

# Who are the sports media darlings?

Emmanuel Rayappa & Dr. Xiaoyi Yang  
Department of mathematics, Creighton University

# Introduction

Sport news is one of the most popular types of news. It not only feeds us the latest game scores but also includes a bunch of professional reviews, prediction, evaluation, and even off-site tidbits.

Therefore, we want to know who those sport news are talking about and whether it is consistent to our general knowledge of the sports figures.

Here is a couple of things we have considered in this study:

- Conducted Sentiment Analysis and some text analysis on Sports Articles.
- Apply Name Entity Recognition Algorithm to detect unique names and places in the articles.
- Develop a way to define the sentimental scores for each person and rank

## Data

- In the study, we focus on football because of its popularity amongst American.
- We found an online source with 156,000 news articles from the Associated Press from 2015.
- Focused on NFL Articles from the sample which had non-sports articles in it.
- The final sample size is 1,726.
- Each news include time stamp, source author, and full news text.
- Processed the files with a function to detect key words relating to football within the articles.

## Methodology

We have considered following tools to help us understand the information in the sport news:

- Word cloud: Visualize the frequency of top words
- Sentimental score : used the syuzhet method; to show the sentimental of words or the news.
- Name Entity Recognition (NER): programmed algorithm to detect unique place and people mentioned in each file.

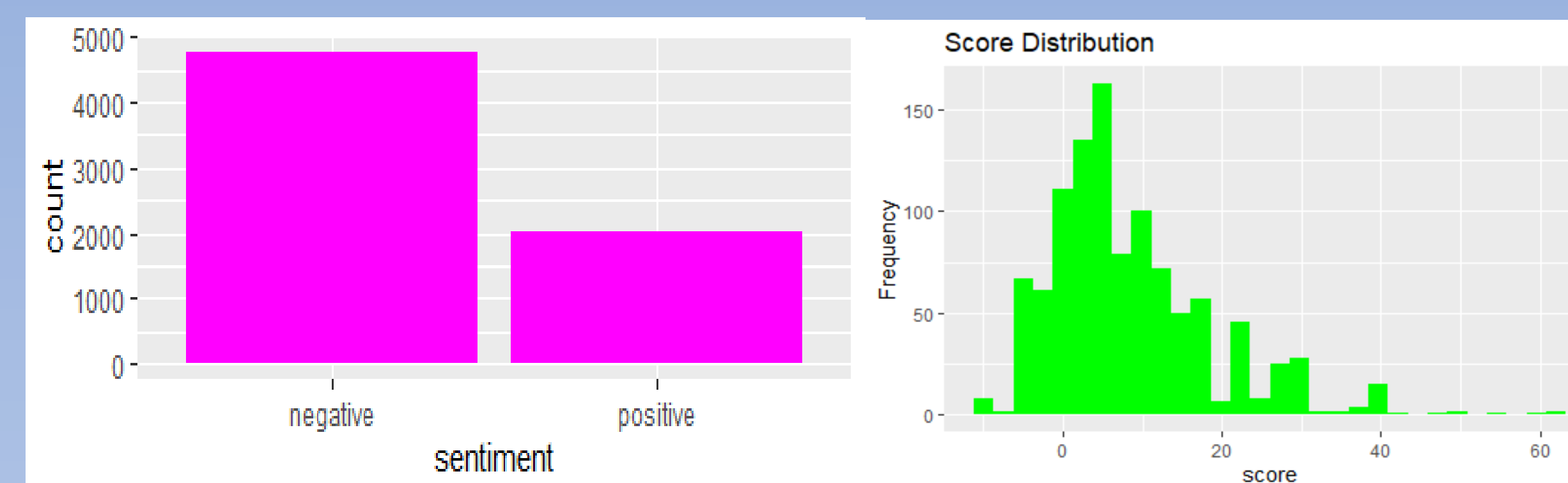
## In general, what the media says?



We first visualize the top 50 words being mentioned in all the news. It seems that besides the game results, like win or lose, injury is one of the major topics in the sport news. One interesting point is that “luck” is also a popular term.

We have also studied the sentimental trends in sport news.

- On the word level, it seems that we have more negative words in general (the figure on the left).
- If we aggregate the word level sentimental scores and get the news level sentimental scores, it seems that we have more positive news (the figure on the right).
- On the new level, the extend of emotion for positive news is stronger than the negative ones.



