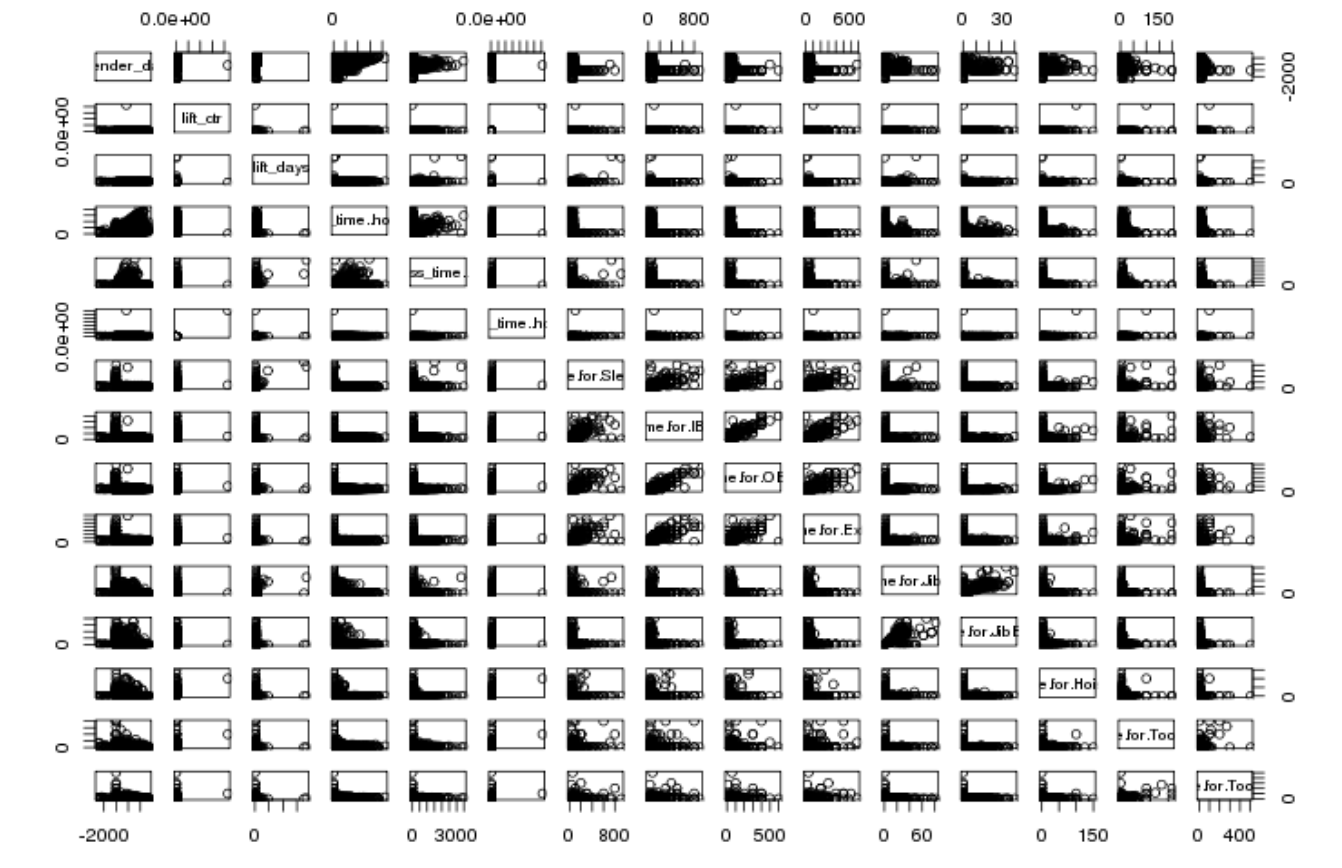
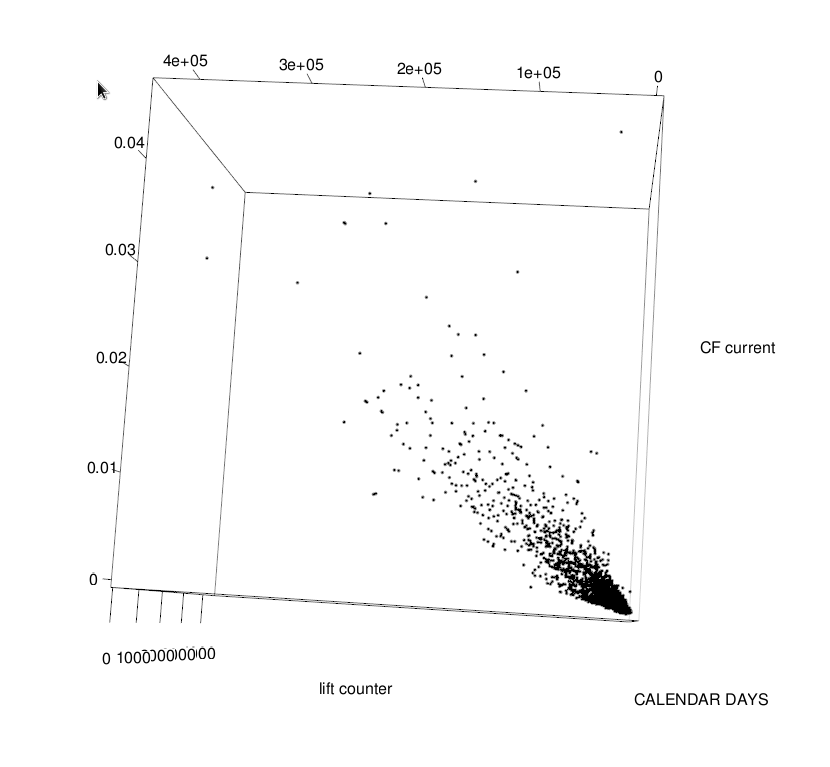
***HIAB report***

