

Strange years

Yohan Duarte

September 18, 2024

Strange years

Here we parse and analyse the french “Deceased persons file” from the INSEE database. We recover thoses datasets using the api of the data.gouv.fr website.

Brief overview of the dataset

The Data contains the following entries:

- Family name
- Name
- Gender
- Date of birth
- Code of the locality of birth
- Name of the locality of birth
- Country of birth
- Date of death
- Code of the locality of birth
- Code of the death certificate

With the years of births and deaths, we can graphs the distributions through time as shown in the figs. 1, 2, 3. We can see the impact of the two world wars in numbers of births during the periods 1915-1919 and 1940-1945.

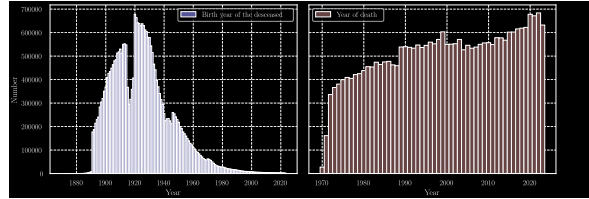


Figure 1: Distribution of the years of birth and death in the dataset

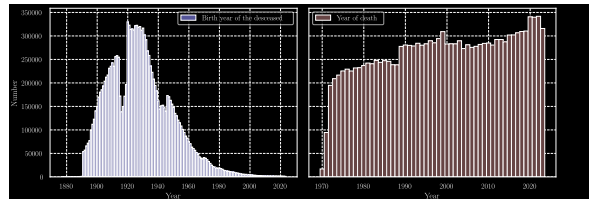


Figure 2: Distribution of the years of birth and death for the male

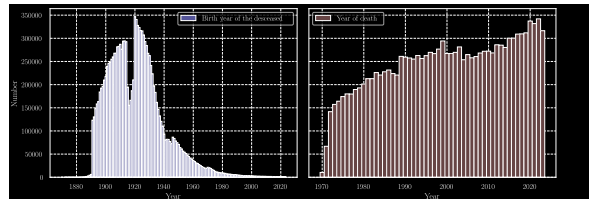


Figure 3: Distribution of the years of birth and death for the female

We can also see the link between the birth year and the year of death, in the figures figs. 4, 5, 6. Those figures show a line when the year of birth match the year of death, we can easily understand that infant mortality is the cause of of that structure in the figure. Another interesting aspect that is highlighted by this figure is the presence of a

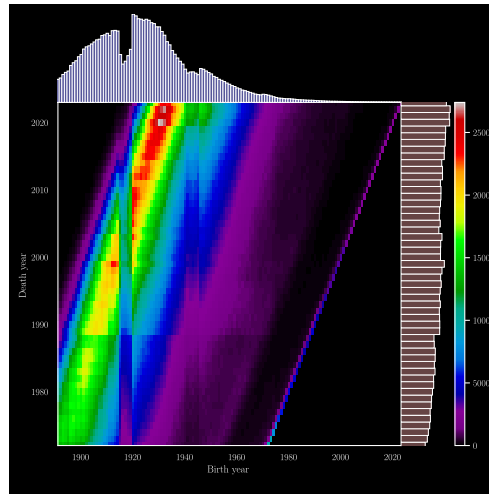


Figure 4: Distributions of year of death given the year of birth

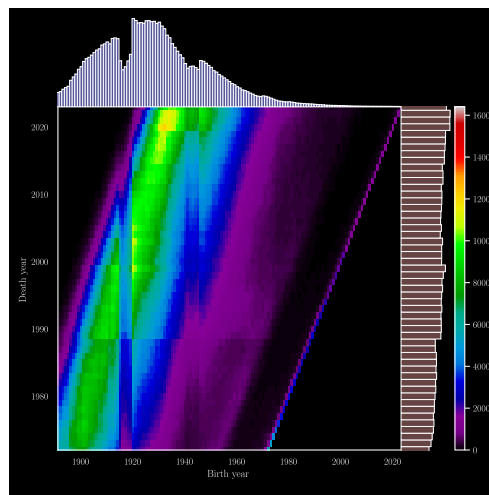


Figure 5: Distributions of year of death given the year of birth for the male

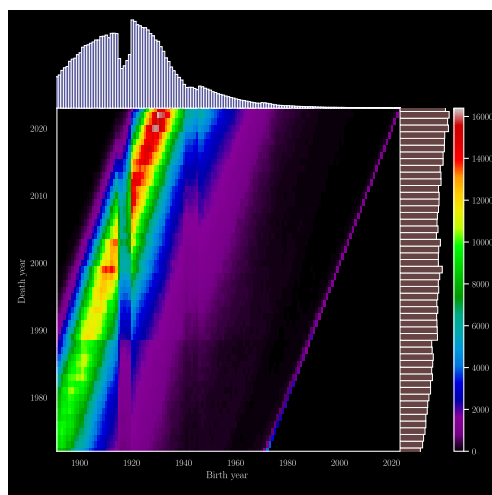


Figure 6: Distributions of year of death given the year of birth for the female