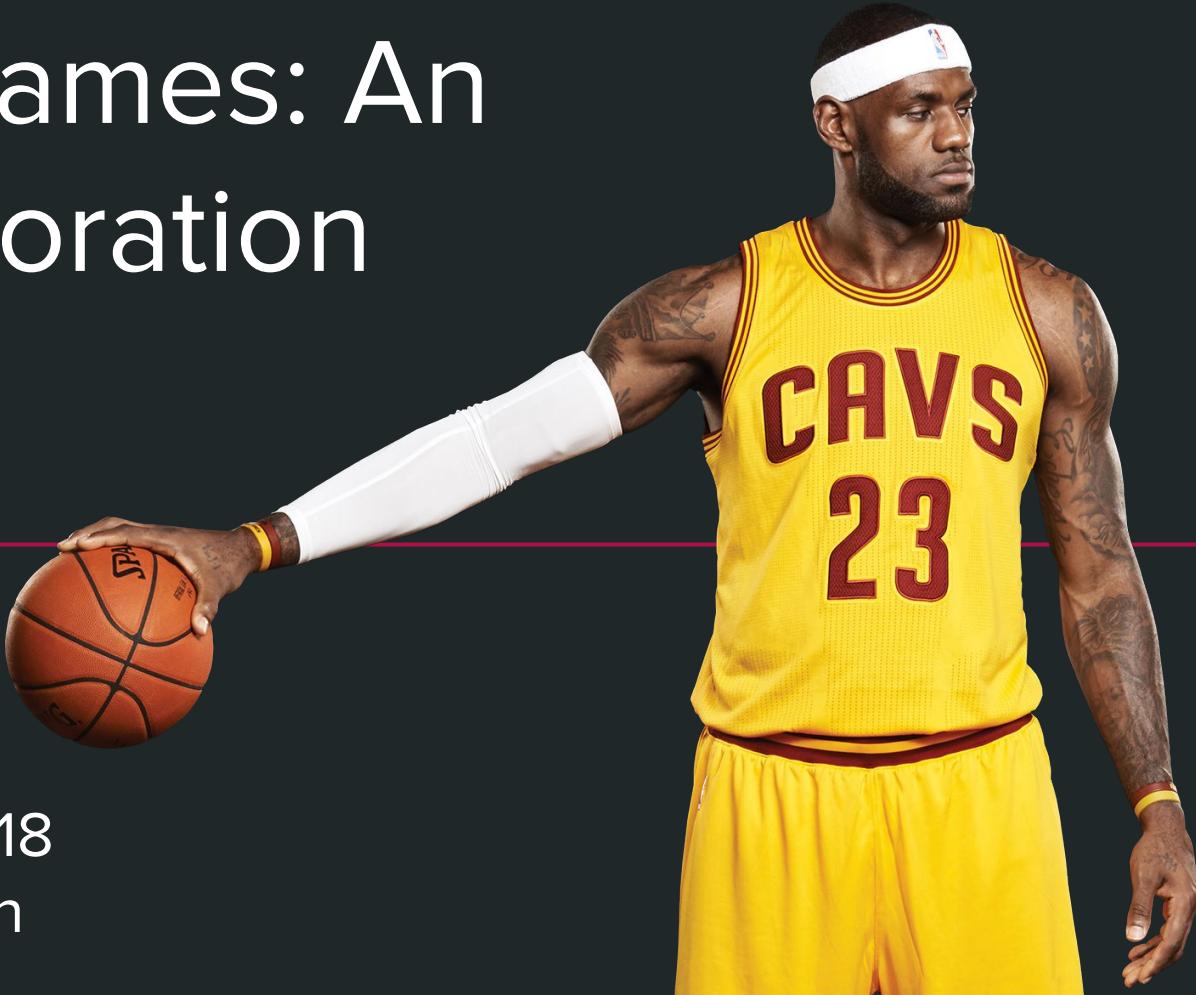
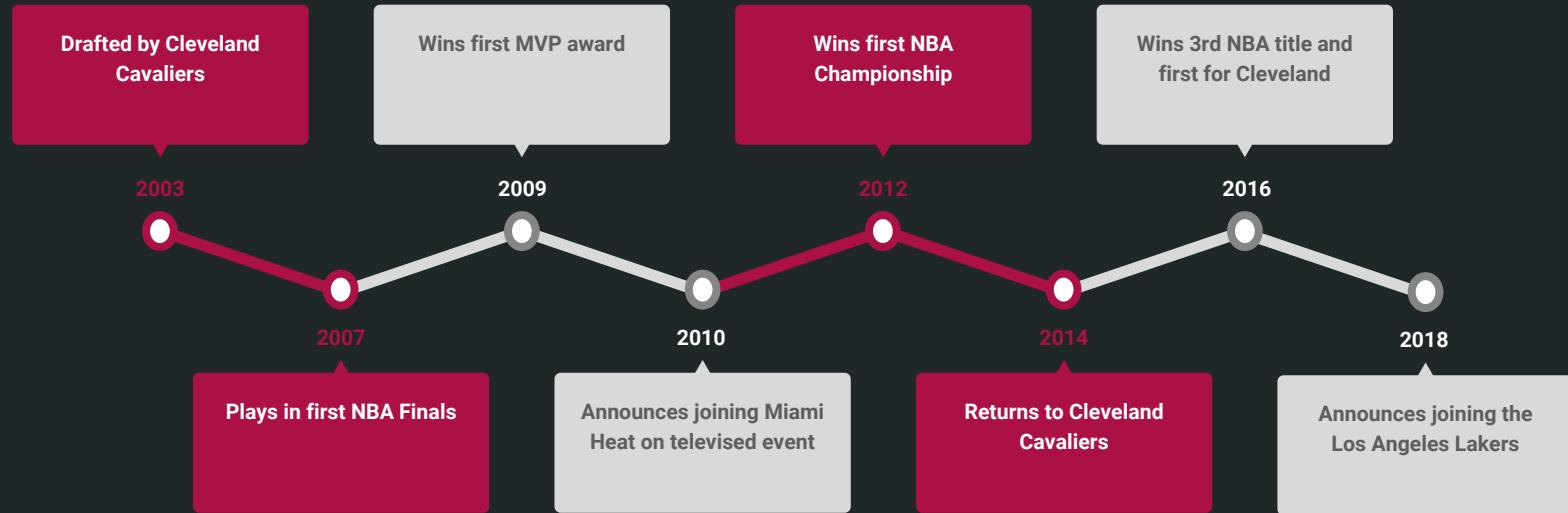


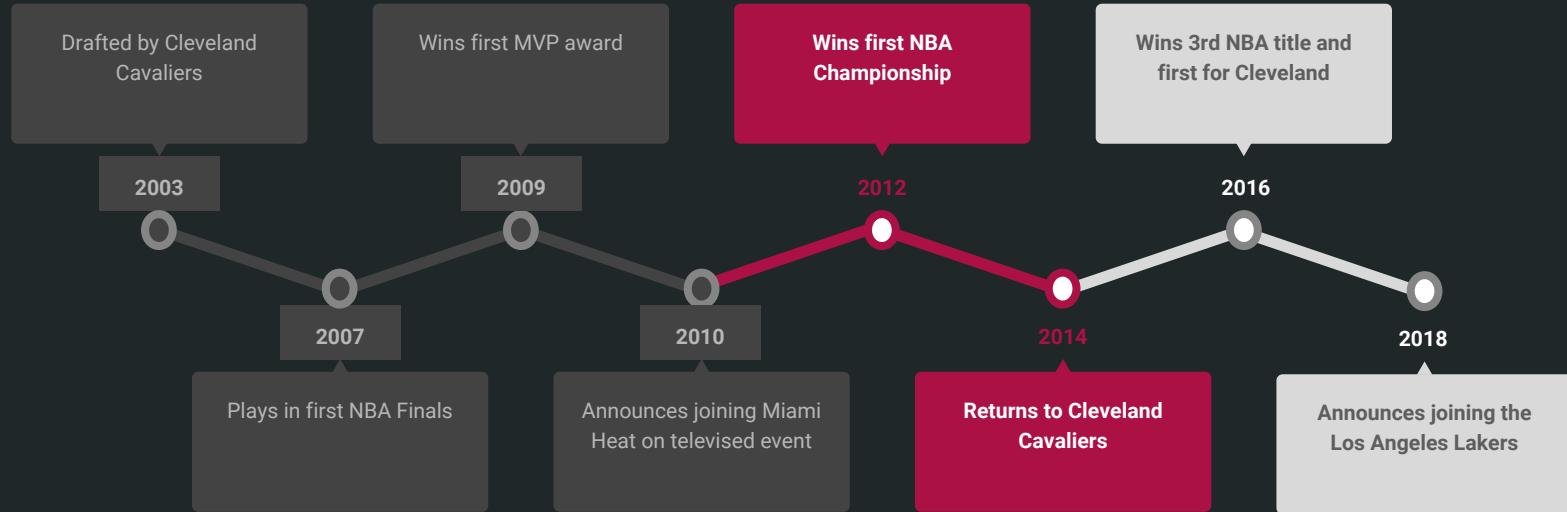
LeBron James: An NLP Exploration



November 16, 2018
Spencer Tollefson



2011 - 2018



Data

/r/nba

45,000 comments

2011-2018

Comment criteria:

- “LeBron” in title of submission
- Top 10 comments per submission
- At least 20 upvotes

Extract comments reflecting “popular opinion”
about LeBron



reddit

LDA Derived Topics

Words

- Best, Player, League

Comment

- lebron james best player in the league since lebron james

Words

- Free, Throw, Ball

Comment

- is he going to go back to his old free throw form instead of pretending to be ray allen

Words

- High, School, League

Comment

- he had his high school games on pay per view i don't know anyone like that

Words

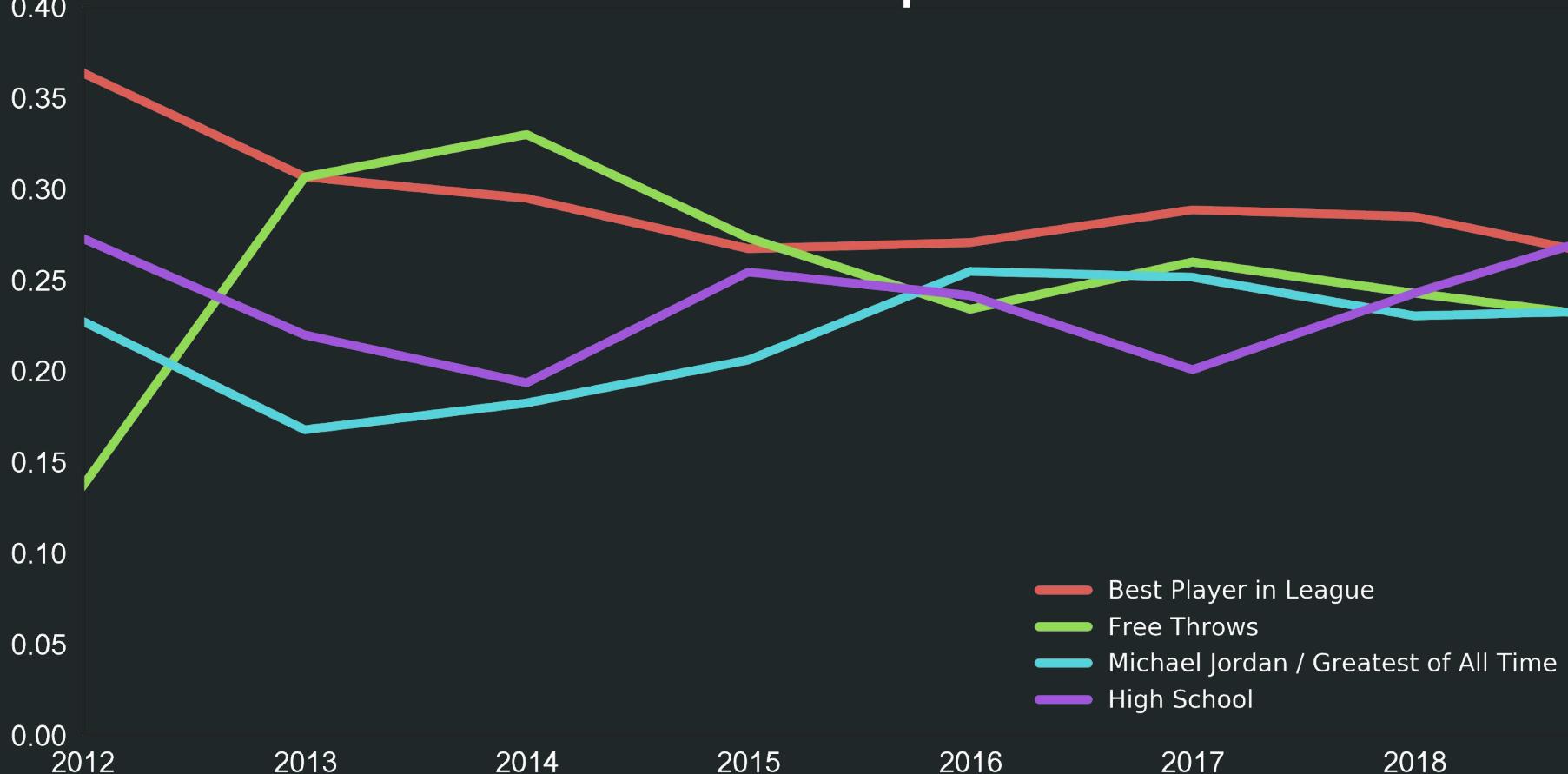
- GOAT, Jordan, Curry

Comment

- it's pretty obvious to me that jordan and lebron are and all time but apparently not to some people