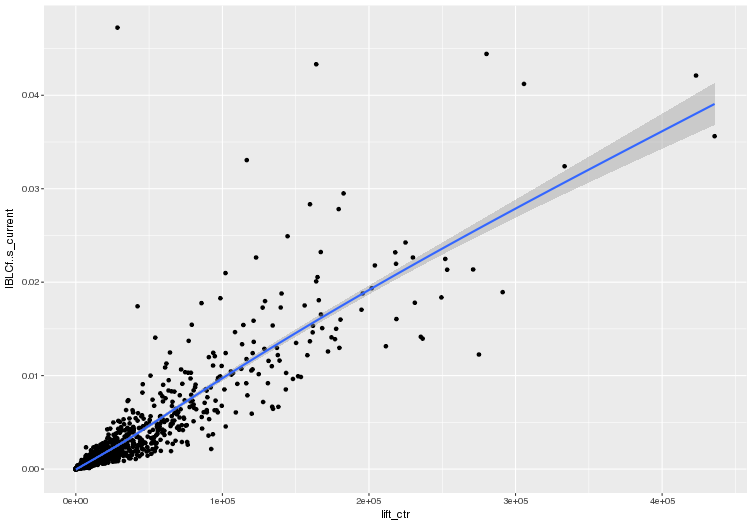
Key graphs that I might include



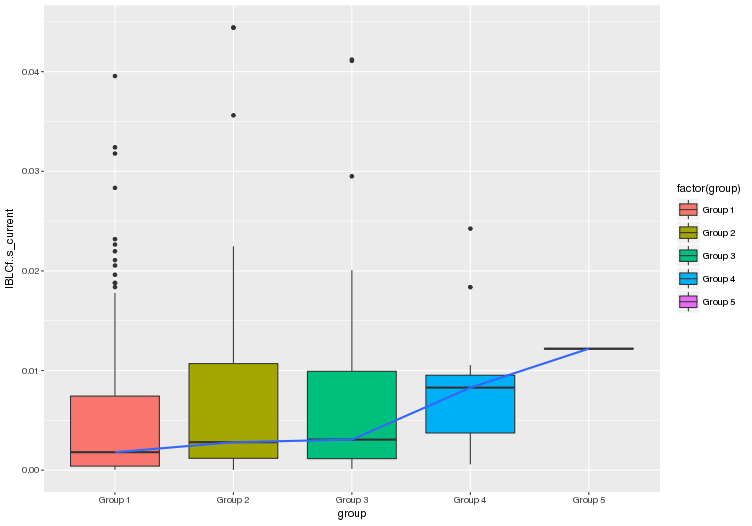
Beginning of any statistical analysis must be accompanied by understanding the data and knowing the variables. This can be done by plotting the numeric variables all together in a single plot in order to understand the behaviour of the variables. This is pretty much essential because there will be variables that will be highly correlated which can then be used to find interesting regression model. In the above image it’s clear that the variables at the centre have a strong correlation with respect to each other and the ones at the corners do not seem to have any correlations among them. So when preparing a model, selecting variables from the diverse ones say from centre of the picture and the middle ones will be useful. The categorical data can be ascertained by the fact that the data points tend to congregate on selected points on x axis.



