An ode to permissionless data science

Hamdan Azhar

New York R Conference

July 14, 2023

@HamdanAzhar

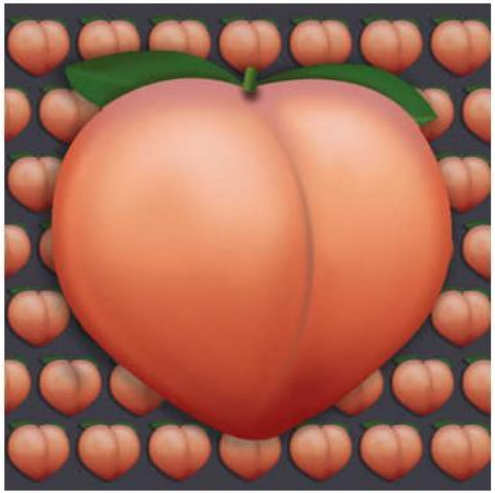


Very Official Study Finds Peach Emoji Most Often Paired With Eggplant

By **Madison Malone Kircher**

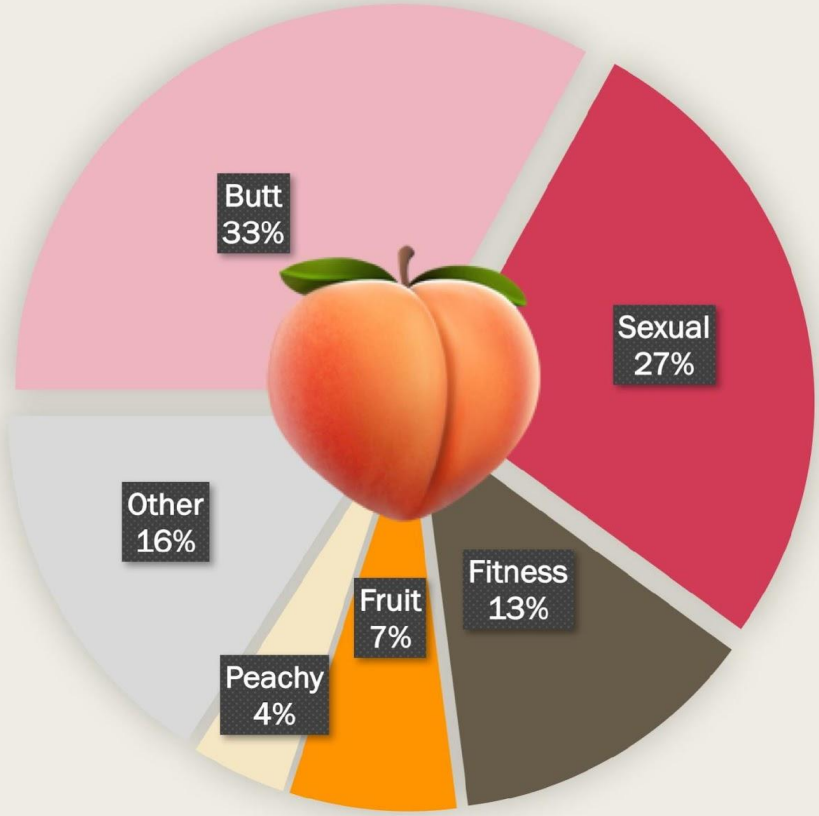
December 16, 2016
1:46 p.m.

Share
Tweet
Share
Share
Email
Comment
Print



The peach emoji has had a tumultuous few months. When Apple rolled out the beta version of iOS 10.2, everybody's favorite G-rated but still semi-explicit pictograph had been redrawn and looked significantly less like a butt. Following the outcries of "think of the sexts!" Apple relented and returned the emoji to its former gluteal glory. Now, the folks over at Emojipedia — the go-to website for everything you ever wanted to know about emoji — have crunched the numbers to find out exactly how many people online are actually using the peach as a substitute for a bum and how many are actually just talking about their favorite Roald Dahl story.

How is the 🍑 emoji used on Twitter?



Source: Analysis of sampled English language tweets mentioning 🍑 (n=100) from Dec. 11, 2016 (Emojipedia)

Most common emojis in tweets with 🍑



Source: Analysis of sampled English language tweets mentioning 🍑 (n=571) from Dec. 11, 2016 (Emojipedia)

MOTHERBOARD

KANYE WEST

A Data Scientist's Emoji Guide to Kanye West and Taylor Swift



MADISON MARGOLIN
Aug 10 2016, 2:32pm





Frequency (Per 1,000 Tweets)

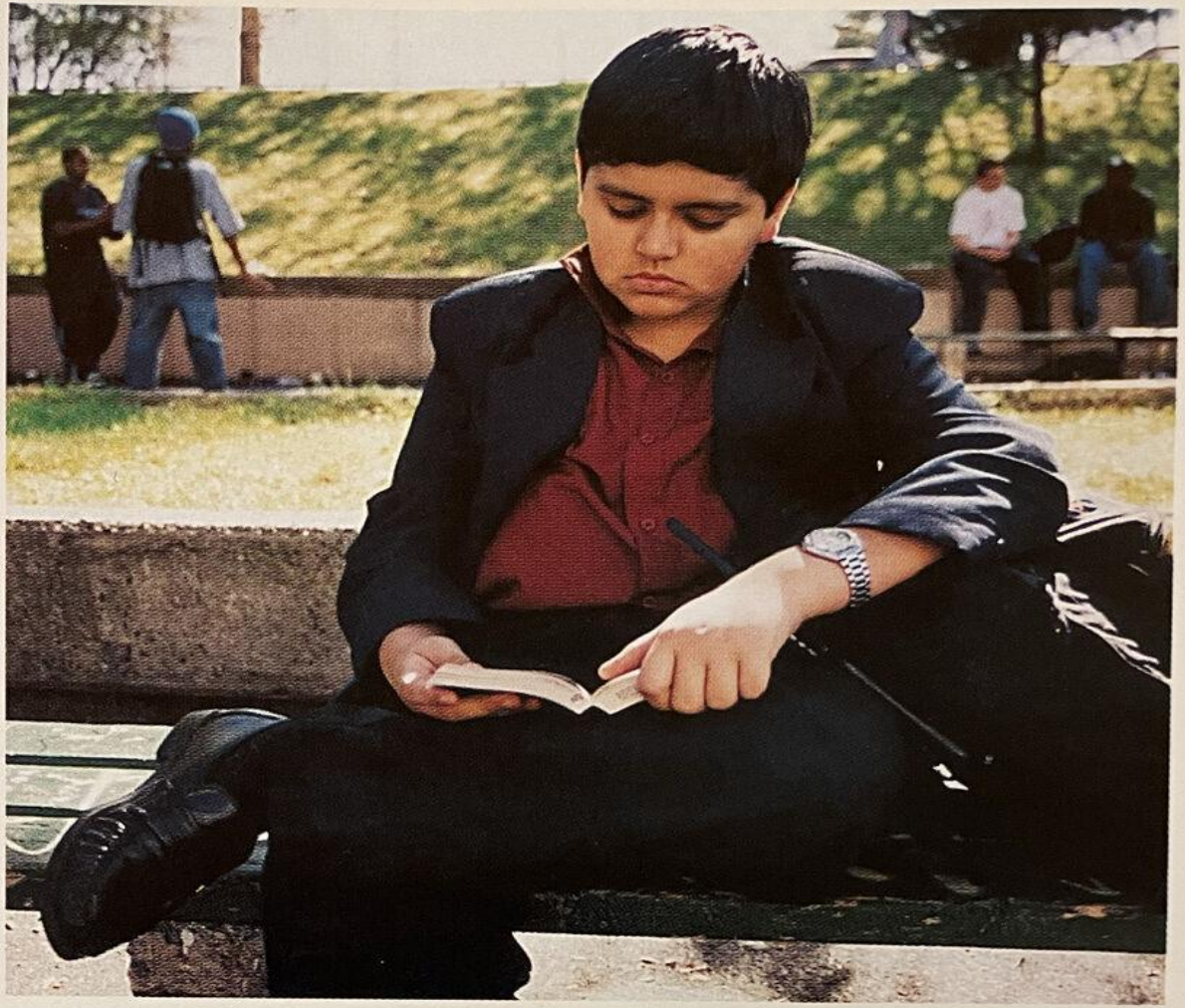
100
90
80
70
60
50
40
30
20
10
0

1 2 3 4 5

1 2 3 4 5

Rank





Data science is quickly becoming just a set of tools – or a set of tools deployed against a handful of very specific problem spaces

“We have become enmeshed in a net of means and have lost sight of the ends…We have the most wonderful instruments and means man has ever had, but we do not stop and ask what they are for.”
- Erich Fromm, 1947

“The best minds of my generation are thinking about how to make people click ads.”
- Jeff Hammberbacher, 2011

The state of data science today

TECHNICAL KNOWLEDGE

Programming Languages:	Python, R, SQL, PySpark, Scala
Specialties:	Regression Analysis, Clustering Analysis, Segmentation Analysis, Decision Trees, Deep Learning, ETL Pipelines, Business Intelligence, AWS, Big Data, Spark
IDEs and Tools:	Tableau, Alteryx, R Studio, Power BI, MS PowerPoint, MS Excel, Git, Jupyter
Data Mining Libraries:	pandas, numpy, scikit-learn, nltk, scipy, matplotlib, seaborn, plotly, ggplot2
Certification:	IBM Data Science Professional, IBM Applied Data Science, Udemy Pyspark, AWS Certified Solutions Architect Associate (P), Tableau Desktop Specialist (P)

TECHNICAL SKILLS

Programming: R, Python, SQL
Data management: SQL Server, Oracle SQL, ETL - Oracle Data Integrator
Data Visualization: Tableau, OBIEE 11g and 12c, Qlik Sense, Cognos, Business Objects
Scheduling: Qlik Nprinting, Cognos Publisher, Oracle BI Publisher
Packages: NumPy, Pandas, matplotlib, SciPy, Scikit-Learn
Tools: Microsoft Office, JIRA, HP Quality Center, Confluence, R Studio, Jupyter, Spyder
Predictive Models: Regressions, Classification Models, Time-Series Analysis and Forecasting Techniques
Machine Learning: Univariate Analysis, Multivariate Analysis, Statistical Modelling, Hypothesis Testing
Expertise: Business Data Modelling, Data Extraction and Cleaning, Excel Reporting, Agile, Waterfall, SDLC

