

EXPLORING COUNTRY MUSIC WITH TOPIC MODELING

DATA ACQUISITION

POPULATE TRACK LIST

- + Scrape Billboard archives with bs4
- + 156,000 tracks, 3620 weeks, 60 years

PULL TRACK LYRICS

- + 15,000 unique songs
- + WikiLyrics API
- + Run script remotely on AWS EC2

scrape_lyrics.py

DATA ACQUISITION

POPULATE TRACK LIST

- + Scrape Billboard archives with bs4
- + 156,000 tracks, 3620 weeks, 60 years

PULL TRACK LYRICS

- + 15,000 unique songs
- + WikiLyrics API
- + Run script remotely on AWS EC2



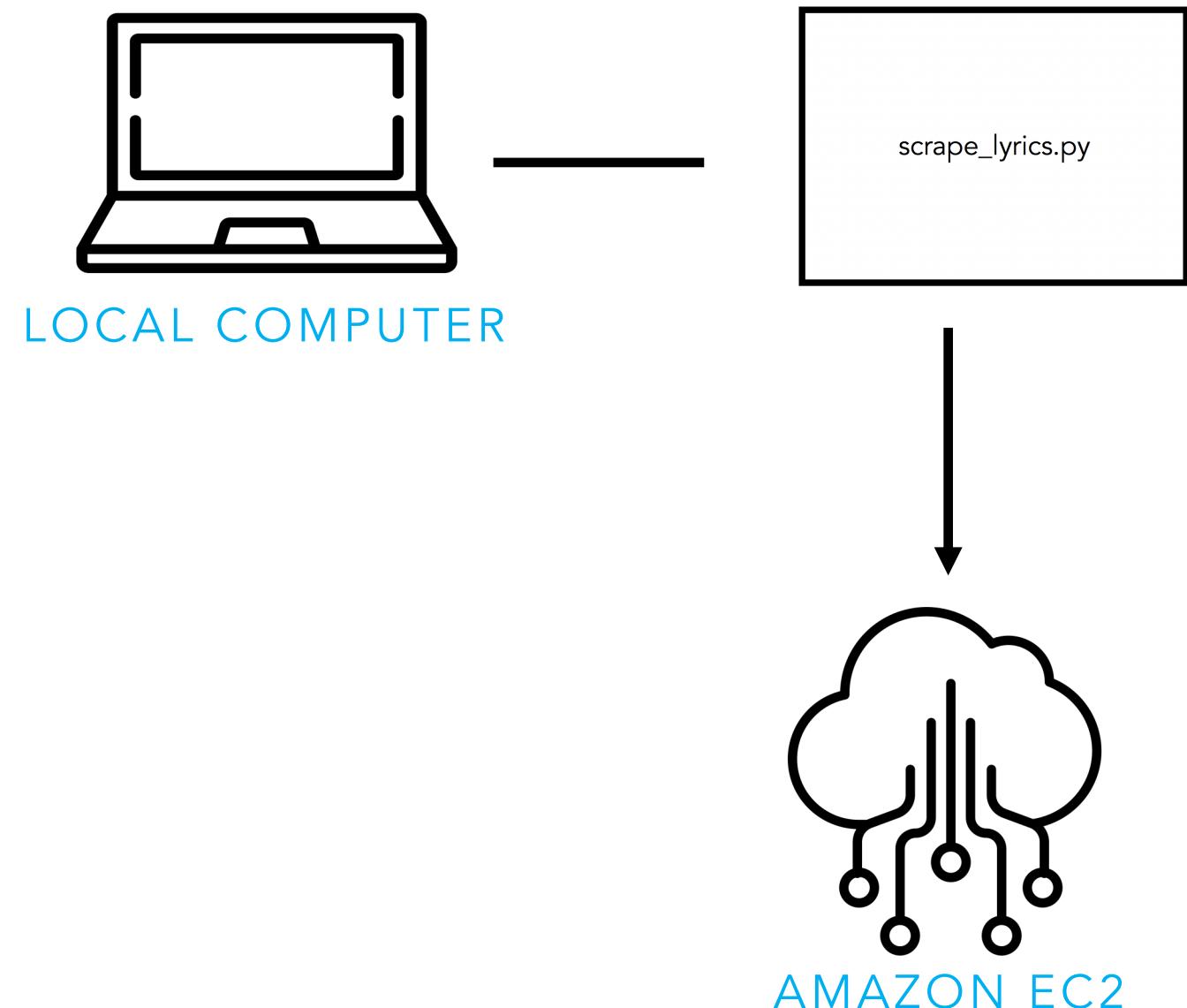
DATA ACQUISITION

POPULATE TRACK LIST

- + Scrape Billboard archives with bs4
- + 156,000 tracks, 3620 weeks, 60 years

PULL TRACK LYRICS

- + 15,000 unique songs
- + WikiLyrics API
- + Run script remotely on AWS EC2



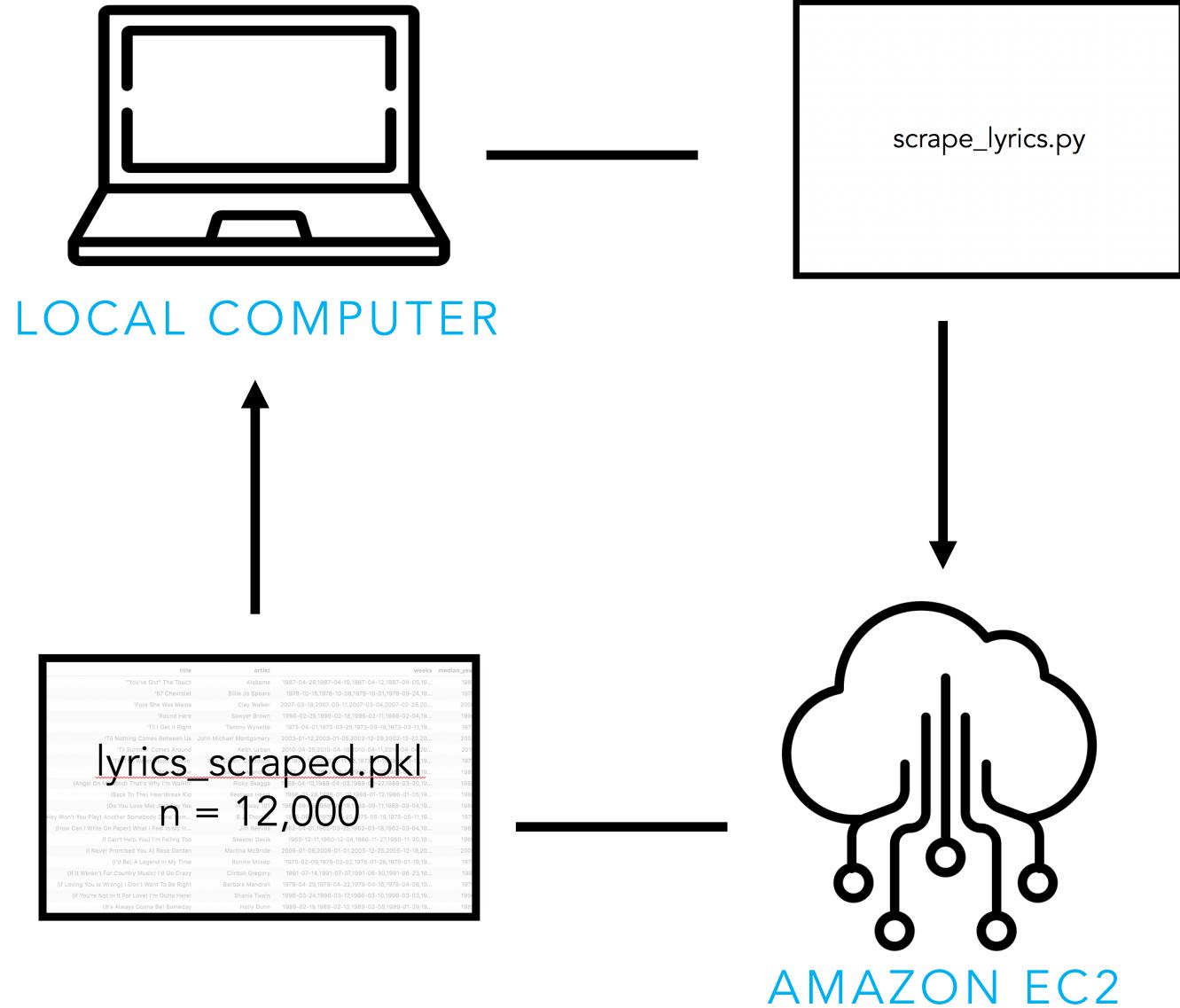
DATA ACQUISITION

POPULATE TRACK LIST

- + Scrape Billboard archives with bs4
 - + 156,000 tracks, 3620 weeks, 60 years

PULL TRACK LYRICS

- + 15,000 unique songs
 - + WikiLyrics API
 - + Run script remotely on AWS EC2



DATA ACQUISITION

POPULATE TRACK LIST

- + Scrape Billboard archives with bs4
- + 156,000 tracks, 3620 weeks, 60 years

PULL TRACK LYRICS

- + 15,000 unique songs
- + WikiLyrics API
- + Run script remotely on AWS EC2

title	artist	weeks	week_count	lyrics
Shine On	Jeff Carson	1998-09-06,1998-08-23	2	Wherever you go, whatever you do,\nDon't ever ...
Good Hands	Troy Olsen	2011-03-27,2011-03-20,2011-03-13,2011-03-06,20...	16	I know your eyes have cried\nToo many tears an...
All I Want For Christmas Is A Real Good Tan	Kenny Chesney	2004-01-04,2003-12-28,2003-12-21,2003-12-14	4	Ooh, don't you think it's a pretty good plan?\...
