**File description:**

*Note: Code is divided into part. All R files use the same data (uploaded in 1 folder). All files require libraries installed in R code from 1 folder.*

1\_DataCleaning

* Installing libraries (importing& organizing& visualization)
* Changing names of variables
* Changing type of variables
* Checking % of NA
* Removing empty or useless variables
* Dealing with outliers

2\_DataVisualStatistics

* Using ready code enabling to put a few graphs on one sheet
* Grouping data (operation on data tables)
* Visualizations with ggplot2 (Citation, First Page; Output type).
* Tackling also results (In % and absolute values) for both REF and RAE
* Comparison of result

3\_CorrelationAndPlots

* Correlation for income exists (0.55)
* Checking correlation between other output stars
* Plotting. Plot grouped by source.
* Seeking for correlation between % of total income per institution from a given sources against result. No correlation.
* Seeking for correlation between total income per institution from a given sources against result. Correlation for: BIS Research Councils, Royal Society, British Academy and Royal Society of Edinburgh (0.79). For other, even if ‘exists’ the values are to small to be reliable for testing.
* There is correlation while grouping by BIS (Business, Innovation, Skills), Industry, Government as a source of income (0.33, 0.59, 0.62)
* Results plotted
* Checking for output type against 4\* result (Taking into account: chapter in books, books, working papers). We can observe correlation for Working Papers (0.6).
* Seeking for correlation between 4\* and number of FTE A staff (0.48) Plots suggest stronger correlation, but there is a high variance.
* Trial to check publishers. For this UOA not enough data.