

***WELCOME***

# Macron v. Le Pen

Lea Balido  
Adam Ricks  
Matt Hwang

8-30-19

A Data Story

A Data Story

# What We Talkin' Bout

**A Brief  
Analysis of Rap in  
the USA**

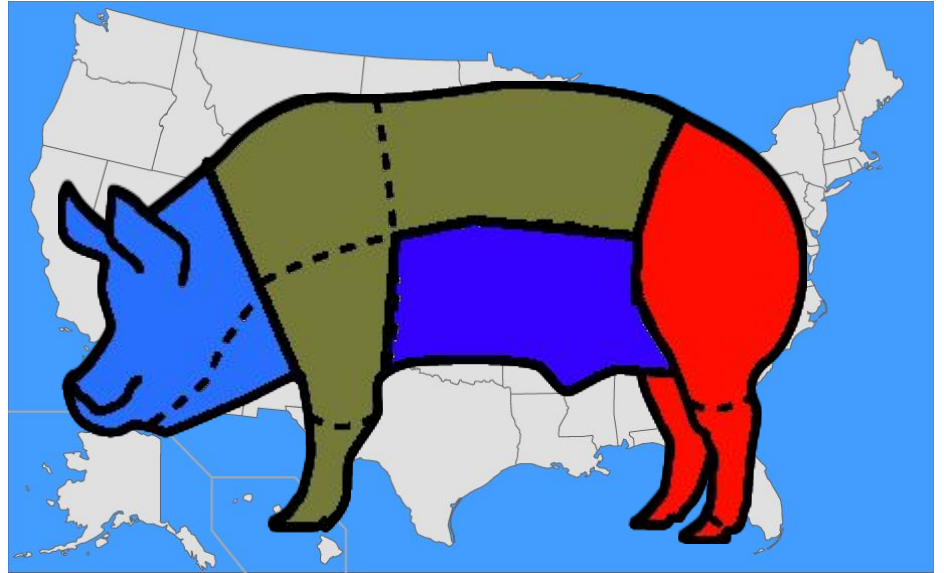


September, Friday the 13th, 2019

**Matt Hwang**

# Rap in the USA

- Although rap music has its origins in NY, it has spread in popularity across the US
- It be generalized and subdivided into four regions
  - West Coast
  - East Coast
  - Mid West
  - Down South




# Ranker.com

-Ranker is a public platform where users can vote and “rank” their opinions on favorite topics

-Used to determine top representatives per region for analysis

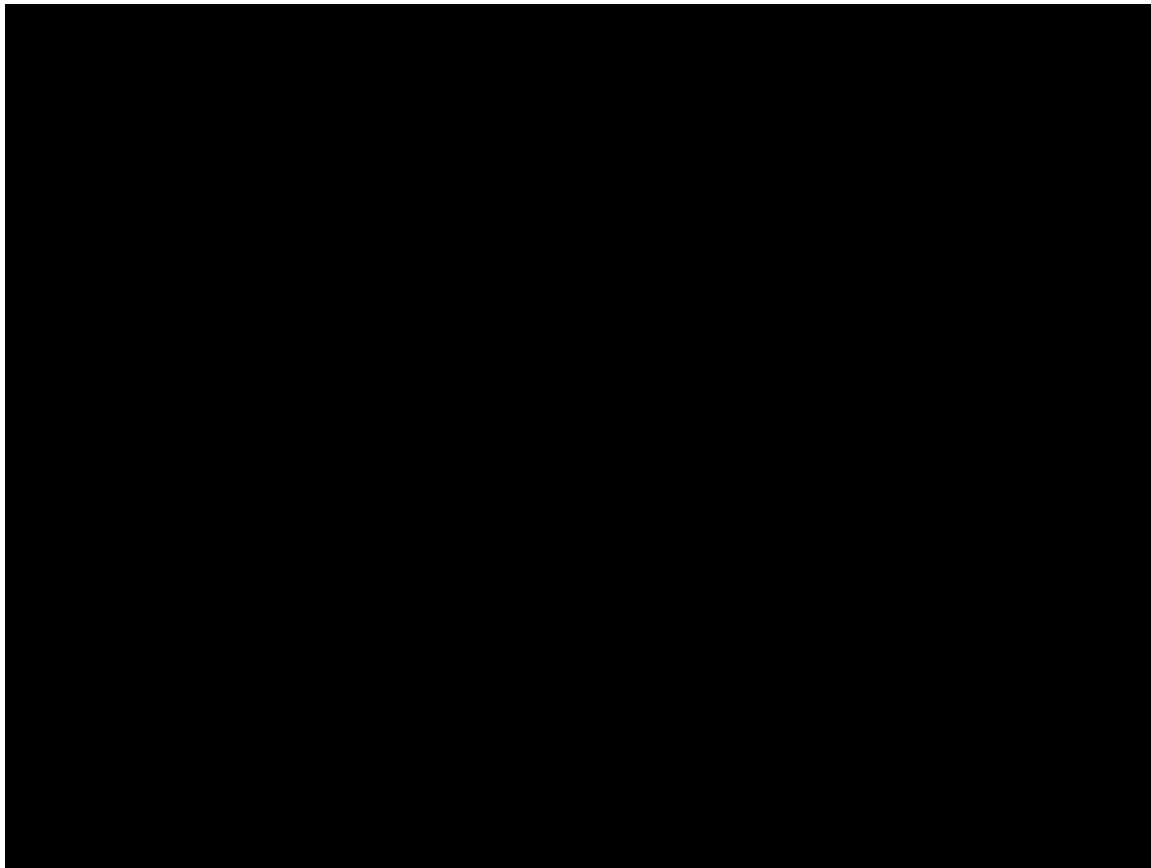
-Chose top 25 artists per region and gathered data via a python webscraping script