Anything For Love	James House	1995-10-22,1995-10-15,1995-10-08	3	NaN
Got The All Overs For You (All Over Me)	Freddie Hart	1973-01-28,1973-01-21,1973-01-14,1973-01-07,19...	16	Love's vibrations I can feel them when I hold ...

lyrics_scraped.pkl

DATA ACQUISITION

POPULATE TRACK LIST

- + Scrape Billboard archives with bs4
- + 156,000 tracks, 3620 weeks, 60 years

PULL TRACK LYRICS

- + 15,000 unique songs
- + WikiLyrics API
- + Run script remotely on AWS EC2

title	artist	weeks	week_count	lyrics
Shine On	Jeff Carson	1998-09-06,1998-08-23	2	Wherever you go, whatever you do,\nDon't ever ...
Good Hands	Troy Olsen	2011-03-27,2011-03-20,2011-03-13,2011-03-06,20...	16	I know your eyes have cried\nToo many tears an...
All I Want For Christmas Is A Real Good Tan	Kenny Chesney	2004-01-04,2003-12-28,2003-12-21,2003-12-14	4	Ooh, don't you think it's a pretty good plan?\...
Anything For Love	James House	1995-10-22,1995-10-15,1995-10-08	3	NaN
Got The All Overs For You (All Over Me)	Freddie Hart	1973-01-28,1973-01-21,1973-01-14,1973-01-07,19...	16	Love's vibrations I can feel them when I hold ...

TEXT PREPROCESSING

TEXT CLEANING PIPELINE

- + Tokenize by word
- + Remove punctuation
- + Lowercase
- + Remove stop words
- + Lemmatize text

TEXT PREPROCESSING

TEXT CLEANING PIPELINE

- + Tokenize by word
- + Remove punctuation
- + Lowercase
- + Remove stop words
- + Lemmatize text



TEXT PREPROCESSING

TEXT CLEANING PIPELINE

- + Tokenize by word
- + Remove punctuation
- + Lowercase
- + Remove stop words
- + Lemmatize text

EXAMPLE LYRICS

SINGER: HANK WILLIAMS JR.