TECHNOLOGY AND SOFTWARE SKILLS

Languages: SQL, PLSQL, PostgreSQL, Python (Seaborn, Bokeh, Plotly, Matplotlib, NumPy, PySpark, PyTorch, TensorFlow, Scikit Learn, BeautifulSoup), JSON, Javascript, R (R- Shinny, Keras, Caret, e1071, RWeka)
BI Tools: Tableau, R, Weka, Minitab, Matlab, PowerBI, Datastage (ETL), Excel (StatTool, Power Queries, Pivot, vlookup), Snowflake, Databricks, Jupyter
Statistics: Time series forecasting Bayes theorem, Hypothesis testing, A/B testing, causation, collinearity, probability, regression
Big Data: NoSQL, MongoDB, OLAP, OLTP, Amazon Web Services (AWS) S3, Hadoop, HDFS, Hive, Spark, Map Reduce, Cassandra, AWS

The original promise of data science

Harvard Business Review

Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and D.J. Patil

From the Magazine (October 2012)

The original promise of data science

“I keep saying the sexy job in the next ten years will be statisticians. People think I’m joking, but who would’ve guessed that computer engineers would’ve been the sexy job of the 1990s?

The ability to take data – to be able to **understand it**, to **process it**, to **extract value from it**, to **visualize it**, to **communicate it** is going to be a hugely important skill in the next decades, not only at the professional level but even at the educational level for elementary school kids, for high school kids, for college kids. Because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it.

I think statisticians are part of it, but it’s just a part. You also want to be able to visualize the data, communicate the data, and utilize it effectively. But I do think those skills – of being able to access, understand, and communicate the insights you get from data analysis – are going to be extremely important. Managers need to be able to access and understand the data themselves.”

- Hal Varian, *McKinsey Quarterly*, January 2009

“Soon, data sets—large, unseemly, raw collections of numbers—began to tell me their stories. Whether these stories concerned demographics, human physiology, or the economy, they shed light on our understanding of the world. I saw firsthand how the language of statistics worked, and the immense power it held.”

- Hamdan Azhar (2008)

Most of data science is deeply human

1. Find a problem to solve
2. Get the data necessary to answer this question
3. Clean the data
4. Analyze the data / build models
5. Tell a compelling story
6. Use the story to drive impact

Human Skills

Technical Skills

Human Skills

** Here, we focus on data science for understanding. There's also data science for prediction, which is gradually becoming more engineering than science and can afford to be more purely technical.*

Stop waiting for someone to “give you a chance” to be a data scientist

The best data science — and the best innovation — has always been permissionless



Hamdan Azhar

Why is this the best time to be a data scientist?

- Data is all around us
- The tools to understand this data are widely available, open source, well documented, and the community is fantastic!

So where do you start?

Start with your passions! The skills you have as a data scientist are not just for you to be great at your job, they can help enable you to pursue your curiosity and answer meaningful questions you care about, and that the world cares about!

PennState

UNIVERSITY OF MICHIGAN

THE UNIVERSITY OF CHICAGO

RON PAUL 2012 RESTORE AMERICA NOW

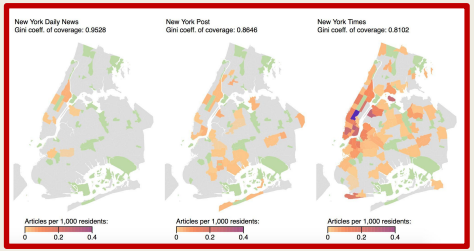
GraphScience

facebook

PRISMOJI

BLOCKCHAIN TECHNOLOGIES CORP

BlackRock®

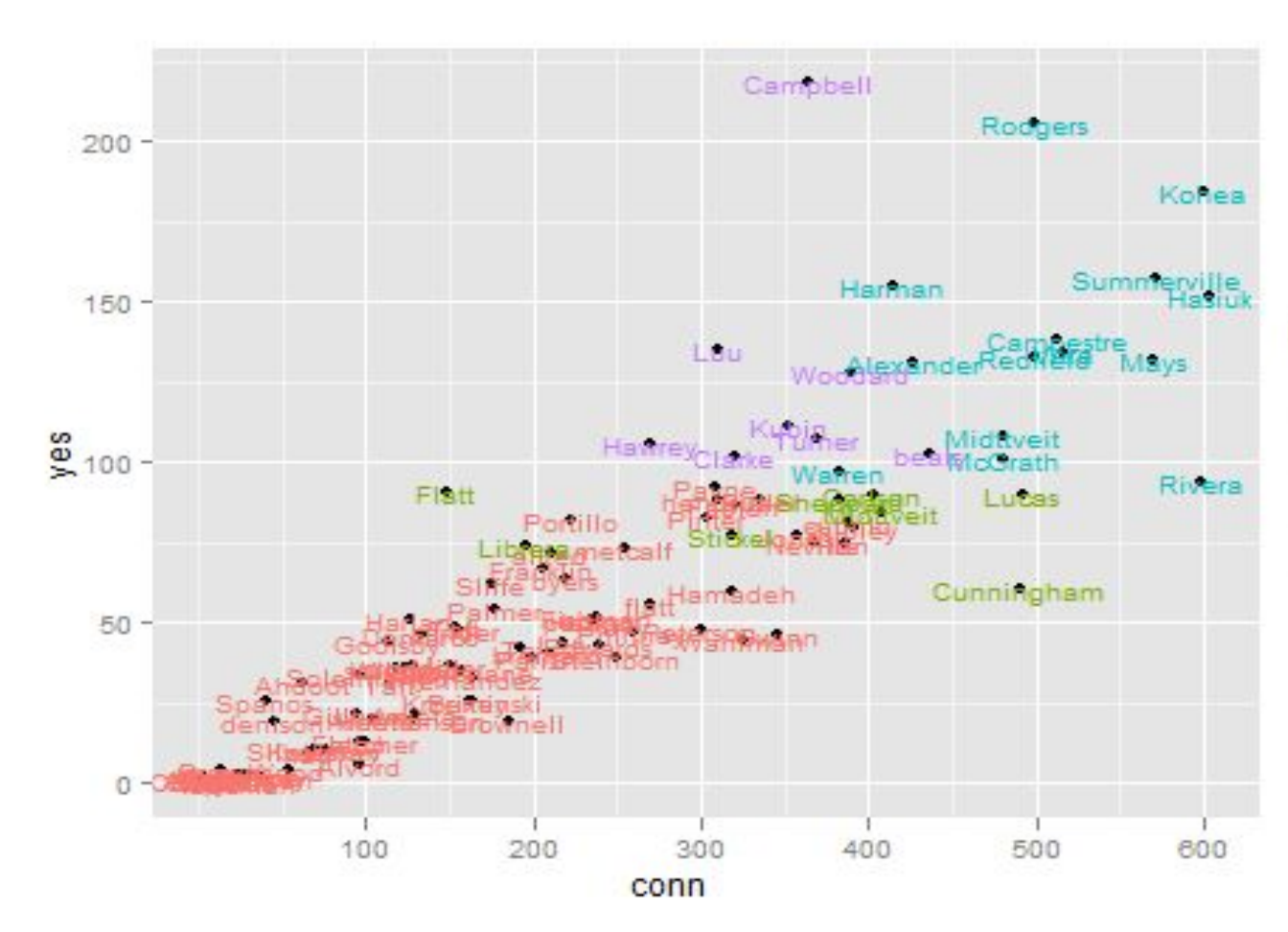





RON PAUL
2012
RESTORE AMERICA NOW
LIBERTY
RON PAUL

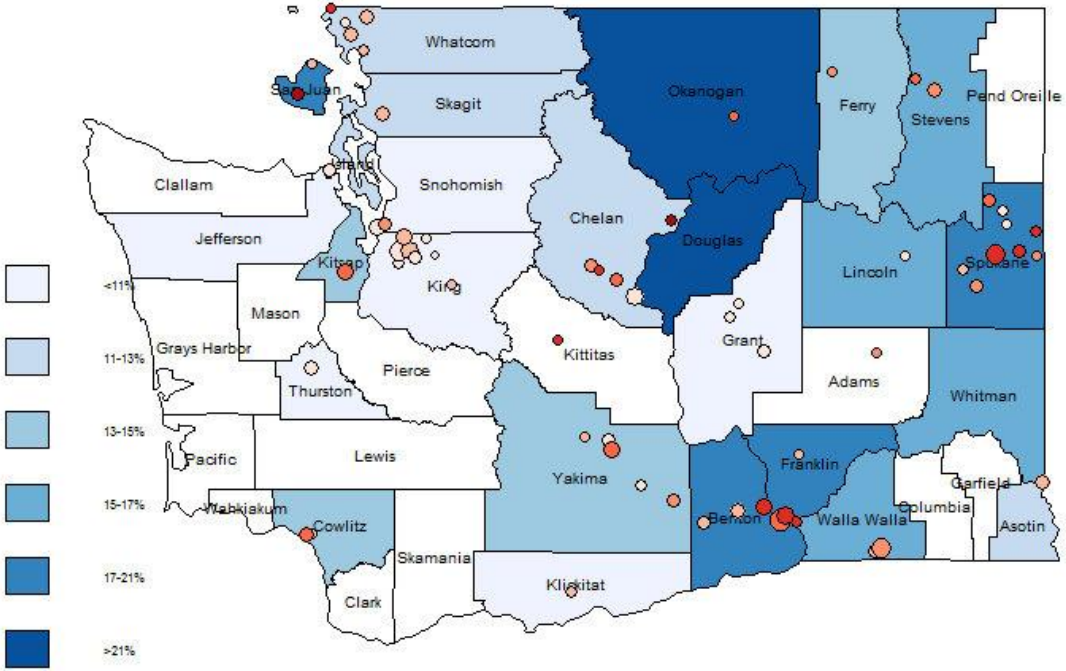

MON Jan 9
710pm
Statistics Central
Grand Total 23,160
Contact Total 12,486
Connect 3,249
Success 2,918
Success % (out of contacts)
70% 92%
Persuasion: total. 10,679
Connect: 1,534
Success 309
20% 30%
RESTRICTED
RON PAUL
2012
RESTORE AMERICA NOW