Proportion
of Total

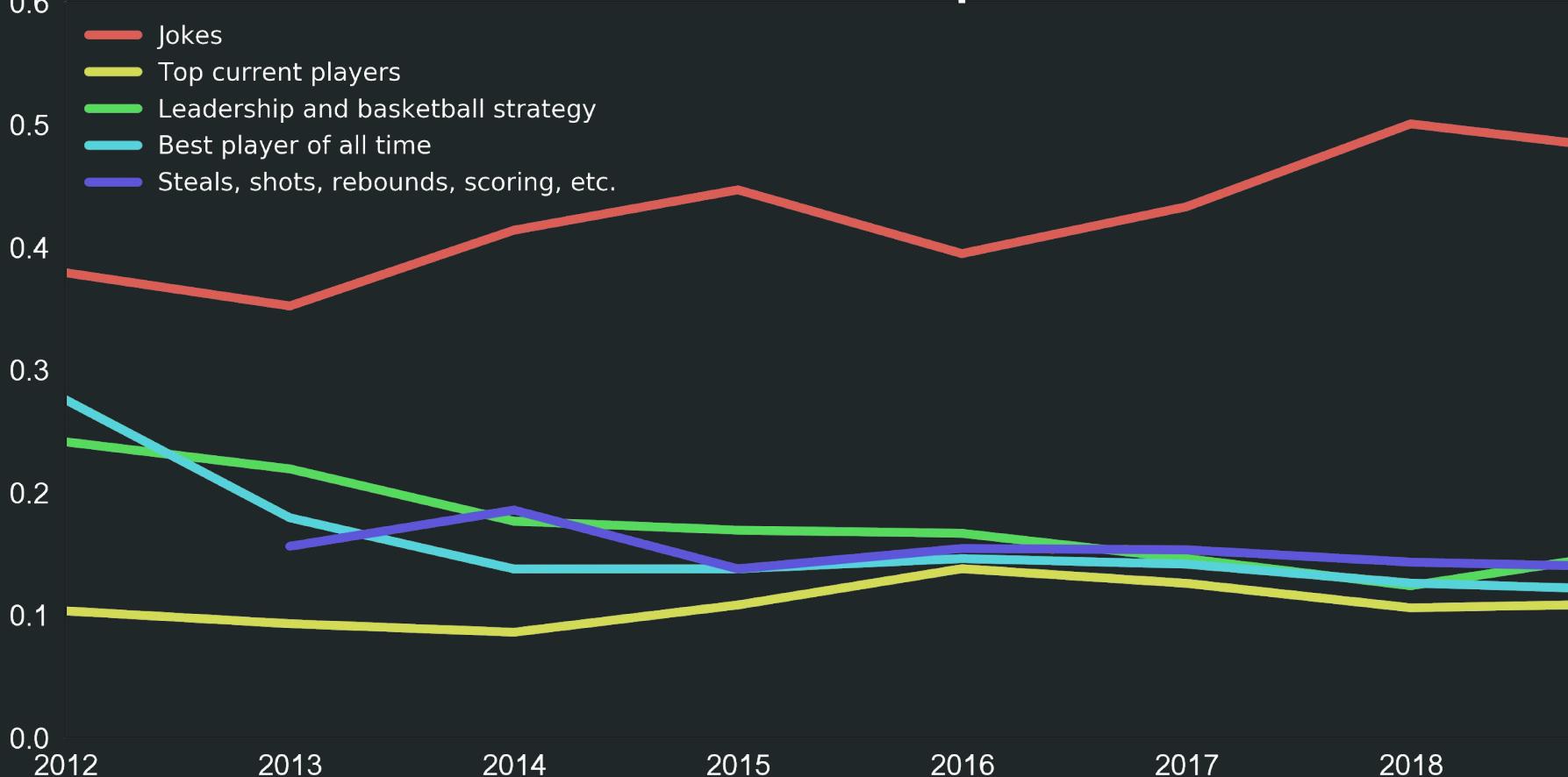
LDA Topics



Proportion
of Total
0.6

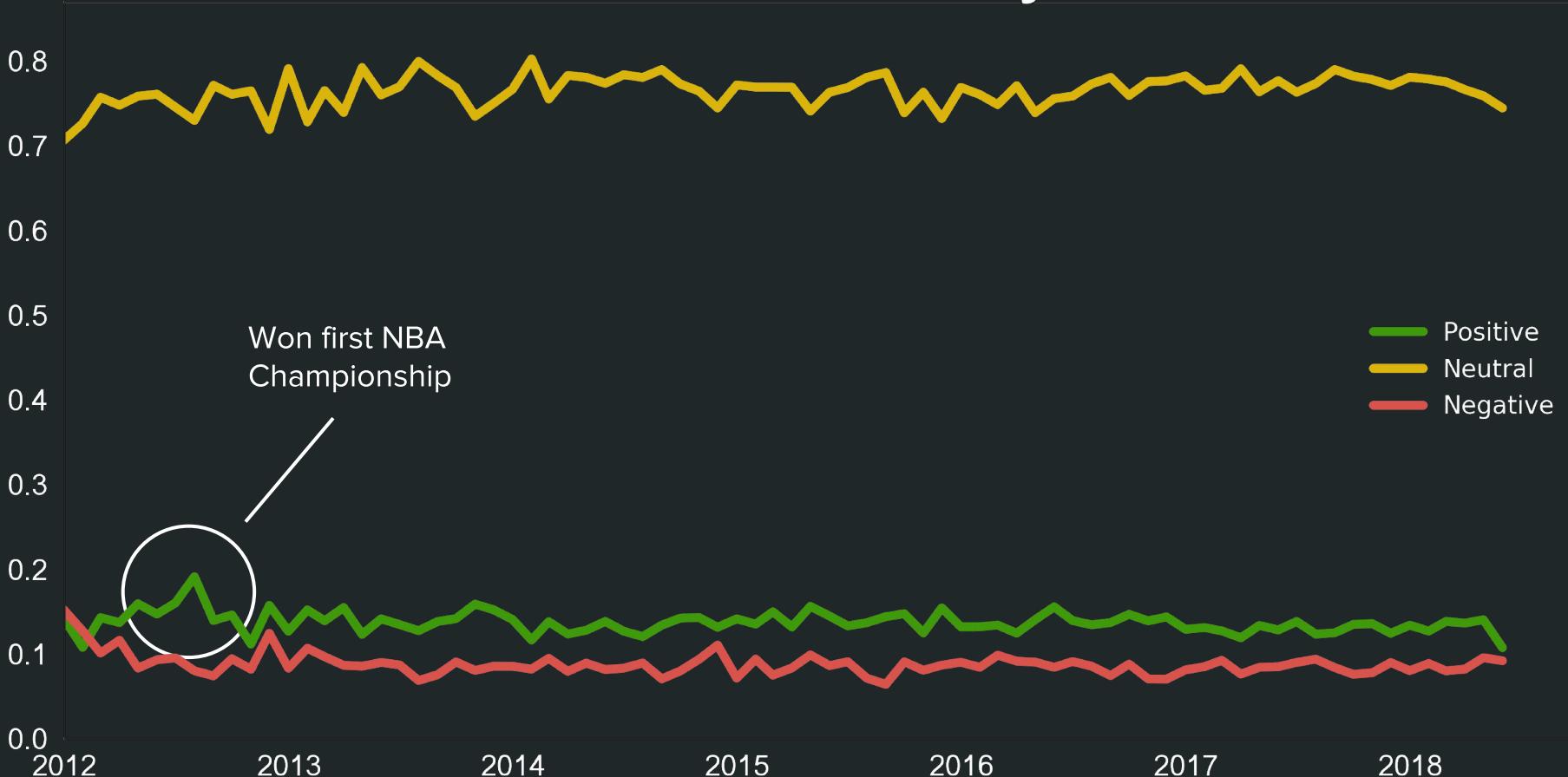
KMeans Topics

- Jokes
- Top current players
- Leadership and basketball strategy
- Best player of all time
- Steals, shots, rebounds, scoring, etc.



Proportion
of Total

Sentiment Analysis



Conclusion

- Derived LDA topics that match intuition
- While generally flat, some topic relative frequency changed with time

And...

Did not see anything too unsurprising

Future Work

- Improve topic extraction
- Compare known real-world events to data
- Change data
 - Restrict to lengthier comments
 - Not only LeBron related

Thank You



spencertollefson.com



spencer@spencertollefson.com



github.com/spencertollefson

Acknowledgements

Vader Sentiment Analyzer

- Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

Appendix

LDA Topics

Words

- Best, Player, League

Comments

- lebron james **best player** in the **league** since lebron james
- maybe if drake became the second **best player** in the league

Words

- Free, Throw, Ball

Comments

- **free throws** are too close for him is there a rule on how far back you can stand...
- is he going to go back to his old **free throw** form instead of pretending to be ray allen

Words

- High, School, League

Comments

- he had his **high school** games on pay per view i don't know anyone like that
- that's what happens when you are immune to every type of injury and came into the nba straight from **high school**

