PROCESS BOOK - CRIME FACTORS PROJECT

Guillame Augustoni and Elena Gazzarrini

OBJECTIVE

Crime is a huge political subject in France. Its causes are also subject of intense political debate. Different factors going from immigration to poverty have been proposed, but evidence seems to be lacking in the political debate.

Usually, maps about crime rate and immigration are displayed together as a conservative political argument to limit immigration into the country, as shown in Figure 1. An example of a study investigating this relation is a 2009 paper by Yu Aoki and Yasuyuki Todo.

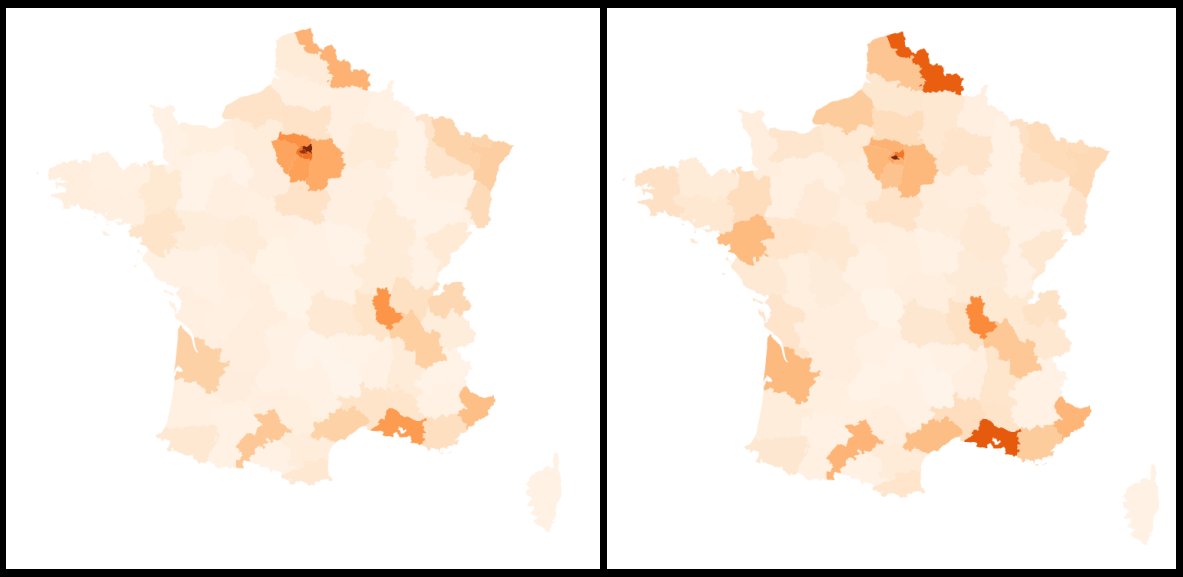


Figure 1. Immigration (left) and crime (right) rates in France, used as a political argument to limit immigration into the nation.

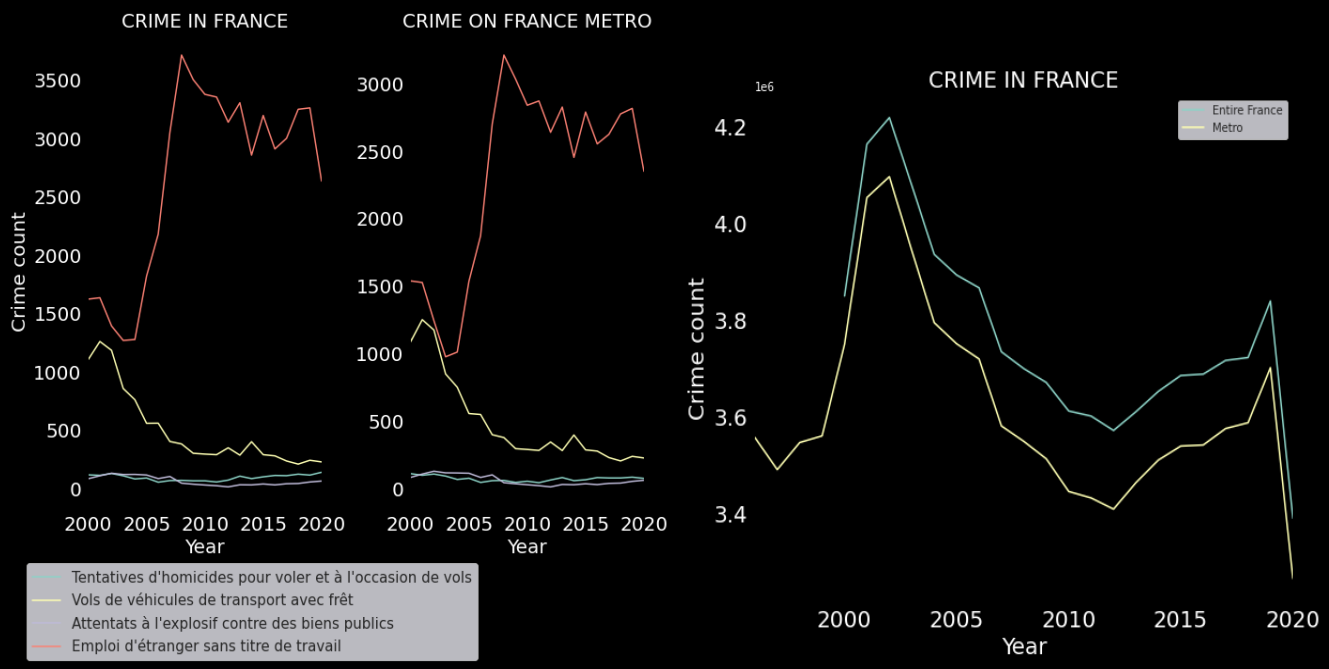
Looking at these maps we may arrive at the conclusion that immigrants are responsible for the crime in France. However, correlation is not causation there may be multiple other factors affecting the crime rate in France.