The screenshot shows the Ranker.com website interface. At the top, the Ranker logo is on the left, and navigation links for 'entertainment', 'music', 'nerdy', 'sports', 'living', 'history', 'culture', 'food', 'channels', and 'shop' are on the right. Below the navigation bar, the main heading is 'The Best Southern Rappers of All Time'. Underneath this, it says 'Originally by Reference' with statistics: '62.3k votes • 5k voters • 203.5k views • 152 items'. The main text block contains an introduction to Southern rap and a list of famous rappers. Below the text, there are filters for 'list ordered by' (all voters), 'rerankers' (7), 'men', 'women', 'age', 'region', and 'bands/musicians'. A search bar and a 'rank your version' button are also present. The list of rappers is displayed in a table-like format with their rank, up/down arrows, vote counts, and a small image of the artist.

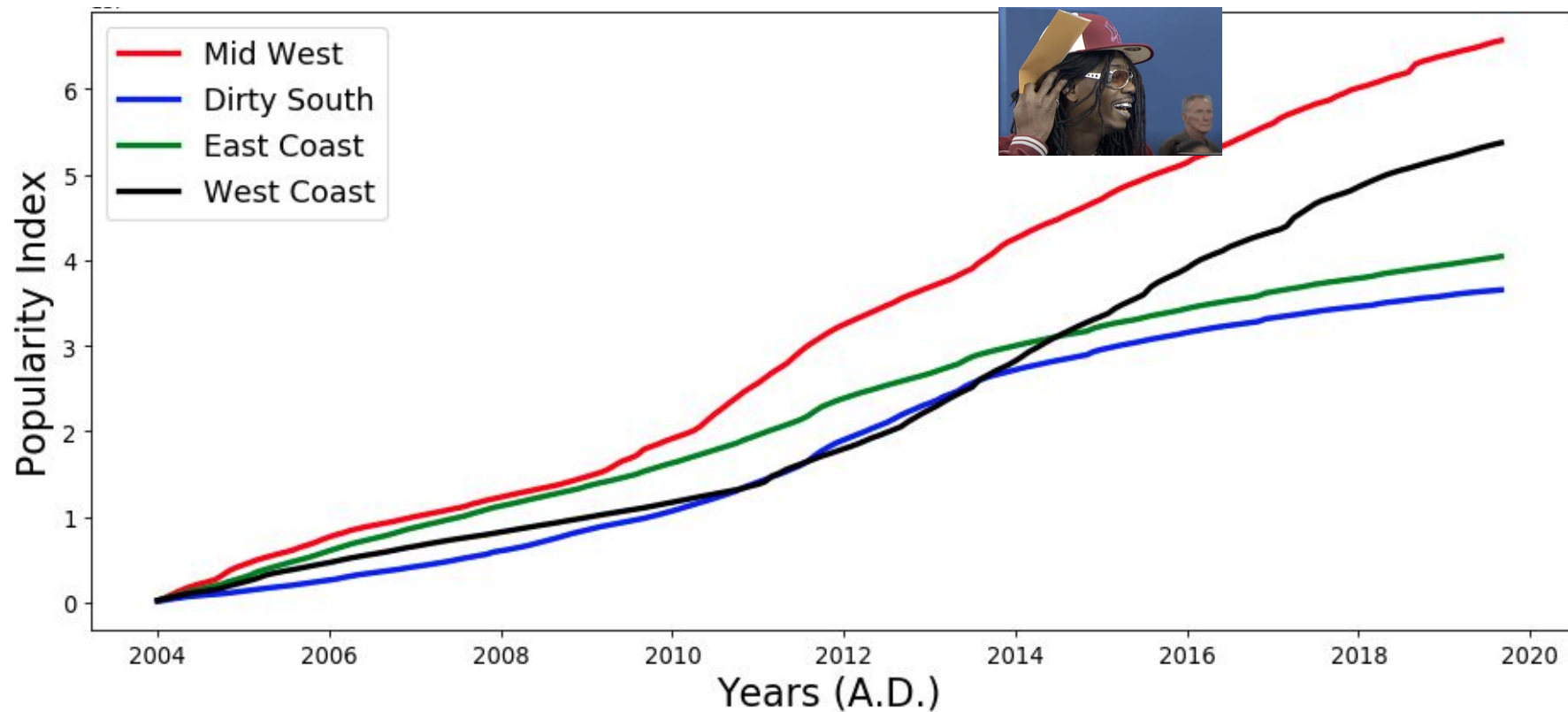
Rank	Artist	Genre	Upvotes	Downvotes
1	Outkast	Alternative hip hop, Dance-pop, Jazz	1,425	414
2	André 3000	Hip hop music, Alternative hip hop, Rock music	1,300	274
3	Scarface	Hip hop music, Horrorcore, Gangsta rap	1,151	307

# Regional Popularity Over Time



Using a popularity metric based on multiple sources the relative regional popularity can be estimated when overlayed on top of historical Google Search Trending Data

# Regional Rap Popularity in America 2004 - Present



# What's Going On In The Mid West?

- Why are they more popular over the artists based in the rest of the USA?
- Are there any obvious features that may correlate to their popularity?
- Is this why Kanye West purchased a ranch in Wyoming?