calling

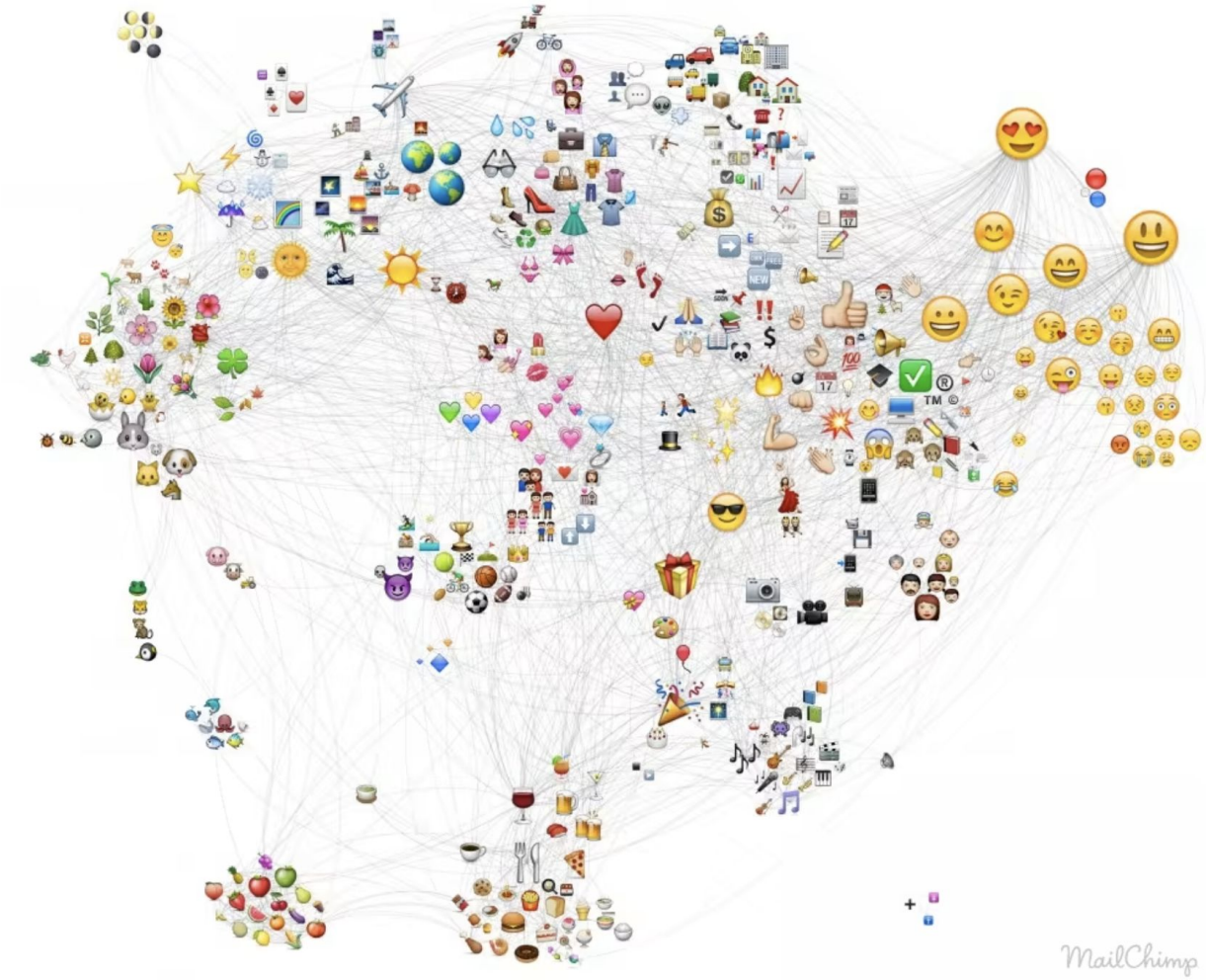

Campbell
Rodgers
Korea
Harman
Sumnerville
Hasiuk
Camestre
Redmond
Mays
Lou
Alexander
Woodard
Midkiff
McGrath
Rivera
Lucas
Warren
Flatt
Portillo
Hamedeh
Cunningham
Clarke
Turner
Brownell
yes
conn
0
50
100
150
200
100
200
300
400
500
600

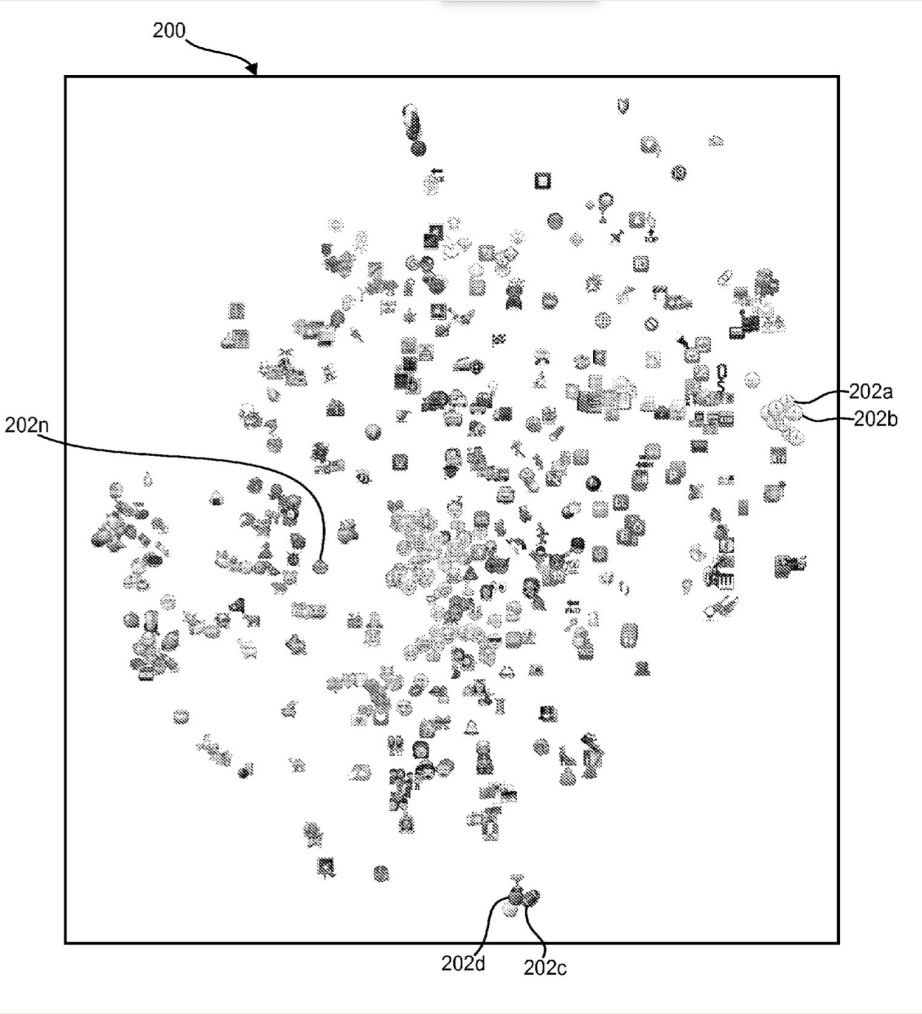

ONPAUL2012.com
COLLETT

2012: RP Support rate, ID data



12,547 answered surveys, Feb. 28, 2012, F


MailChimp


200
202a
202b
202n
202d
202c

US010546015B2

(12) **United States Patent**
Dimson et al.

(10) **Patent No.: US 10,546,015 B2**
(45) **Date of Patent: Jan. 28, 2020**

(54) **DETERMINING AND UTILIZING CONTEXTUAL MEANING OF DIGITAL STANDARDIZED IMAGE CHARACTERS**

(71) Applicant: **Facebook, Inc.**, Menlo Park, CA (US)

(72) Inventors: **Thomas Frederick Dimson**, Stanford, CA (US); **Bogdan State**, Menlo Park, CA (US); **Shankar Kalyanaraman**, Menlo Park, CA (US); **Hamdan Azhar**, Menlo Park, CA (US)

(73) Assignee: **Facebook, Inc.**, Menlo Park, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 1092 days.

(21) Appl. No.: **14/956,148**

(22) Filed: **Dec. 1, 2015**

(65) **Prior Publication Data**

US 2017/0154055 A1 Jun. 1, 2017

(51) **Int. Cl.**
G06F 16/583 (2019.01)

(52) **U.S. Cl.**
CPC ***G06F 16/5838*** (2019.01)

(58) **Field of Classification Search**
CPC G06F 16/5846; G06F 16/5838; G06F 16/5854; G06F 16/5866; G06F 16/9535; G06F 16/951; G06F 16/90344; G06F 16/2425; G06F 16/583
USPC 707/769, 772
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

8,442,940	B1 *	5/2013	Faletti	G06F 17/2785 707/610
2009/0028445	A1 *	1/2009	Wu	G06K 9/723 382/225
2009/0285492	A1 *	11/2009	Ramanujapuram	G06K 9/4671 382/209
2010/0185978	A1 *	7/2010	Laurie	G06F 9/451 715/802
2016/0224684	A1 *	8/2016	Corlett	G06F 3/04817

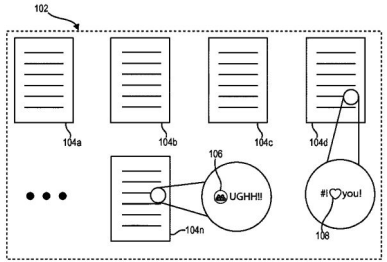
* cited by examiner

Primary Examiner — James Trujillo
Assistant Examiner — Fariborz Khoshnoodi
(74) *Attorney, Agent, or Firm* — Keller Jolley Preece

(57) **ABSTRACT**

One or more embodiments of the disclosure include a character management system that determines and utilizes contextual meaning with regard to standardized image characters. In particular, the character management system determines contextual meaning with regard to standardized characters based on the manner in which the standardized characters are utilized in a repository of electronic documents. Moreover, the character management system can utilize the contextual meaning to search, recommend, analyze, and translate with regard to the use of standardized image characters.

