Week 4 - Assignment Instructions

In week 4 we have looked at how combining internal and external measures can provide a more complete look at the athlete, the training and/or the recovery situation. In the workbook, we have had a look at game data from Catapult Sports devices worn by the UM Women’s Field Hockey team. We have evaluated the relationship between player heart rate (an internal measure) and the Catapult playerload (an external measure). In addition, we have established that there is a reasonably strong relationship between the two of them -- but that it is not consistent across different team members. This was the main reason to evaluate this relationship.

It is disappointing that internal measures are more prone to artifact and or interference. However, it appears that for some of the athletes in this dataset the real problem may have been that they just don’t like to wear the HR monitor -- and so they didn’t. This (of course) led us to 100% missing data for HR for a few of the athletes. However, we have focused on those that had data in order to appreciate the types of analyses that can be performed.

For this assessment we will use the same functions and methods to answer some questions from the same dataset.

**Scenario:**

We have explored (in the workbook) the Field Hockey game from Sept 30th, 2018… and we have found that a very limited number of athletes had good heart rate data. In addition, for those athletes with good HR data, we found considerable difference in the heart rates they maintained on the field -- even when adjusting for the same player load (i.e. external effort). For this weeks assessment, let’s explore the data from a different date to 1) compare the number of players with reasonable HR data (<10% zero heart rate values); 2) evaluate whether it is the same players that have good heart rate values -- and the same players with 100% missing data; and 3) evaluate how similar the relationship is between different games regarding the individual player using simple linear regression calculations.

These are the steps that you need to take to complete the assignment:

1. Load the datafile (FH.csv)
2. Create a slice of the dataframe to use for this assessment based on the date of the game: August 19, 2018.
3. Determine the total number of HR data points that were recorded with a value of zero and note the total number of rows in the dataset for the date you have selected (August 19)
4. Determine the % of the missing HR data for each individual athlete and keep only the athletes that have less than 10% of their HR data missing for the remainder of this assignment.
5. Calculate the playerload “delta” (player load change) and be sure you eliminate any negative PLdelta scores.
6. Determine the correlation between the Heart Rate and the rolling playerload change value over 30, 60 and 90 seconds and note the pearson correlation “r” value for this date.
7. Use the groupby, plot.scatter method to calculate the relationship between the rolling playerload change over 60 seconds versus the heart rate values. Note which of the graphs looks especially peculiar.
8. Determine the slope and y-intercept for (only) the athletes that had good data on both days. There should be 6 in total. Compare and contrast the values between the two days. Identify the top 2 players on both days.