TITLE: FAMILY TRADITION

Country music singers
Have always been a real close family
But lately some of my kinfolks
Have disowned a few others and me
I guess it's because I kind of changed my direction
Lord, I guess I went and broke their family tradition

They get on me and want to know, Hank
Why do you drink?
Hank, why do you roll smoke?
Why must you live out the songs that you wrote?
Over and over
Everybody makes my prediction
So if I get stoned, I'm just carrying on an old family
tradition

TEXT PREPROCESSING

TEXT CLEANING PIPELINE

- + Tokenize by word
- + Remove punctuation
- + Lowercase
- + Remove stop words
- + Lemmatize text

EXAMPLE LYRICS

SINGER: HANK WILLIAMS JR.

TITLE: FAMILY TRADITION

```
['Country', 'music', 'singers', 'Have', 'always', 'been', 'a',  
'real', 'close', 'family', 'But', 'lately', 'some', 'of', 'my',  
'kinfolks', 'Have', 'disowned', 'a', 'few', 'others', 'and',  
'me', 'I', 'guess', "it's", 'because', 'I', 'kind', 'of', 'changed',  
'my', 'direction', 'Lord', 'I', 'guess', 'I', 'went', 'and',  
'broke', 'their', 'family', 'tradition', 'They', 'get', 'on', 'me',  
'and', 'want', 'to', 'know', 'Hank', 'Why', 'do', 'you',  
'drink', '?', 'Hank', 'why', 'do', 'you', 'roll', 'smoke', '?',  
'Why', 'must', 'you', 'live', 'out', 'the', 'songs', 'that', 'you',  
'wrote', '?', 'Over', 'and', 'over', 'Everybody', 'makes', 'my',  
'prediction', 'So', 'if', 'I', 'get', 'stoned', "I'm", 'just',  
'carrying', 'on', 'An', 'old', 'family', 'tradition' ]
```

TEXT PREPROCESSING

TEXT CLEANING PIPELINE

- + Tokenize by word
- + Remove punctuation
- + Lowercase
- + Remove stop words
- + Lemmatize text

EXAMPLE LYRICS

SINGER: HANK WILLIAMS JR.

TITLE: FAMILY TRADITION

```
['country', 'music', 'singers', 'have', 'always', 'been', 'a',
'real', 'close', 'family', 'but', 'lately', 'some', 'of', 'my',
'kinfolks', 'have', 'disowned', 'a', 'few', 'others', 'and', 'me',
'i', 'guess', 'because', 'i', 'kind', 'of', 'changed', 'my',
'direction', 'lord', 'i', 'guess', 'i', 'went', 'and', 'broke',
'their', 'family', 'tradition', 'they', 'get', 'on', 'me', 'and',
'want', 'to', 'know', 'hank', 'why', 'do', 'you', 'drink', 'hank',
'why', 'do', 'you', 'roll', 'smoke', 'why', 'must', 'you', 'live',
'out', 'the', 'songs', 'that', 'you', 'wrote', 'over', 'and',
'over', 'everybody', 'makes', 'my', 'prediction', 'so', 'if', 'i',
'get', 'stoned', 'just', 'carrying', 'on', 'an', 'old', 'family',
'tradition']
```

TEXT PREPROCESSING

TEXT CLEANING PIPELINE

- + Tokenize by word
- + Remove punctuation
- + Lowercase
- + **Remove stop words**
- + Lemmatize text

EXAMPLE LYRICS

SINGER: HANK WILLIAMS JR.

TITLE: FAMILY TRADITION

```
['country', 'music', 'singers', 'always', 'real', 'close', 'family',
'lately', 'kinfolks', 'disowned', 'others', 'guess', 'kind',
'changed', 'direction', 'lord', 'guess', 'went', 'broke',
'family', 'tradition', 'get', 'want', 'know', 'hank', 'drink',
'hank', 'roll', 'smoke', 'must', 'live', 'songs', 'wrote',
'everybody', 'makes', 'prediction', 'get', 'stoned',
'carrying', 'old', 'family', 'tradition', 'proud', 'name',
'although', 'kind', 'music', 'mine', 'exactly', 'stop', 'think',
'put', 'position', 'get', 'stoned', 'sing', 'night', 'long',
'family', 'tradition']
```

TEXT PREPROCESSING

TEXT CLEANING PIPELINE

- + Tokenize by word
- + Remove punctuation
- + Lowercase
- + Remove stop words
- + **Lemmatize text**

EXAMPLE LYRICS

SINGER: HANK WILLIAMS JR.

TITLE: FAMILY TRADITION

```
['country', 'music', 'singer', 'always', 'real', 'close', 'family',
'lately', 'kinfolk', 'disowned', 'others', 'guess', 'kind',
'changed', 'direction', 'lord', 'guess', 'went', 'broke',
'family', 'tradition', 'get', 'want', 'know', 'hank', 'drink',
'hank', 'roll', 'smoke', 'must', 'live', 'song', 'wrote',
'everybody', 'make', 'prediction', 'get', 'stoned', 'carrying',
'old', 'family', 'tradition', 'proud', 'name', 'although', 'kind',
'music', 'mine', 'exactly', 'stop', 'think', 'put', 'position',
'get', 'stoned', 'sing', 'night', 'long', 'family', 'tradition']
```

BEFORE

Country music singers
Have always been a real close family
But lately some of my kinfolks
Have disowned a few others and me
I guess it's because I kind of changed my direction
Lord, I guess I went and broke their family tradition

They get on me and want to know, Hank
Why do you drink?
Hank, why do you roll smoke?
Why must you live out the songs that you wrote?
Over and over
Everybody makes my prediction
So if I get stoned, I'm just carrying on an old family
tradition

