Webscraping the WSJ

Joseph van Bemmelen

Project Background

How do newspapers differ?

How often do WSJ articles come out?

What news areas does the WSJ focus on?

Can we tell which headline is from which newspaper? *(NY Post vs. WSJ)*

Is there a way to quantify the political leanings of different newspapers?

THE WALL STREET JOURNAL.

U.S. Edition ▾ | November 15, 2016 | Today's Paper

Y VB ▾ WSJ+

What's News →

Christie Ally Exits Trump Team as Carson Spurns Offer

Former House Intelligence panel Chairman Mike Rogers has left President-elect Trump's transition team amid a power consolidation that has pushed out key figures. Ben Carson declined an offer to become the next Health and Human Services secretary. 438

- **House GOP Nominates Ryan for Speaker**
- **House Democrats' Leadership Vote Delayed**
- **Trump Draws Criticism Over Bannon Pick**

Trump's Businesses Bring Potential Conflicts of Interest

Donald Trump will enter the White House with more potential conflicts of interest and less transparency about his finances than any recent president. 297

- **Firms Bet on Which Trump Will Govern**

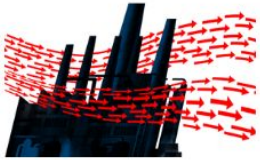
Wal-Mart Tells Workers: Don't

Forest Fires Spread by Drought Strike Southern Appalachians

A prolonged drought in the Southeast is sparking dozens of wildfires in the southern Appalachian Mountains, threatening homes and taxing firefighters as they evacuate people from the area.

Syrian Regime, Aleppo Rebels Gird for Battle

Syrian President Bashar al-Assad's forces and rebels in Aleppo are hunkered down and massing on either side of the five-mile-long front line dividing the city, preparing for intensified battles. The rebels are outmatched by the air power of Syria and Russia, but regime ground forces haven't



THE GREAT UNRAVELING

The Places That Made Donald Trump President

PROPERTY REPORT

Trouble Brewing in Commercial Real Estate

Markets

U.S. EUROPE ASIA FX RATES FUTURES

1D 5D 3M 6M 1Y

DJIA	18903.68	34.99	0.19%
S&P 500	2179.14	14.94	0.69%
Nasdaq	5283.02	64.62	1.24%
Russell 2000	1301.86	3.26	0.25%
DJ Total Mkt	22635.19	146.27	0.65%