20 Claims, 14 Drawing Sheets



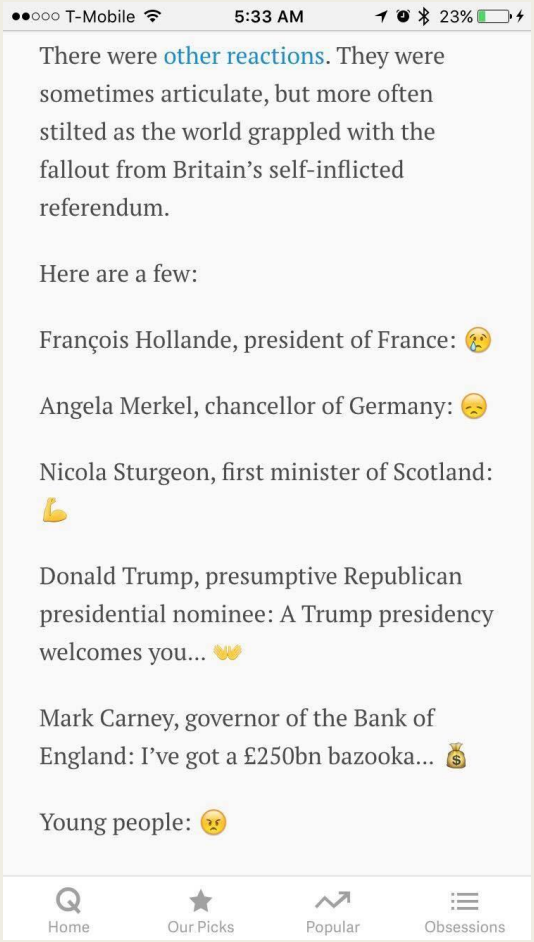


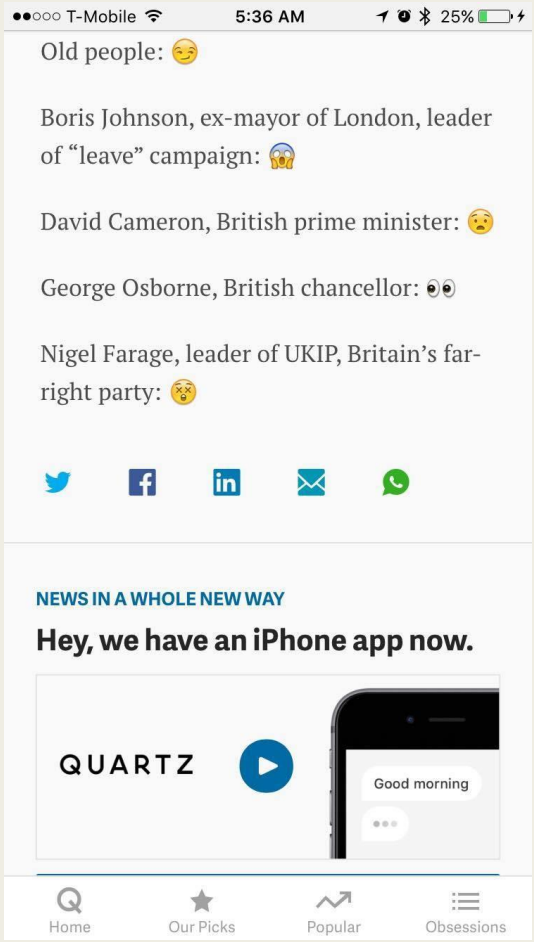
The world's reaction to Brexit, in emoji

What people really mean, decoded.

qz.com

Like Comment Share





My reaction to this article, in emoji



MOTHERBOARD

UK

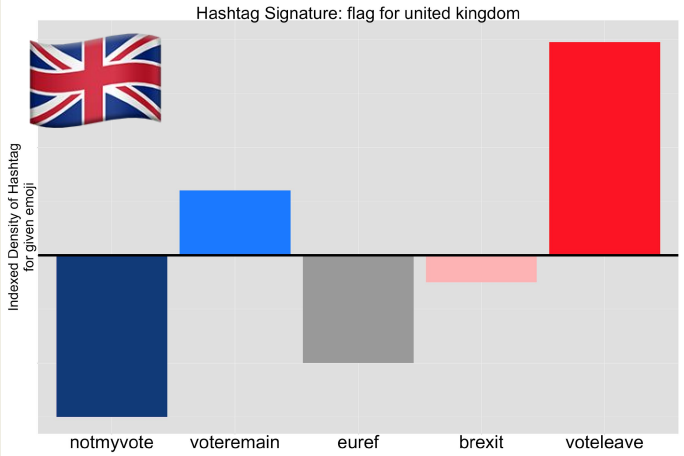
Here Are the Most Popular Emojis From the #Brexit Reaction

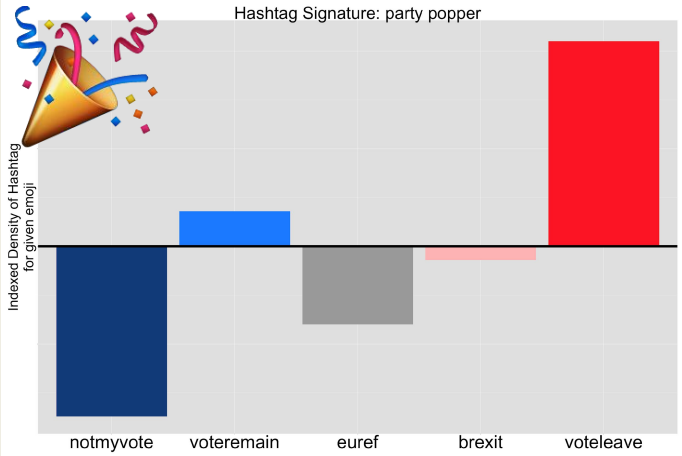
HA

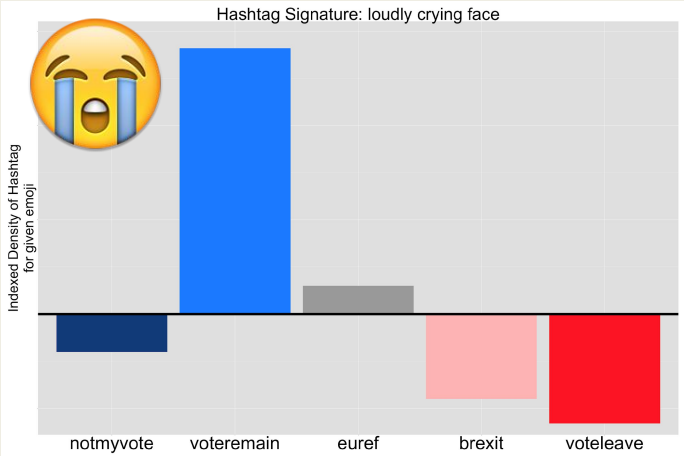
HAMDAN AZHAR

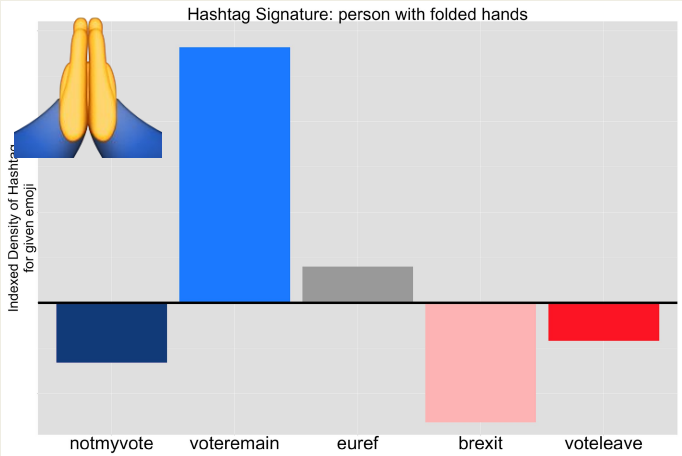
Jun 30 2016, 9:34am

emoji	emoji name	brexit rank	general rank	brexit index*	general index*	overindex**
😂	face with tears of joy	1	1	100	100	
🇬🇧	flag of united kingdom	2	363	87	0.2	400x
👍	thumbs up sign	3	18	26	11	2.3x
👏	clapping hands sign	4	45	24	6	3.9x
❤️	heavy black heart	5	3	21	45	
😭	loudly crying face	6	7	17	29	
😔	pensive face	7	13	14	18	
😩	weary face	8	11	13	22	
😢	crying face	9	27	12	9	1.3x
🙈	see-no-evil monkey	10	24	12	9	1.3x


Hashtag Signature: flag for united kingdom
Indexed Density of Hashtag for given emoji
notmyvote
voteremain
euref
brexit
voteleave


Hashtag Signature: party popper
Indexed Density of Hashtag for given emoji
notmyvote
voteremain
euref
brexit
voteleave


Hashtag Signature: loudly crying face
Indexed Density of Hashtag for given emoji
notmyvote
voteremain
euref
brexit
voteleave


Hashtag Signature: person with folded hands
Indexed Density of Hashtag for given emoji
notmyvote
voteremain
euref
brexit
voteleave

