Fangraphs Future Role & Sentiment Analysis
By: John Henry Hawes

The goal of this quick study was to see if sentiment analysis from the sentimentr package had any correlation to future scout grades.

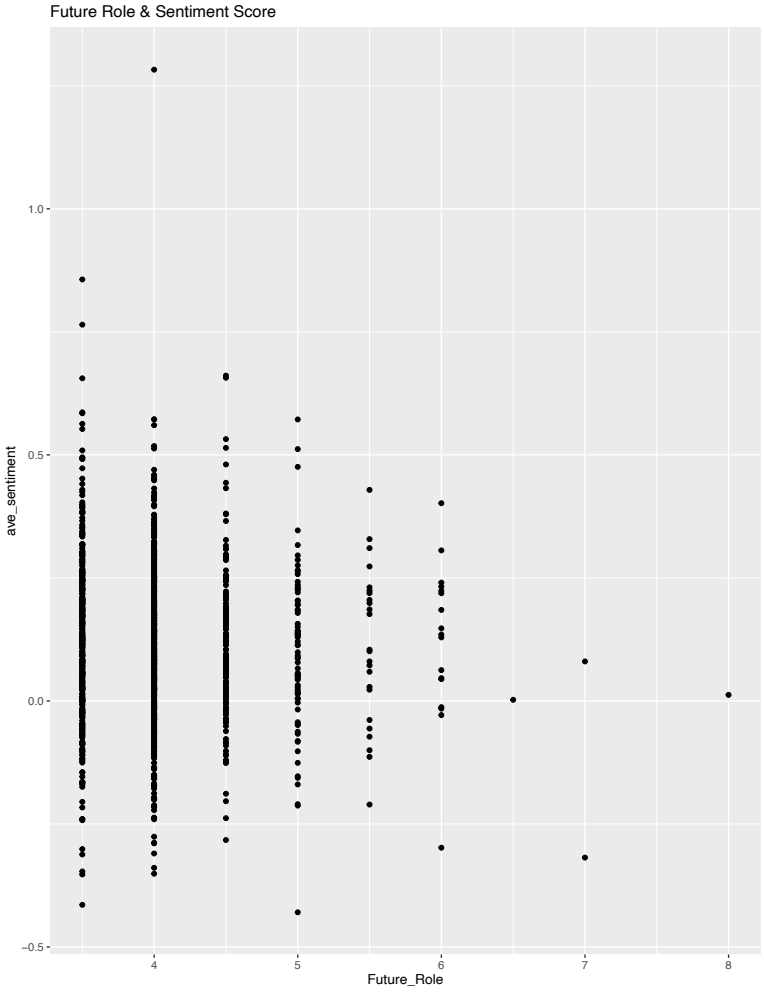
I started by scraping all 30 teams top prospects reports from Fangraphs 2020 Reports. I have attached one of the scripts just for viewing purposes but the script was pretty straight forward thanks to the read_html function in R.

Once all of our reports were scraped and bind together. I then ran each report through the sentiment_by function to yield the average sentiment score: ave_sentiment. For a quick refresher the sentimentr package scores words and sentences based off positive and negative statements defined in the sentimentr library. Note this is not a baseball specific library but still wanted to see if the normal library would yield any correlation.

The plot to the right shows the average sentiment score for each player(y-axis) and then the assigned future role grade by Fangraphs (x-axis).

As we can see scores are all over the place for all role grades, as there is little to no correlation between sentiment score and future role grade.

Going forward, creating a baseball specific library for the sentiment scoring would be more beneficial as it would score words and sentences that the baseball world actually views as positive or negative. But to conclude the basic sentimentr package yields no correlation to Future Role.



```
#install.packages("rvest")
##library(rvest)


fg_wbpg <- read_html("https://blogs.fangraphs.com/top-33-prospects-boston-red-sox/")



P_name <- fg_wbpg %>%
  html_nodes(".align-L") %>%
  html_text()

P_name


P_age <- fg_wbpg %>%
  html_nodes(".align-L+ td") %>%
  html_text()

P_age

P_pos<- fg_wbpg %>%
  html_nodes(".sortable td:nth-child(5)") %>%
  html_text()


P_pos

P_role<- fg_wbpg %>%
  html_nodes(".sortable td:nth-child(7)") %>%
  html_text()

P_role

P_report <-fg_wbpg %>%
  html_nodes("#tool-caps-prospects-list p:nth-child(1)") %>%
  html_text()

P_rk<- fg_wbpg%>%
  html_nodes(".sortable td:nth-child(1)") %>%
  html_text()

P_lvl<- fg_wbpg%>%
  html_nodes(".sortable td:nth-child(4)") %>%
  html_text()



BOS<- data.frame(Name=P_name,Rk=P_rk, Age=P_age,ORG="BOS", Highest_LVL=P_lvl,
POS=P_pos, FV=P_role, Report= P_report)
```