Nov 15 '16, 3:27 PM EST MARKETS →

Opinion →

Europe's Trump Panic

Review & Outlook

A Columnist's Responsibility

By Bret Stephens | Global View

The Electoral College Is Anything But Outdated

By Larry P. Arnn | Commentary

What's News: Business & Finance

41 min ago

What's News: Business & Finance

41 min ago

What's News: World-Wide

41 min ago

What's News: World-Wide

41 min ago

High-School Students Stage Anti-Trump Protest in Washington

By Allison Kite 54 min ago

Winning numbers drawn in 'Take 5' game

55 min ago

Page 1 of 1259

A-HED

Candy Corn Lovers Will Eat Candy Corn Anything—No Matter What It Tastes Like

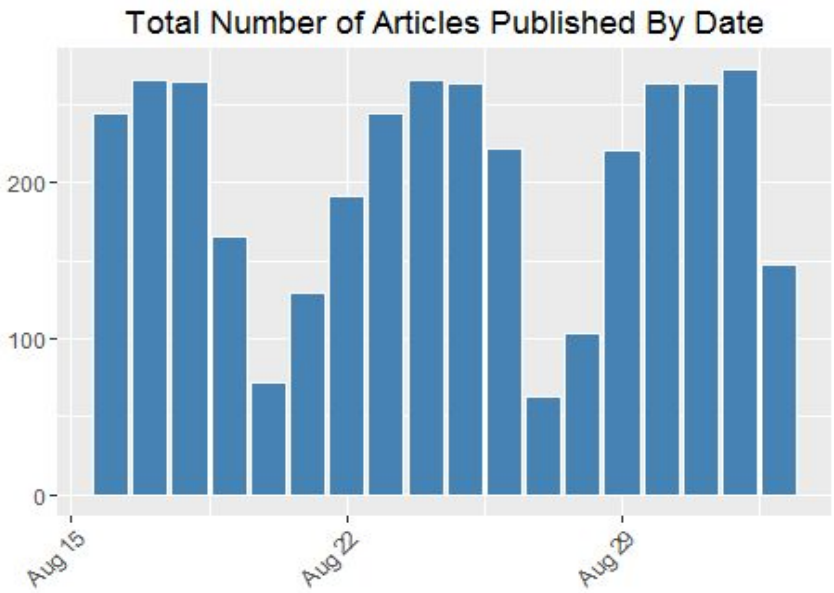
Confectioners often have no idea what candy corn should taste like; 'eating an antique candlestick'

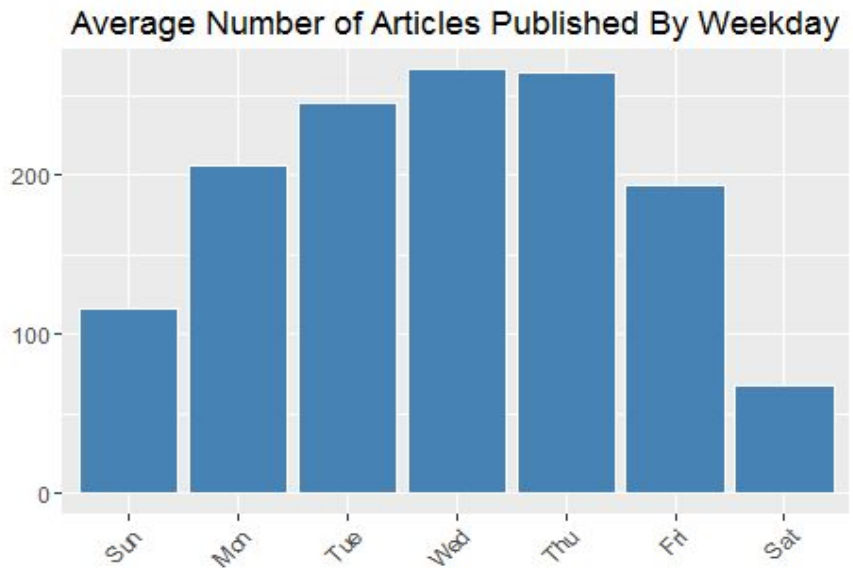
The sugary sweetness of candy corn is invading everyday treats like Oreos and M&M's. The Wall Street Journal taste-tested some of the latest candy corn offerings. Video/Photo: Rob Alcaraz/The Wall Street Journal

