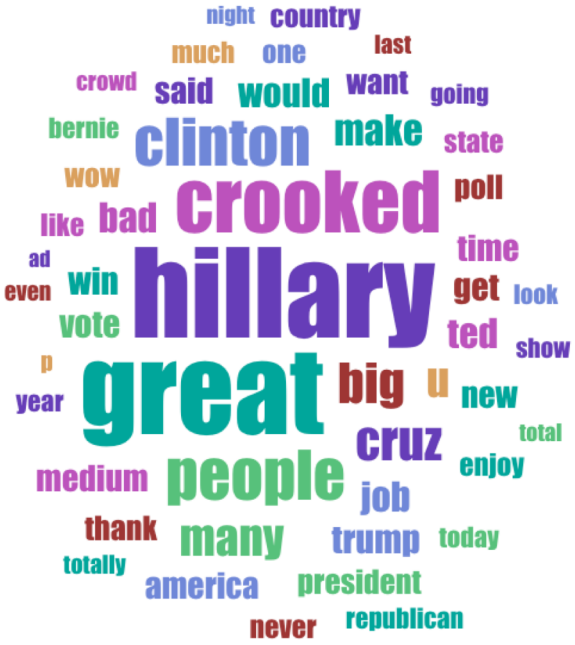
KNOW YOUR PROBLEM

2016 TWITTER WORD COMPARISON

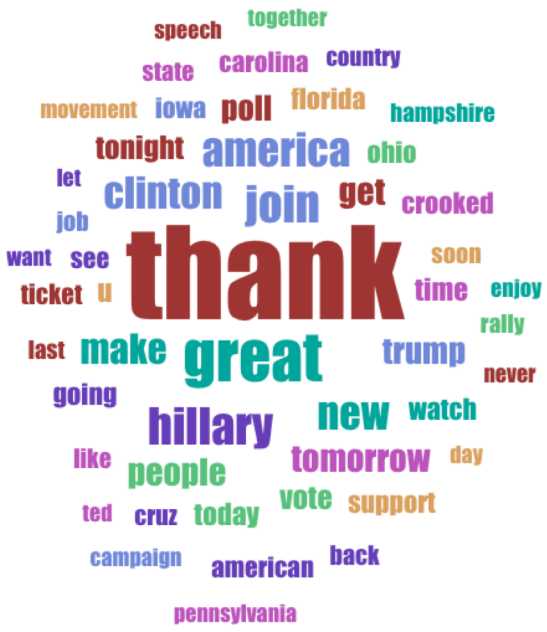
Top Words	Uses
hillary	230
great	223
crooked	156
people	125
clinton	120
big	99
cruz	97
many	90
u	85
make	75

Export: Raw Formatted

1 2 3 4 5 ...



Android



Non-Android

Top Words	Uses
thank	530
great	244
hillary	167
join	163
america	143
new	136
clinton	132
make	119
get	100
people	99

Export: Raw Formatted

1 2 3 4 5 ...

KNOW YOUR PROBLEM

Donald J. Trump
@realDonaldTrump

There was No Collusion with Russia (except by the Democrats). When will this very expensive Witch Hunt Hoax ever end? So bad for our Country. Is the Special Counsel/Justice Department leaking my lawyers letters to the Fake News Media? Should be looking at Dems corruption instead?

1:43pm · 2 Jun 2018 · Twitter for iPhone

Donald J. Trump
@realDonaldTrump

Landing in Las Vegas now for a Make America Great Again Rally supporting @DeanHeller and @DannyTarkanian. Also doing interview there with @seanhannity live on @FoxNews. Big crowd, long lines. Will be great! #MAGA

9:27pm · 20 Sep 2018 · Twitter for iPhone

Donald J. Trump
@realDonaldTrump

The Failing New York Times wrote a story that made it seem like the White House Councel had TURNED on the President, when in fact it is just the opposite - & the two Fake reporters knew this. This is why the Fake News Media has become the Enemy of the People. So bad for America!

8:06am · 19 Aug 2018 · Twitter for iPhone

Donald J. Trump
@realDonaldTrump

Join me this Saturday in Wheeling, West Virginia at 7pmE! Tickets: donaldjtrump.com/rallies/wv-sep...