The aim of this project is to investigate the correlation between France's crime rates at a departmental level and other parameters describing the nation's wellbeing, such as education, poverty, immigration and unemployment. To show the impact that different factors have on crime rates we employ linear models. Interactive maps and plots are developed to allow the user to select the factors, change the parameters and see their evolution in time.

DATA COLLECTION

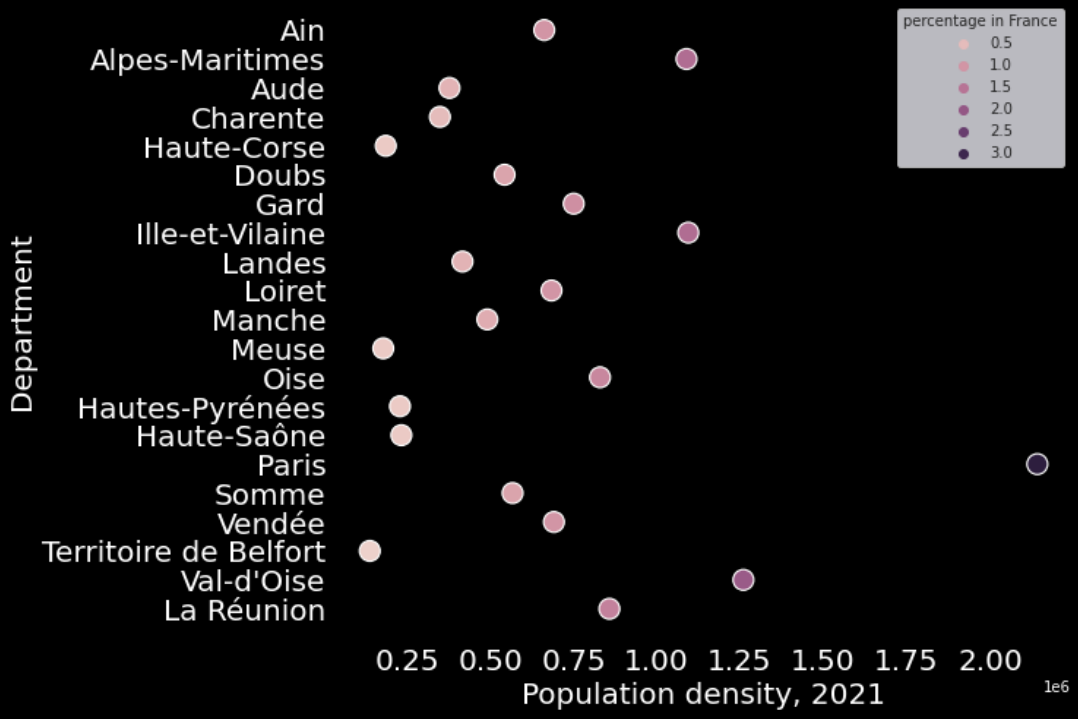
One of the main challenges throughout the project was collecting coherent data. The main dataset is the crime dataset, which contains a number of mensual crimes reported to the police per departement for 106 categories of crime (examples include "Règlements de compte entre malfaireurs", "Infractions à la législation sur les chèques", and "Autres délits économiques et financiers"). This dataset includes data on crime back to 1996 for most departments however, for overseas France some years are missing. Our website will focus on mainland France as overseas France has unique circumstances which will biais the data. Linear models will use crime data from 2017 as most of the other indicators come from the years 2016-2017. Data is mensual but has been grouped by year by us for easier analysis.

The data is in the form of an excel spreadsheet with one sheet per departement, this form of data is being transformed to a more usable .csv format. An initial graph is plotted to see the crime rate increase from year 2000 to 2020. The plot on the right shows the metropolitan crime rate, which we see makes up most crime cases in France.