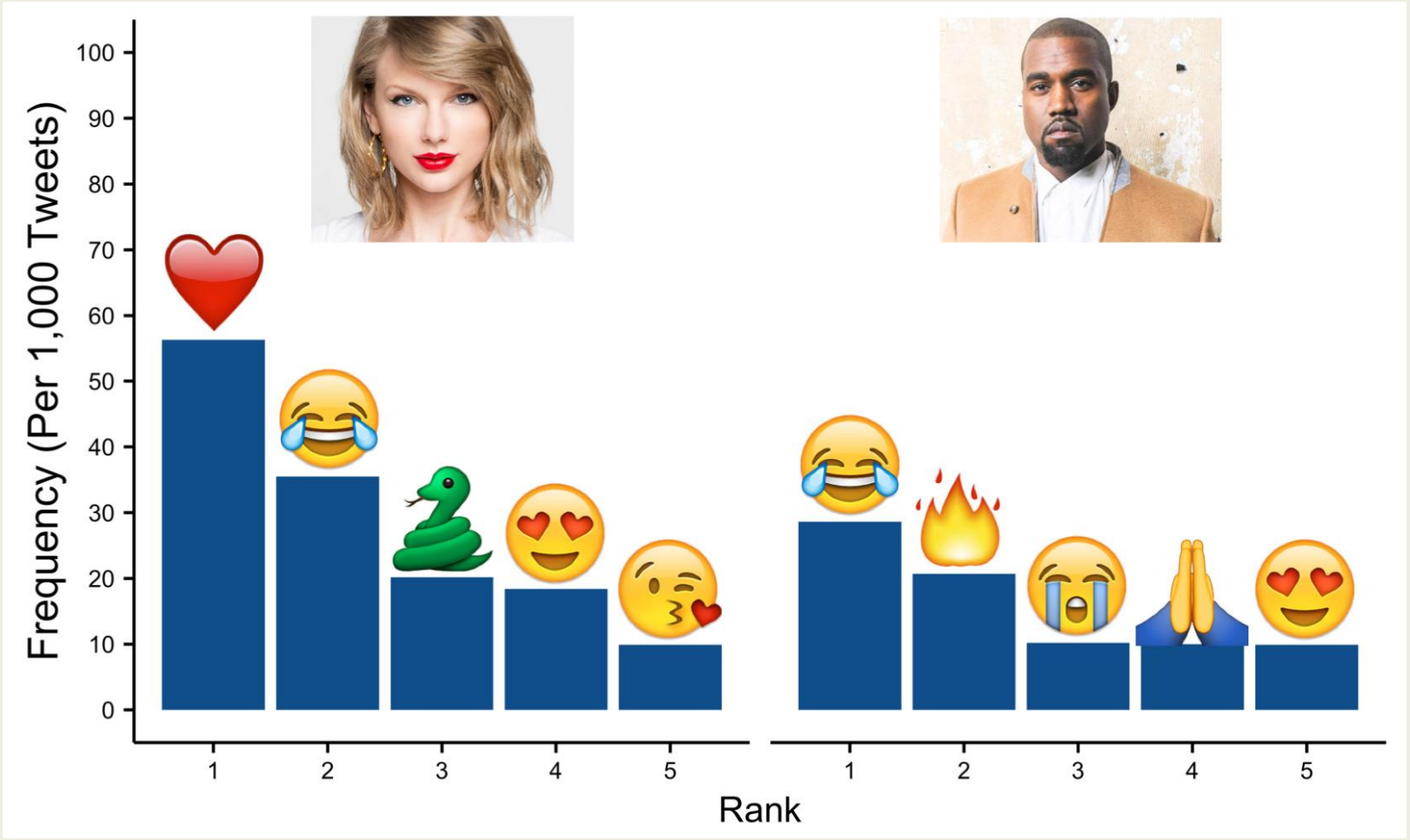
MOTHERBOARD

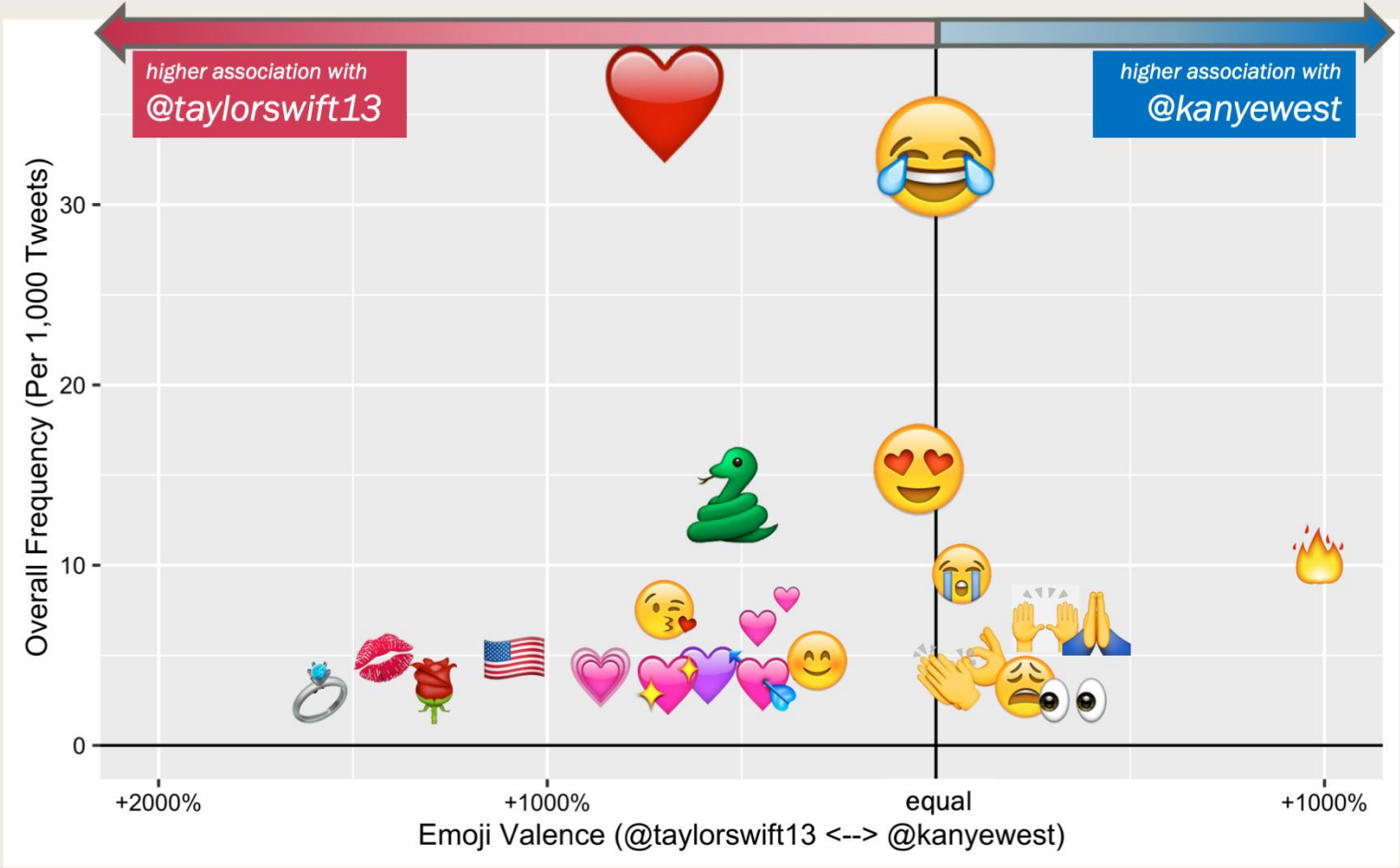
KANYE WEST

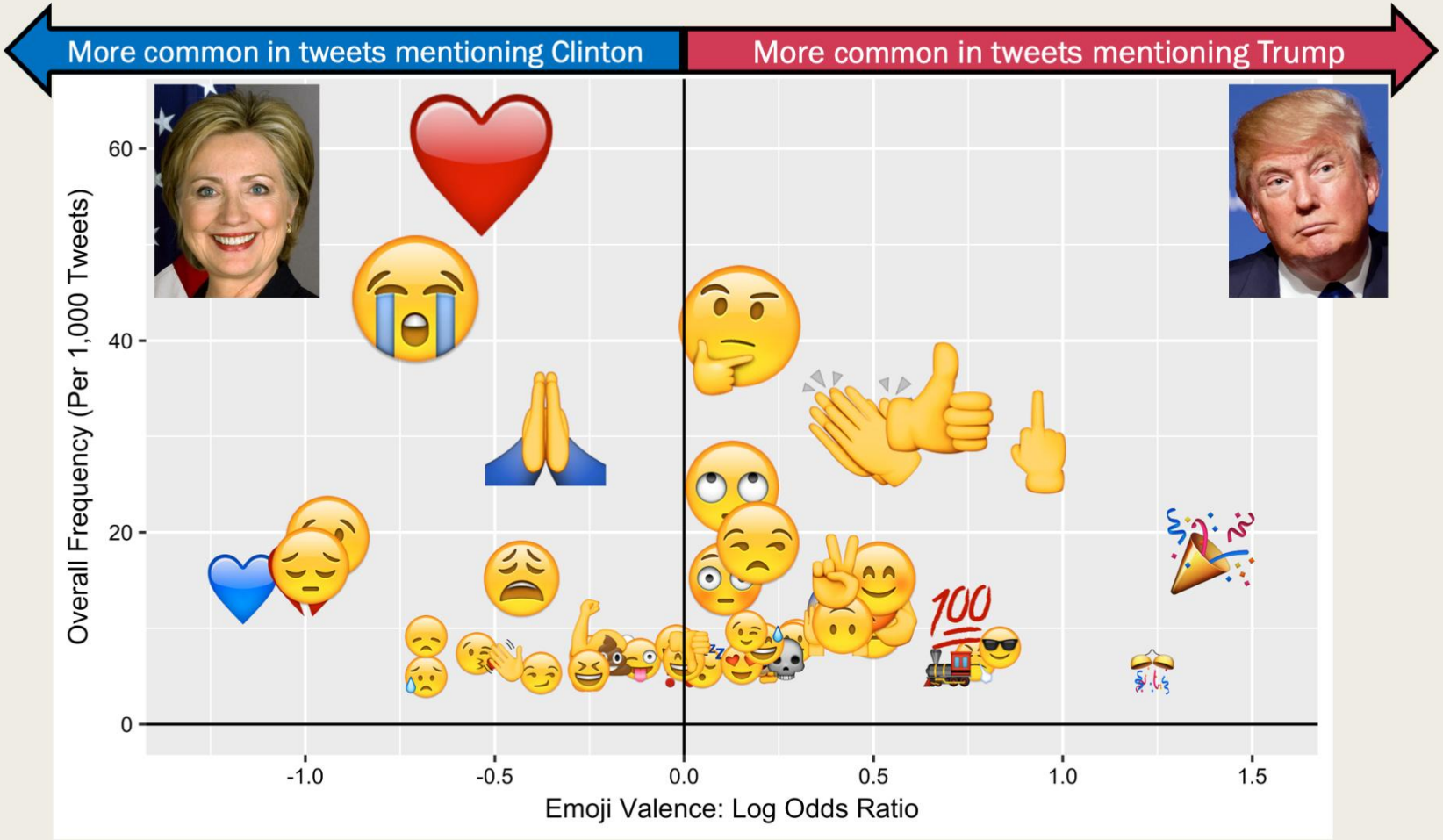
A Data Scientist's Emoji Guide to Kanye West and Taylor Swift