Words

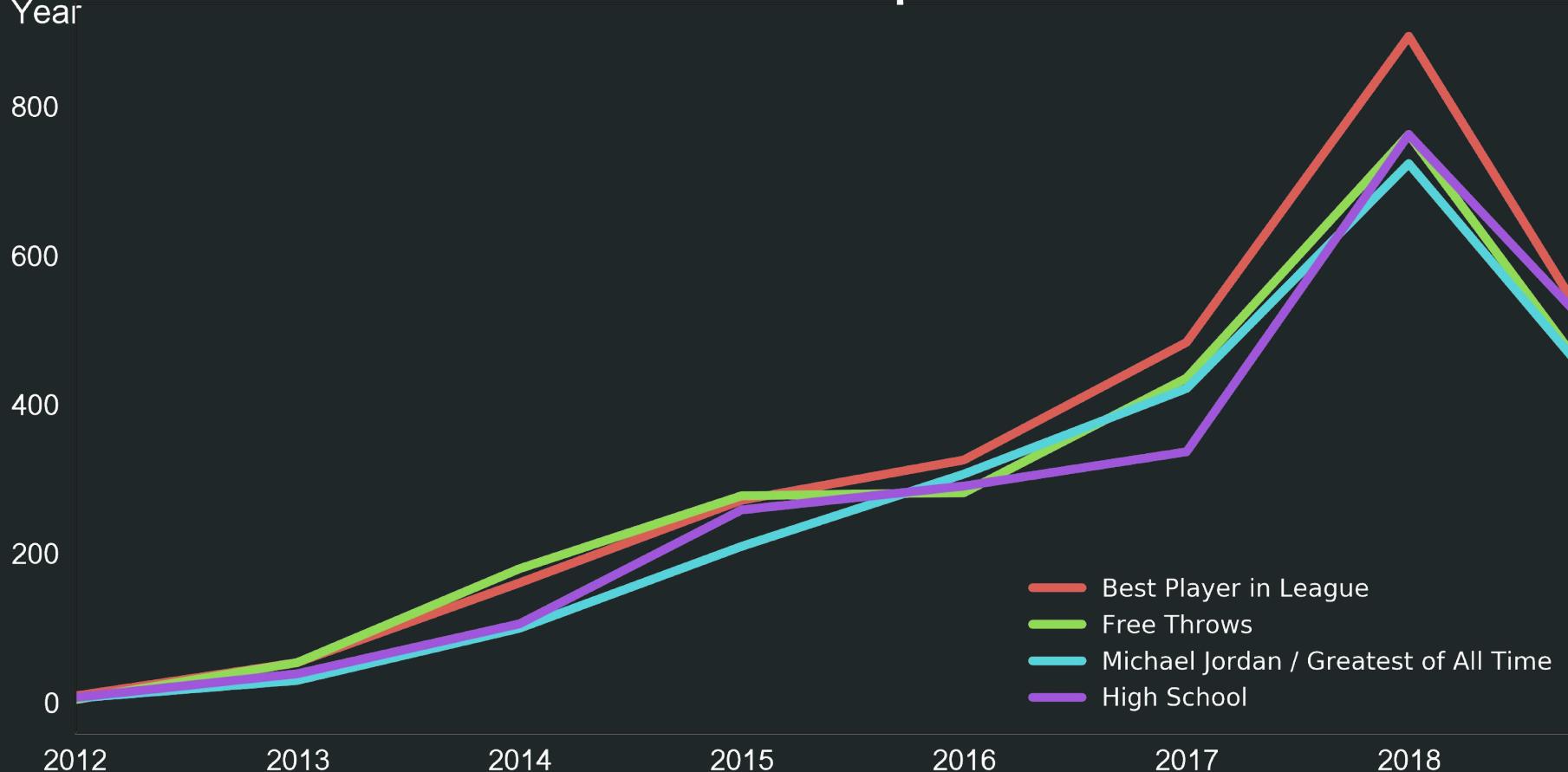
- GOAT, Jordan, Curry

Comments

- **jordan** would've had million by now all of his followers would have at least million **jordan goat**
- it's pretty obvious to me that **jordan** and lebron are and all time but apparently not to some people

Count
per
Year

LDA Topics

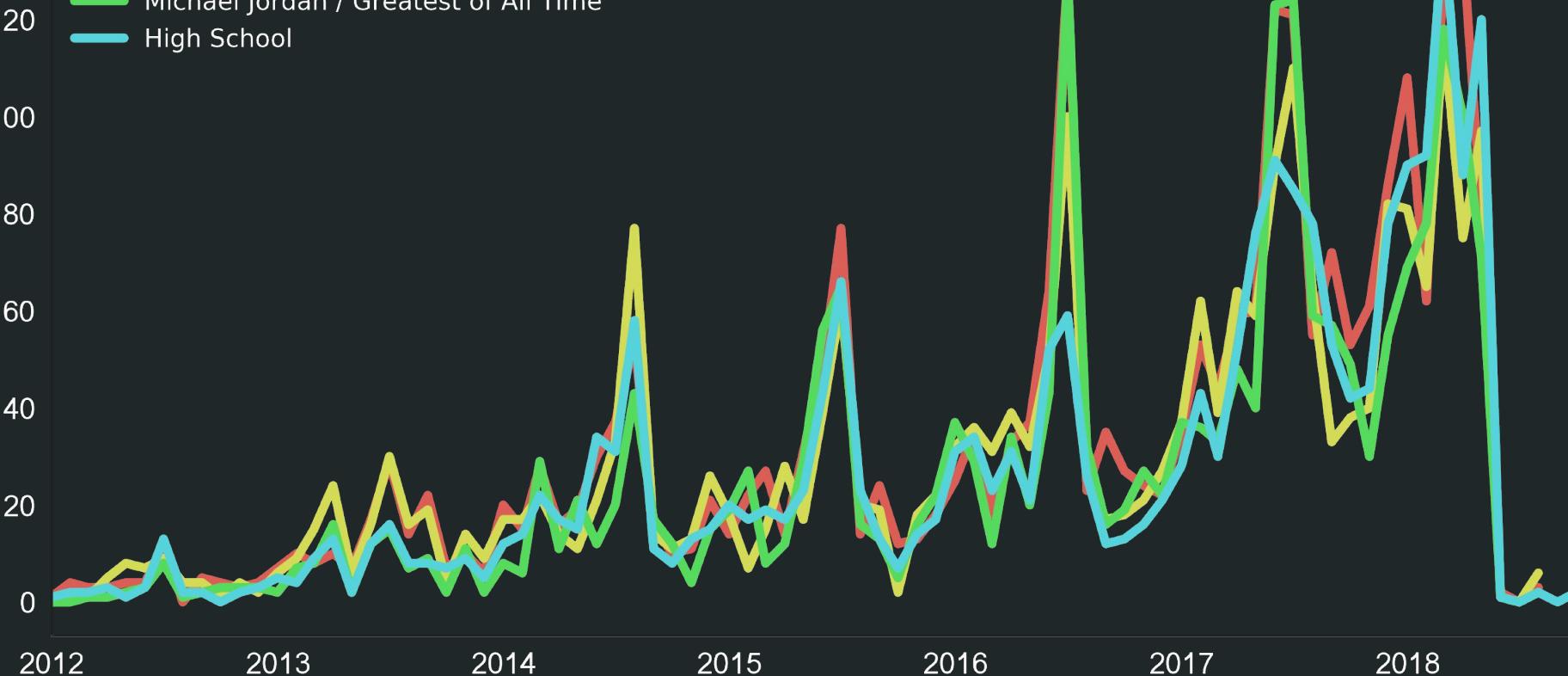


LDA Topics by Month

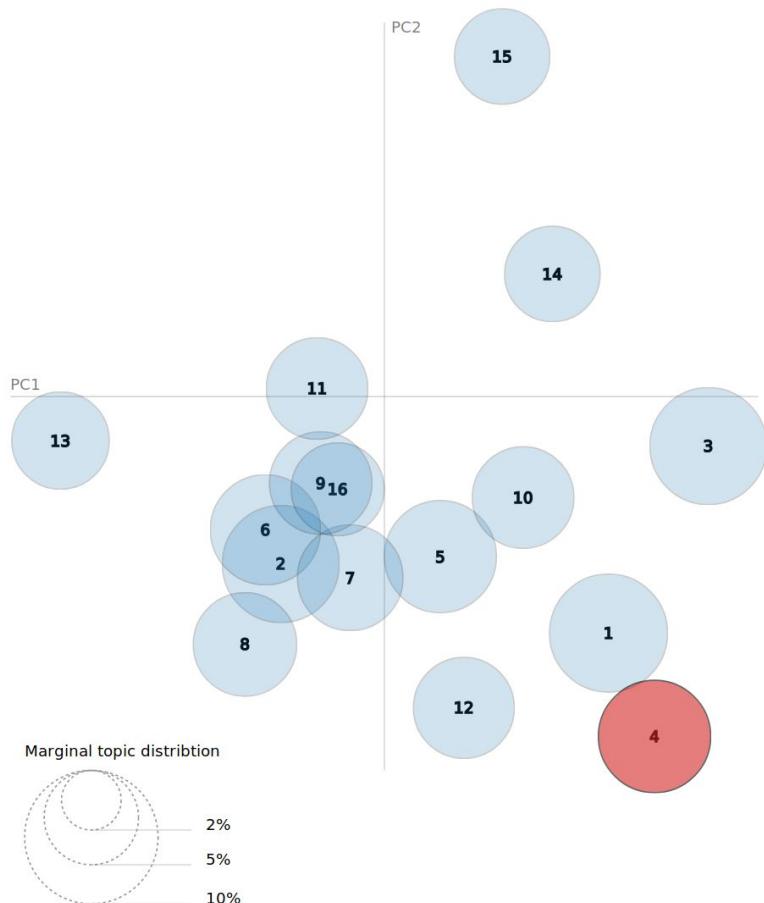
Appendix

Count
per
Month

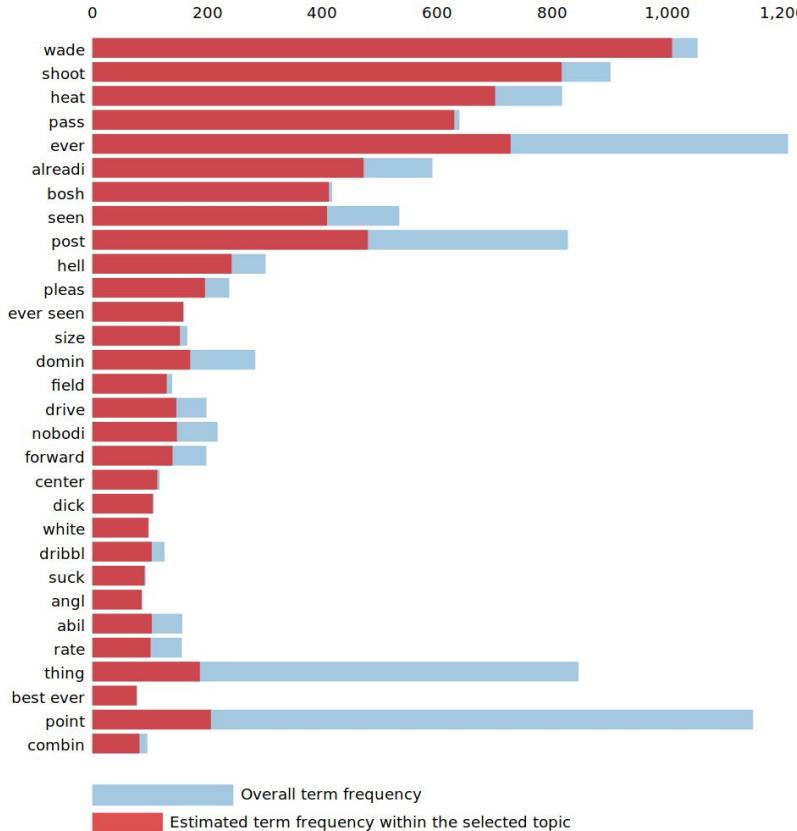
- Best Player in League
- Free Throws
- Michael Jordan / Greatest of All Time
- High School



Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 4 (7.1% of tokens)

