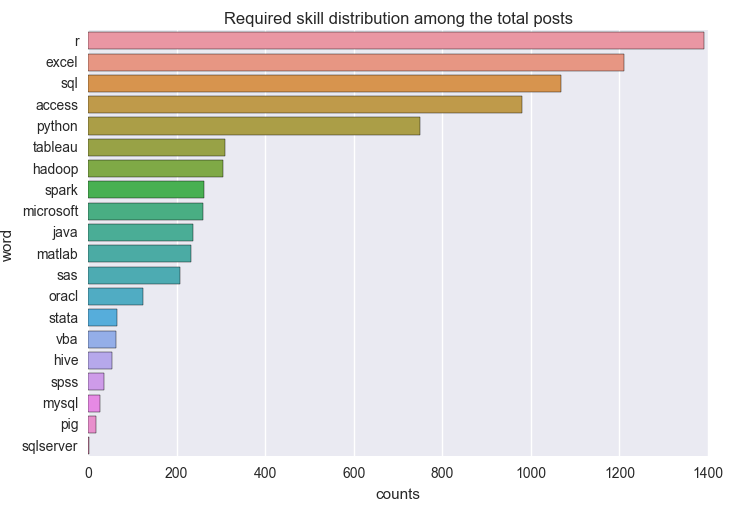
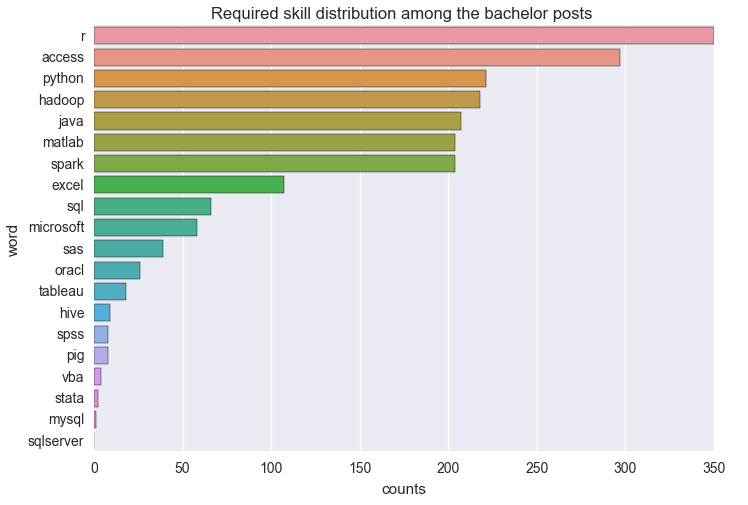
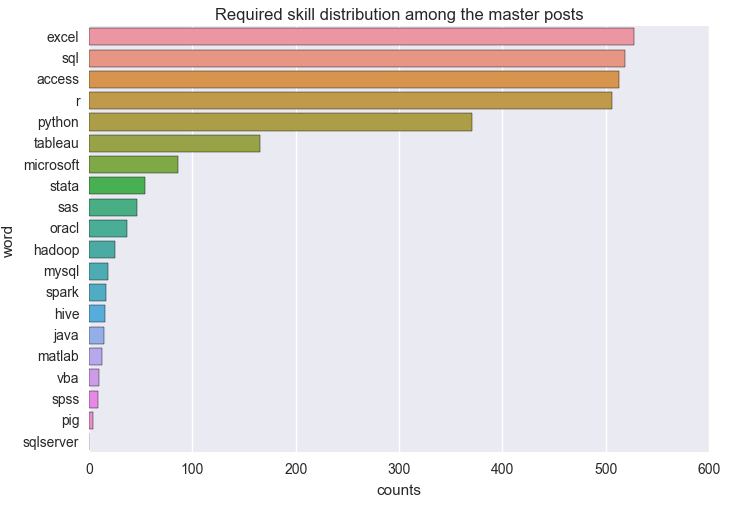
Barplot analysis