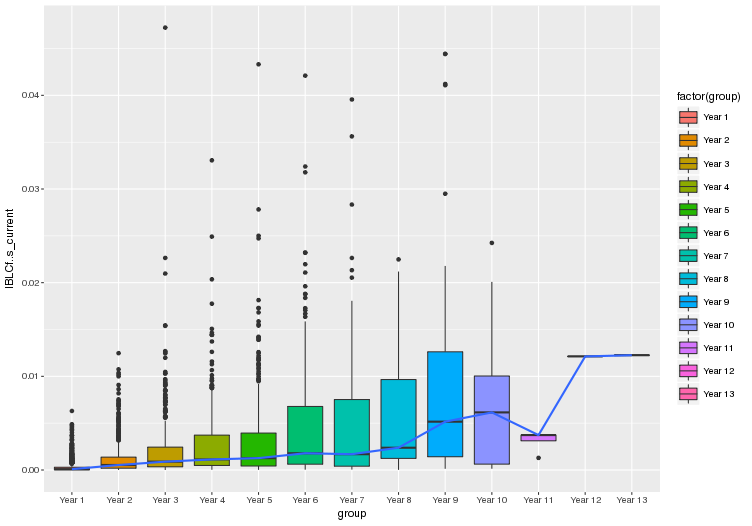
The next step includes finding out whether there is a correlations among variables that are under consideration. This gives a better insight over whether  a  concerned variable is correlated to two or more variables. These 3 D plots also provide a better insight for what variables that are supposed to be considered in case any regression model is being built for getting better fit lines. In the above graph it became quite clear that there is a strong correlation between CF current (Damage) and the lift counter which was useful in the to analyse the cause and effect in depth.



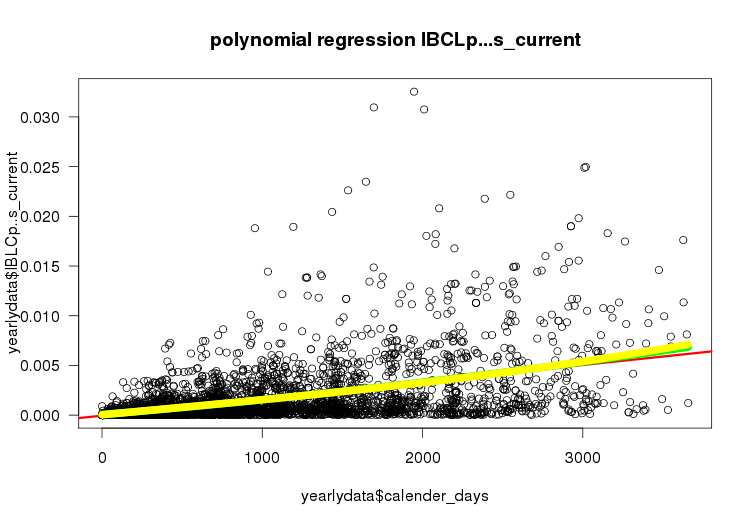
The graph showing relationship between damage and lift current. The 3 D plot was the reason why this was plotted. Most of the scatter plots are usually represented with a regression line and the line along with the grey area (confidence interval) shows the best fit curve for the given plot.