Let's find out with DATA





# Lyrical Analysis

- A potential indication of popularity may stem from the **topic** of discussion
- A database of all song lyrics for the top 3 artists per region (from Ranker) was created by scraping a lyrics website (AZ-Lyrics.com) with a python webscraping tool and some VPN magic

felt guilty 🤔



## West Coast

- 2Pac  
- Snoop Dogg  
- Kendrick  
Lamar

## East Coast

- Jay-Z  
- 50 Cent  
- Biggie  
Smalls

## Mid West

- Kanye  
- Eminem  
- Big Sean

## Dirty South

- J.Cole  
- 2 Chainz  
- Lil' Wayne

# Lyrical Analysis Cont.

Web Harvest Yielded:

- 12 Artists
- 3,440 songs
- 1,000,000+ words
- One Love
- Filtered HTML Jargon and removed STOP / filler words from analysis



# Results - Started From The Bottom

Now we're here

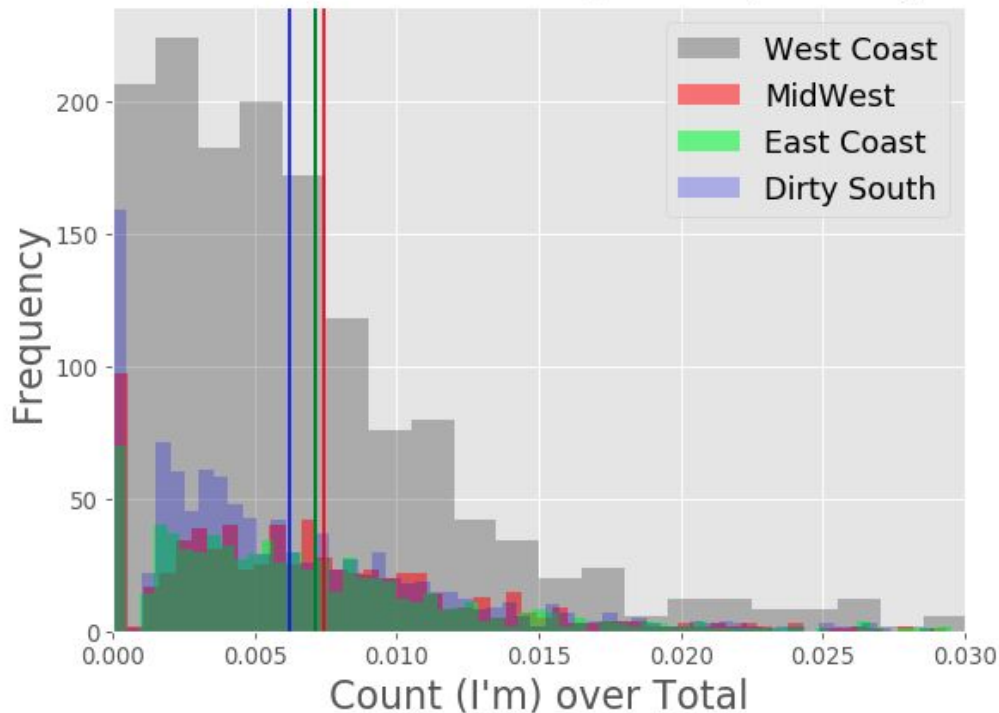
- Top word across all artists and regions is “Im” signaling the artists’ tendencies to speak about themselves and their stories and also hobbies and interests.
- Also a large mixture of less than nice words floated to the top.

Is the frequency of talking about themselves or frequency of cursing indicative of popularity?

Word	Count	artist
im	4831	Lil' Wayne
im	3412	Eminem
im	2691	Lil' Wayne
im	2662	50 Cent
im	2566	Snoop Dogg
im	1980	2Pac
im	1950	Lil' Wayne
im	1770	2 Chainz
im	1737	Jay-Z
im	1695	50 Cent
im	1611	Big Sean
yeah	1530	Lil' Wayne
im	1519	2Pac
im	1434	Lil' Wayne
money	1427	Lil' Wayne
im	1425	Snoop Dogg

# MidWest Vs The Rest

Mean 'I'm' Divided by Total Words (cleaned) in Songs Per Region



- Artists from the Midwest talk about themselves the most
- Mean self reference per song divided by total words per song
- 0.0075% calculated value for MW on average
  - Once per 135 words

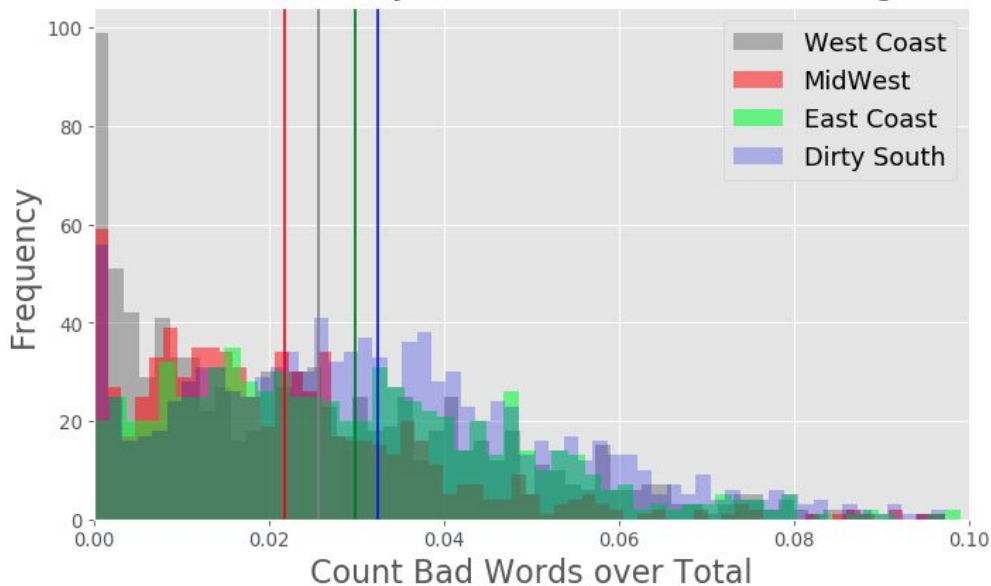
# Choice Words 🤔

- Mid West manners run deep
- Calculated by running lyrics through a list of banned Google words