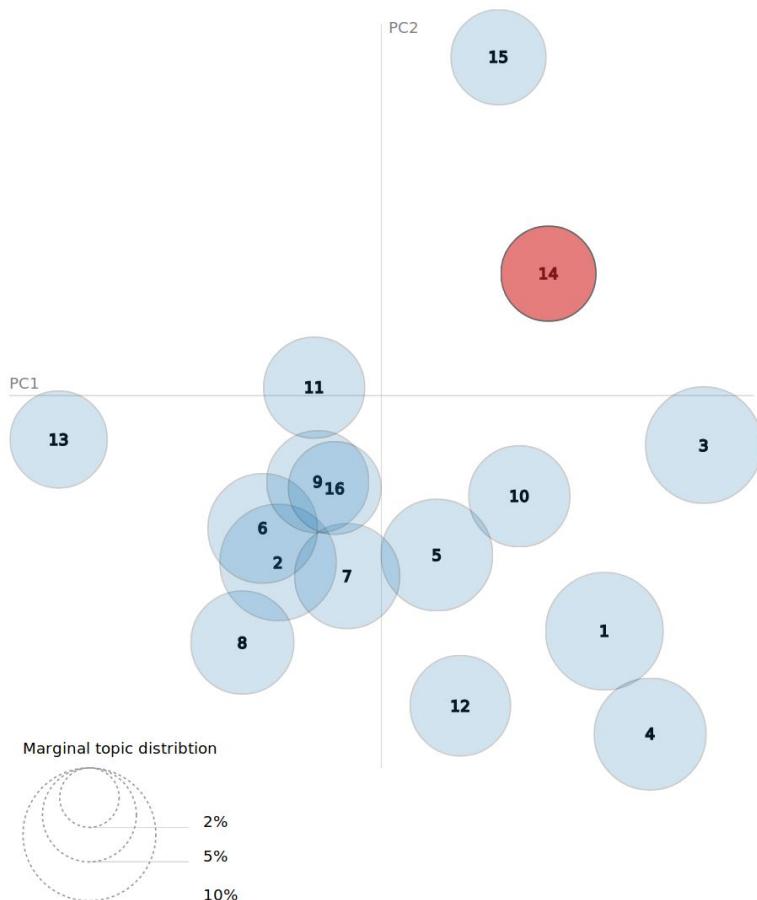


Overall term frequency

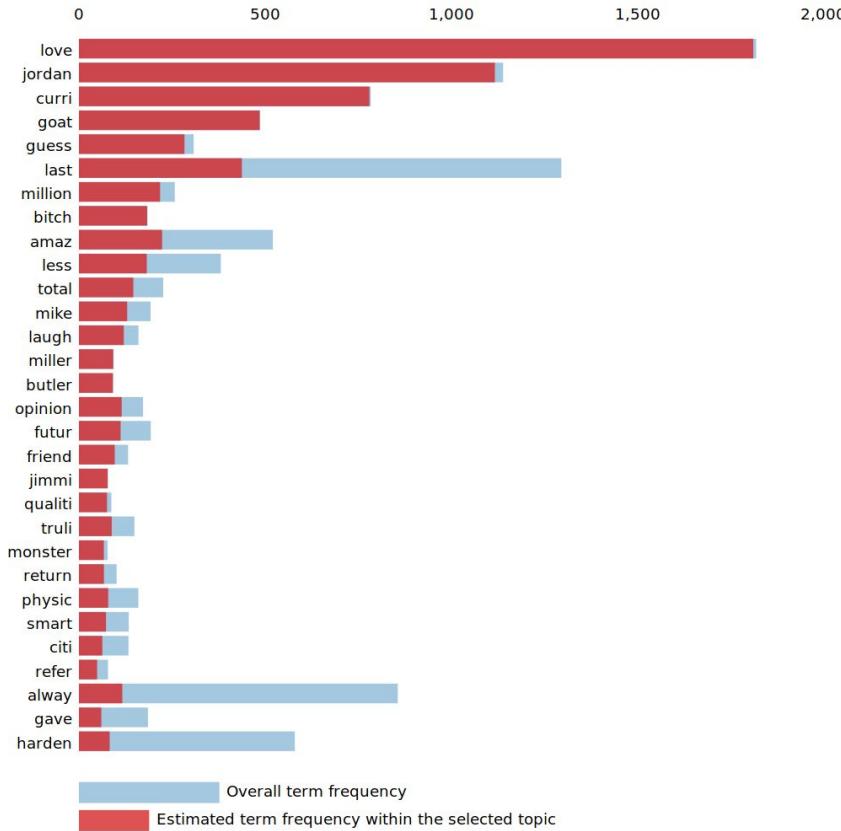
Estimated term frequency within the selected topic

1. $\text{salinity}(\text{term } w) = \text{frequency}(w) * [\sum_t p(t | w) * \log(p(t | w)/p(t))]$ for topics t ; see Chuang et. al.
2. $\text{relevance}(\text{term } w | \text{topic } t) = \lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

Intertopic Distance Map (via multidimensional scaling)

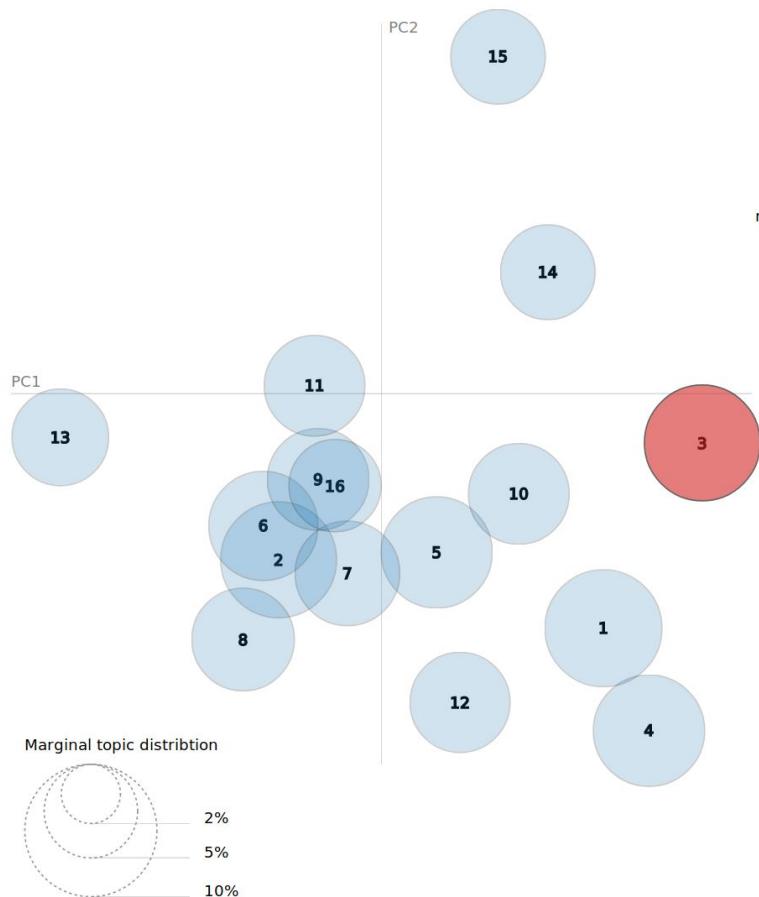


Top-30 Most Relevant Terms for Topic 14 (5.1% of tokens)

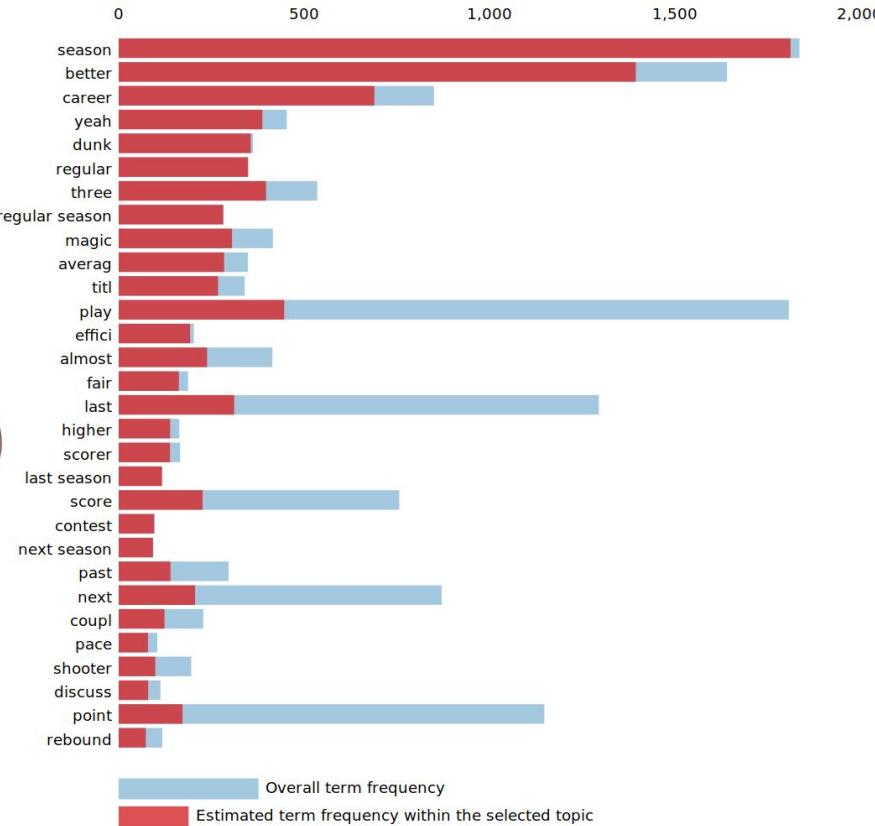


1. $\text{salency}(\text{term } w) = \text{frequency}(w) * [\sum_t p(t | w) * \log(p(t | w)/p(t))]$ for topics t; see Chuang et. al.
2. $\text{relevance}(\text{term } w | \text{topic } t) = \lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

Intertopic Distance Map (via multidimensional scaling)

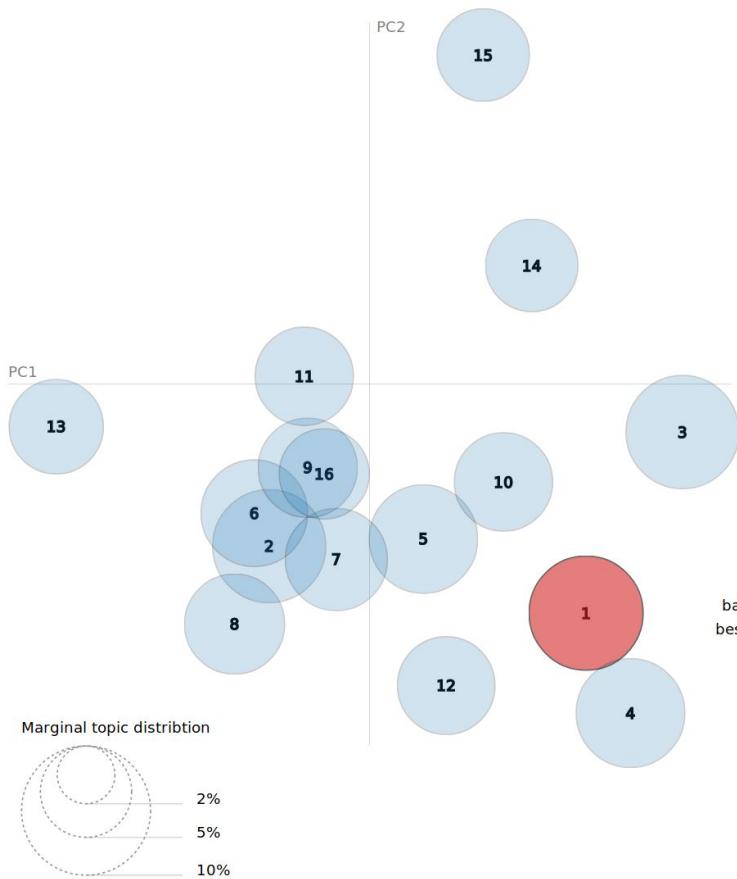


Top-30 Most Relevant Terms for Topic 3 (7.7% of tokens)

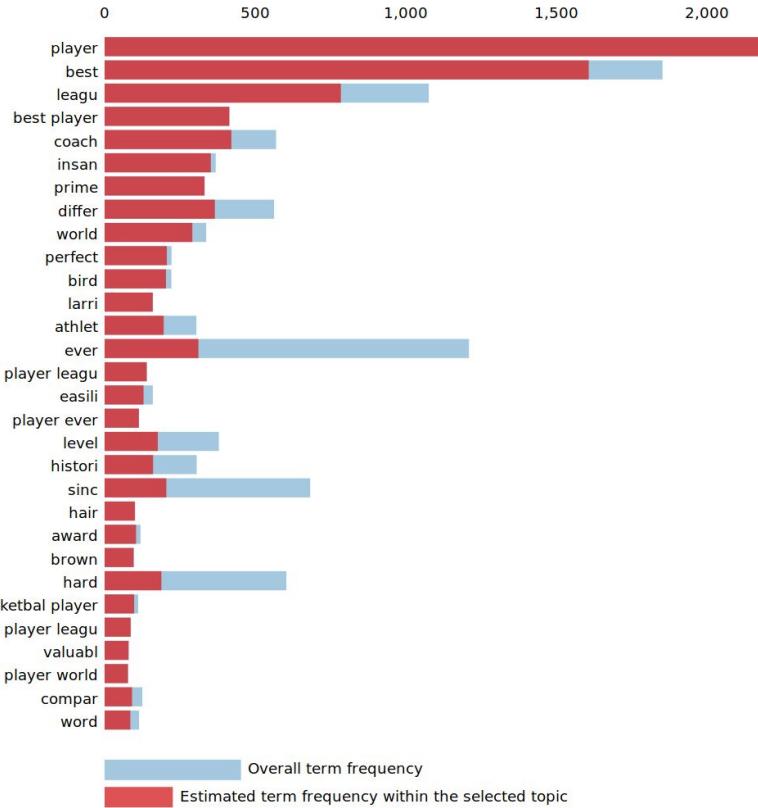


1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al
 2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