By ANNIE GASPARRO
Oct. 28, 2016 10:59 a.m. ET

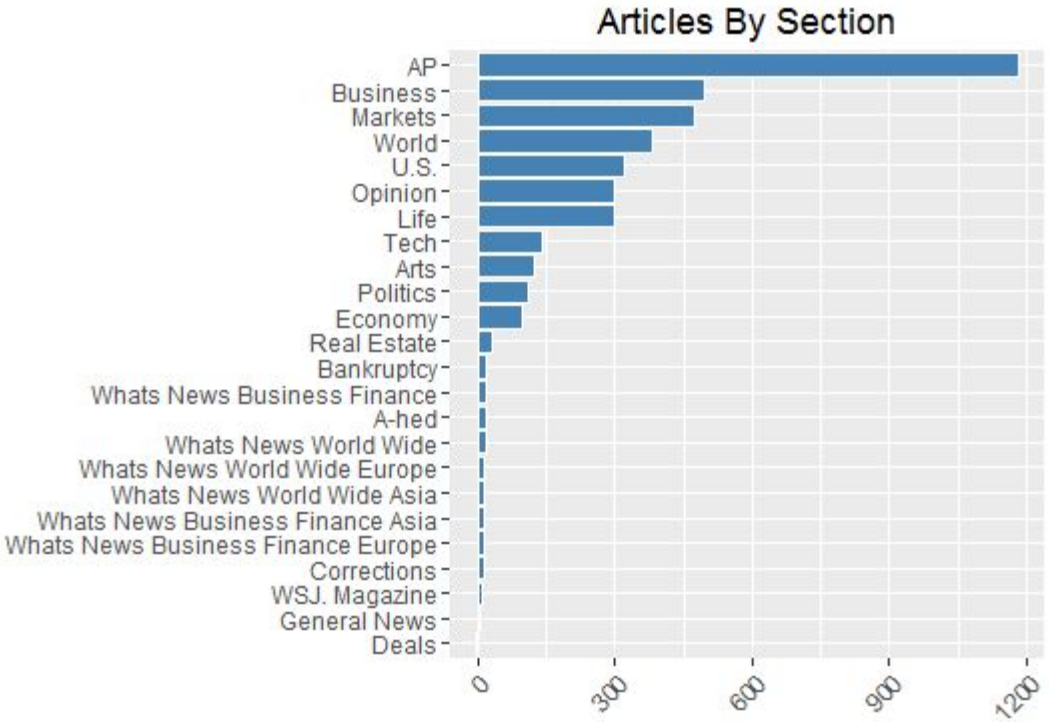
49 COMMENTS

Article Frequency

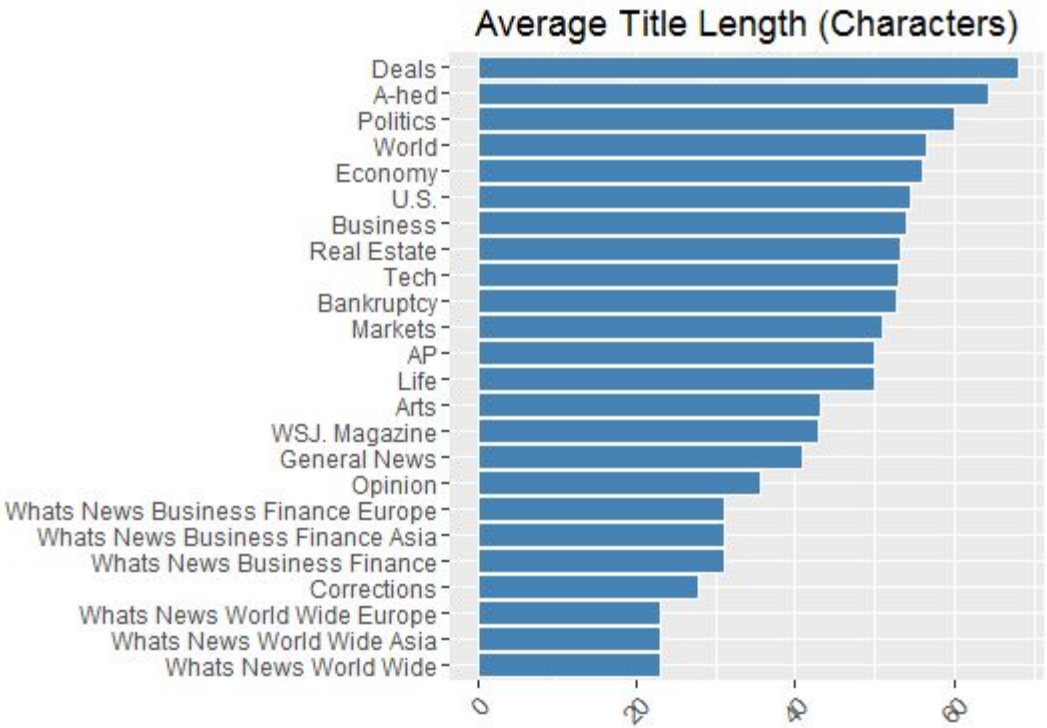




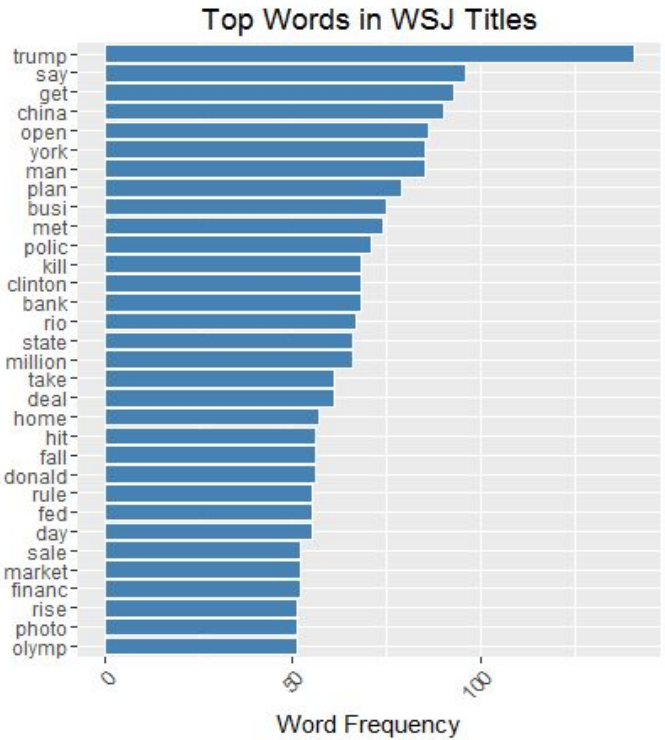