## Who Swears the LEAST

1st Place	Mid West	2.19%	😇
2nd Place	West Coast	2.57%	😬
3rd Place	East Coast	2.99%	😐
Dead Last	Dirty South	3.24%	🚫

Mean Bad Words divided by Total Words (cleaned) in Songs Per Region



# Boot Strap

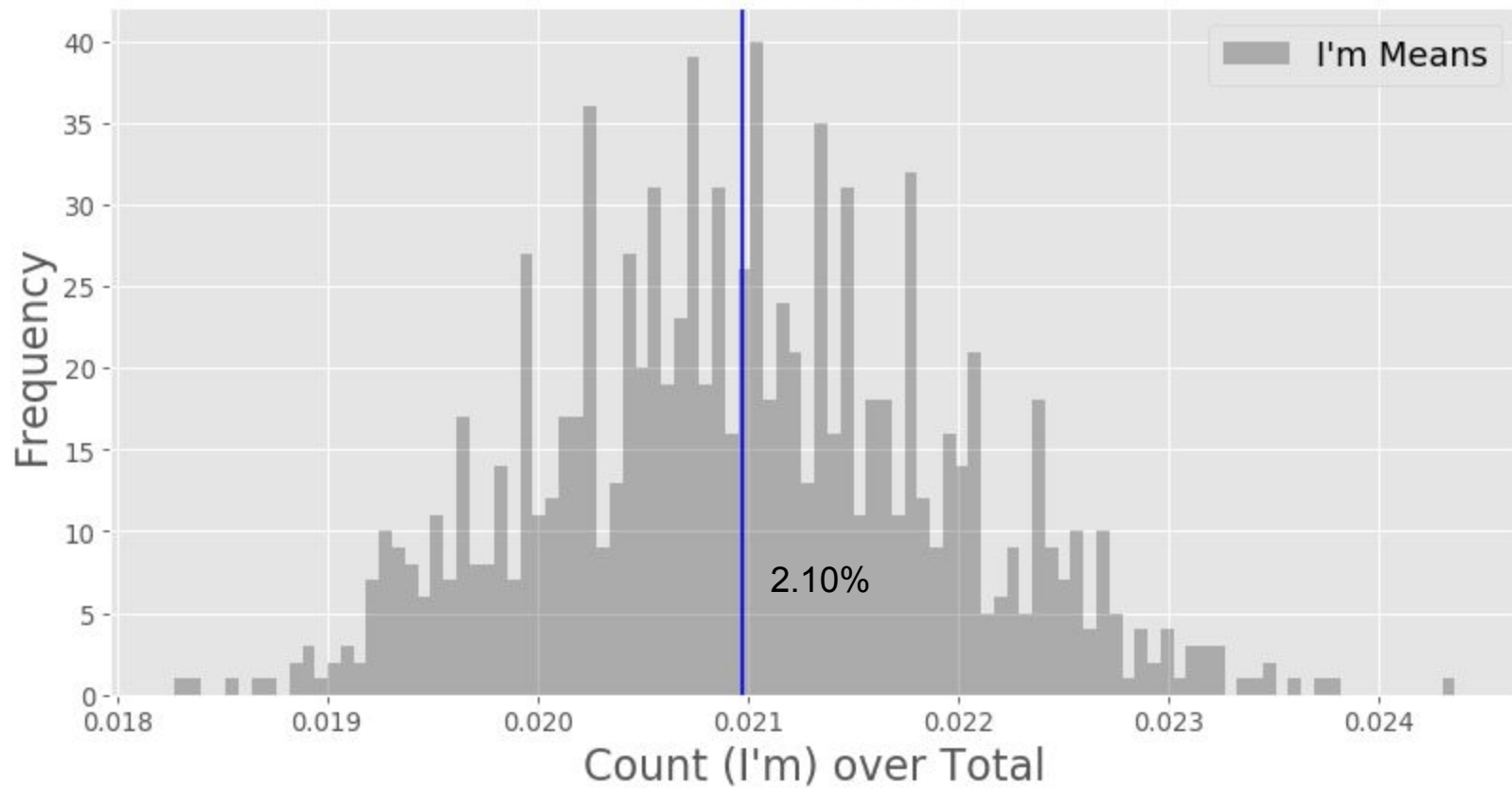
Let's increase improve the quality of our sampling with Bootstrapping!