Intertopic Distance Map (via multidimensional scaling)



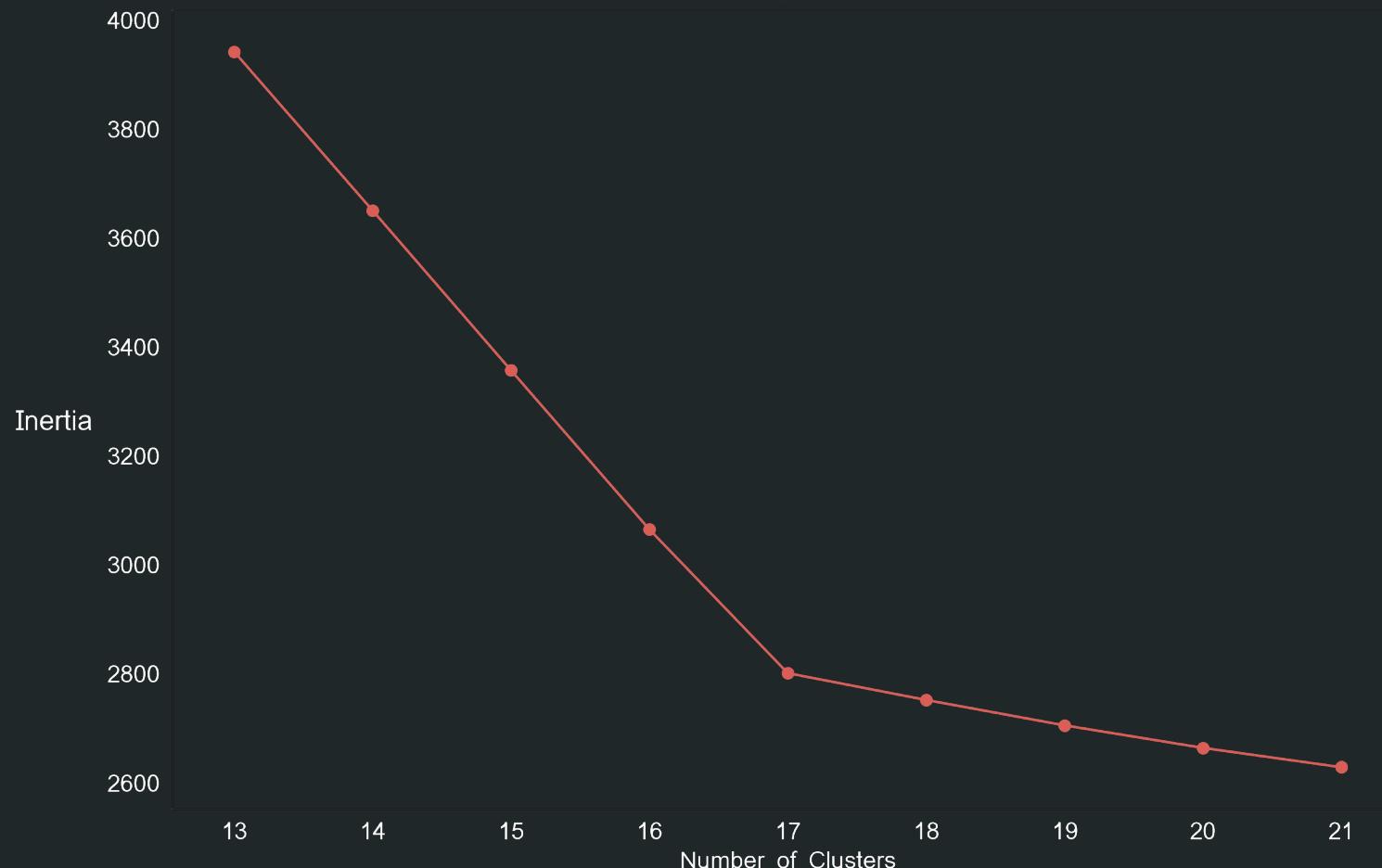
Top-30 Most Relevant Terms for Topic 1 (7.8% of tokens)



1. $\text{salency}(\text{term } w) = \text{frequency}(w) * [\sum_t p(t | w) * \log(p(t | w)/p(t))]$ for topics t ; see Chuang et. al.
2. $\text{relevance}(\text{term } w | \text{topic } t) = \lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