AFTER

['country', 'music', 'singer', 'always',
'real', 'close', 'family', 'lately',
'kinfolk', 'disowned', 'others',
'guess', 'kind', 'changed', 'direction',
'lord', 'guess', 'went', 'broke',
'family', 'tradition', 'get', 'want',
'know', 'hank', 'drink', 'hank', 'roll',
'smoke', 'must', 'live', 'song', 'wrote',
'everybody', 'make', 'prediction',
'get', 'stoned', 'carrying', 'old',
'family', 'tradition', 'proud', 'name',
'although', 'kind', 'music', 'mine',
'exactly', 'stop', 'think', 'put',
'position', 'get', 'stoned', 'sing',
'night', 'long', 'family', 'tradition']

TOPIC MODEL: NON-NEGATIVE MATRIX FACTORIZATION WITH TFIDF VECTORIZATION

MODEL PARAMETERS

- + n_topics = 5
- + max_df = .6
- + min_df = .2
- + l1_ratio = .3
- + alpha = .05

TOPIC MODEL: NON-NEGATIVE MATRIX FACTORIZATION WITH TFIDF VECTORIZATION

MODEL PARAMETERS

- + n_topics = 5
- + max_df = .6
- + min_df = .2
- + l1_ratio = .3
- + alpha = .05

TOPIC 1

time, know, gone, home, think
heart, mind, long, cause, life, old

TOPIC 2

forever, heart, want, true, know
make, need, let, feel, world

TOPIC 3

song, sing, play, future, music,
hear, sad, write, memory, life

TOPIC 4

baby ,girl, wanna, gonna, good
right, need, crazy, feel, country

TOPIC 5

tonight, day, night, break, better,
shine, today, worry, alright, fine

TOPIC MODEL: NON-NEGATIVE MATRIX FACTORIZATION WITH TFIDF VECTORIZATION

MODEL PARAMETERS

- + n_topics = 5
- + max_df = .6
- + min_df = .2
- + l1_ratio = .3
- + alpha = .05

TOPIC 1

TOPIC 2

TOPIC 3

TOPIC 4

TOPIC 5

time, know, gone, home, think
heart, mind, long, cause, life, old

forever, heart, want, true, know
make, need, let, feel, world

song, sing, play, future, music,
hear, sad, write, memory, life

baby, girl, wanna, gonna, good
right, need, crazy, feel, country

tonight, day, night, break, better,
shine, today, worry, alright, fine

TOPIC MODEL: NON-NEGATIVE MATRIX FACTORIZATION WITH TFIDF VECTORIZATION

MODEL PARAMETERS

- + n_topics = 5
- + max_df = .6
- + min_df = .2
- + l1_ratio = .3
- + alpha = .05

IS COUNTRY DEPRESSING?

TOPIC 1

time, know, gone, home, think
heart, mind, long, cause, life, old

TOPIC 2

forever, heart, want, true, know
make, need, let, feel, world

TOPIC 3

song, sing, play, future, music,
hear, sad, write, memory, life

TOPIC 4

baby, girl, wanna, gonna, good
right, need, crazy, feel, country

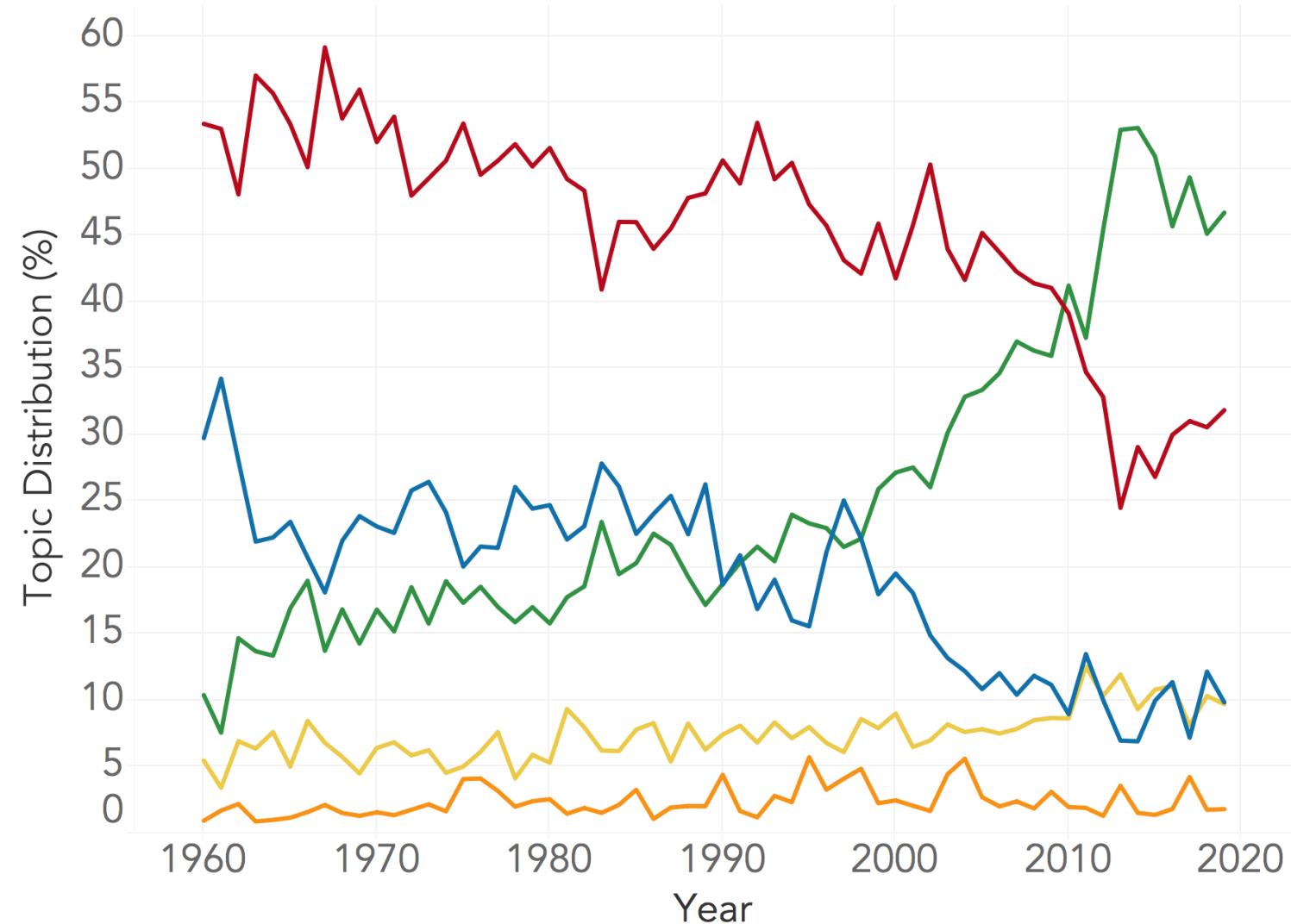
TOPIC 5

tonight, day, night, break, better,
shine, today, worry, alright, fine

NOT ANYMORE?