Originally I took top 3 per region and analyzed their discography

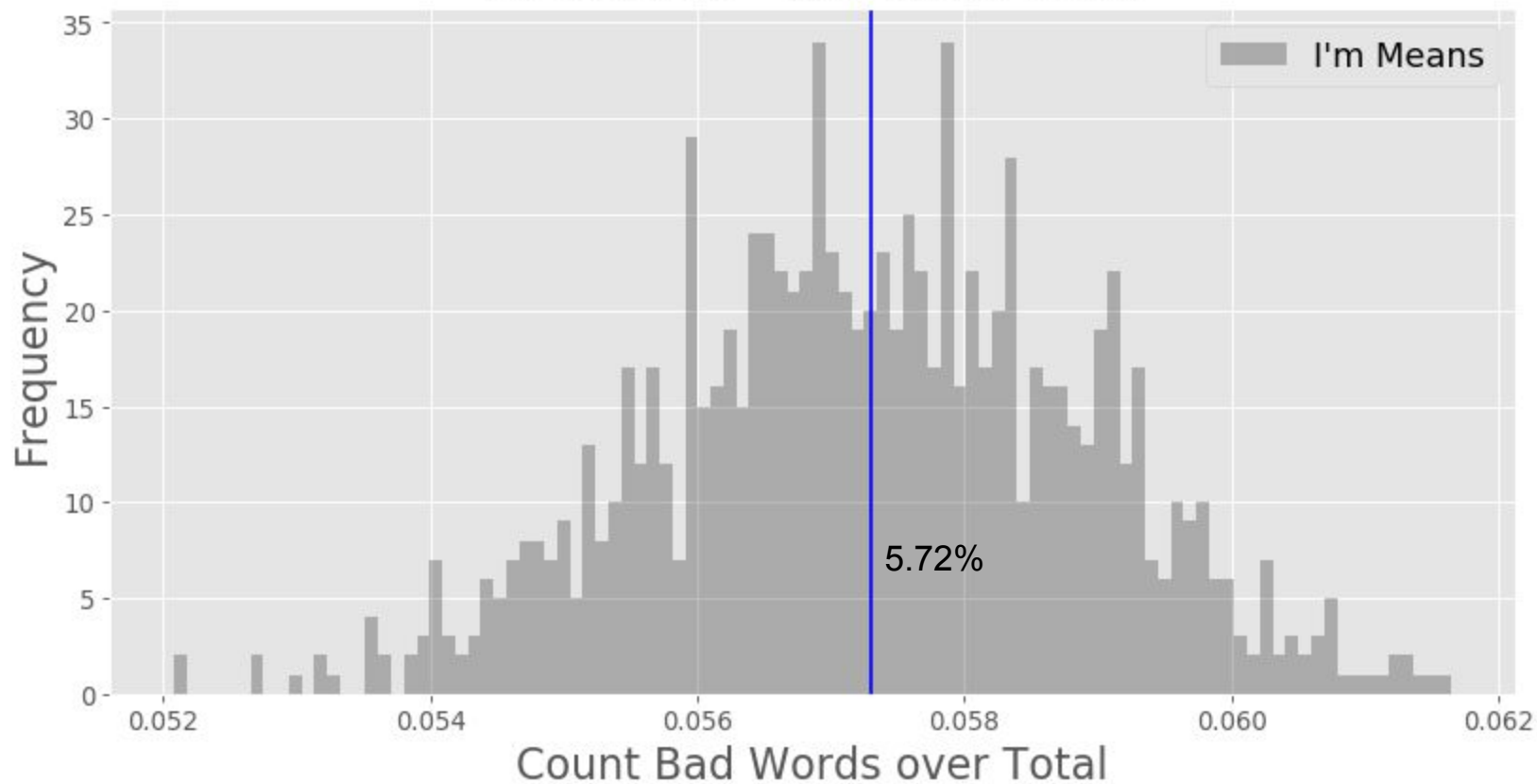
What could be more representative of the population is taking random 4 songs from top 25 artist per region as an initial sample (stratified random sampling) then running it 1000x times to determine mean values.

**DISCLAIMER - I did not finish and only bootstrapped the MidWest and also took a different approach that produced different numbers**

