Problem Set 4

Steven Plaisance

February 2023

Though scraping is not required to access this dataset, I believe the NFL's Big Data Bowl competition datasets are some of the most interesting public football data available. This data provides raw latitude and longitude values for every player on the field, sampled at 10 Hz or 10 observations per second.

In terms of scrapeable data, NFL contract data via OverTheCap.com or Spotrac.com offers lots of potential in terms of target variables. After all, the market decides a player's worth in the NFL, and these websites store years and years worth of market data. This can be easily paired with data scraped from Football Reference or pulled through nflfastR API to get started on an NFL player performance model.

1. mydf$date is of class character.

2. df1 is of class tbl_df, while df is of class spark_df

3. The column names of df and df1 are the same.