****

Damage occurring at various stages of crane’s life (calendar days). Number of days less than 2000 aren’t considered. Each interval consists of 500 Days. The line connecting these boxplots are the median values of each group and gradually increase as time progresses.



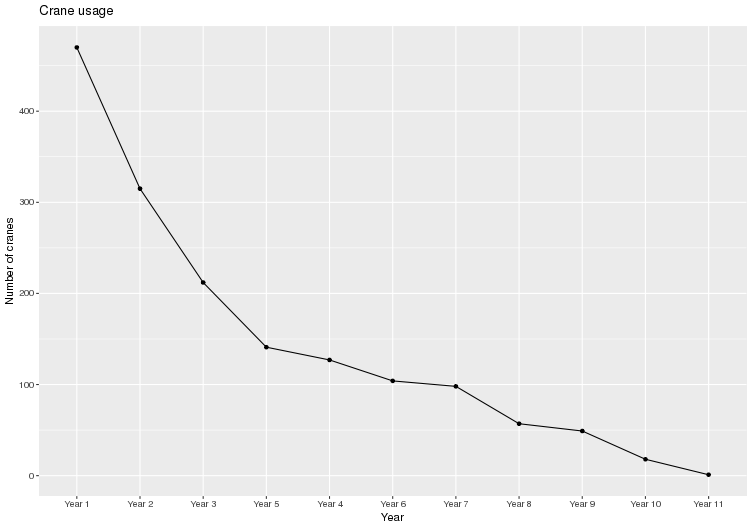
Analysis of damage based upon the calendar year. It can be seen that as the year progresses the dmage increases as well but not to an extent where there’s a considerable amount of increase. It’s only in the year 9 that the jump is a bit noticeable.



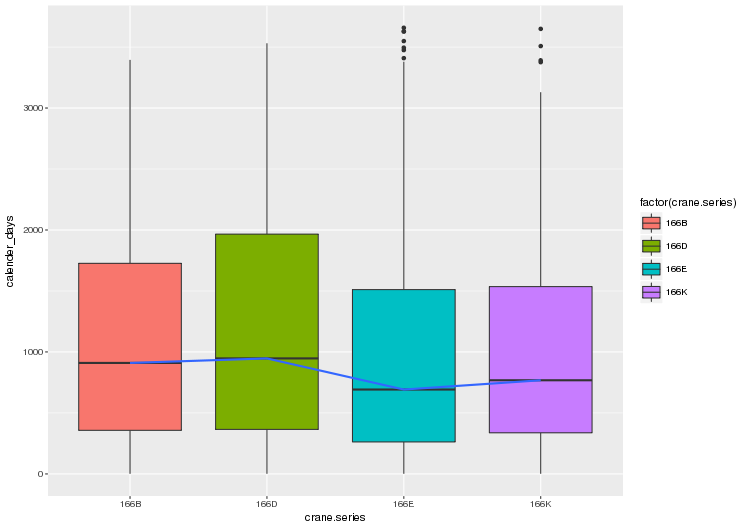
Damage vs calendar days was studied again, The task here was to find out which model suits the best. A comparison of linear fit (red line), quadratic fit (green line) and cubic fit (yellow line). Tough the three fits were found to be extremely close to each other, nonce of them could actually explain the data and the r squared value was too less.

