

Clusta' Rhymes

Presented by:

Categorizing Hip-Hop styles through
semantic and prosodic clustering

Kelsey Glenn

Metis, San Francisco
Summer, 2020




“You ain’t got the answers” –Kanye West

Discovery and Understanding

- Music is **inundated** and **topheavy**
- Hip-Hop is diverse and difficult to classify

Overview: Workflow



Data Collection & Preprocessing



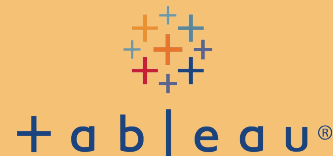
Topic Modeling & Clustering



Deployment & Demonstration



+ a b | e a u[®]



Part I: The Data

Data Collection



BeautifulSoup

Preprocessing & Feature Engineering

TF-IDF [Vec][tor][iz][ation]

—— bars ——
syl-la-bles
{unique, words}

Non- Negative Matrix Factorization

Part II: Topic Modeling

Topic Area I: “Flexin on Em”

Bragging
(‘b****s’, ‘d**k’, ‘pull’, ‘money’)

1st/2nd Person Pronouns
(‘im’, ‘em’, ‘yall’, ‘ya’, ‘ima’)

Part II: Topic Modeling

Topic Area II: “Beerbongs and Bentleys”

Interjections

(‘yeah’, ‘ooh’, ‘woo’, ‘skrrt’, ‘yuh’)

Materialistic Nouns

(‘diamonds’, ‘racks’, ‘bag’, ‘bands’, ‘drip’)

Non-
Negative
Matrix
Factorization

Part II: Topic Modeling

Topic Area III: “Talk Dirty”

Emotion

('love', 'baby', 'girl', 'feel')

Intimacy

('night', 'heart', 'hold', 'body', 'eyes')

Non-
Negative
Matrix
Factorization

Part II: Topic Modeling

Topic Area IV: “Lyrical Miracle”

“Catchall”

Few Interjections

Diverse Vocabulary

Non-
Negative
Matrix
Factorization

K-Means

Features:

Topic Probabilities

Syllabic Rate

Unique Word Rate

Clusters:

