

Recitation Week 5 Problems

Today imagine you are a Premiere League soccer coach and you want to know which players to consider buying / trading for next season. You will use data analysis skills you have learned to provide support for your argument. To begin, import the data from the player_data.csv file.

- 1) *Skip this if it takes too much time* Clean the data in each variable (you may want to use the substr() function) - remove any unnecessary spaces, replace any non-standard characters with an underscore "_".
- 2) (a) Which player had the most total points over the season? (hint: check out the arrange() function). (b) Who had the highest points per minute? (hint: you will need to calculate points per minute).
- 3) Which team had the most points over the season? To find this, you will need to (a) import the team_lookup.csv file and merge it onto your data, then (b) summarize or "aggregate" your data to the team level (as opposed to the player level).
- 4) *Skip if you are running out of time* Come up with a metric to measure players' performances controlling for what team they were on. Who scored best? (e.g. you may want to calculate who scored highest above their team average...).
- 5) Which players would you draft / purchase? Are these the most expensive players? If not why do you think they may be undervalued?