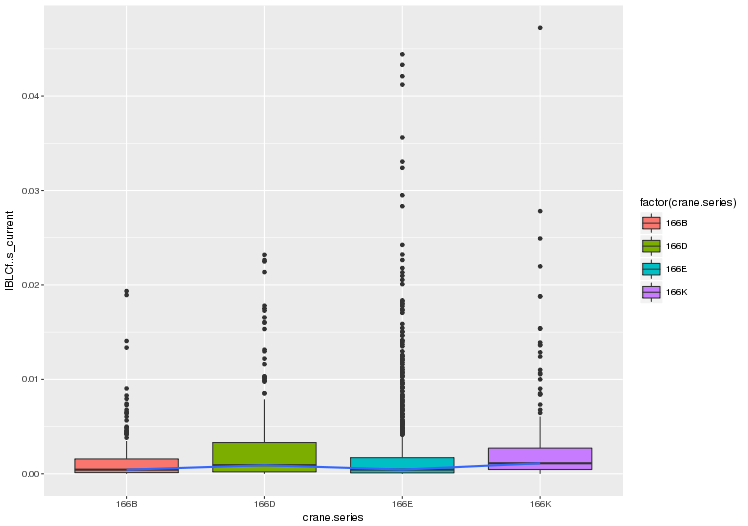
Summary of model:

Model 1: IBLCp..s\_current ~ calender\_days  
Model 2: IBLCp..s\_current ~ calender\_days + I(calender\_days^2)  
Model 3: IBLCp..s\_current ~ calender\_days + I(calender\_days^3)  
Since the R-squared value was too less, the line of fit was not seen feasible at this stage of the project. The fit for the 3 degrees of equations were approximately the same.

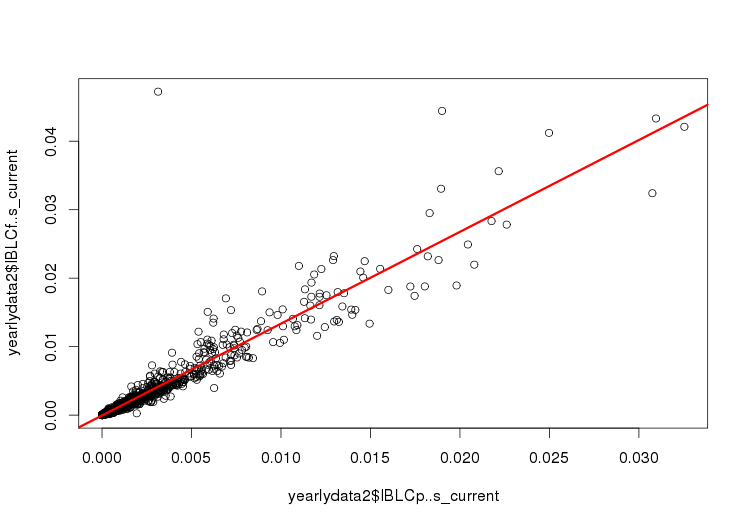


Distribution of crane usage over theyears/ The analysis found that there is a steep decrease in the first five years after which there is a gradual decrease.





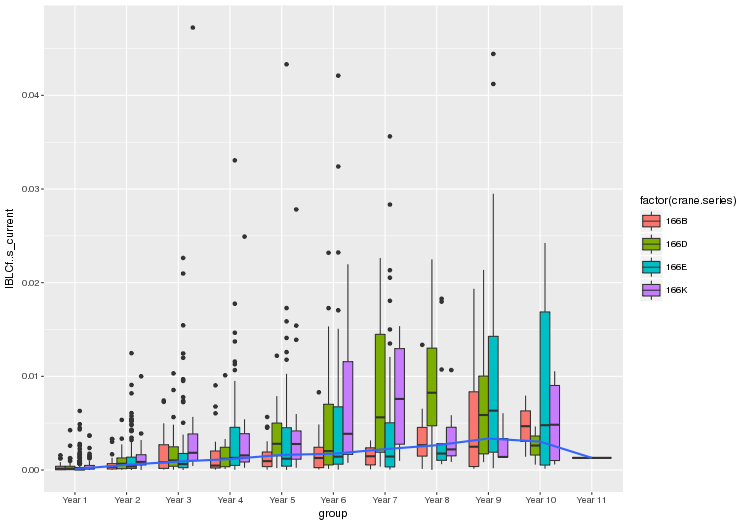
To get further insights, the extent of damage had to be checked across different crane models and crane series (shown above is only crane series). The first boxplots indicate usage across different crane series and the second box plot gives a brief idea of the damage incurred by different crane series. It can be seen that the ones that are most often used tend to have a higher damage rate (166 B series in this regard). This plot provided a crucial information regarding lift counter and damage being highly correlated.



