

Appendix

Here we provide tables that give a much more detailed breakdown of FOOTBALL. For instance, we show breakdowns by position (Table 1), time period (Table 2), and race (Table 4). Additionally, we fully specify the transcript and roster collection process.

Position	white	nonwhite	Total
QB	54.0k	17.2k	71.3k
RB	3.1k	42.8k	45.8k
WR	6.5k	38.1k	44.6k
DB	2.5k	32.1k	34.6k
LB	4.4k	14.0k	18.4k
DE	3.0k	9.2k	12.2k
TE	6.1k	5.4k	11.448
DT	1.3k	6.7k	8.0k
OT	2.7k	3.3k	6.0k
K	5.7k	279	5.9k
OG	1.9k	2.1k	4.0k
P	3.6k	219	3.8k
C	1.4k	254	1.6k
LS	92	0	92
OL	27	38	65
DL	11	51	62

Table 1: Number of mentions in FOOTBALL by position. There are 267,778 player mentions in total.

Years	NFL	NCAA	Total
1960-1969	5	0	5
1970-1979	53	50	103
1980-1989	36	76	112
1990-1999	57	106	163
2000-2009	105	194	299
2010-present	345	428	773

Table 2: Number of games in FOOTBALL by decade

Years	Mentions by Decade		Total
	nonwhite	white	
1960-1969	1.0k	641	1.6k
1970-1979	9.4k	9.2k	18.6k
1980-1989	7.3k	11.1k	18.4k
1990-1999	9.5k	18.9k	28.4k
2000-2009	18.7k	35.8k	54.5k
2010-present	50.2k	95.9k	146.1k

Table 3: Number of mentions in FOOTBALL by decade

League	nonwhite	white
NFL	137.4k	84.0k
NCAA	34.2k	12.1k
Total	171.6k	96.1k

Table 4: Mentions in FOOTBALL by race and league

Years	NFL		NCAA	
	nonwhite	white	nonwhite	white
1960-1969	75	40	0	0
1970-1979	189	198	18	39
1980-1989	185	291	15	85
1990-1999	136	419	27	130
2000-2009	265	761	76	312
2010-present	481	1.3k	116	419

Table 5: Number of distinct labeled players

Years	% white (dataset)	% white (NFL)
1960-1969		61.0
1970-1979		47.1
1980-1989		38.4
1990-1999		29.9
2000-2009		40.5
2010-present		28.1

Table 6: Percentage of mentions by race and decade in FOOTBALL compared with the actual race distribution of players in the NFL by decade (Lapchick, 2017). The distributions are very similar. *Lapchick (2017) does not provide data for this time span.

Transcript collection: additional details

We collect YouTube videos from the following YouTube channels, which aggregate football games *Mark Jones, NFL Full Games 2018 / 2019, Adrián GTZ Montoya, NFL Full Games, Bart Simpson, Bryan Mears, NFL, Alex Roberts, Sports, Danger zone, Nittany 96*. (Some channels publish dedicated lists of football games, amid other content. Other channels post football games only).

After downloading videos, we identified the teams playing and the year of each video by matching strings in video titles. In isolated cases when string matching failed, we manually identified the teams and year from the video itself.

We remove stop words and entities from our original text before processing mentions using the NLTK English stopwords list.

Roster collection: additional details

We collected rosters for all NFL teams from 1960 to the present from footballdb.com. Because some players have the same name, we used the player name-position pair (e.g. Tom Brady, QB) to identify a unique entity in our dataset.

We collect NCAA rosters from the college football page sports-reference.com, downloading rosters for the 290 available schools in the NCAA from 1869 to the present.

We lower case all names on rosters. We also remove periods, apostrophes and hyphens within names (e.g., Odell Beckham, Jr., Ra'shede Hageman, Dominique Rodgers-Cromartie).

In total, including the mentions for which we could not acquire racial metadata, we gathered a total of 545,232 labeled and non labeled player mentions, 265,879 NCAA mentions and 279,353 NFL mentions. The higher number of NFL mentions (despite fewer NFL games) is due to incomplete roster information for NCAA teams (some years are incomplete, some years missing altogether). For the players we were able to acquire race labels for, 137,428 nonwhite and 84,038 white mentions are collected from NFL games and 34,196 nonwhite and 12,116 white mentions are collected NCAA games.

References

Marfatia Lapchick. 2017. The 2017 racial and gender report card: National football league.