- TOPIC 1** time, know, gone, home, think
heart, mind, long, cause, life, old
- TOPIC 2** forever, heart, want, true, know
make, need, let, feel, world
- TOPIC 3** song, sing, play, future, music,
hear, sad, write, memory, life
- TOPIC 4** baby ,girl, wanna, gonna, good
right, need, crazy, feel, country
- TOPIC 5** tonight, day, night, break, better,
shine, today, worry, alright, fine

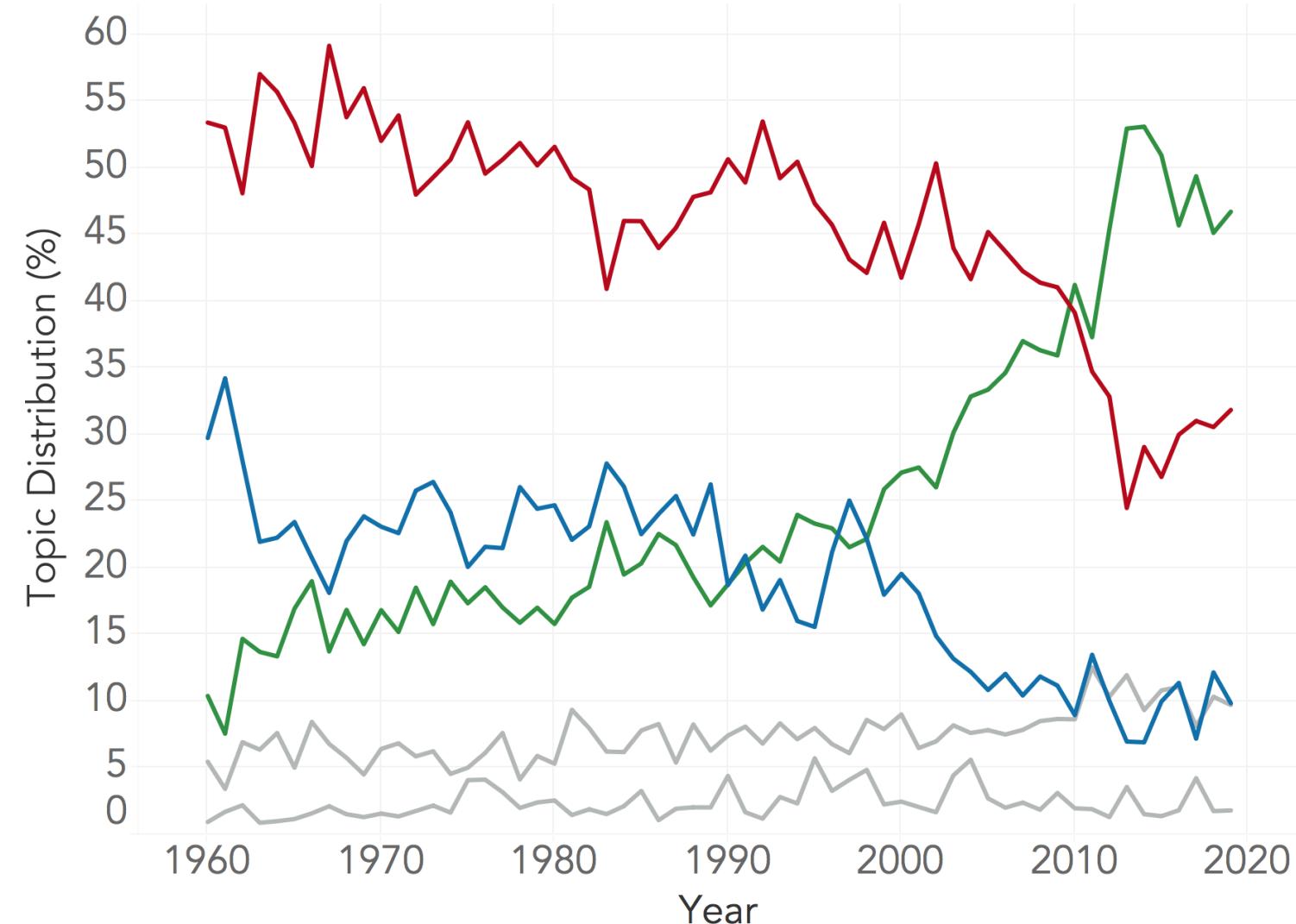
COUNTRY SONG TOPICS HAVE SHIFTED OVER 60 YEARS



NOT ANYMORE?