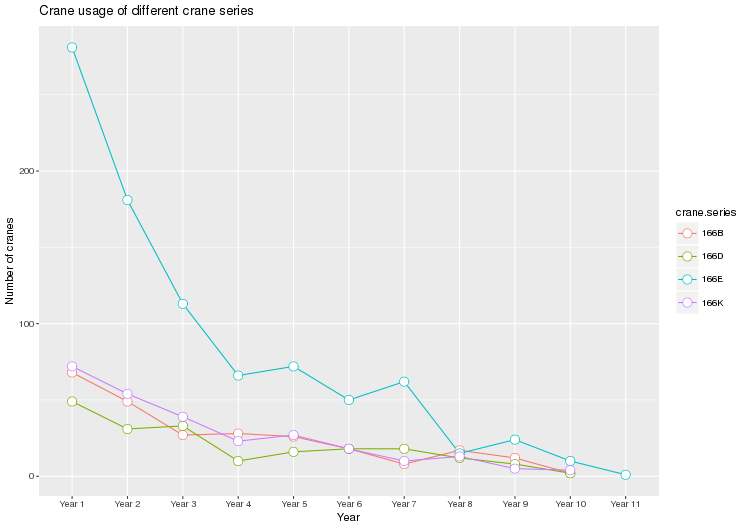
The perfect fit for the above data set between 2 aforementioned variables