4:48pm · 26 Sep 2018 · Twitter for iPhone

RESEARCH AND REFERENCES

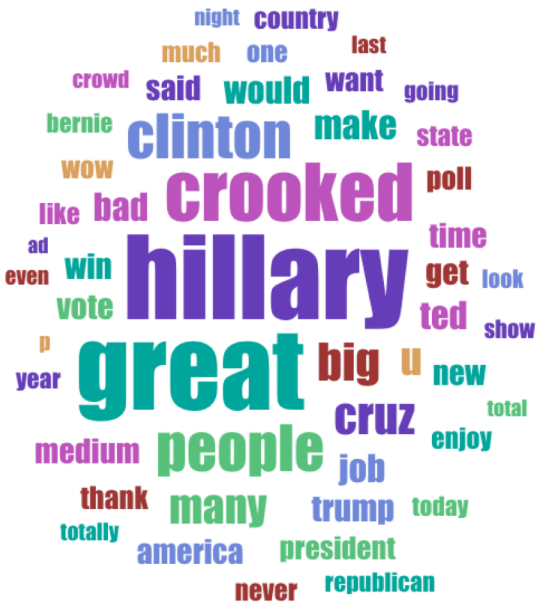
Source	Description
Variance Explained, "Text analysis of Trump's tweets confirms he writes only the (angrier) Android half"	David Robinson's original analysis of Android vs. iPhone tweets from @realdonaldtrump
The New York Times Magazine, "The Man Behind the President's Tweets"	Profile of Dan Scavino, Donald Trump's Director of Social Media, with discussion of others being able to tweet as @realdonaldtrump, by Robert Draper
The New York Times, "A Homebody Finds the Ultimate Home Office"	Analysis of Donald Trump's typical routine when arriving in the White House, including use of an Android phone, by Maggie Haberman
Android Central, "Which Android Phone does Donald Trump use?"	Analysis of Donald Trump's likely cell phone model based on public photos, by Alex Dobie
Boston Globe, "Inside the Trump Tweet Machine"	Background on White House staff members composing tweets to mimic Donald Trump's style, by Annie Linskey

Most Common Android Words

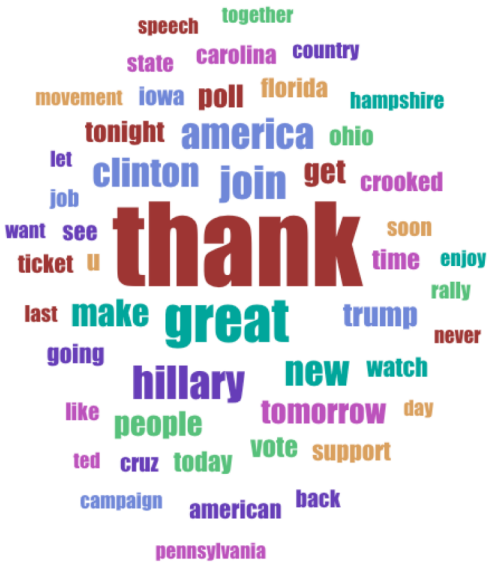
Top Words	Uses
hillary	230
great	223
crooked	156
people	125
clinton	120
big	99
cruz	97
many	90
u	85
make	75

Export: Raw Formatted

1 2 3 4 5 ...



