Initially, I populated Tobin’s Q metric for the samples in the dataset. Please refer to the code and the output file. Based on this, I calculated correlation between R&D expenditure and Tobin’s Q metric. Correlation between R&D expenditure and tq came out to be: -0.01. Thus, the results suggested no correlation between them. Since there are patterns in data like common operating industry of firms and time frame of analysis, it is better to find correlation between R&D expenditure and firm’s performance by grouping based on these parameters. As a result of data grouping and analysis, the variable ‘**corr\_by\_time\_industry**’ states correlation segregated by financial year and industry. The results suggest that there is indeed a huge difference in correlation based on industry of operation and financial year. For example, firms with industry code **337910** and **423450** showed high correlation of **1.0** and **0.97** respectively between R&D expenditure and performance in year 2010 whereas firms with industry code **423690** and **333242** showed very poor correlation of **0.33** and **0.19** respectively in the same financial year.  
Note: Some of the correlation values are “not interpretable” and internally represent NaN because the values do not vary.  
Formula for calculating correlation:   
  
 **cor(i,j) = cov(i,j)/[stdev(i)\*stdev(j)]**

If the values of the ith or jth variable do not vary, then the respective standard deviation will be zero and so will the denominator of the fraction. Thus, the correlation will be NaN.

Additional note: Please refer to the code and output file attached.