## Booty Strap



## Booty Strap - Bad Words Mean





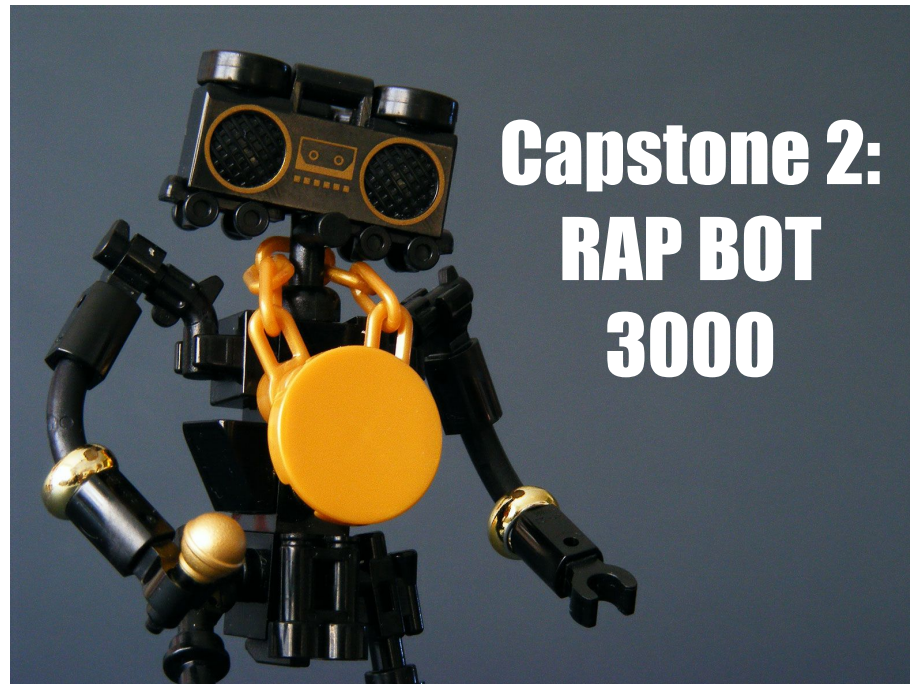
# Small Fish Big Pond

Without context it is hard to determine a definitive strong positive correlation

However it is interesting to take note of these trends

Would love to run further analysis on the gathered lyrical datasets with NLP

Potential for Machine Learning

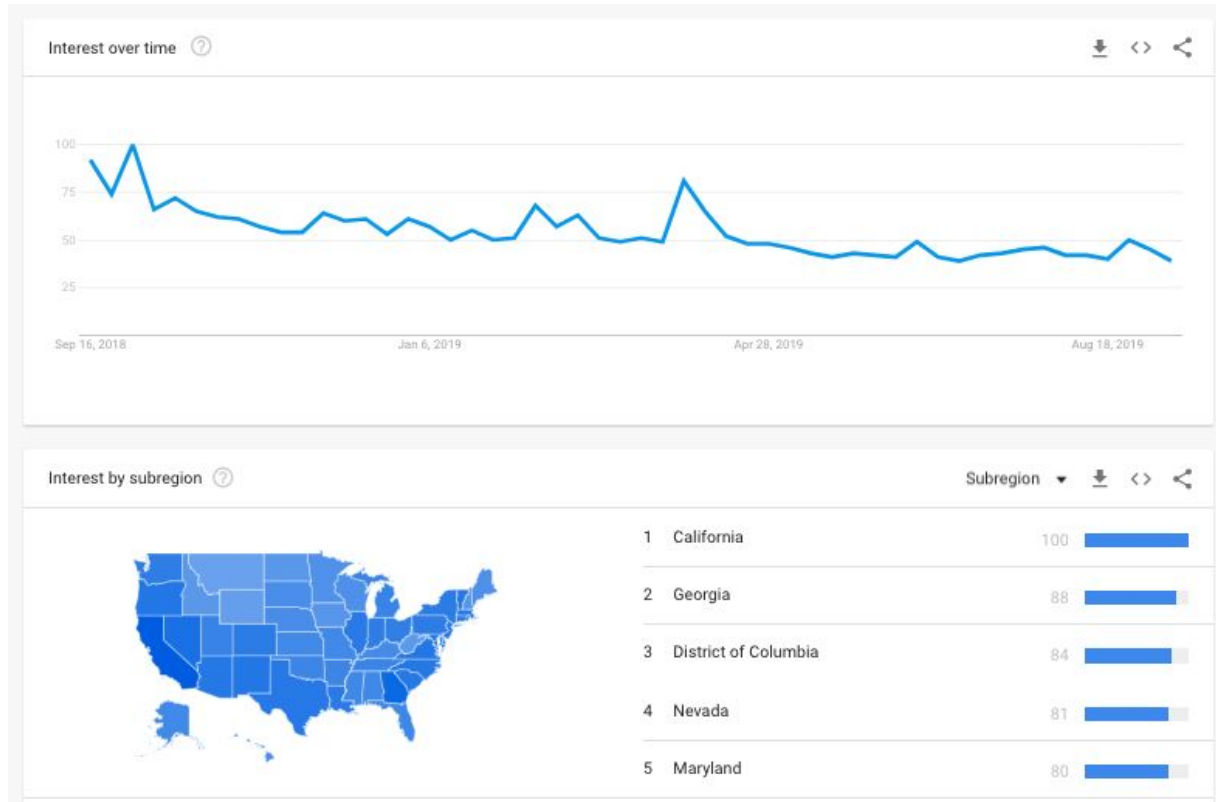


# The End



Thanks for coming  
to my TED Talk 🙏

# Extra Slides - Google Trends



Category: All categories	
Month	Scarface: (Unit
2004-01	38
2004-02	32
2004-03	28
2004-04	25
2004-05	21
2004-06	19
2004-07	19
2004-08	22
2004-09	24
2004-10	25
2004-11	23
2004-12	24
2005-01	24
2005-02	24
2005-03	25
2005-04	23
2005-05	27
2005-06	25
2005-07	22
2005-08	19
2005-09	20
2005-10	29
2005-11	31

	MidWest_Spotify_Totals	Dirty_South_Spotify_Totals	West_Coast_Spotify_Totals	East_Coast_Spotify_Totals	date
0	246620	122764	215113	233593	2004-01-01
1	535356	279395	435139	483267	2004-02-01
2	920519	404058	621445	719941	2004-03-01
3	1317393	537931	798758	933802	2004-04-01
4	1656751	650154	982855	1173643	2004-05-01
5	1950342	735140	1123947	1381209	2004-06-01
6	2194538	806722	1252312	1552101	2004-07-01
7	2405427	872273	1362015	1744674	2004-08-01
8	2659478	942744	1496787	1938362	2004-09-01
9	3145252	1023320	1669152	2146739	2004-10-01
10	3712830	1111326	1935171	2456835	2004-11-01
11	4084066	1220215	2145827	2698273	2004-12-01



# Extra Slides - AZ-Lyrics

AZ A B C D E F G H I J K L M N O P Q R S T U V W X Y Z #

Q Search



**FIND HISTORICAL**  
Hundreds Available! Only at the W

Like 4

### 21 Savage Lyrics

▶ Play "A Lot"

on Apple Music

▶

"a lot"  
(feat. J. Cole)

[21 Savage:]  
I love you  
Turn my headphone down a little bit, yeah  
For so many reasons  
Yeah, yeah, yeah, yeah, yeah (I do), yeah, yeah yeah yeah  
Yeah, yeah, ah, ah, whoa, whoa, whoa, yeah

How much money you got? (Straight up)  
How much money you got? (Straight up)  
How much money you got? (Straight up)  
How much money you got? (A lot)