Areas of Focus

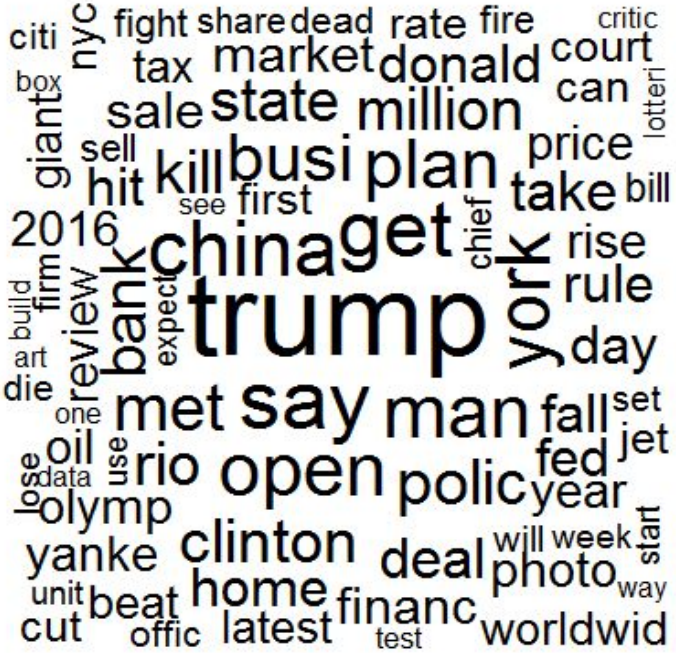


Length of Titles

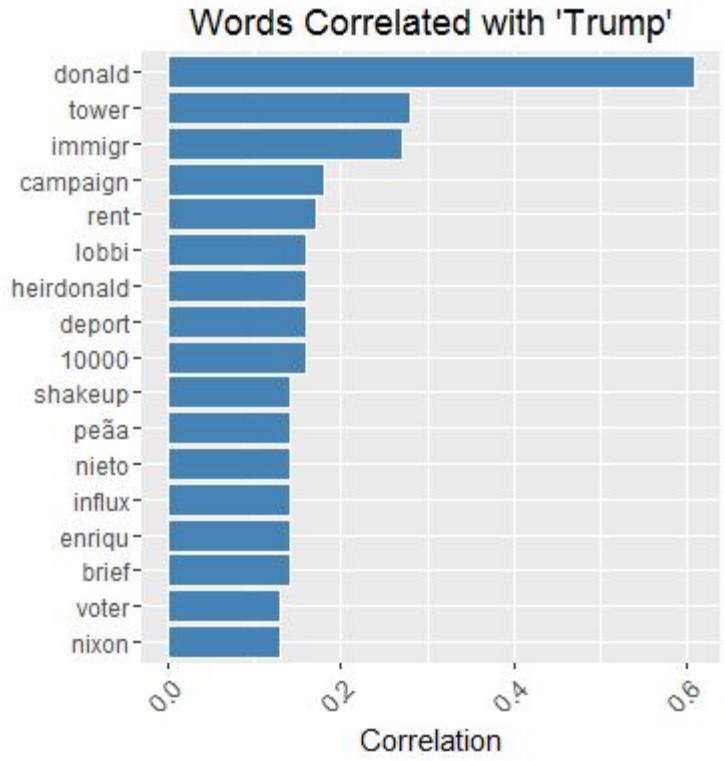


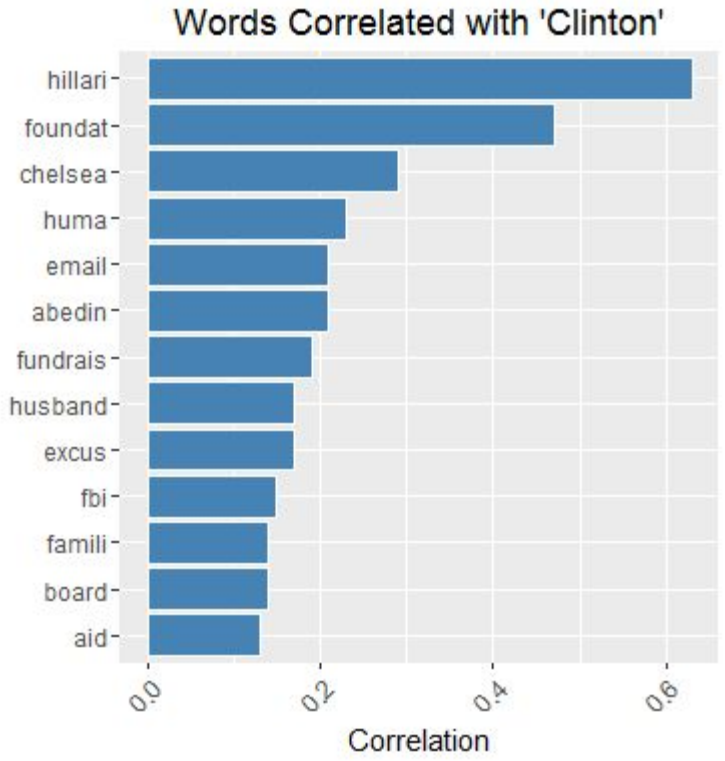
Top Words


Top Words in WSJ Titles
trump
say
get
china
open
york
man
plan
busi
met
polic
kill
clinton
bank
rio
state
million
take
deal
home
hit
fall
donald
rule
fed
day
sale
market
financ
rise
photo
olymp
0
50
100
Word Frequency