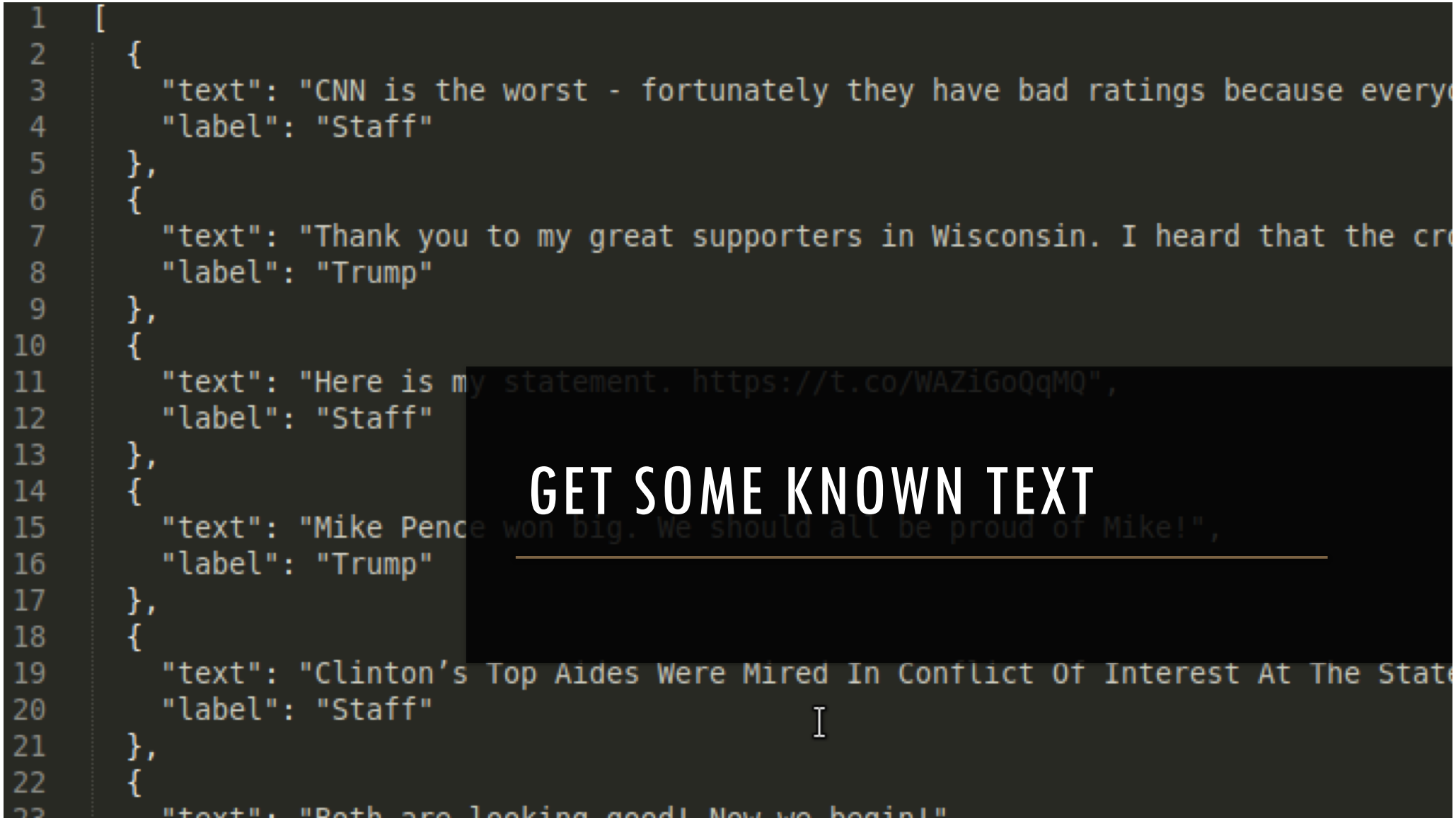
Top iOS/Web Words



Top Words	Uses
thank	530
great	244
hillary	167
join	163
america	143
new	136
clinton	132
make	119
get	100
people	99

Export: Raw Formatted

1 2 3 4 5 ...


[
{
"text": "CNN is the worst - fortunately they have bad ratings because every
"label": "Staff"
},
{
"text": "Thank you to my great supporters in Wisconsin. I heard that the cr
"label": "Trump"
},
{
"text": "Here is my statement. https://t.co/WAZiGoQqMQ",
"label": "Staff"
},
{
"text": "Mike Pence won big. We should all be proud of Mike!",
"label": "Trump"
},
{
"text": "Clinton's Top Aides Were Mired In Conflict Of Interest At The State
"label": "Staff"
},
{
GET SOME KNOWN TEXT

