US Names

In this assignment you will do a basic analysis on names of children born in the US between 1880 and 2014. The analysis will be embedded into a small framework that is accessible through a web service.

The data set of US names is available on Kaggle:

`https://www.kaggle.com/kaggle/us-baby-names`

We will use the file `NationalNames.csv`. You need (to create) a Kaggle account to download these data.

You are free to use the programming language, analysis tools, and web framework of your choice. The only requirements are that we can read your code and run the final framework on an Ubuntu 16.04 system with Linux kernel 4.4.0, Python 2.7.11/3.5.1, Java 1.8.0, and GCC 5.3.1. As a suggestion, you could build the framework in Python, doing analysis with the Pandas library and using the Flask web framework to process analysis requests and publish the results.

Code Repository

Please use a Git repository for your code and share it with us by publishing on GitHub. (Please do not add the data set, only the code.) We would like to see incremental commits at sensible checkpoints (e.g., at the completion of a feature) to see how the code has evolved. Create a `README.md` file, in which you describe what the code does and how to use the framework.

Analysis

Analyse the data for a combination of two arbitrary names, which will later be specified by a user. Take into account the facts that some names do not appear every year and that names may appear twice for a given year. We would like you to extract

- counts for the specified names as a function of the year (create plots later)
- the mean and standard deviation of the yearly counts for each name

Web Application

Make the analysis available as a web application. A user must be able to enter two names in their web browser and click on a "run analysis" button to analyse the data for the specified names. When the analysis finishes, the results are also shown in the browser. Make sure you describe in your `README.md` file how to start the web server you are using and how to use the web interface. Please show the following results:

- the mean and standard deviation of the yearly counts for each name
- plots of the counts for the specified names as a function of the year
- a scatter plot of the yearly counts for the two names, to visualize their correlation