Club & Trap

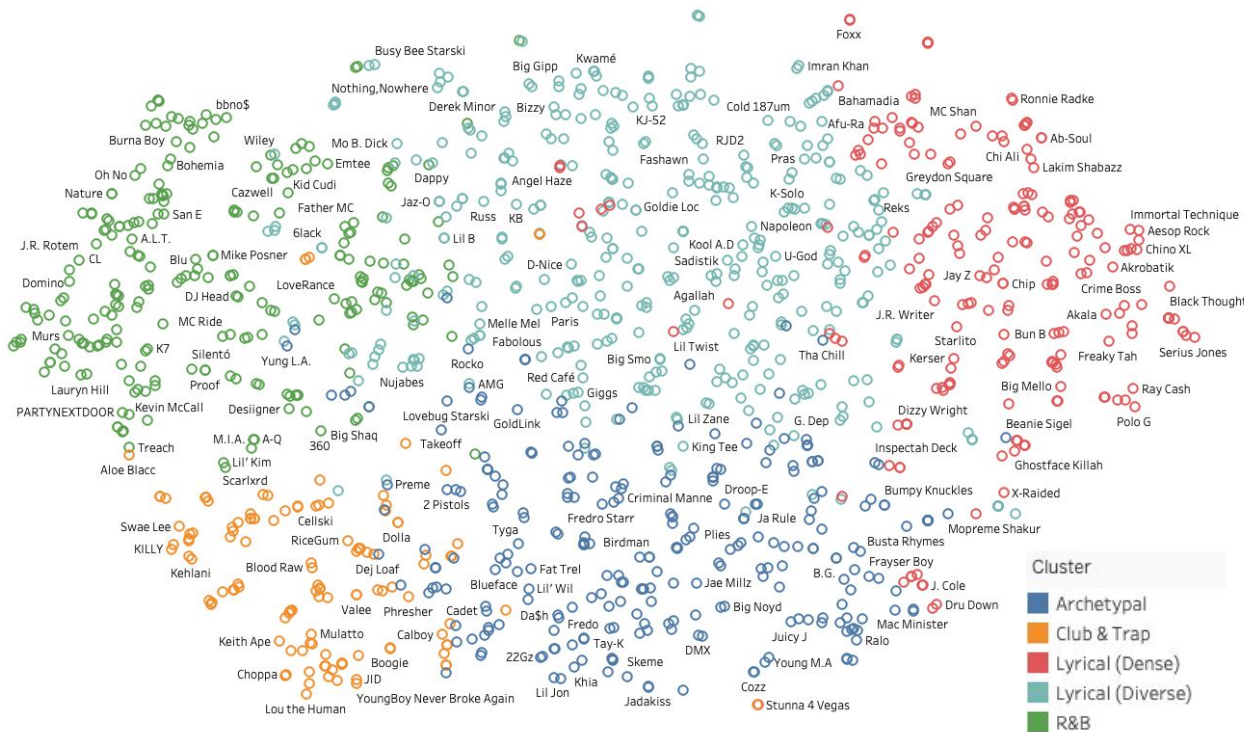
R&B

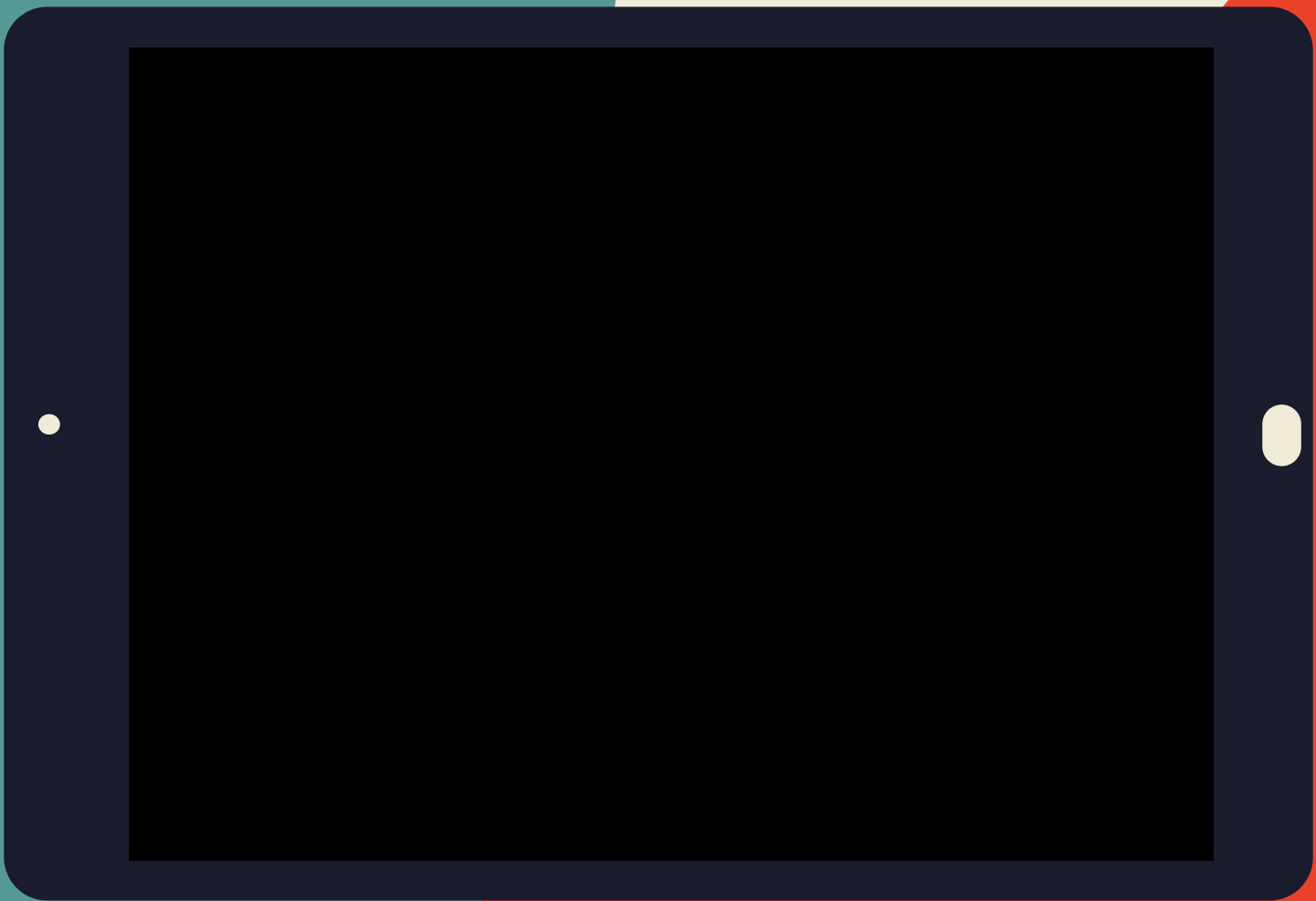
Archetypal

Lyrical : Dense

Lyrical: Diverse

Part II: Clustering





Future Work

- Expand scrape and artist attribution accuracy
- Explore methods for incorporating rhyme data
- Apply combined methods like LDA2Vec to apply embeddings to extract topic probabilities



THANKS!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**



APPENDIX

Initial Artists in Data:
1443

After PreProcessing:
1260

Total Songs in Data:
17566

Total Word Count:
~43 million

APPENDIX

Top Words by Topic

Warning: Contains profanity

Topic 0

im, like, dont, just, got, know, aint, cause, man, life, time, thats, make, em, say, shit, way, right, ill, think, ya, feel, yo, people, gonna, gotta, youre, let, tell, day, need, better, said, wanna, mind, yall, ive, come, look, world, black, good, really, little, hard, new, live, trying, real, god

Topic 1

nigga, niggas, bitch, fuck, shit, aint, got, bitches, money, like, ass, im, gon, hoes, em, dont, lil, pussy, real, fuckin, bout, know, hit, gang, thats, yo, dick, big, yall, ho, fuckin, young, want, ya, boy, ain, man, don, pull, talkin, bad, wit, wanna, run, cause, broke, gotta, ima, block, make

Topic 2

yeah, ayy, uh, ooh, im, bitch, oh, got, woo, like, huh, woah, know, hey, ima, okay, gon, tryna, shawty, ah, diamonds, lil, skrrt, racks, aint, nah, bad, dont, pull, just, big, bag, need, yuh, fuck, whoa, gang, money, bands, wanna, uhhuh, party, new, look, wrist, way, ho, gettin, drip, thats

Topic 3

love, baby, girl, oh, know, want, dont, wanna, just, youre, let, need, feel, tell, way, come, don, time, make, night, right, ill, say, ooh, heart, got, cause, good, tonight, said, gone, really, leave, ain, things, ive, thing, away, hold, think, body, like, wont, mind, shes, babe, aint, bad, eyes, home

K-Means Clustering

Scree Plot

