데이터과학 Homework 2

*Jupyter notebook에 과제 해결과정을 자세히 기록하여 해당 notebook (.ipynb) 파일을 제출합니다.

All questions given below uses the births.csv file attached in the assignment (also can be find in the Python Datascience Notebook folder).

1. (sorting) Sort the births data in an ascending order using the following methods and record the running time using **%timeit**.
   (a) Selection sorting
   (b) Insertion sorting
   (c) Merge sorting
   (d) Bubble sorting

   Then compare the outcome with np.sort command. Give your reasoning on which of the methods above is the most effective algorithm for sorting large data (considering both time and stability). You do not need to print the whole data array when answering this question.

2. Using sorting and indexing methods of your choice (including partitioning and argsort) answer the following questions.
   (a) Calculate the number of entries with births between 7000-10000 and find the indices of minimum, maximum values of this range.
   (b) Select the data for May 20th of each year and create an array containing only those values. Align the entries in an ascending order and give the minimum and maximum values.

3. Draw a histogram of births by hand and compare the running time with the NumPy routines. Which is more efficient in visualizing the data?