NAÏVE

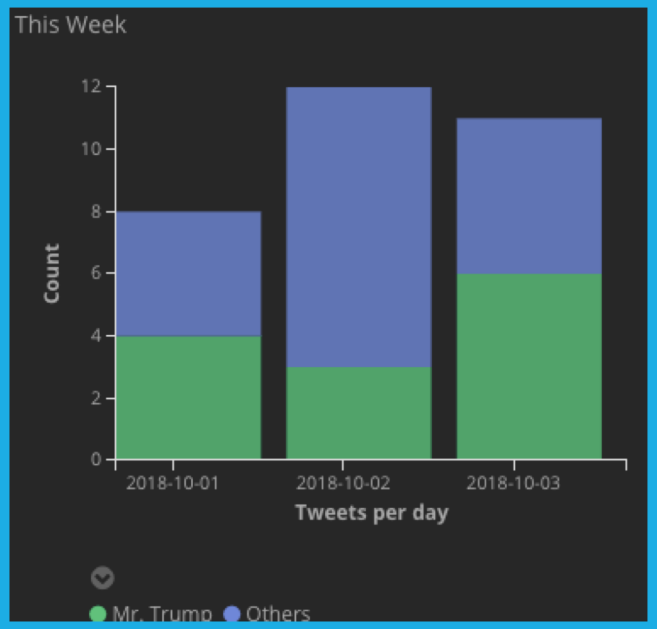
Everything is independent

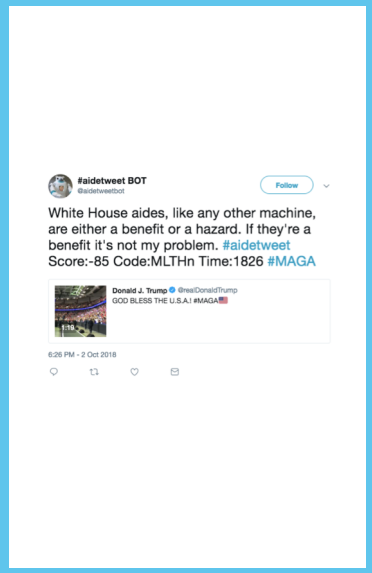
BAYES

Inputs (history) predict outputs (future)

DEMO (2)

WHAT TO DO NOW?





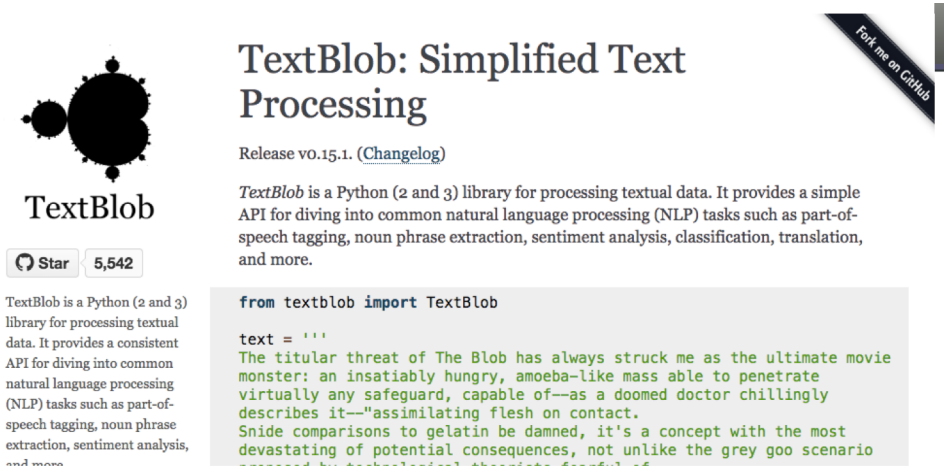
FINAL DISCLAIMERS

Please keep in mind:

- No magic.
- TextBlob isn't perfect
  - Slow
  - Overfit
  - Best practices
- Naïve Bayes isn't perfect
- It's all just probabilities

