Exam Data Engineering :

Meteorite data recordings

Mario Veruete

March 2021

You are going to work on a dataset about meteorite landing sites in the world containing the following informations: Name, ID, NameType, Classification, Mass, Fall, Year, Coordinates.

The dataset can be dowloaded here or directly from the NASA's website.

I expect a report for April 1st 2021 (pdf file only) containing a detailed description of the computations you did in order to answert the questions. The best is if you use LaTeX but this is not mandatory. You will post your code on a Github repository that you will share with me. Please put the link to your repository in the pdf file.

Remark: The data may need to be treated before doing a particular computation (e.g. delete missing values,...).

1 Questions

1. Make an histogram of the mass distribution of meteorites. Do it again for the meteorites having a mass less or equal to 50 000 grams.

2. Make a plot of the number of meteorites as a function of time (by year). Find a linear fit ($y = ax + b$) that approximates the trend of the curve. Using this function say what would be the number of landing meteorites next year. Is this approach of prediction scientifically robust?

3. We will concentrate now in the case of Oman. Create a plot of this country with different points representing the spatial distribution of the landing sites.

4. Propose a distribution (uniform, gaussian, cobinaison of different ones...) in order to describe the distribution of meteorite landing sites in Oman.

5. Based on that distribution compute the probability that a meteorite land in the circle of center {latitude,longitude}={18.9644, 53.9555} and radius equals to 100 Kilometers.