- TOPIC 1** time, know, gone, home, think
heart, mind, long, cause, life, old
- TOPIC 2** forever, heart, want, true, know
make, need, let, feel, world
- TOPIC 3** song, sing, play, future, music,
hear, sad, write, memory, life
- TOPIC 4** baby ,girl, wanna, gonna, good
right, need, crazy, feel, country
- TOPIC 5** tonight, day, night, break, better,
shine, today, worry, alright, fine

COUNTRY SONG TOPICS HAVE SHIFTED OVER 60 YEARS



LYRICS EXAMPLE

SINGER: JIM REEVES

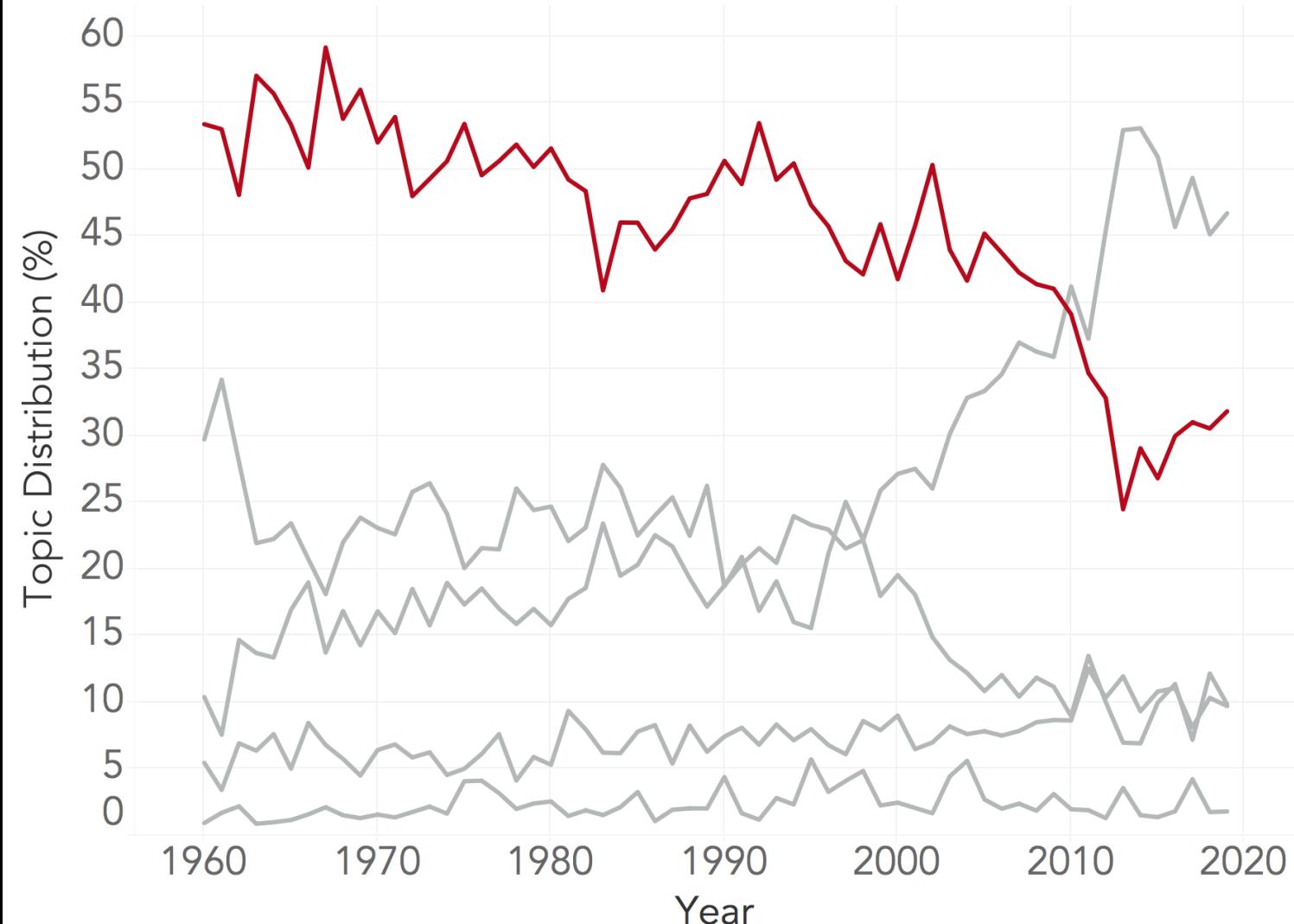
TITLE: I'M GETTIN' BETTER

YEAR: 1960

I talked with an old friend,
He asked about you
I managed to tell him
With just a tear or two
I'm gettin' better, I'm gettin' better

I know I'll never be the same
I was standin' too close to the flame
But it doesn't hurt so much
To hear your name
I think I'll make it, I'm gettin' better