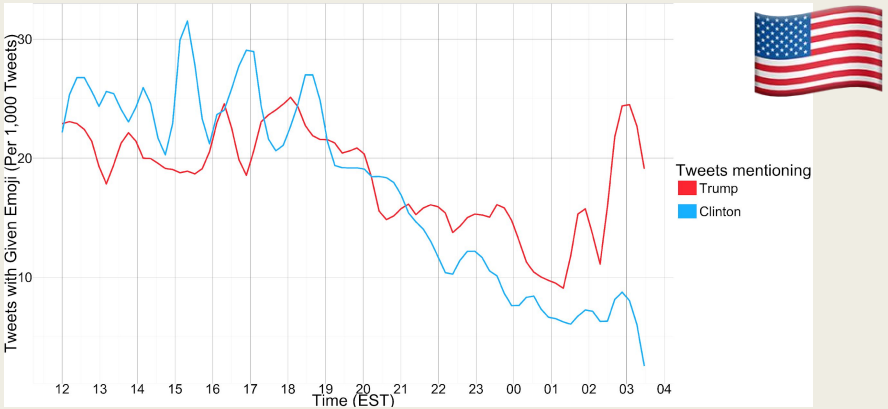


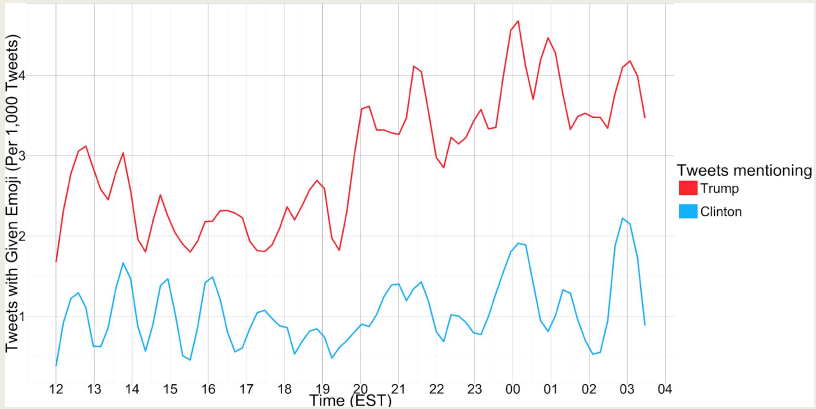
MADISON MARGOLIN
Aug 10 2016, 2:32pm

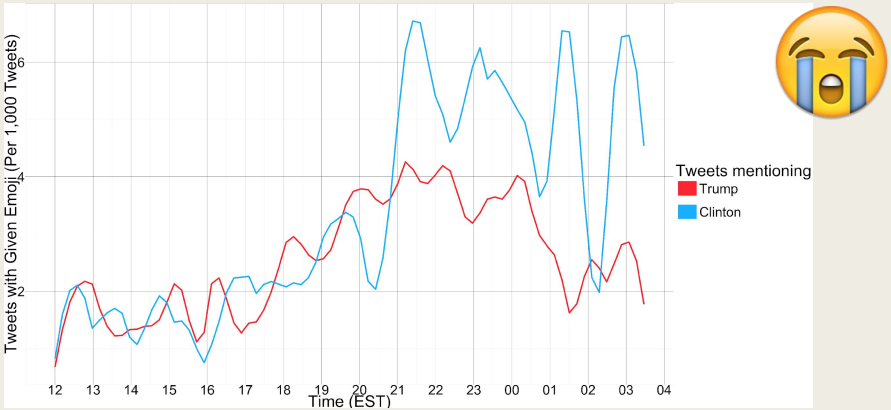

Frequency (Per 1,000 Tweets)
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
1
2
3
4
5
Rank

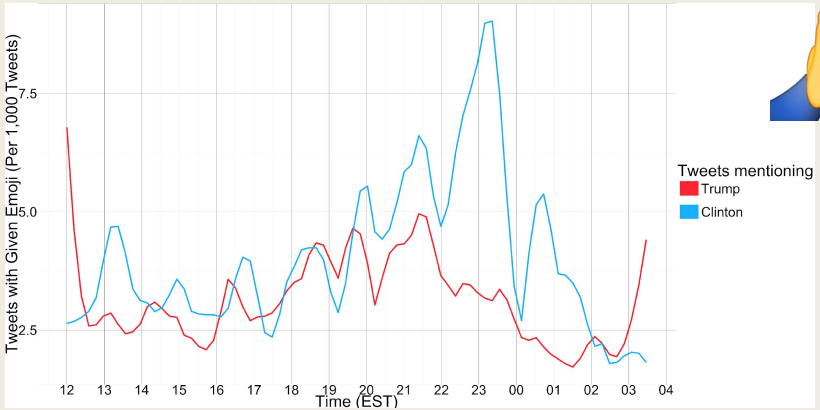

higher association with
@taylorswift13
higher association with
@kanyewest
Overall Frequency (Per 1,000 Tweets)
30
20
10
0
+2000%
+1000%
equal
+1000%
Emoji Valence (@taylorswift13 <--> @kanyewest)


More common in tweets mentioning Clinton
More common in tweets mentioning Trump
Overall Frequency (Per 1,000 Tweets)
60
40
20
0
-1.0
-0.5
0.0
0.5
1.0
1.5
Emoji Valence: Log Odds Ratio


Tweets with Given Emoji (Per 1,000 Tweets)
30
20
10
12 13 14 15 16 17 18 19 20 21 22 23 00 01 02 03 04
Time (EST)
Tweets mentioning
Trump
Clinton

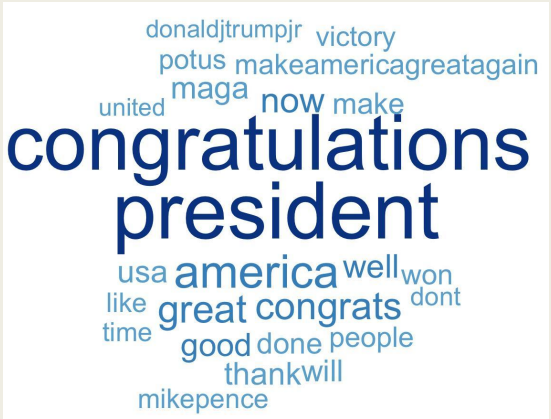

Tweets with Given Emoji (Per 1,000 Tweets)
4
3
2
1
12 13 14 15 16 17 18 19 20 21 22 23 00 01 02 03 04
Time (EST)
Tweets mentioning
Trump
Clinton


Tweets with Given Emoji (Per 1,000 Tweets)
6
4
2
12 13 14 15 16 17 18 19 20 21 22 23 00 01 02 03 04
Time (EST)
Tweets mentioning
Trump
Clinton


Tweets with Given Emoji (Per 1,000 Tweets)
7.5
5.0
2.5
12 13 14 15 16 17 18 19 20 21 22 23 00 01 02 03 04
Time (EST)
Tweets mentioning
Trump
Clinton

The divided emoji states of America





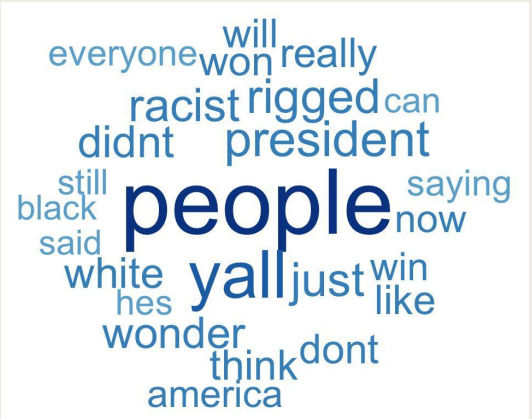
Yeah wow @realDonaldTrump you real worked hard AWESOME JOB!! THANK YOU !! 👌👍 Please tell Obama ....YOUR FIRED..!!👈





I'm not scared let's just get these four years over with. Fuck trump, fuck the people who voted him and the people who didn't vote at all 🖕🏽

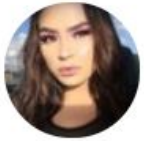




How can I go to work in the morning knowing the people I work with voted for trump. 🤔 I just can't.

Source: Analysis of sampled Election Night tweets, from Nov. 9, 2016 12:00am – 3:30am EST. Word clouds shown for tweets mentioning Hillary Clinton or Donald Trump as well as the given emoji above. Randomly selected tweet shown for each emoji.