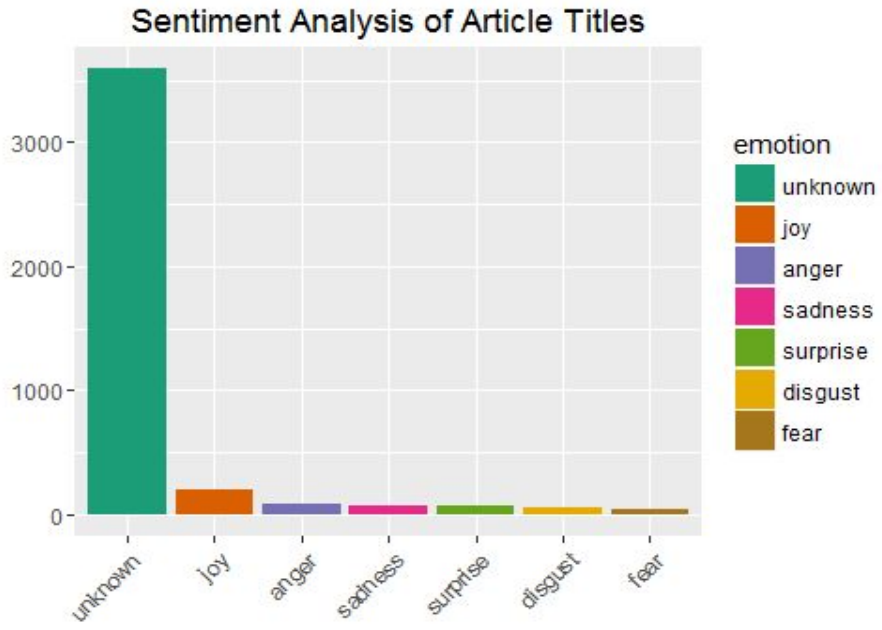

citi
nyc
fight
share
dead
rate
fire
critic
box
tax
market
donald
court
giant
sale
state
million
can
lotteri
sell
busi
plan
price
hit
kill
see
first
take
bill
2016
china
get
chief
rise
firm
bank
york
rule
build
review
expect
trump
art
day
die
one
met
say
man
fall
set
oil
use
rio
open
polic
fed
jet
lose
data
year
olymp
yanke
clinton
deal
will
week
start
photo
unit
beat
home
financ
way
cut
offic
latest
test
worldwid

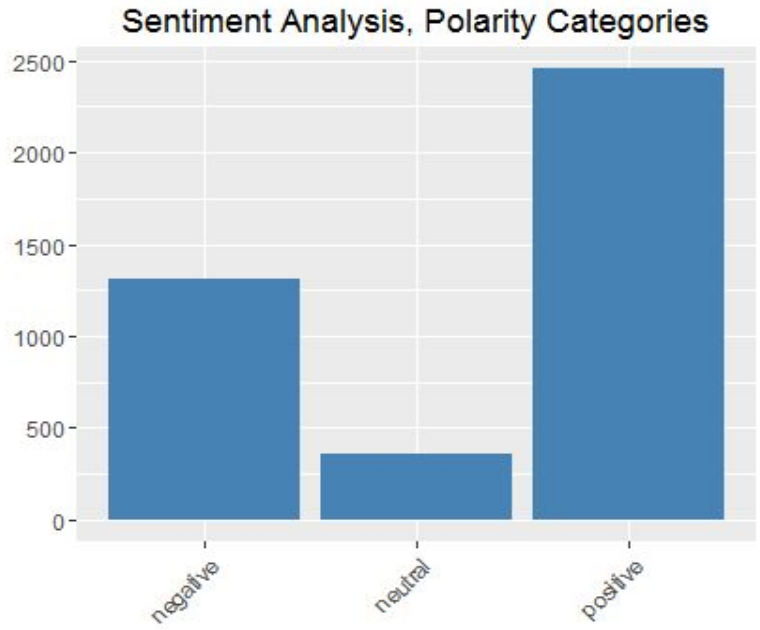
Word Correlation





Text Analysis





Future Steps

- Use selenium to scrape comment count to measure an article's popularity and the full text of the articles
- Download articles from the same timeframe from NY Post, NY Times, others
- Perform additional text analysis on the newspapers to compare what different newspapers focus on, quantify possible political leaning (especially editorials)
- See which topics have the most comments, most popular
- Develop an algorithm to determine which newspaper an article came from

Thank you!