TOPIC 1 time, know, gone, home, think
heart, mind, long, cause, life, old



LYRICS EXAMPLE

SINGER: LUKE COMB

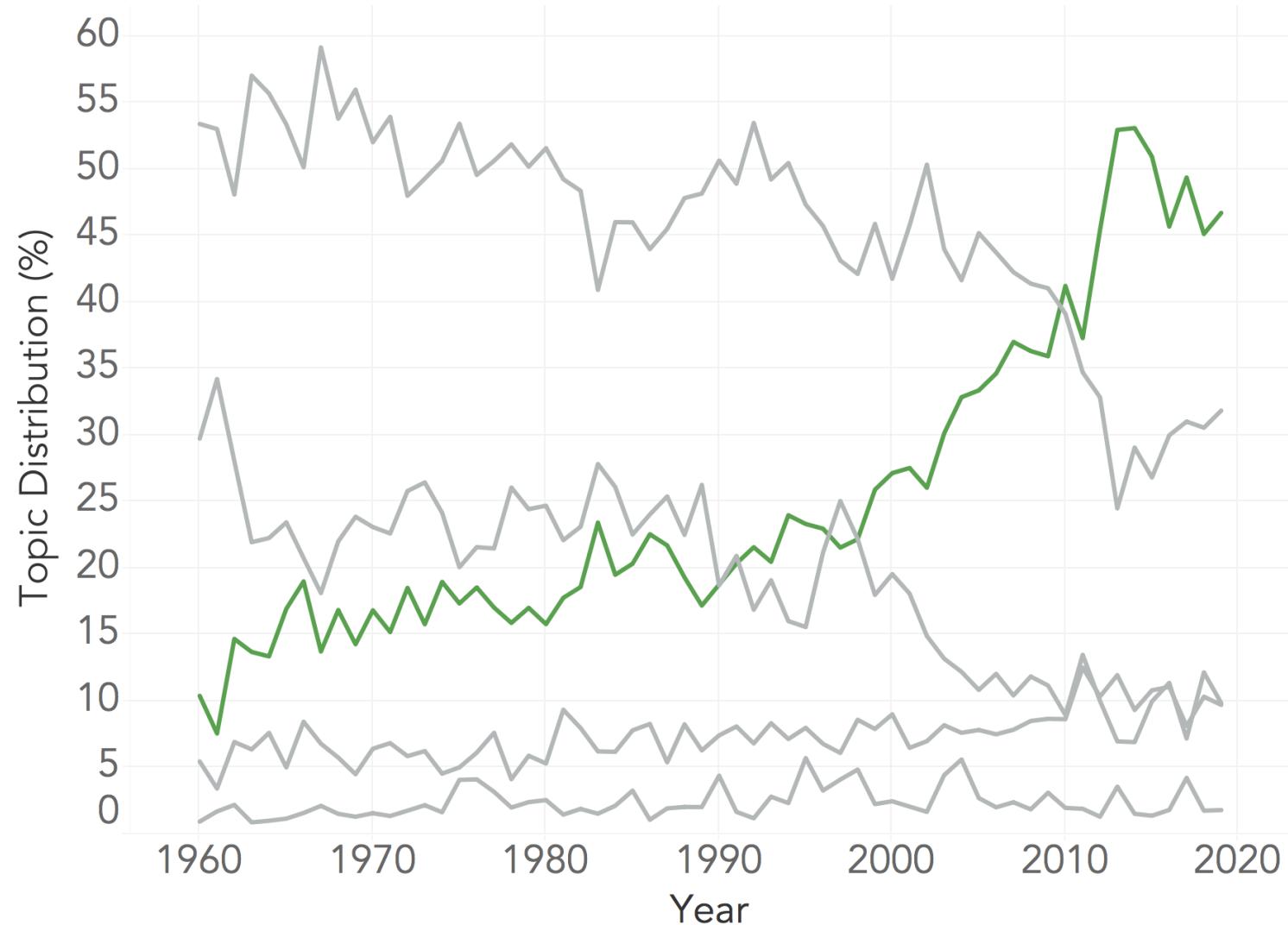
TITLE: A LONG WAY

YEAR: 2019

Six beers, five guys
And if I remember right
It was a Wednesday when we got our buzz
A second-hand six pack
Half smoked, soft pack
Light it up, boys take a drag

One girl, and 45 bucks
was all it took that night
Hit the steak house, make-out in the movies
Way up in the back
Credits roll, on pace
To make it all the way to second base
Tried to play it cool like it was nothing new
And she didn't have clue

TOPIC 4 baby, girl, wanna, gonna, good
right, need, crazy, feel, country



TOPIC MODEL: ANCHORED COREX WITH COUNT VECTORIZATION

PARAMETER TUNING

n_topics = 5

anchor strength = 10

anchors =

TOPIC MODEL: ANCHORED COREX WITH COUNT VECTORIZATION

PARAMETER TUNING

n_topics = 5

anchor strength = 10

anchors = [LOVE]

SEMI-SUPERVISED LEARNING WITH ANCHORS PROVIDES CLEAN TOPICS

TOPIC 1	LOVE, town, country, truck, boy, old, beer, heart, big, car
TOPIC 2	sky, wind, sun, river, summer, tree, rain, moon, field, star
TOPIC 3	said, daddy, man, year, wife, kid, son, went, took, came
TOPIC 4	yeah, got, little, gonna, girl, baby, wanna, hey, like, gotta
TOPIC 5	red, white, west, pulled, born, mud, american, band, song, people

TOPIC MODEL: ANCHORED COREX WITH COUNT VECTORIZATION

PARAMETER TUNING

n_topics = 5

anchor strength = 10

anchors = [LOVE]

[FAMILY]

SEMI-SUPERVISED LEARNING WITH ANCHORS PROVIDES CLEAN TOPICS

- TOPIC 1 LOVE, town, country, truck, boy, old, beer, heart, big, car
- TOPIC 2 FAMILY, daddy, kid, son, man, year, wife, dad, mama, school
- TOPIC 3 said, took, looked, saw, went, walked, heard, came, told, turned
- TOPIC 4 yeah, got, little, gonna, girl, baby, wanna, like, gotta, hey
- TOPIC 5 ride, road, wind, summer, sun, radio, tree, song, sky, red

TOPIC MODEL: ANCHORED COREX WITH COUNT VECTORIZATION

PARAMETER TUNING

n_topics = 5

anchor strength = 10

anchors = [LOVE]

[FAMILY]

[SUMMER]

SEMI-SUPERVISED LEARNING WITH ANCHORS PROVIDES CLEAN TOPICS

TOPIC 1

LOVE, town, country, truck, boy,
old, beer, heart, big, car

TOPIC 2

FAMILY, daddy, kid, son, man,
year, wife, dad, mama, school

TOPIC 3

SUMMER, truck, country, boy, big,
car, road, hot, dog, city

TOPIC 4

yeah, got, little, gonna, girl,
baby, wanna, like, hey, gotta

TOPIC 5

song, ride, guitar, texas, cowboy,
red, wind, mile, star, highway

TOPIC MODEL: ANCHORED COREX WITH COUNT VECTORIZATION

PARAMETER TUNING

n_topics = 5

anchor strength = 10

anchors = [LOVE]

[FAMILY]

[SUMMER]

[GUITAR]

SEMI-SUPERVISED LEARNING WITH ANCHORS PROVIDES CLEAN TOPICS

TOPIC 1

LOVE, town, country, truck, boy,
old, beer, heart, big, car

TOPIC 2

FAMILY, daddy, kid, son, man,
year, wife, dad, mama, school

TOPIC 3

SUMMER, truck, country, boy, big,
car, road, hot, dog, city

TOPIC 4

GUITAR, country, old, cowboy, city,
song, boot, texas, rock, radio

TOPIC 5

yeah, got, gonna, baby, wanna,
gotta, hey, like, bout, right

TOPIC MODEL: ANCHORED COREX WITH COUNT VECTORIZATION

PARAMETER TUNING

n_topics = 8

anchor strength = 10

anchors = [LOVE]

