Application on Linear Regression : NBA

Goal : -Become familiar with the use and the interpretation of a data frame  
 -Become familiar with the Linear Regression method  
 -Calculate SSE, RMSE, R2  
 -Make Predictions

Part 1

1. Load train data “NBA\_train.csv” using read.csv( ) function. Name your variable NBA   
   *Our data contains stats from all teams in season since 1980*
2. Explore your data using str( ), summary( )  
   *#Some variables start with X cause R doesn’t accept that variables start with a number*
3. How many times a team won 11 games? list these teams? (use table & subset function)
4. How many times a team with 43 wins didn’t make it to the playoffs (use table function)   
   How many wins a team need (at least) to be in the playoffs?
5. Add point difference variable to our data, Call it PtsDiff
6. Create a plot of teams wins according to PtsDiff using plot()  
   *#So we can see that linear regression will be a good fit to our model*
7. Create ur linear regression model using lm( dep ~ indep1 +indep2+.. , data=..), name it WinsReg
8. Explore it using summary()
9. What is our linear regression equation? How many games a team with 200 PtsDiff will win?
10. Using abline() function show the regression line on your plot
11. How many points difference do we need to be in the playoffs?

Part 2

1. Create a linear model to predict points scored using X2PA + X3PA + FTA + AST + ORB + DRB + TOV + STL + BLK, name it PointsReg
2. Which variables are significant?
3. Calculate the sum of squared errors SSE (You can access to the errors vector using PointsReg$residuals). Is it interpretable?
4. Calculate the root mean squared errors RMSE (it’s like the average error we make) .  
   (use nrow(NBA) to get the number of observations)
5. What is the less significant variable in our model?   
   Create PointsReg2 without the use of this variable.   
   Compare the R squared value
6. Create the best possible model and compare it to the initial one that uses all the variables

Part 3 – Predictions for 2012/2013

1. Load the test set, name your variable NBA\_test
2. Make predictions ( for season 2012/2013) using predict( model ,newdata= dataframe ) and the testing
3. Calculate SSE, SST , R2