## Define people level sentimental score

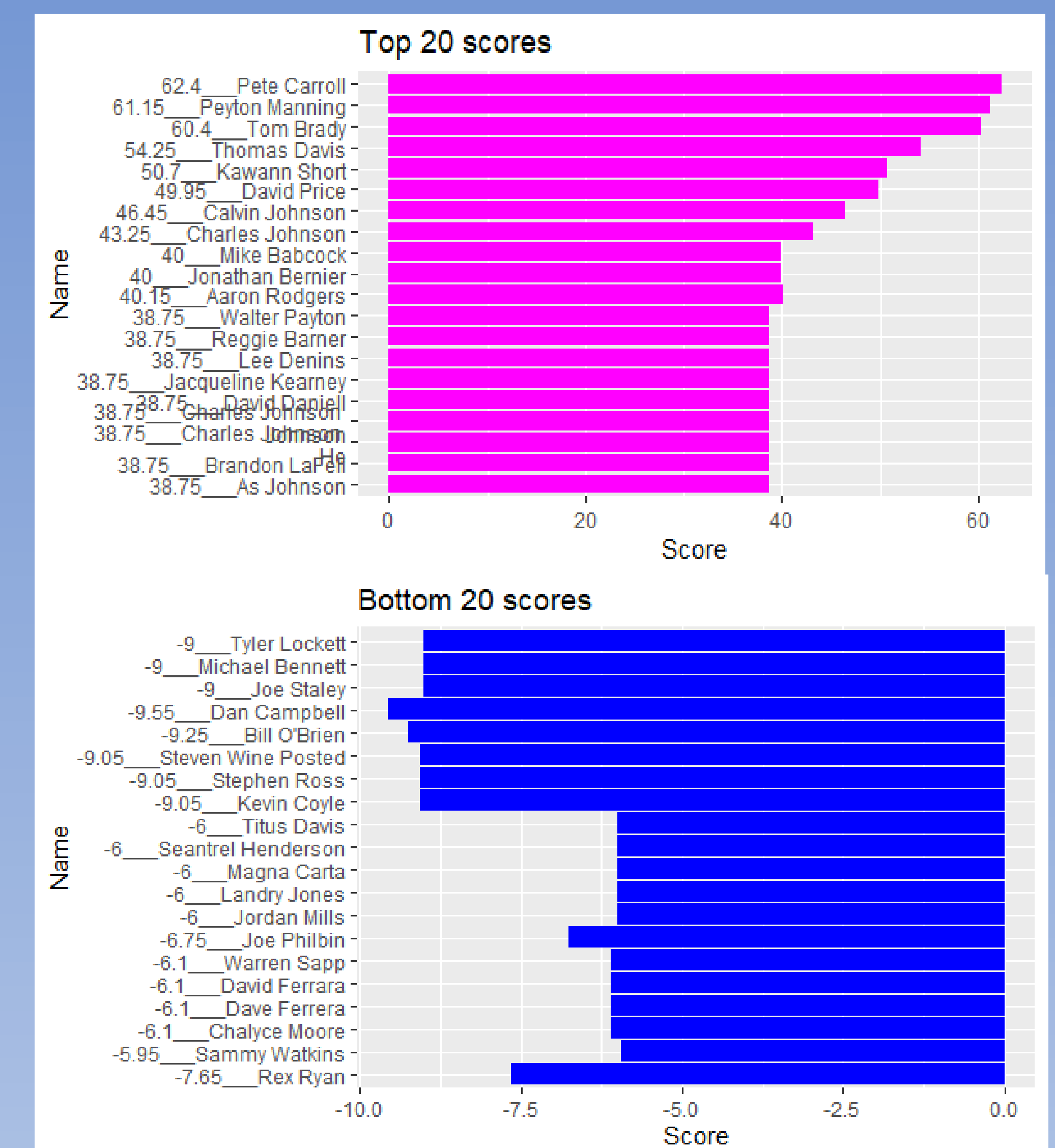
In the next step we use NER to extract names from each news. We also define the sentimental score for each person is the sum of sentimental score of the news the person has being mentioned.

We have also considered to define the location sentimental score by the same way but NER accuracy for location is much lower than the person names.

## Most and Least popular

After calculating the sentimental score for each person, we list the top 20 and bottom 20 in our data:

- Top 20: Thomas Davis, a Linebacker for the Carolina Panthers, enjoyed a breakout year and made his first Pro-Bowl that year.
- Bottom 20: Sammy Watkins, a Wide Receiver for the Buffalo Bills, struggled with injuries in 2015 and then criticized fans who mocked him online.



## Conclusion and Future works

- We have studied the common topics and general sentimental trends of sport news
- We have also designed the pipeline to extract the people who are most favored or least favored by the media
- For future experiment and research, we would like to use a larger sample size and focus on a different sport.
- Moreover, we can also generate networks from the news to identify who are usually mentioned together