@HamdanAzhar



Val
@itsvaleriee___

Follow

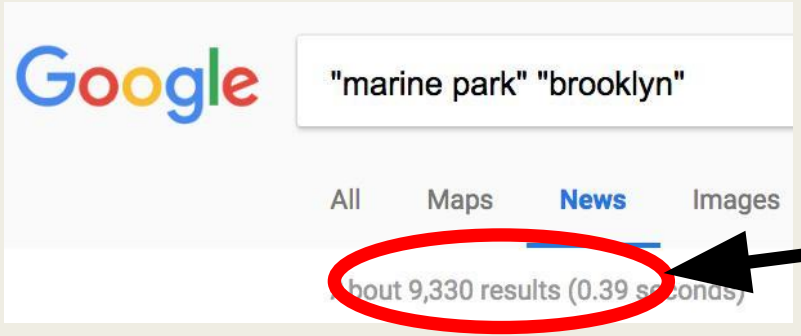
How can I go to work in the morning knowing the people I work with voted for trump. 🤔 I just can't.

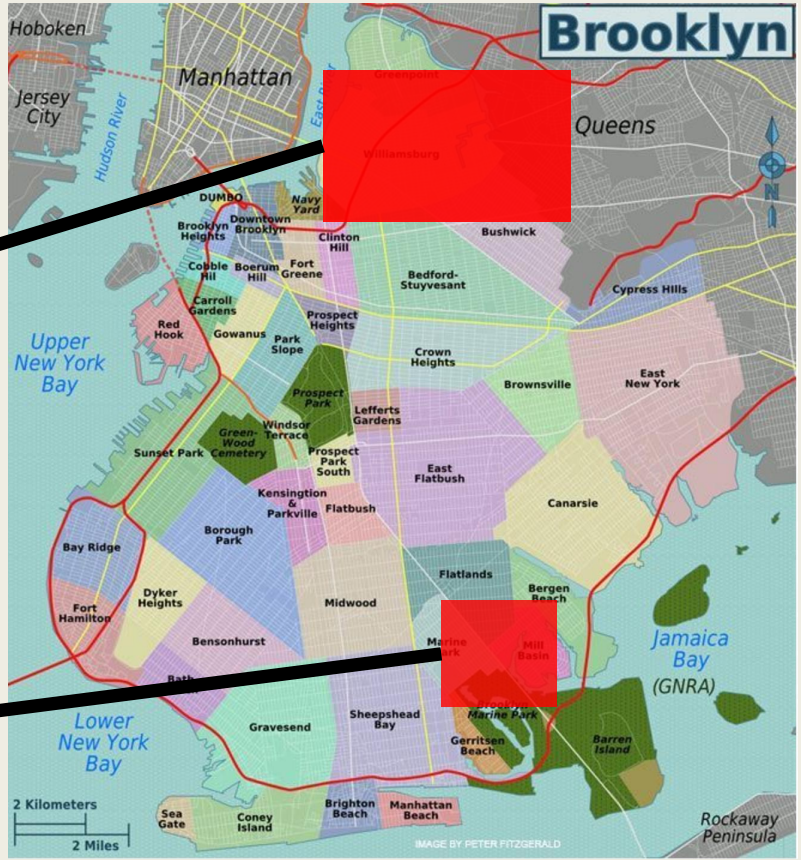
1:41 AM - 9 Nov 2016



MISINFOCON
МИСИНФОКОН
Trust, Verification, Fact Checking & Beyond


Google
"williamsburg" "brooklyn"
All
News
Maps
Images
bout 114,000 results (0.32 econds)


Google
"marine park" "brooklyn"
All
Maps
News
Images
bout 9,330 results (0.39 seconds)


Brooklyn
Hoboken
Manhattan
Jersey City
Hudson River
Queens
Williamsburg
Greenpoint
DUMBO
Navy Yard
Brooklyn Heights
Downtown Brooklyn
Clinton Hill
Bushwick
Cobble Hil
Boerum Hill
Fort Greene
Bedford-Stuyvesant
Cypress Hills
Carroll Gardens
Red Hook
Gowanus
Park Slope
Prospect Heights
Upper New York Bay
Crown Heights
Brownsville
East New York
Prospect Park
Lefferts Gardens
Windsor Terrace
Green-Wood Cemetery
Sunset Park
Prospect Park South
East Flatbush
Kensington & Parkville
Flatbush
Canarsie
Borough Park
Bay Ridge
Flatlands
Dyker Heights
Fort Hamilton
Midwood
Bergen Beach
Bensonhurst
Jamaica Bay (GNRA)
Mill Basin
Marine Park
Lower New York Bay
Gravesend
Sheepshead Bay
Gerritsen Beach
Barren Island
2 Kilometers
2 Miles
Sea Gate
Coney Island
Brighton Beach
Manhattan Beach
Rockaway Peninsula
IMAGE BY PETER FITZGERALD

Williamsburg



Source: Analysis of newspaper headlines from the Gothamist (2015-2017)

Our latest installment of Quick Bites brings us to Greenpoint for boiled peanuts, smoked chicken, and pinball. THE VIBE Opening a dive bar as a dive bar is a tricky undertaking, with too many newcomers erring too...

Williamsburg Deli Hides A 'Secret' Agave Bar Behind A Freezer Door



La Superior owner Felipe Mendez expands on his Cerveceria Havemeyer restaurant in Williamsburg next week, opening up a new space devoted to Mexican pantry items, mezcal and music. La Milagrosa Agave Bar and Listening Room (opening Wednesday)...

Portland's Metal Pizzeria Sizzle Pie Opens Next Week In Williamsburg



Their motto is "Death To False Pizza"—that's how you know Sizzle Pie doesn't fuck around. The Portland-born pizzeria moves into Williamsburg next Monday, transforming the former Grandma Rose's pizzeria into a pie slinging joint that'd send most...

Brooklyn's Best Pop-Up Dinner Party Returns Next Month



(Photo by Louise Palmberg courtesy Oxalis) Last year, a pop-up restaurant started by Restaurant Daniel alum Nico Russell and some of his colleagues enjoyed a sold-out run at two different Brooklyn locations. Oxalis began last summer...

There's Now A Harry Potter-Themed Pasta Restaurant In Williamsburg



Take 2016's biggest dining trend, add one part burgeoning

Marine Park



Articles about "marinepark"

Brooklyn Man Allegedly Threatened Nigerian Priest With A Bat In Marine Park Hate Crime

Uriroghene Okrokoto A Roman Catholic priest from Nigeria says that a man threatened him with a baseball bat in Marine Park earlier this month, called him a racial epithet and yelling at him to "go back...



Man Unloading Christmas Trees From Truck Fatally Run Over In Brooklyn

image via Google Maps A truck driver was arrested last night after allegedly running over another man in Brooklyn. According to police, 56-year-old Darren Danks was behind the wheel of a truck parked near Avenue U...



FDNY Inspector Arrested For Allegedly Trying To Bribe Buildings Department Inspector

The cash and courtesy card an FDNY inspector allegedly tried to pass to a Buildings Department inspector, left, and the FDNY inspector's Brooklyn house, right. (DOI/Google) Police arrested an FDNY fire prevention inspector yesterday after they...

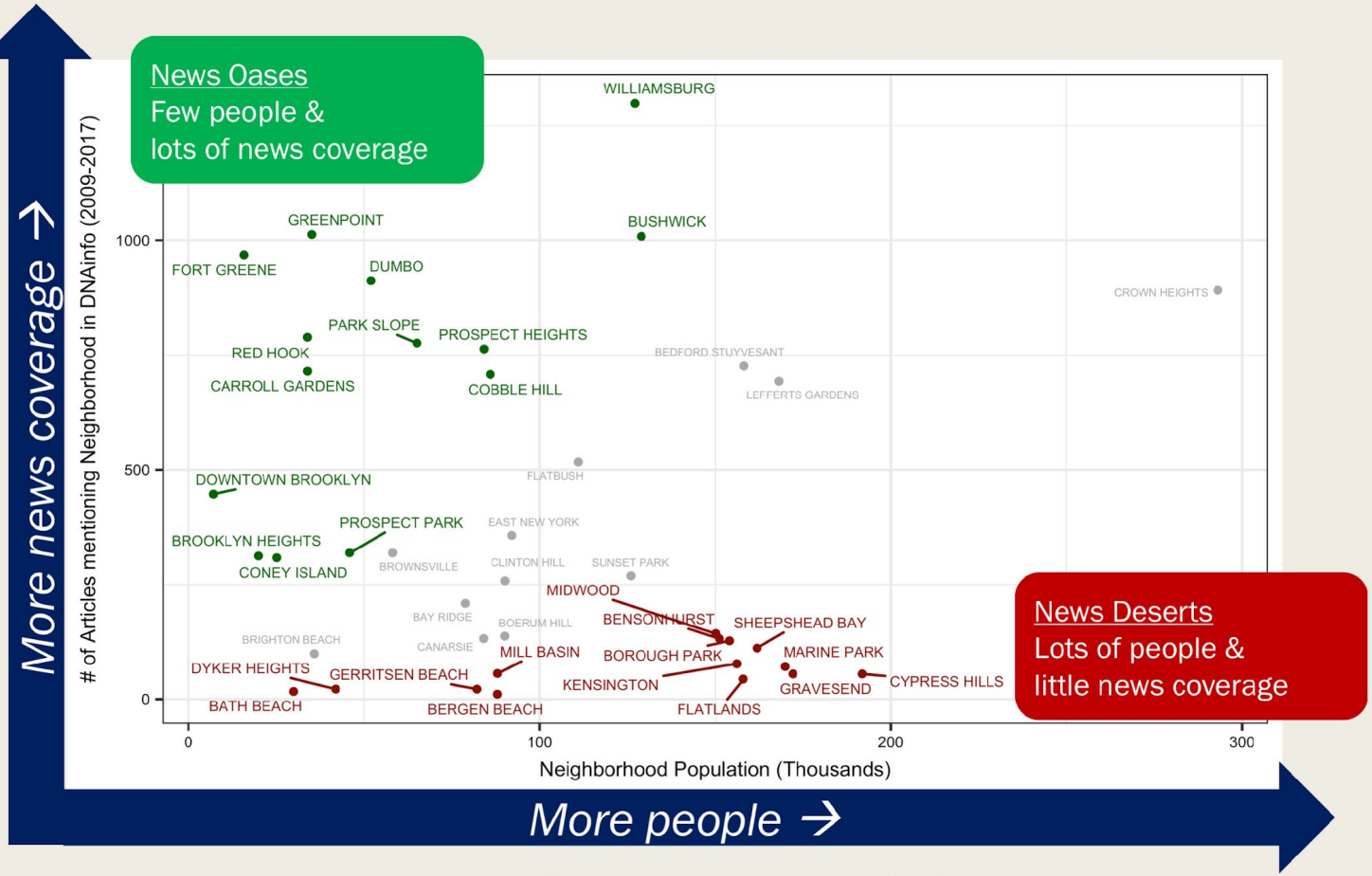


NYPD Is Searching For This Compulsive Swastika Scrawler

via NYPD The NYPD is tracking an industrious bigot suspected of etching swastikas all over Brooklyn's Marine Park neighborhood. Just after 4 a.m. on Monday, April 25th, the man used a white marker to draw swastikas...