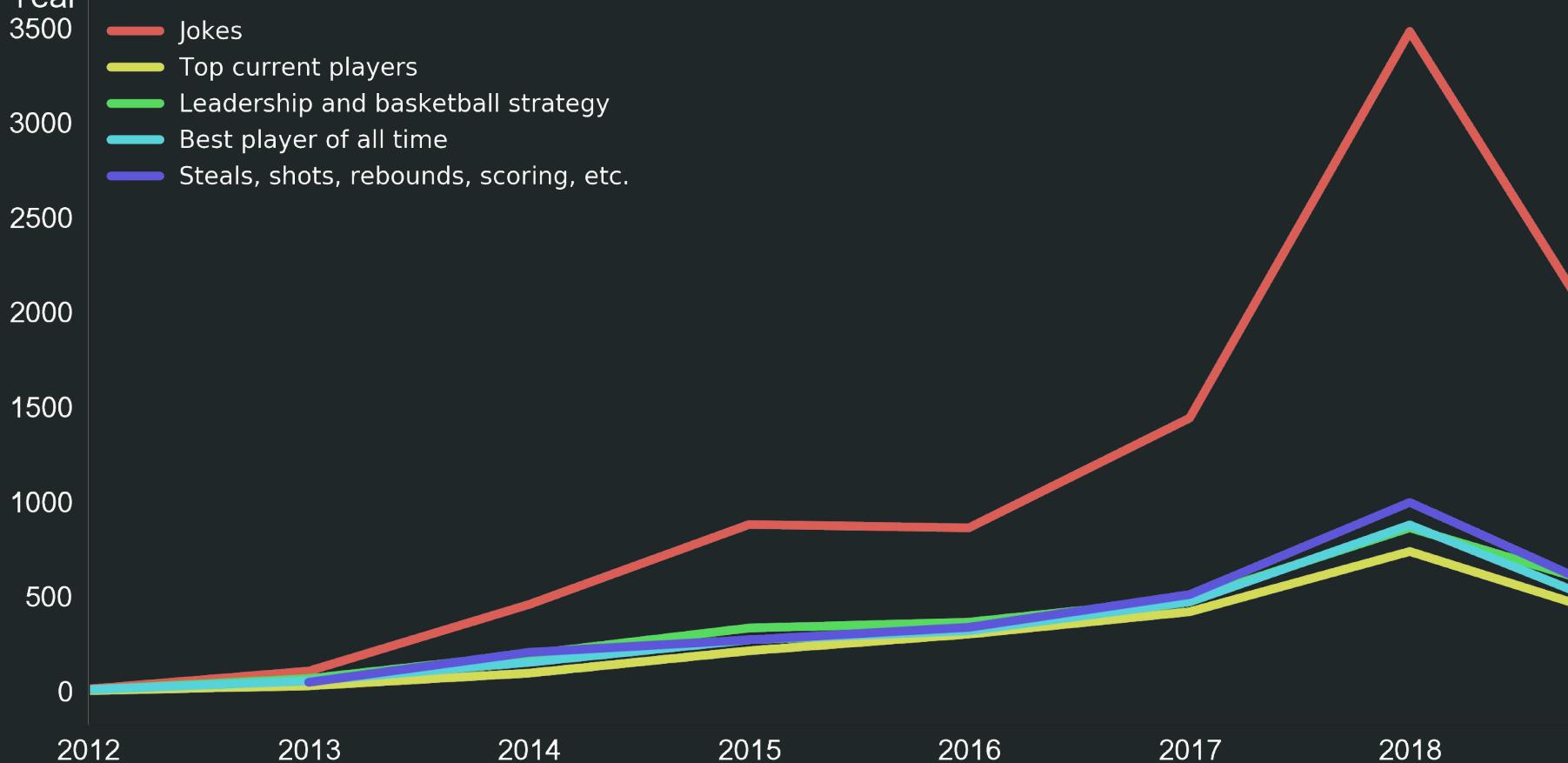
KMeans Clusters

KMeans Inertia



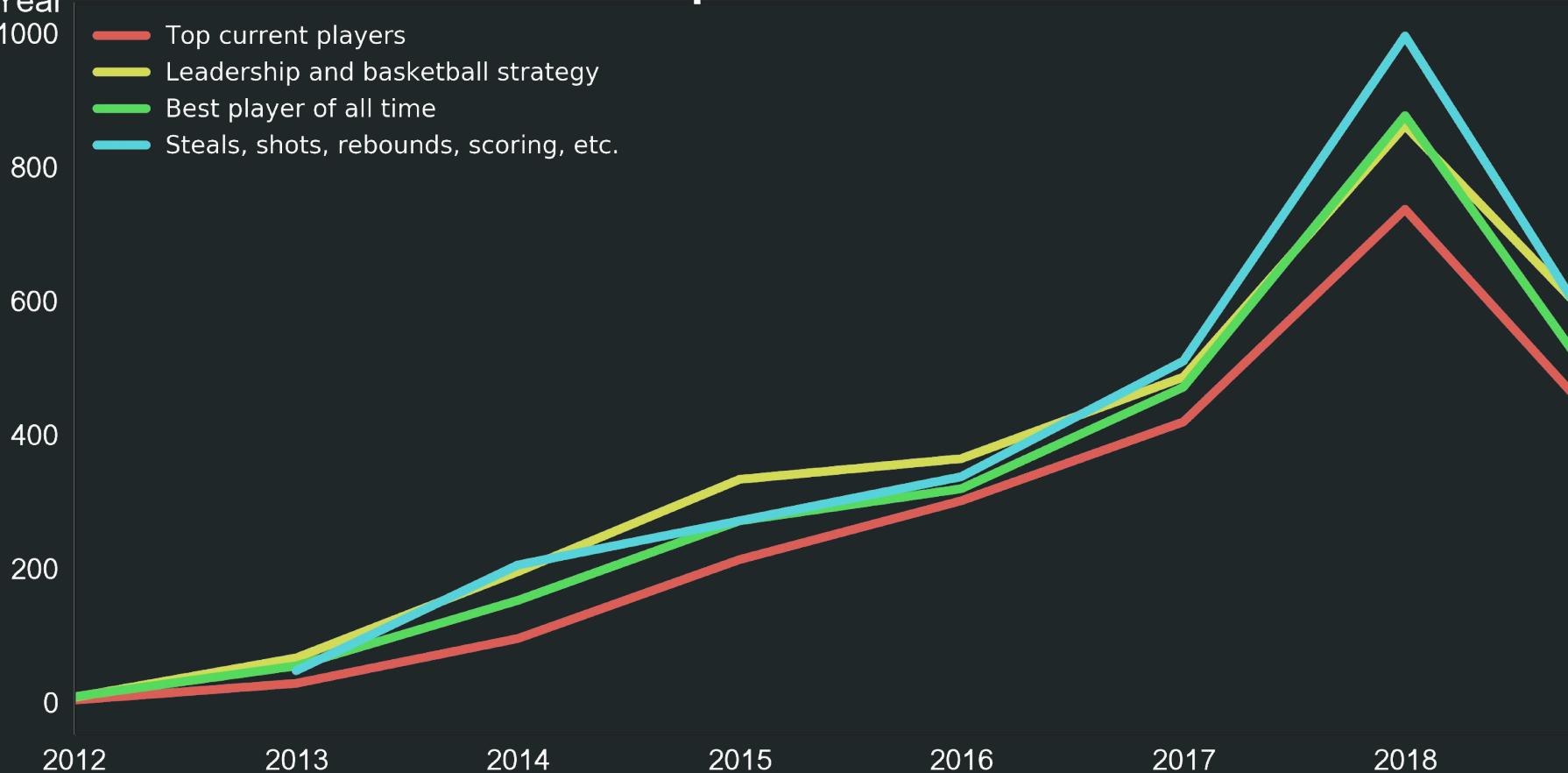
Count
per
Year

KMeans Topics



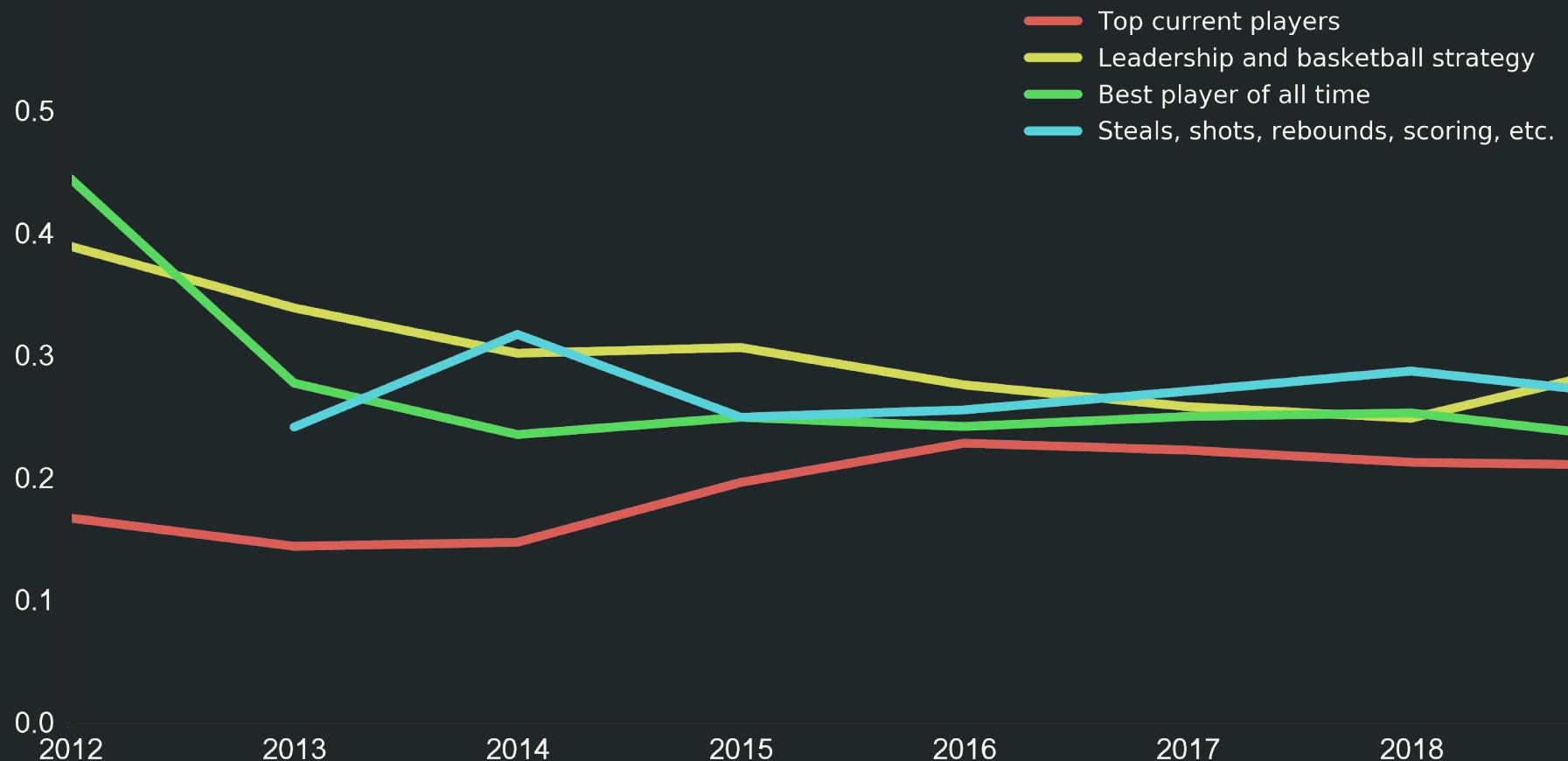
Count
per
Year

KMeans Topics without 'Jokes'



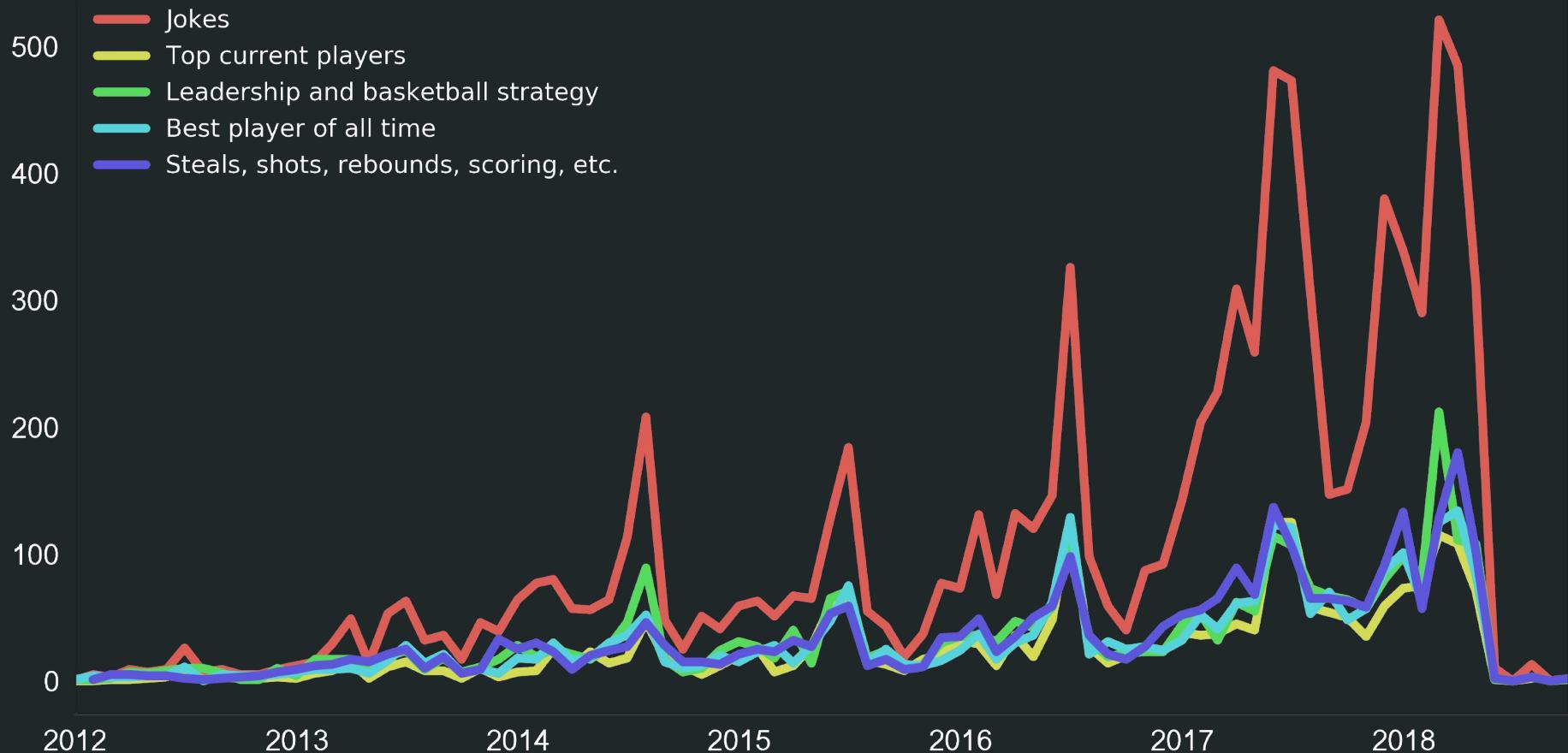
Proportion
of Total
0.6

KMeans Topics without 'Jokes'