***Y = 1.3408537571657 \* x + -5.2176258924594e-05***

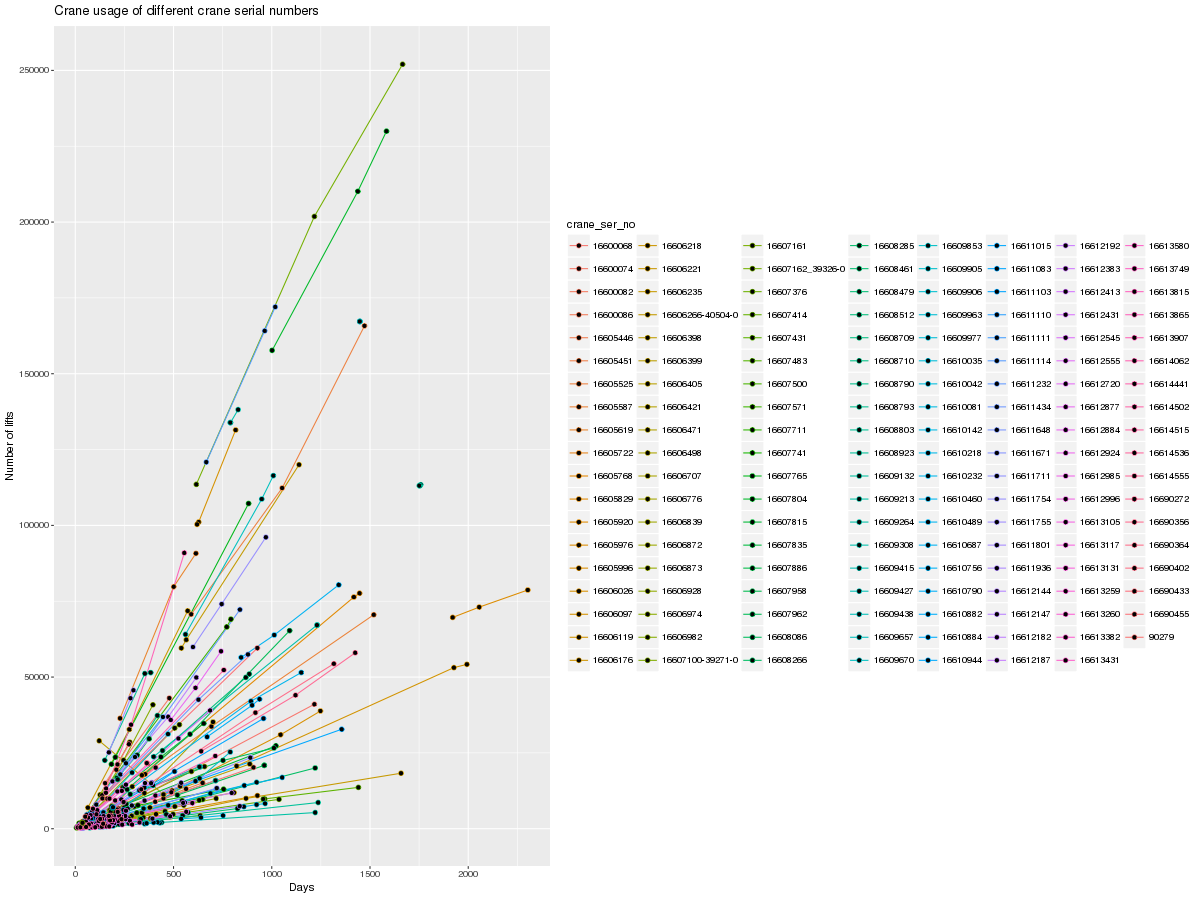
Line of fit approximately 90 %

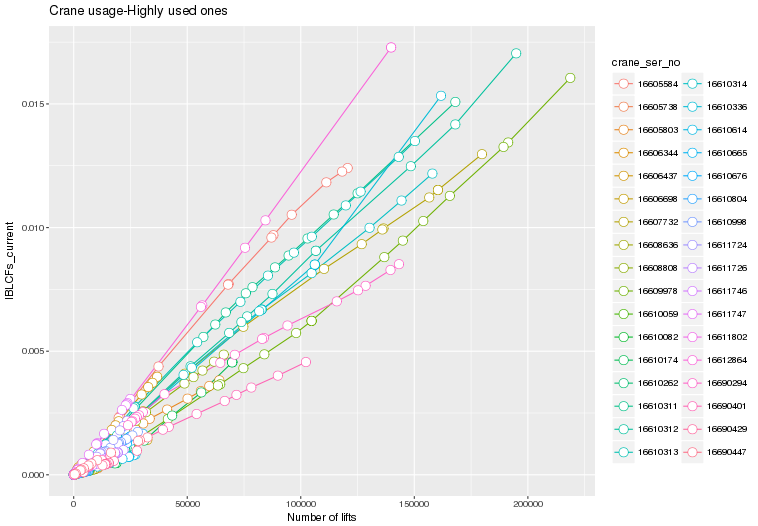


Another series of box plot indicating damage occouring as time progresses. This provided a great insight as well as it could be seen that the median damage across all the crane series were well below 0.01 though some models had damages way greater than the average ones.



Usage of cranes across different models. The above graph was plotted since it became essential to ascertain if any particular crane series had been used quite more often than the others. As can be seen, more of 166E cranes were utilized and other 3 model series being used approximately the same.





The final 2 graphs had to be plots involving the lift counters. A detailed plot of the number of lifts with respect to time grouping by crane serial numbers had to be essential since the characteristics of any individual crane behaving non linearly had to be studied. As can be seen with the graph most of the cranes had to undergo maintenance within the first 500 days of usage. The detailed in depth analysis of the same can be found in the log file.

The crane usage vs number of lifts had to be the ground breaking graph of this project as the expected S-curve didn’t turn up as intended to be. All of the characteristic curves had linear trajectory with slopes varying mildly.