**Workshop questions**

1. How are the performance measures distributed? Start by using stargazer() and then plot the density of two variables, goals and disposals.
   1. Is it good or bad for a measure to have high variance (think in terms of the trade-off between sensitivity and noise)?
   2. How can thinking in terms of changes help us understand the properties of performance measures? What can we learn by looking at disposals in t-1 and disposals in t?
2. In general, how do the individual measures correlate with team success (as measured by margin)? Start by aggregating the data set to the team-game level, then use cor(). Next, use scatter plots - plot() - to examine the relationships between margin and: 1) running distance; 2) disposals; 3), and tackles.
   1. Focusing on the empirical relationship between tackles and margin - would it be a good idea to place a *negative* weight on tackles when evaluating players’ performance? Why/why not?
   2. In general, do these associations tell us that certain measures (or the actions that drive these measures) ‘cause’ teams to win? Why/why not?
3. In economics, the standard production function, y = f(x), tells us that increased inputs, x, lead to increased outputs, y. Should we adjust the measures in our data set for differences in inputs (or playing opportunities), and, if so, why/why not?
   1. What is one way we could hold constant differences in inputs to evaluate a player’s productivity? Think about how we might be able to use one of the variables in our dataset to scale the other measures...
4. In practice, workers’ production functions are rarely independent. That is, one worker’s output often serves as an input into another worker’s production function: y2=f(x2,y1) where y1=f(x1,y2). Do we expect interdependencies in performance in our setting?
   1. To test this, let’s explore how performance measures are correlated across playing positions. How might forwards’ performance depend on midfielders’ output? How can we show this in the data?
   2. How do these interdependencies complicate the process of evaluating the performance of an individual player?
   3. How might we be able to filter out these interdependencies? Think about an empirical model that allows us to identify ‘abnormal’ performance (i.e., performance above or below what we would expect given teammates’ outputs)...
5. Do random shocks - like weather - introduce noise into the measures? Why might this be a bad thing?
   1. How does this noise show up in our measures? Let’s use plot(density()) to look at how wet weather changes the distribution of two performance measures: disposal efficiency and tackles.
   2. How can we filter out this noise? Think about whether this is a common shock and what this means for differences in performance across players within a given game (the idea of ‘abnormal’ performance is again useful).
   3. Do we always want to filter out these shocks (or, do we want the players to expect that we will always filter out these shocks)?