How much money you got? (A lot)  
How many problems you got? (A lot)  
How many people done doubted you? (A lot)  
Left you out to rot? (A lot)  
How many pray that you flop? (A lot)

```
{ "2Chainz": [{"artist": "2chainz", "track": "fork", "lyrics": "\n      <i>\n      [Intro\n<br/>\n      <i>\n      [Hook:] \n      </i>\n      <br/>\n      I had a dream that rap w\nI woke up on the block, had to hit it with the fork\n      <br/>\n      Skrrr, skrrr, skrrr, skrrr: hit it with the fork\n      <br/>\n      ain't bein' sold\n      <br/>\n      So much money on me, it won't even fold\n      <br/>\n      it won't even fold\n      <br/>\n      So much money on me, it won't even fold\n      <br/>\n      me, it won't even fold\n      <br/>\n      So much money on me, it won't even fold\n      <i>\n      [Verse 1] \n      </i>\n      <br/>\n      I got Medusa on my sneakers\n      \"nice to meet ya\" \n      <br/>\n      100K for a feature, hundred K's at my leisure\n      at your people\n      <br/>\n      I be higher than a eagle\n      <br/>\n      When I'm\n      <br/>\n      Free my nigga Segel\n      <br/>\n      I am ridin' on a jet, headin' to tha\n      Soon as I land I be in that Testarossa\n      <br/>\n      If I die tonight, you gon' see\n      <br/>\n      I'm the man in my city, same thing in South Dakota\n      <br/>\n      Man I\n      show you how I do it\n      <br/>\n      I drink red bitches, I don't drink Red Bulls\n      tried to give me wings, but I already had some\n      <br/>\n      I'm all that and then
```

```
{ "Discography": [{"artist": "2chainz", "track": "fork", "lyrics": "\n      I had a dream that rap wouldnt work      I woke up on the block had to hit it wit\n      skrrr skrrr skrrr hit it with the fork      Skrrr skrrr skrrr skrrr hit it with t\n      work records aint bein sold      So much money on me it wont even fold\n      even fold      So much money on me it wont even fold      So much money on\n      So much money on me it wont even fold      I got M\n      My dick up like nice to meet ya      100K for a feature hundred Ks at my leisure\n      people      I be higher than a eagle      When Im sipping on that codeine\n      I am ridin on a jet headin to that Costa      Soon as I land I be in that Testaro\n      tonight you gon see some flicks in Ghosta      Im the man in my city same thing i\n      Im running up that check show you how I do it      I drink red bitches I dont dri\n      they tried to give me wings but I already had some      Im all that and then some\n      And its boomin
```

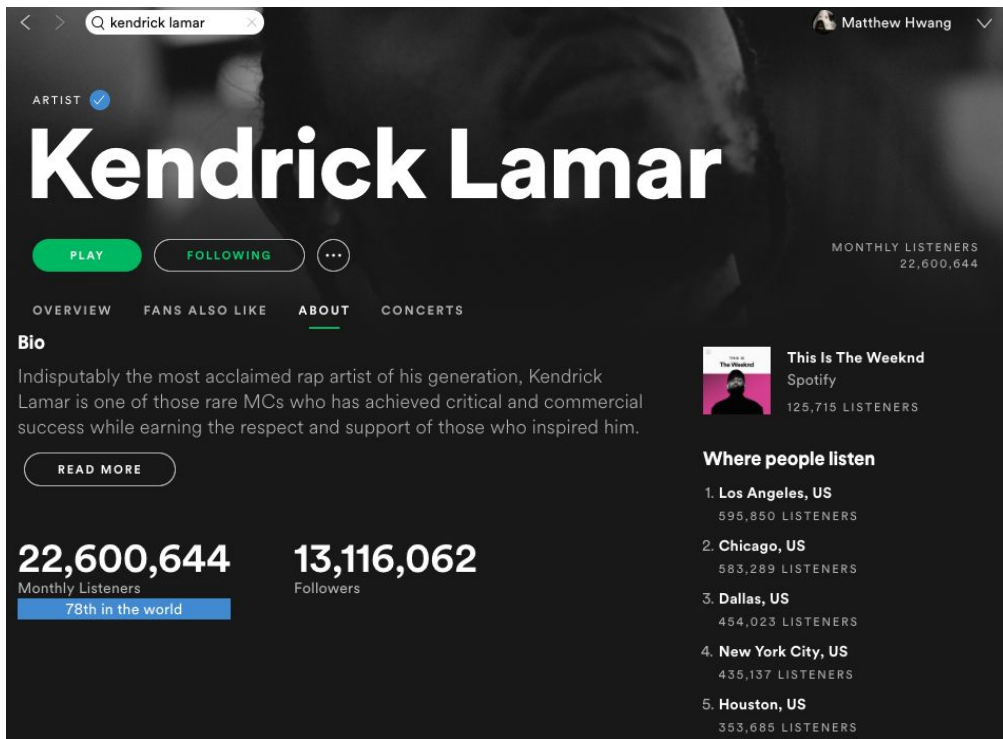
# Webscraping Workflow

```
In [22]: #WEB TO LETTER LINKS
def get_letter_links(website_url):
    soup = BeautifulSoup(requests.get(website_url).content, 'html.parser') #get that shit
    links = []
    home_url = "https:"
    for each in soup.find_all('a'):
        links.append(home_url + each.get("href")) #store that shit
    return links[1:28] #gets only nav links, return that shit

#LETTER LINKS TO ARTIST LINKS
def get_artist_links(letter_url):
    soup = BeautifulSoup(requests.get(letter_url).content, 'html.parser') #get that shit
    links = []
    home_url = "https://www.azlyrics.com/"
    for each in soup.find_all('a'):
        links.append(home_url + each.get("href")) #store that shit
    return links[28:-8] #takes out nav links and footer links, return that shit

#ARTIST LINKS TO TRACK LINKS
def get_track_links(artist_home_url):
    name = artist_home_url.split('/')[4].split('.html')[0]
    soup = BeautifulSoup(requests.get(artist_home_url).content, 'html.parser') #get that shit
    links = []
    home_url = "https://www.azlyrics.com"
    for each in soup.find_all('a'):
        links.append(each.get("href").replace("../", home_url)) #store that shit
    return {name:links[30:-8]} #takes out nav links and footer links, return that shit as a dict for JSON reasons
```

