News

The Untold Story

Objective

- Explore news article data to better understand what is portrayed to the public
- Find underlying structure in the articles
- Cluster the data together for analysis
- See if one can see biases appear

The Data

Scraping text from articles such as:

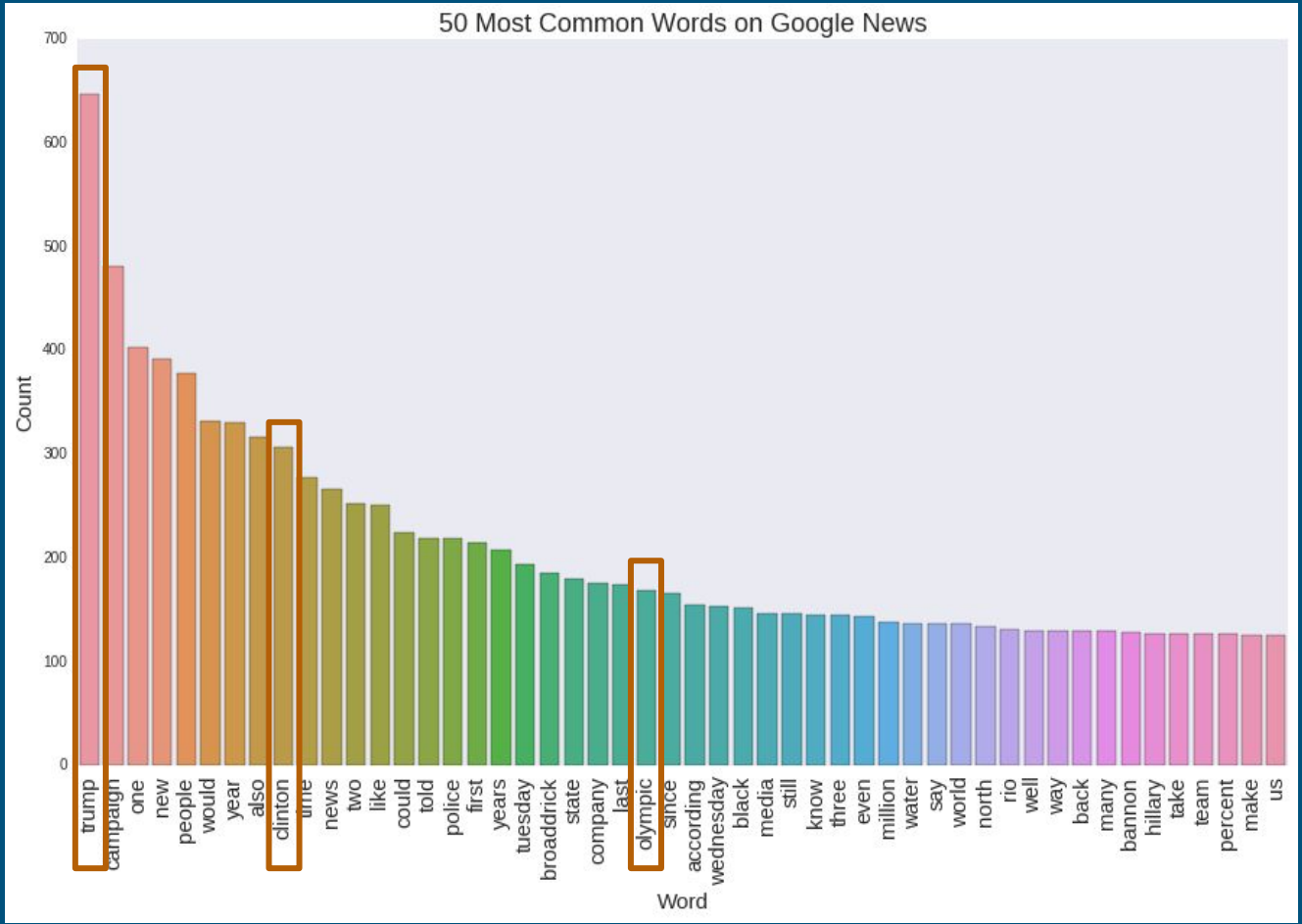
SundayReview | OP-ED COLUMNIST

A Playboy for President

Ross Douthat AUG. 13, 2016

Donald Trump in 2003 with Victoria Silvstedt, 1997 playmate of the year, left, and his future wife, Melania Knauss, at a Playboy event. Richard Corkery/NY Daily News Archive, via Getty Images

IN a different campaign or era, it would have been a race-altering moment; in this one, it was barely a scandal. There was Melania Trump, the potential first lady of the United States, posing stark naked in '90s-era photos published by the New York Post — and then in the next day's edition, canoodling lipstick-lesbian style in bed. Yet the press yawned, her husband's latest outrage overshadowed it, and it only stayed a story because the date of the photos raised questions about the future Mrs.