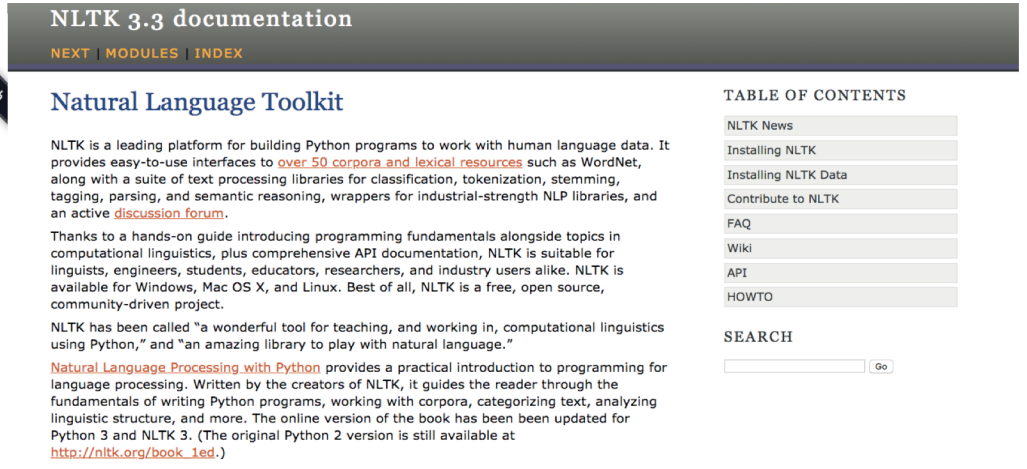
TextBlob Documentation

https://textblob.readthedocs.io/



NLTK Documentation

https://www.nltk.org/



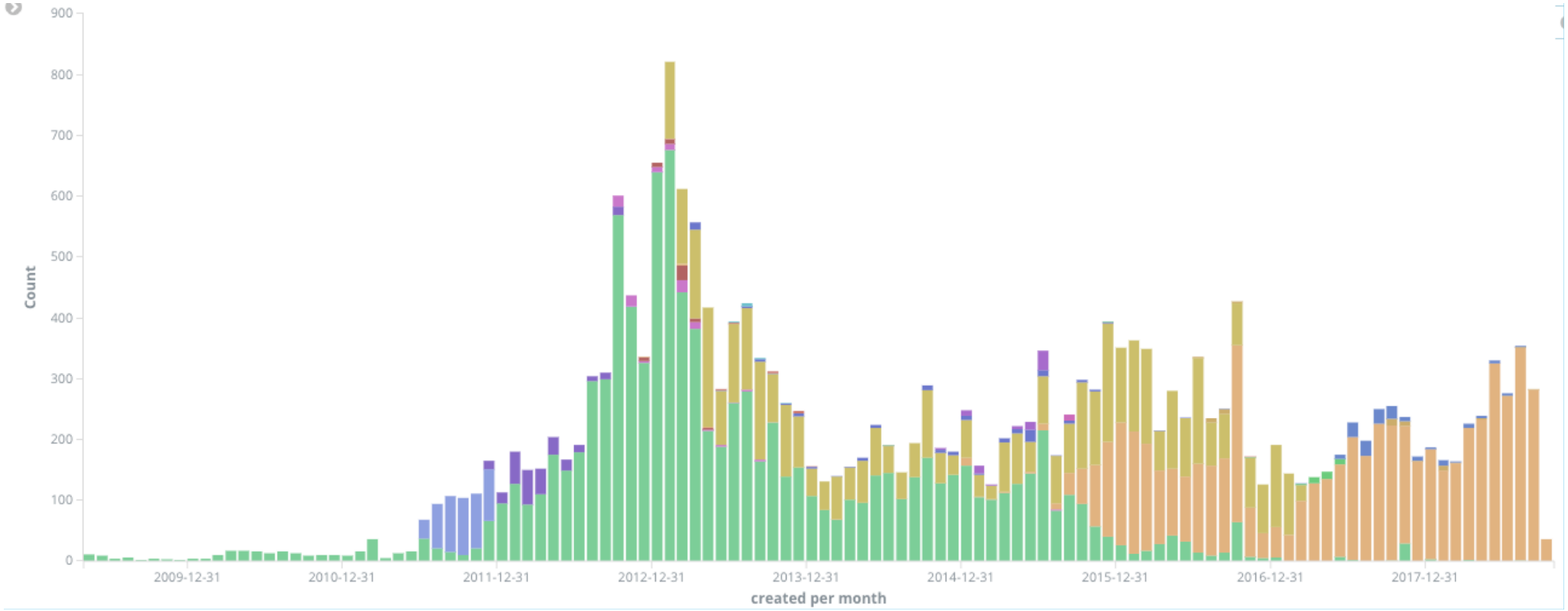
RESOURCES

POLITICAL TEXT ANALYSIS

Source	Description
@trumpornotbot	Twitter bot created by Andrew McGill of the Atlantic, using Naïve Bayes to determine the likely author of Donald Trump's tweets
TrumpTweetTrack.com @trumptweettrac	Full archive of Donald Trump's tweets analyzed with likely authors and various charts and graphs, including real-time analysis as tweets are posted using gradient boosting
TrumpTwitterArchive.com @realtrumptweet	Full archive of Donald Trump's tweets, updated regularly, in a fully searchable and mobile-friendly web page (along with several other politicians)
State of the Union	Online archive of State of the Union speeches with various text analysis and graphs
Factba.se @factbasefeed	Analysis of tweets, speeches, videos, depositions, and a wide variety of other topics regarding Donald Trump as well as Congress

THANKS TO STEVE LORIA

Maintainer of TextBlob and several other projects
@sloria1 | stevenloria.com



THANK YOU!

@daveklee