# Extra Slides - Spotify / Sound Cloud



Spotify artist page for Kendrick Lamar. The page features a large profile picture of Kendrick Lamar, a search bar at the top with "kendrick lamar" entered, and a navigation bar with "ARTIST" selected. The main title "Kendrick Lamar" is prominently displayed. Below the title are buttons for "PLAY", "FOLLOWING", and a menu icon. To the right, it shows "MONTHLY LISTENERS 22,600,644". The navigation bar includes "OVERVIEW", "FANS ALSO LIKE", "ABOUT" (selected), and "CONCERTS". The "Bio" section describes Kendrick Lamar as the most acclaimed rap artist of his generation. A "READ MORE" button is present. The "Where people listen" section lists top cities: Los Angeles, US (595,850 listeners), Chicago, US (583,289 listeners), Dallas, US (454,023 listeners), New York City, US (435,137 listeners), and Houston, US (353,685 listeners). A small album cover for "This Is The Weeknd" is also visible.

Q kendrick lamar

Matthew Hwang

ARTIST

# Kendrick Lamar

PLAY FOLLOWING

MONTHLY LISTENERS 22,600,644

OVERVIEW FANS ALSO LIKE ABOUT CONCERTS

## Bio

Indisputably the most acclaimed rap artist of his generation, Kendrick Lamar is one of those rare MCs who has achieved critical and commercial success while earning the respect and support of those who inspired him.

READ MORE

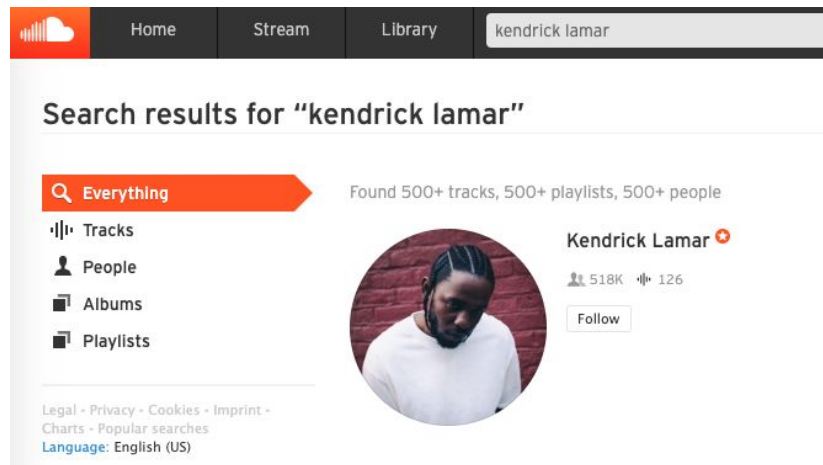
22,600,644 Monthly Listeners  
78th in the world

13,116,062 Followers

This Is The Weeknd Spotify 125,715 LISTENERS

### Where people listen

1. Los Angeles, US  
595,850 LISTENERS
2. Chicago, US  
583,289 LISTENERS
3. Dallas, US  
454,023 LISTENERS
4. New York City, US  
435,137 LISTENERS
5. Houston, US  
353,685 LISTENERS



SoundCloud search results for "kendrick lamar". The page shows a navigation bar with "Home", "Stream", "Library", and a search bar containing "kendrick lamar". The search results section is titled "Search results for 'kendrick lamar'" and shows "Found 500+ tracks, 500+ playlists, 500+ people". A filter bar is set to "Everything". On the left, there are links for "Tracks", "People", "Albums", and "Playlists". On the right, there is a profile card for "Kendrick Lamar" with a circular profile picture, a "Follow" button, and statistics: "518K" followers and "126" tracks. At the bottom, there are links for "Legal", "Privacy", "Cookies", "Imprint", "Charts", and "Popular searches". The language is set to "English (US)".

Home Stream Library kendrick lamar

## Search results for "kendrick lamar"

Found 500+ tracks, 500+ playlists, 500+ people

Everything

- Tracks
- People
- Albums
- Playlists

Kendrick Lamar

518K 126

Follow

Legal - Privacy - Cookies - Imprint - Charts - Popular searches  
Language: English (US)

# Extra Slides - Profanity Check

[RobertJGabriel](#) / [Google-profanity-words](#)

[Watch](#) 6 [Star](#) 65 [Fork](#) 42

[Code](#) [Issues](#) 0 [Pull requests](#) 0 [Actions](#) [Projects](#) 0 [Wiki](#) [Security](#) [Insights](#)

Full List of Bad Words and Top Swear Words Banned by Google as they closed it

[profanity](#) [profanity-detection](#) [profanityfilter](#) [side-project](#) [google](#)

[3](#) commits [2](#) branches [0](#) releases [1](#) contributor [Apache-2.0](#)

[Branch: master](#) [New pull request](#) [Create new file](#) [Upload files](#) [Find File](#) [Clone or download](#)

**RobertJGabriel** Push changes

Latest commit a24e30c on Apr 25, 2017

<a href="#">.github</a>	Push changes	2 years ago
<a href="#">.gitignore</a>	Initial commit	2 years ago
<a href="#">LICENSE</a>	Initial commit	2 years ago
<a href="#">README.md</a>	Push changes	2 years ago
<a href="#">list.txt</a>	Create list.txt	2 years ago

[README.md](#)

## Full List of Bad Words and Top Swear Words Banned by

<https://github.com/RobertJGabriel/Google-profanity-words>