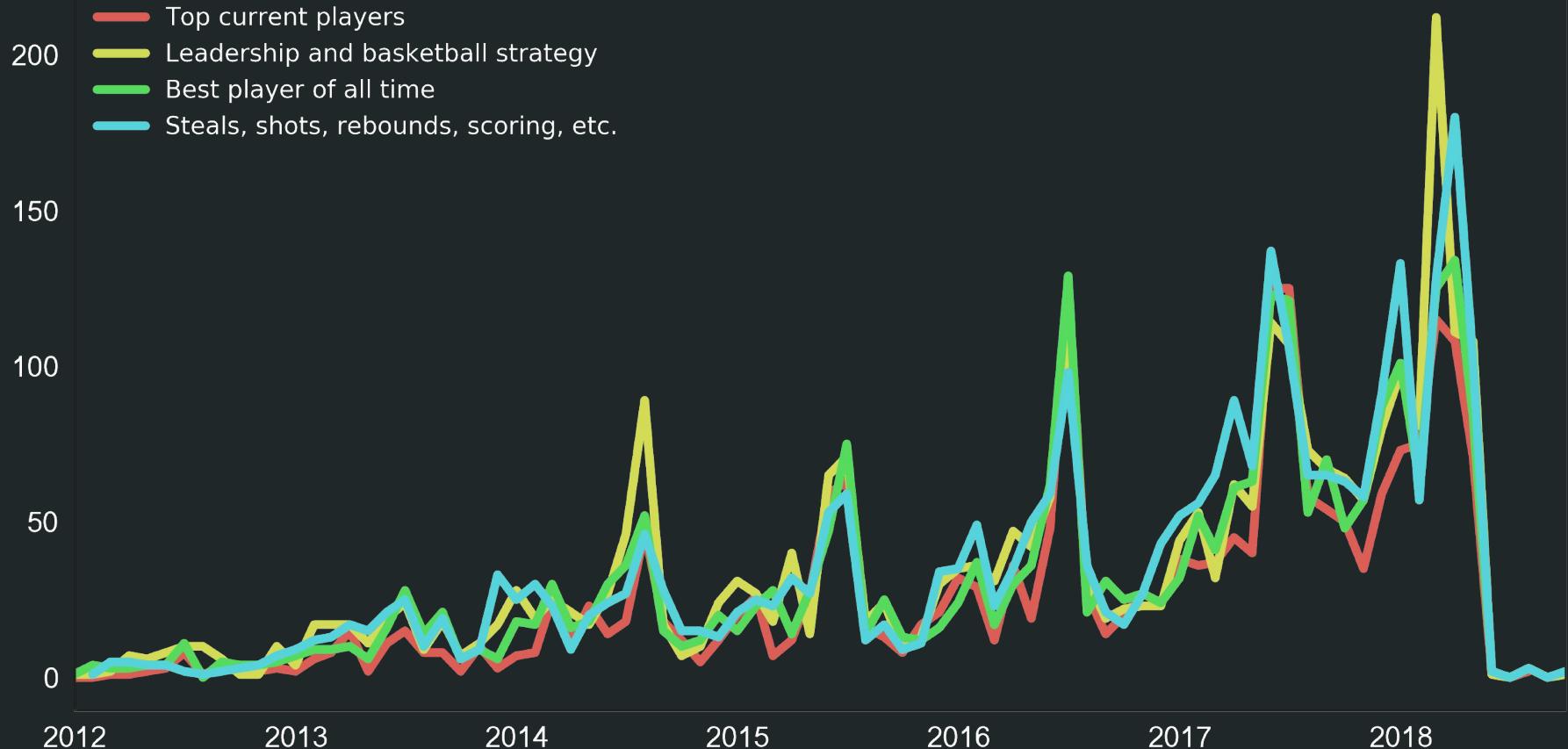
Count
per
Month

KMeans Topics by Month



Count
per
Month

KMeans Topics without 'Jokes' by Month



Proportion
of Total

0.6

0.5

0.4

0.3

0.2

0.1

0.0

2012

2013

2014

2015

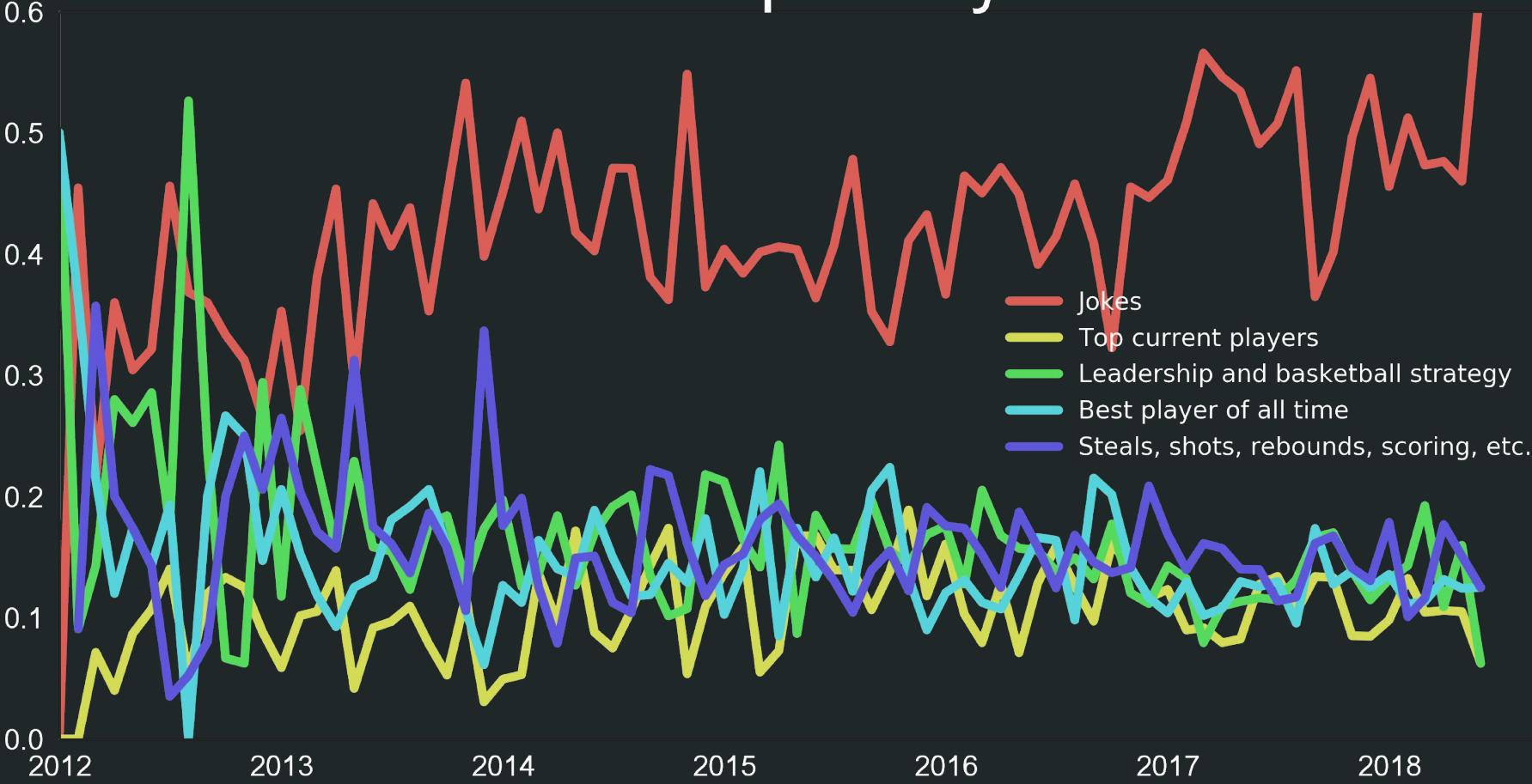
2016

2017

2018

KMeans Topics by Month

- Jokes
- Top current players
- Leadership and basketball strategy
- Best player of all time
- Steals, shots, rebounds, scoring, etc.

