**7PAM2000 Applied Data Science 1  
Assignment 1: Visualisation**

**Introduction to the data set**

The following analysis is done on the following data set: “The Best NBA Players, According To RAPTOR”. It is downloaded from the following website.

<https://projects.fivethirtyeight.com/nba-player-ratings/>

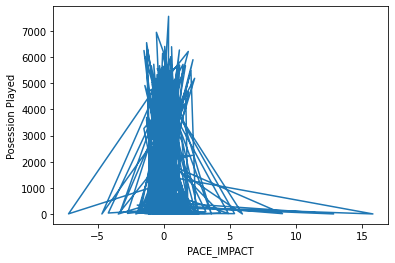
This data set contains the statistics of players in the NBA. The player data is tracked and each player’s individual plus-minus performance is determined. There are multiple datasets in this source. I chose RAPTOR data for every player in the latest season.

The data set has multiple parameters like player name, raptor box offense, raptor box defence, pace impact, etc. My focus is the following parameters. The labels along with their explanation are given below.

1. poss- Possessions played.
2. pace\_impact - Player impact on team possessions per 48 minutes.
3. raptor\_defense - Points above average per 100 possessions added by a player on defense, using both box and on-off components.
4. raptor\_offense - Points above average per 100 possessions added by a player on offense, using both box and on-off components.

Line Plots:

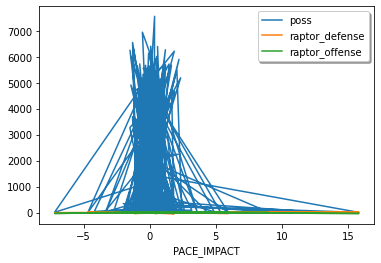
First a line plot is plotted in between pace\_impact and poss. It is shown in the following figure.



Note the graph is zig zag and no meaningful interpretation can be obtained between pace impact and possession played. Now multiple plots are plotted for the following data.

1. Pace\_impact vs poss
2. Pace\_impact vs raptor defense
3. Pace\_impact vs raptor offense

The plot is shown in the following figure.

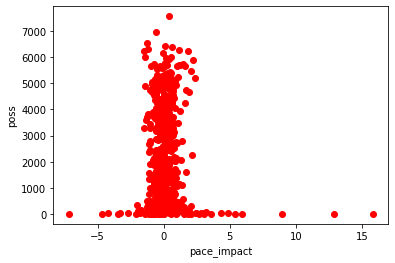


As it is obvious that the graph is clutterd and no meaningful interpretation of the data can be done here. To avoid this scatter plots are plotted.

Scatter Plot:

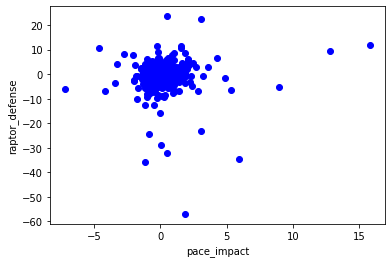
A scatter plot plots each data as a point(dot) in the graph. There will not be any line joining all the plots. Pace\_impact vs raptor offense

The plot for Pace\_impact vs raptor offense is shown in the following figure



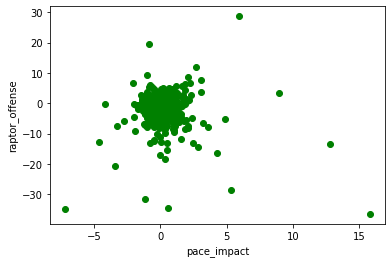
From this it can observed that the possession is more when the pace\_impact is neither strong nor weak. A weak pace\_impact reduces the possession and so does a strong pace\_impact.

The plot for Pace\_impact vs raptor defense is shown in the following figure



From this scatter plot, it can be inferred that pace\_impact can have stronger or poorer rafter defense. For example, pace\_impact at 0 has raptor\_defense at 10 and -10. This indicates a range of values raptor\_defense can take for some values of pace impacts.

The plot for Pace\_impact vs raptor offense is shown in the following figure

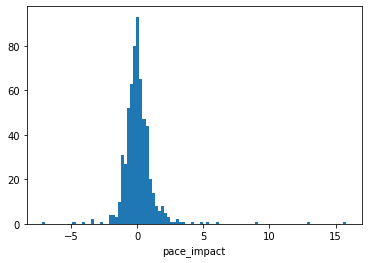


This is also similar to pace\_impact vs raptor defense. pace\_impact can have stronger or poorer rafter offense. For example, pace\_impact at 0 has raptor\_offense at 20 and -20. This indicates a range of values raptor\_offense can take for some values of pace impacts.

Histograms:

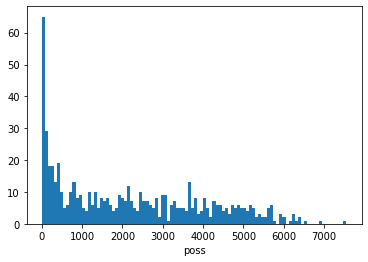
A histogram is a graph that shows the frequency distributions of the values that the variable takes. It shows how many times each value occurred. This information cannot be inferred from line plots and scatter plots.

The histogram plot for Pace\_impact is shown in the following figure



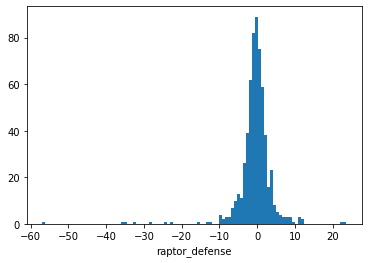
It shows that pace\_impact with value 0 and around it has occurred more times.

The histogram plot for ‘poss’ is shown in the following figure



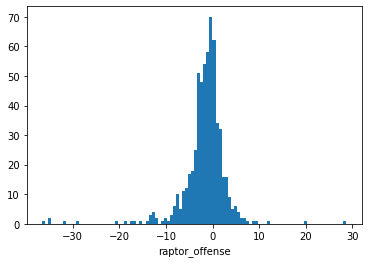
It can be inferred that possession with values 0 occurred more times and it decreased as possession value increased.

The histogram plot for raptor defense is shown in the following figure



It can be inferred that raptor defense points with 0 have occurred more times. The frequency count of raptor\_defense values with higher and lower values is less.

The histogram plot for raptor offense is shown in the following figure



It can be inferred that raptor offense points with 0 have occurred more times. The frequency count of raptor\_offense values with higher and lower values is less. This is similar to the histogram of raptor\_defense