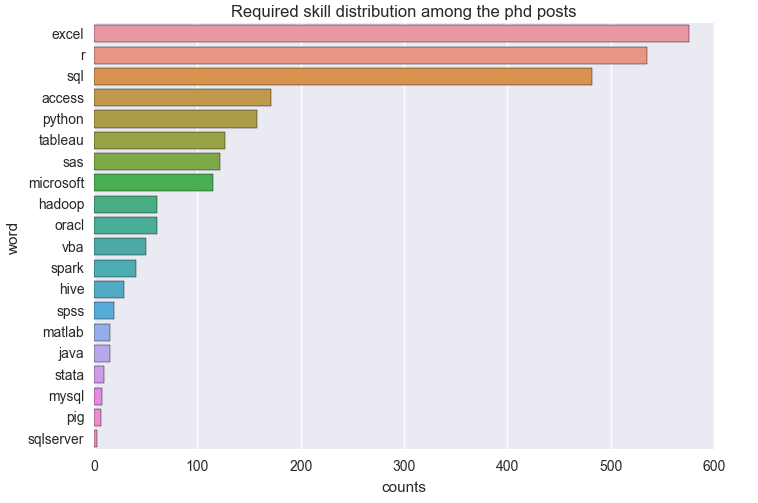
This graph shows the required skill distribution among the total posts. In the graph, the horizontal axis stands for the “counts”, which means the total number of every skill needed in the total 2445 “Data Analyst” posts. On the other hand, the vertical axis means the skills’ name that is required among the total posts. As the different skill distributed in the graph, we can find that the most used skill is “r”, and then the second most required skill is “excel”. And the next most required skill are “sql”, “access”. And then “python” is also used in a wide range. While the “tableau”,”hadoop”, “spark”, “microsoft”, “java”, “matlab”, “sas”, “oracl” are not used as often as the other skills we mentioned above. In the end, “stata”, “vba”, “hive”, ”spss”, “mysql”, ”pig”, “sqlserver” are required less, in some word, they are almost not required except some special cases. These are the total “Data Analyst”posts’ required skills.



This graph shows the “Data Analyst” required skills distribution among the bachelor degree. In the graph, the horizontal axis stands for the “counts”, which means the total number of every skill needed in the bachelor level “Data Analyst” posts. On the other hand, the vertical axis means the skills’ name that is required among the total posts. As directly showed in the graph, we can find that the most used skill is “excel”, and then the second most required skill is “r”. And the next most required skills are “sql”, “access”. And then “python” is also used in a wide range. While the “tableau”,”hadoop”, “spark”, “microsoft”, “java”, “matlab”, “sas”, “oracl” are not used as often as the other skills we mentioned above. In the end, “stata”, “vba”, “hive”, ”spss”, “mysql”, ”pig”, “sqlserver” are required less, in some word, they are almost not required except some special cases. There are not too many differences between the bachelor degree and total statistics. The biggest difference is that the most required skill is “r”, which is obvious more than other skills. But “excel” is not required as more as other degree required. This may because of the undergraduate students can arrive at the proficient “r” level.



This graph shows the “Data Analyst” required skills distribution among the master degree. In the graph, the horizontal axis stands for the “counts”, which means the total number of every skill needed in the master level “Data Analyst” posts. On the other hand, the vertical axis means the skills’ name that is required among the total posts. As directly showed in the graph, we can find that the most used skill is also “excel”, and then the second most required skill is “sql”. And the next most required skill are “r”, “access”. And then “python” is also used in a wide range. While the “tableau”,”hadoop”, “spark”, “microsoft”, “java”, “matlab”, “sas”, “oracl” are not used as often as the other skills we mentioned above. In the end, “stata”, “vba”, “hive”, ”spss”, “mysql”, ”pig”, “sqlserver” are required less, in some word, they are almost not required except some special cases. There are not too many differences between the master degree and total statistics. The most big difference is that the most required skill becomes into “excel” rather than “r”, this may because of the master level students acquire more kinds of skills not only be limited into “r” programming language.



This graph shows the “Data Analyst” required skills distribution among the phd degree. In the graph, the horizontal axis stands for the “counts”, which means the total number of every skill needed in the phd level “Data Analyst” posts. On the other hand, the vertical axis means the skills’ name that is required among the total posts. As directly showed in the graph, we can find that the most used skill is also “excel”, and then the second most required skill is “r”. And the next most required skill are “sql”, “access”. And then “python” is also used in a wide range. While the “tableau”,”hadoop”, “spark”, “microsoft”, “java”, “matlab”, “sas”, “oracl” are not used as often as the other skills we mentioned above. In the end, “stata”, “vba”, “hive”, ”spss”, “mysql”, ”pig”, “sqlserver” are required less, in some word, they are almost not required except some special cases. There are not too many differences between the phd degree and total statistics. The most big difference is that the most required skill becomes into “excel” rather than “r”, but “r” also has a high frequency among these required skills. This may because of the “Data Analyst ” posts that is suitable for the phd students are more target to the “excel” and “r”, which can give the phd students a hint that “excel” and “r” play an important role in the “Data Analyst” posts.



This graph shows the term frequency that corresponding to “Data Analyst” posts required skills. In the word frequency graph, the size of the word stand for the appeared frequency of each skill. The bigger size showed, the more the skill is required. As intuitively showed in the graph, the most required skill is “R\_PROGRAMMING”, which is also called “r”. And then the second most required skill is “excel”. And the next most required skill are “sql”, “access”. And then “python” is also used in a wide range. While the “tableau”,”hadoop”, “spark”, “microsoft”, “java”, “matlab”, “sas”, “oracl” are not used as often as the other skills we mentioned above. In the end, “stata”, “vba”, “hive”, ”spss”, “mysql”, ”pig”, “sqlserver” are required less, in some word, they are almost not required except some special cases. This is corresponding to our conclusion in the bar plots at first, while this graph is more intuitive.