50 Most Common Words on Google News
Count
Word
0
100
200
300
400
500
600
700
trump
campaign
one
new
people
would
year
also
clinton
time
news
two
like
could
told
police
first
years
tuesday
broaddrick
state
company
last
olympic
since
according
wednesday
black
media
still
know
three
even
million
water
say
world
north
rio
well
way
back
many
bannon
hillary
take
team
percent
make
us

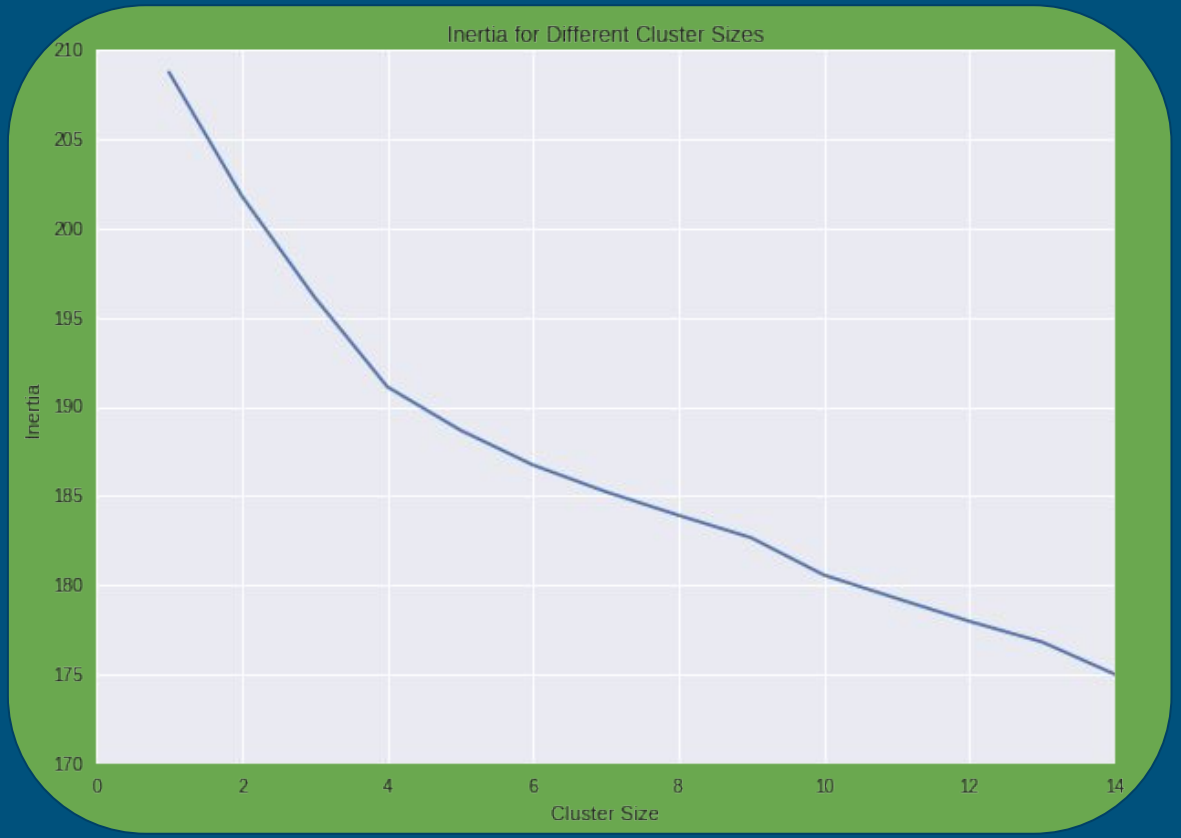
Clusters

K-Means with TF-IDF

- Cleaning
- Stop Words
- Document Frequency
- Max Features
- Number of Clusters

Inertia

Help to evaluate the number of clusters to use



Clusters

Google

Cluster 1: gold, rio, olympic, biles, jennings, walsh, beach, nbc, best, time

Cluster 2: new, year, one, company, people, would, percent, also, case, million

Cluster 3: police, fire, hickey, syria, olympic, people, wednesday, milwaukee, government, rio

Cluster 4: trump, campaign, clinton, bannon, manafort, conway, republican, ailes, breitbart, candidate

Cluster 5: north, korea, south, country, family, reported, officials, according, year, another

Analyzing Political Bias

Comparing Parties

DONALD TRUMP

Polarity: 0.093

Subjectivity: 0.348

Count: 646 mentions

NYT Adjectives: Angry, Volatile, Presidential, Incendiary

HILLARY CLINTON

Polarity: 0.067

Subjectivity: 0.416

Count: 306 mentions

NYT Adjectives: Partial, Revolutionary, Comfortable

Flask Demo

Thank You!