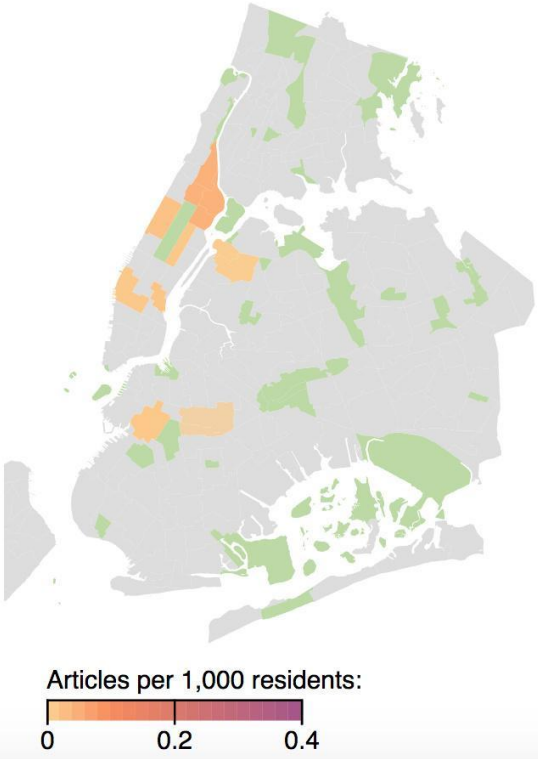


Source: Analysis of newspaper headlines from the Gothamist (2015-2017)

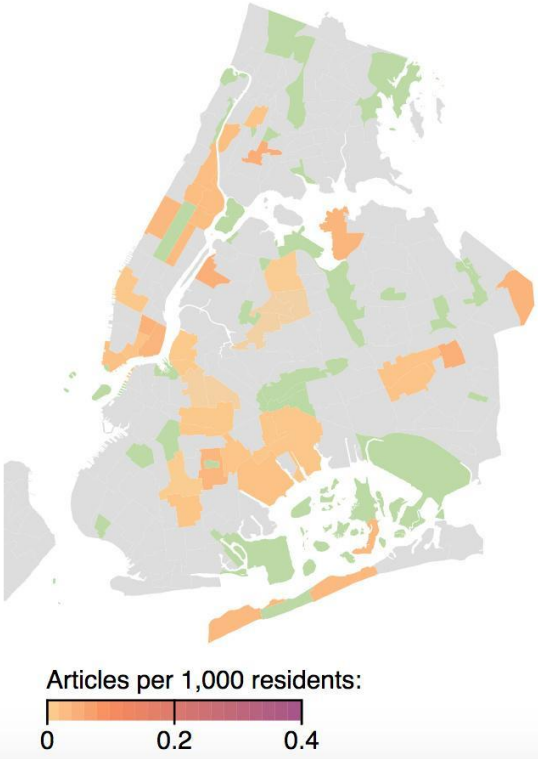


Source: *Analysis of 15,887 articles from DNAinfo mentioning 40 Brooklyn neighborhoods, covering 2009-2017*

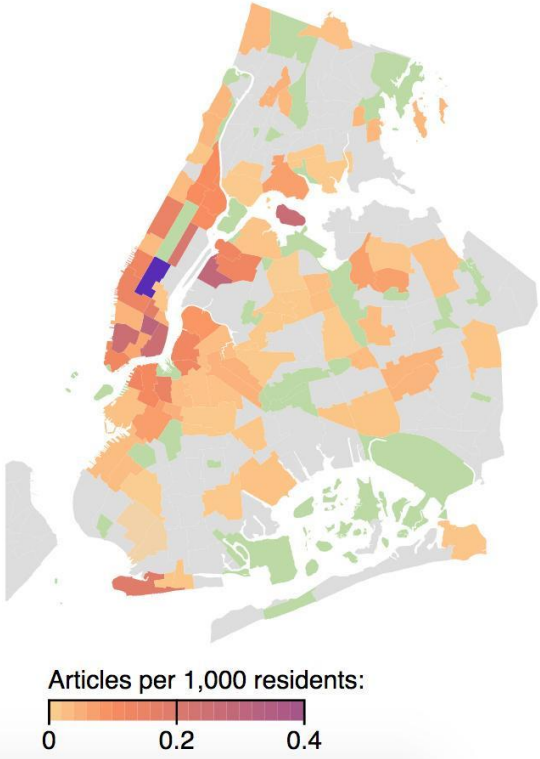
New York Daily News
Gini coeff. of coverage: 0.9528



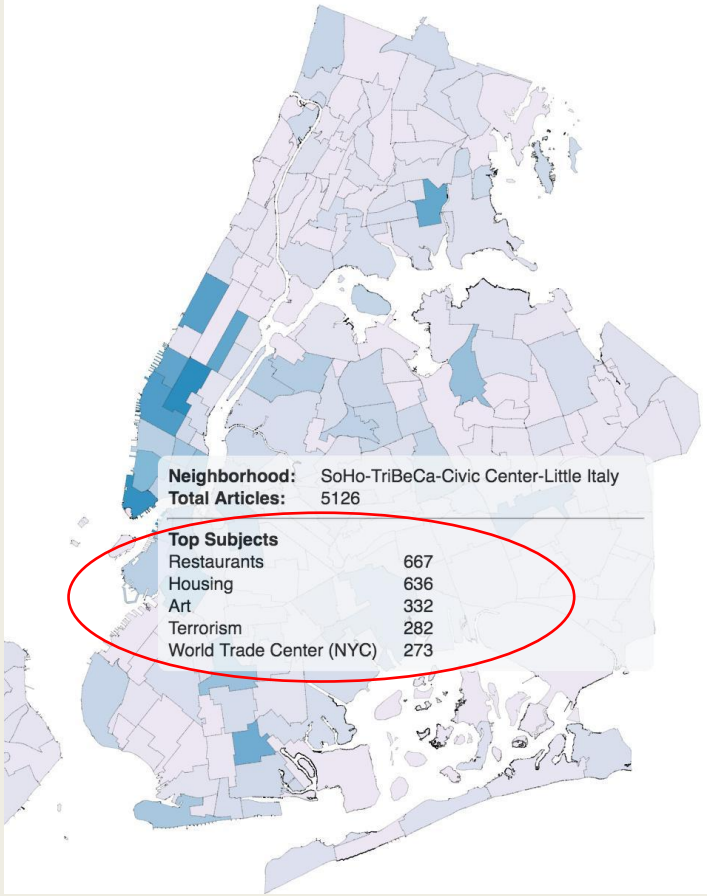
New York Post
Gini coeff. of coverage: 0.8646



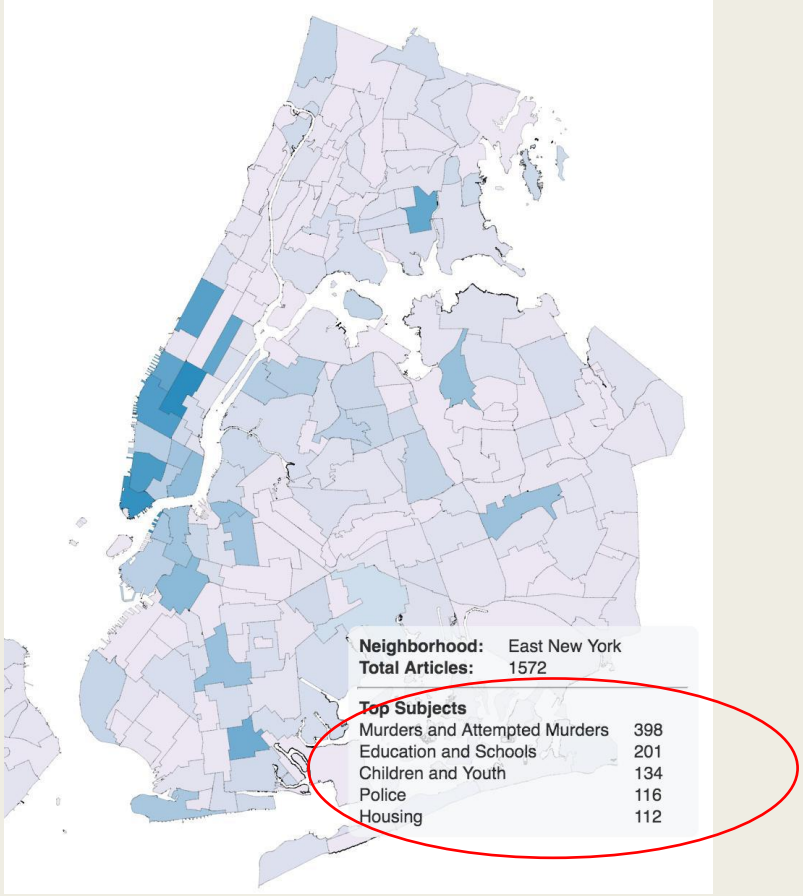
New York Times
Gini coeff. of coverage: 0.8102



New York Times by Neighborhood



New York Times by Neighborhood



Takeaways

The best way to become a data scientist is to do data science.

Start with something you're passionate about!

Build some cool stuff.

Share it with the world.

You got this!

Thank you!



hamdan.azhar@gmail.com

@HamdanAzhar