[FAMILY]

[SUMMER]

[GUITAR]

[BABY]

[BEER]

[GOD]

TOPIC 1	LOVE, heart, beer, bar, town, drink, car, true, em, home
TOPIC 2	FAMILY, daddy, said, kid, son, man, boy, wife, year, mama
TOPIC 3	SUMMER, road, sky, ride, field, white, tree, hot, sun, south
TOPIC 4	GUITAR, country, song, old, cowboy, town, city, radio, texas, play
TOPIC 5	BABY, crazy, kiss, ooh, oh, honey, let, need, babe, resist
TOPIC 6	BEER, got, little, truck, boy, car, big, hey, bar, em
TOPIC 7	GOD, man, son, church, died, thank, hand, war, daughter, jesus yeah, girl, gonna, like, wanna, gotta, right, bout, night, drive
TOPIC 8	

EXPLORING LOVE IN COUNTRY MUSIC

PARAMETER TUNING

n_topics = 7

anchor strength = 5

anchors = [LOVE]
[LOVE]
[LOVE]
[LOVE]

TOPIC 1

TOPIC 2

TOPIC 3

TOPIC 4

TOPIC 5

TOPIC 6

TOPIC 7

LOVE, heart, dog, sun, train,
wind, tree, hold, west, round

LOVE, lover, sweet, locked, touch,
stacked, live, power, world, self

LOVE, truck, country, car, little,
town, old, big, road, money

LOVE, yeah, bar, drink, got,
beer, song, whiskey, bottle, hit

baby, wanna, gonna, girl, right,
gotta, like, make, crazy, ya

daddy, said, son, boy, man,
kid, mama, year, wife, dad

hair, jimmy, elvis, mary, clean,
suit, rich, senorita, king, la

CONCLUSION

COUNTRY != BRO COUNTRY

GOING FORWARD

- MODEL INDIVIDUAL ARTISTS
- VISUALIZE SIMILARITY
- SENTIMENT ANALYSIS
- MORE DATA, MORE FEATURES
- WORD2VEC

NMF

- TOPIC 1** time, know, gone, home, think
heart, mind, long, cause, life, old
- TOPIC 2** forever, heart, want, true, know
make, need, let, feel, world
- TOPIC 3** song, sing, play, future, music,
hear, sad, write, memory, life
- TOPIC 4** baby ,girl, wanna, gonna, good
right, need, crazy, feel, country
- TOPIC 5** tonight, day, night, break, better,
shine, today, worry, alright, fine

COREX

- TOPIC 1** **LOVE**, town, country, truck, boy,
old, beer, heart, big, car
- TOPIC 2** **FAMILY**, daddy, kid, son, man,
year, wife, dad, mama, school
- TOPIC 3** **SUMMER**, truck, country, boy, big,
car, road, hot, dog, city
- TOPIC 4** **GUITAR**, country, old, cowboy, city,
song, boot, texas, rock, radio
- TOPIC 5** yeah, got, gonna, baby, wanna,
gotta, hey, like, bout, right

APPENDIX

TOPIC MODEL: ANCHORED COREX WITH COUNT VECTORIZATION

PARAMETER TUNING

n_topics = 5

anchor strength = none

anchors = none

BASE MODEL – NO ANCHORING

- | | |
|---------|---|
| TOPIC 1 | boy, daddy, old, kid, son,
big, town, money, love, mama |
| TOPIC 2 | wanna, drink, little, gonna, truck,
girl, hey, baby, like, gotta |
| TOPIC 3 | summer, tree, sun, bass, wild,
hair, hay, train, air, sky |
| TOPIC 4 | country, beer, song, car, road,
radio, guitar, bar, ride, cowboy |
| TOPIC 5 | said, man, came, year, went,
took, wife, looked, saw, turned |

TOPIC MODEL: ANCHORED COREX WITH COUNT VECTORIZATION

PARAMETER TUNING

n_topics = 5

anchor strength = none

anchors = none

COREX PROVIDES CLEAN DEFINED TOPICS

FAMILY	boy, daddy, old, kid, son, big, town, money, love, mama
YOUNG LOVE	wanna, drink, little, gonna, truck, girl, hey, baby, like, gotta
LAND	summer, tree, sun, bass, wild, hair, hay, train, air, sky
POP/BRO COUNTRY	country, beer, song, car, road, radio, guitar, bar, ride, cowboy
MISC	said, man, came, year, went, took, wife, looked, saw, turned

TOPIC MODEL: NON-NEGATIVE MATRIX FACTORIZATION WITH TFIDF VECTORIZATION

MODEL PARAMETERS

- + n_topics = 5
- + max_df = .6
- + min_df = .2
- + l1_ratio = .3
- + alpha = .05

TOPIC 1

TOPIC 2

TOPIC 3

TOPIC 4

TOPIC 5

time, know, gone, home, think
heart, mind, long, cause, life, old

forever, heart, want, true, know
make, need, let, feel, world

song, sing, play, future, music,
hear, sad, write, memory, life

baby, girl, wanna, gonna, good
right, need, crazy, feel, country

tonight, day, night, break, better,
shine, today, worry, alright, fine

TOPIC MODEL: NON-NEGATIVE MATRIX FACTORIZATION WITH TFIDF VECTORIZATION

MODEL PARAMETERS

- + n_topics = 5
- + max_df = .6
- + min_df = .2
- + l1_ratio = .3
- + alpha = .05

TOPIC 1

time, know, gone, home, think
heart, mind, long, cause, life, old

TOPIC 2

forever, heart, want, true, know
make, need, let, feel, world

TOPIC 3

song, sing, play, future, music,
hear, sad, write, memory, life

TOPIC 4

baby, girl, wanna, gonna, good
right, need, crazy, feel, country

TOPIC 5

tonight, day, night, break, better,
shine, today, worry, alright, fine