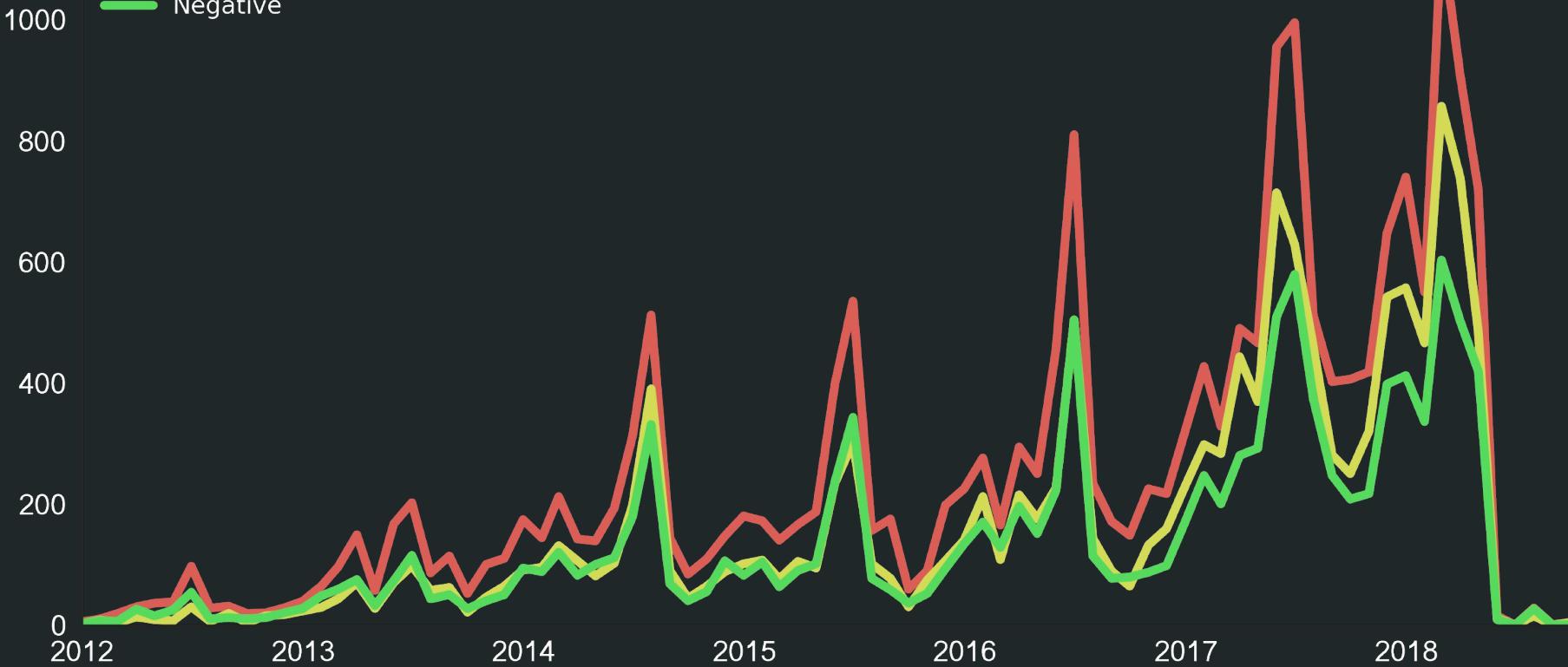


Sentiment Analysis

Sentiment Analysis

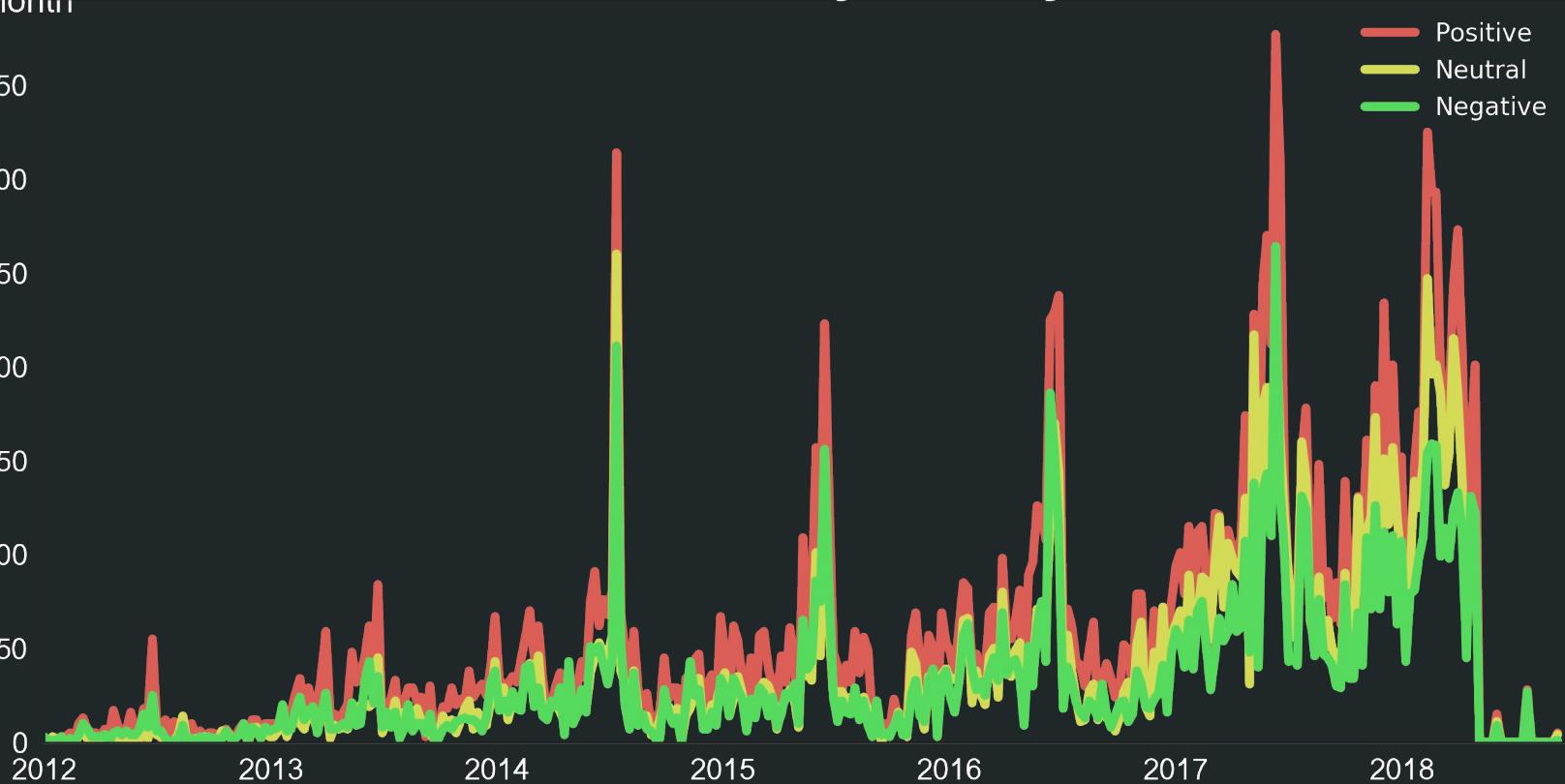
Count
per
Month

- Positive
- Neutral
- Negative



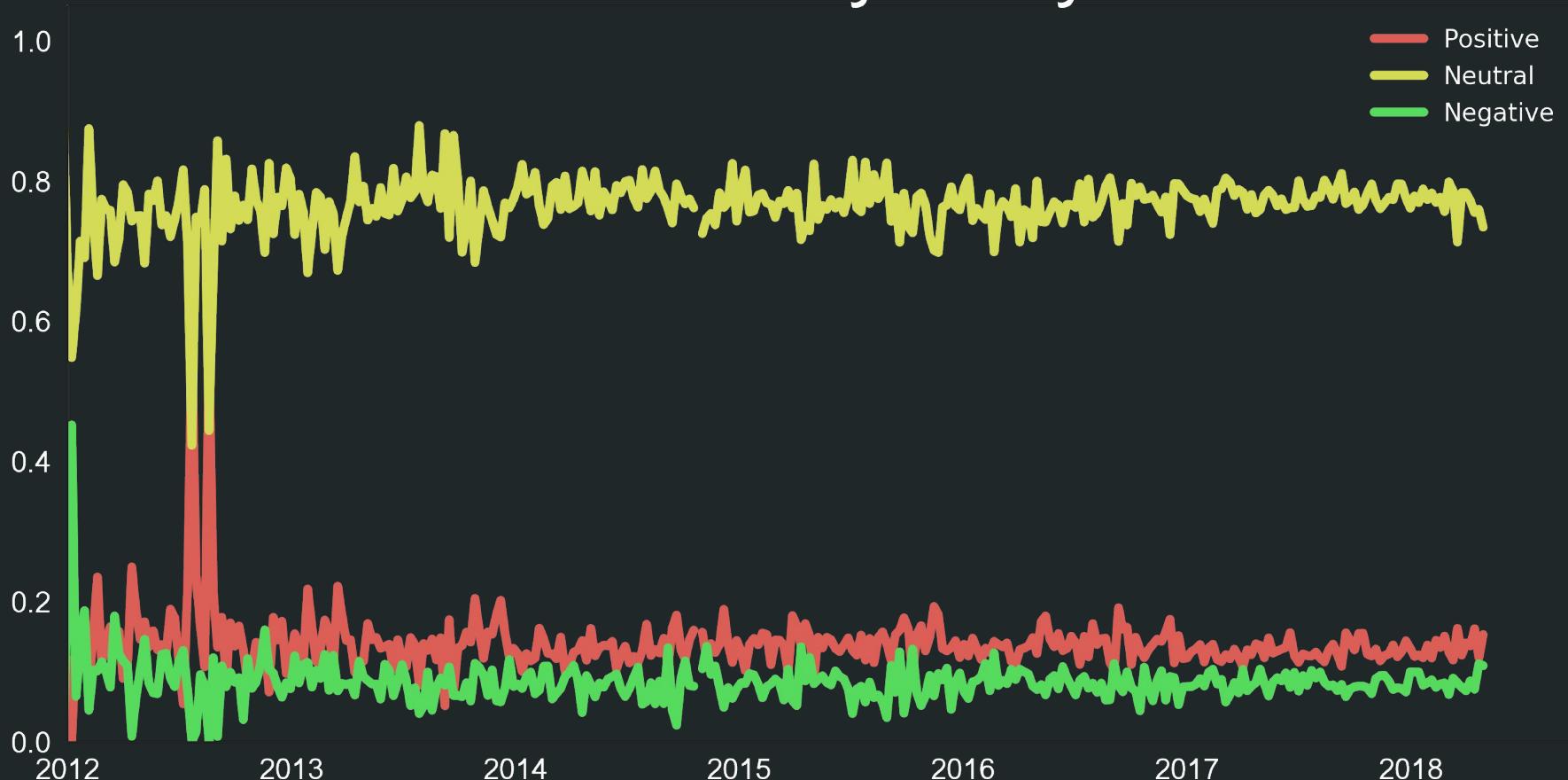
Count
per
Month

Sentiment Analysis by Week



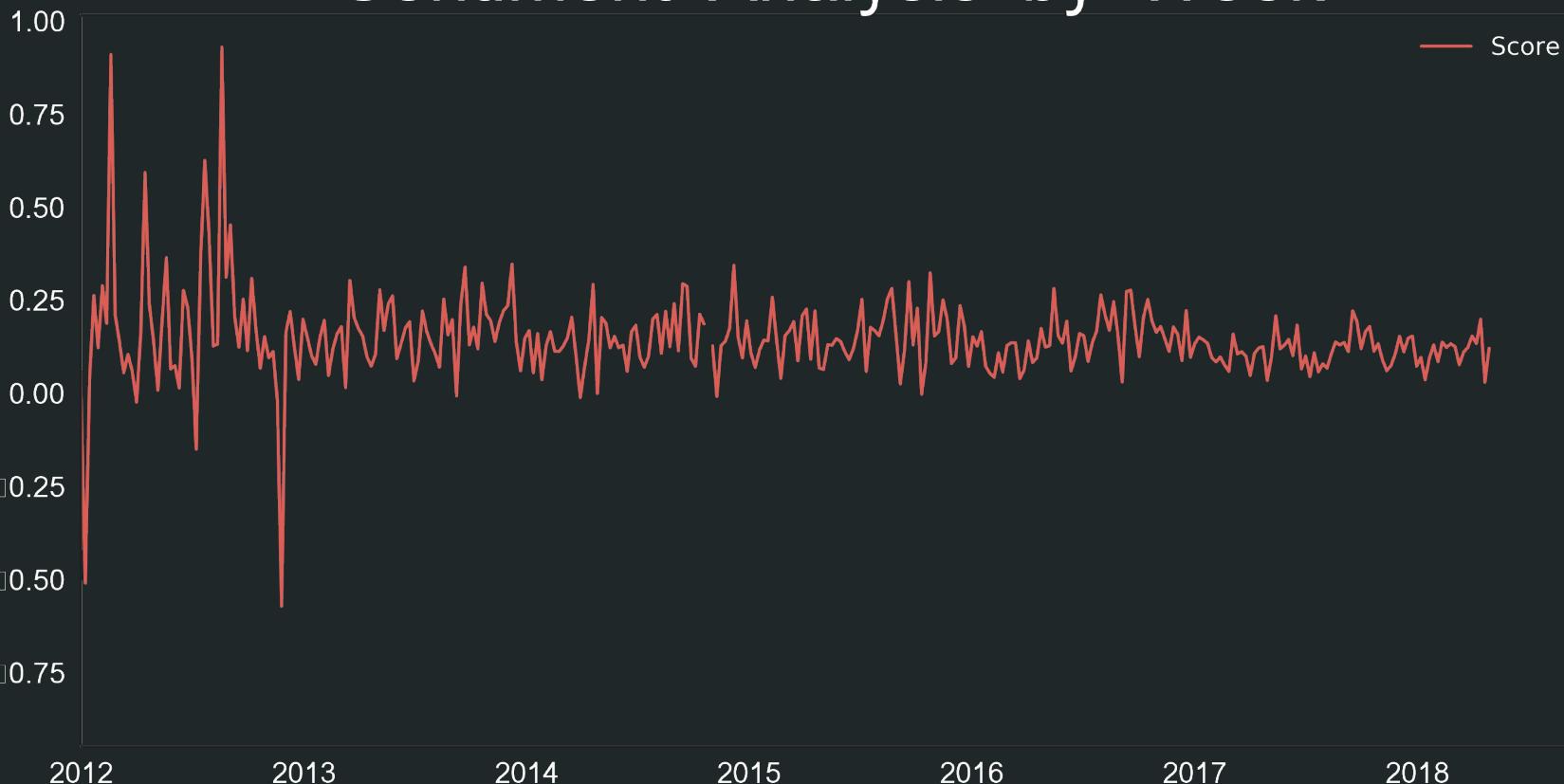
Proportion
of Total

Sentiment Analysis by Week



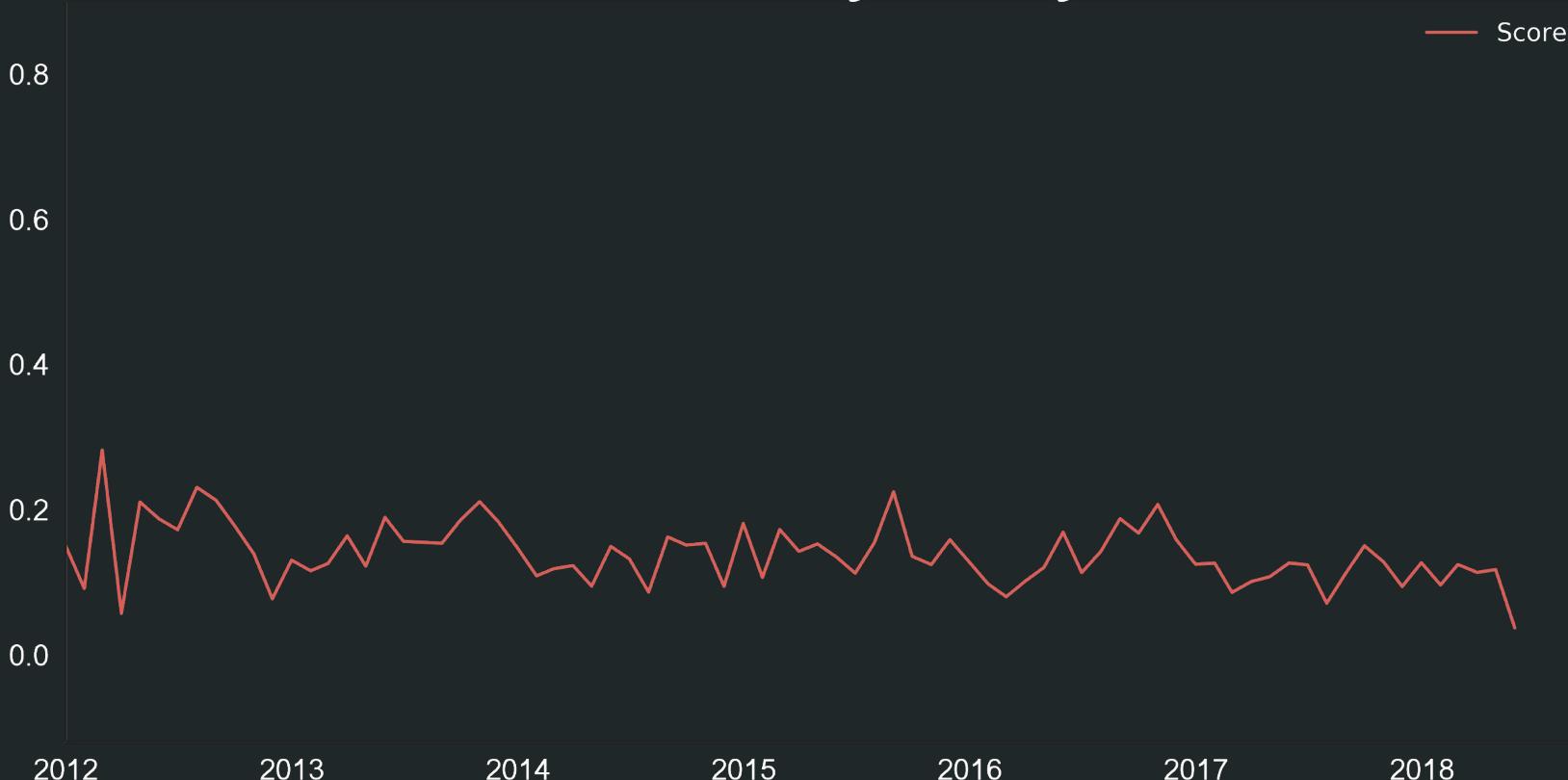
Proportion
of Total

Sentiment Analysis by Week



Proportion
of Total

Sentiment Analysis by Month





[–] sj2011 1614 points 1 year ago

I got to watch both Michael Jordan and LeBron James play basketball. Life's okay.



[–] wowheypaperboy 743 points 1 year ago

Lebron just pulled off an All Star Game move in the NBA Finals.



[–] jjkiller26 132 points 3 months ago

Is Lebron gonna run for president

[–] Tyrion_Wayne 503 points 9 months ago

LEBRON HAS NOT EVEN BEGUN TO PEAK



[–] itsjoeydavis 2730 points 6 months ago

Every year we argue back and forth about who the mvp is. It's Lebron. It's Lebron every year.