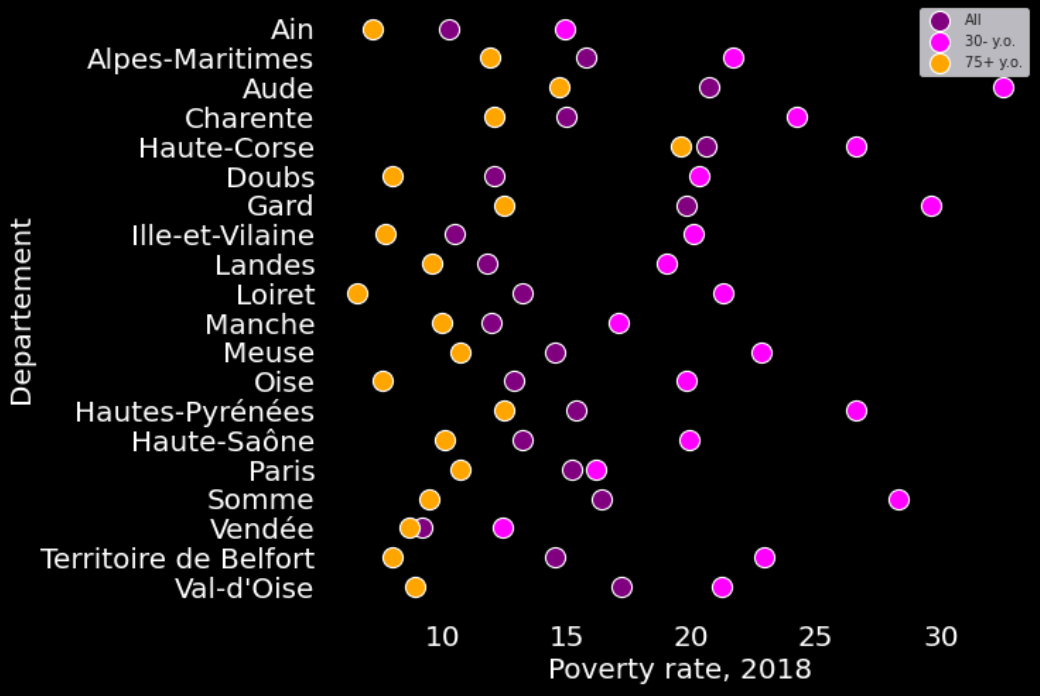


We import separate individual Excel files from the same website in order to be consistent all the way throughout. We start an exploratory analysis by plotting the data on jupyter notebooks, employing pandas and seaborn packages.

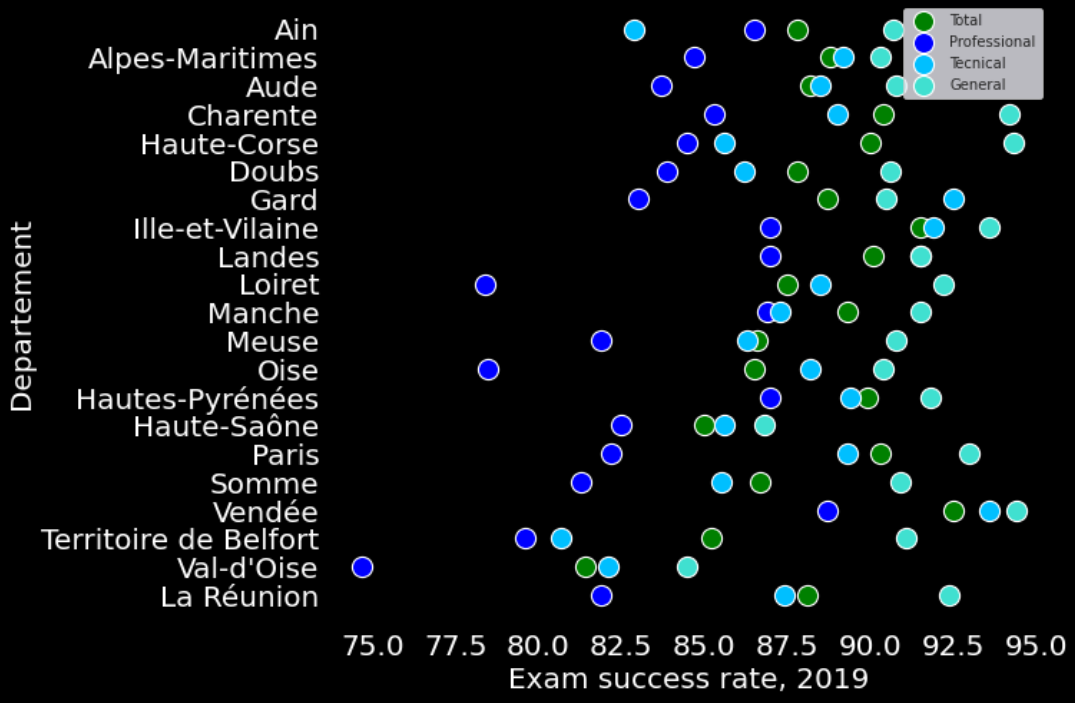
- POPULATION density in 2021 fro some specific departments, with Paris having of course the highest density



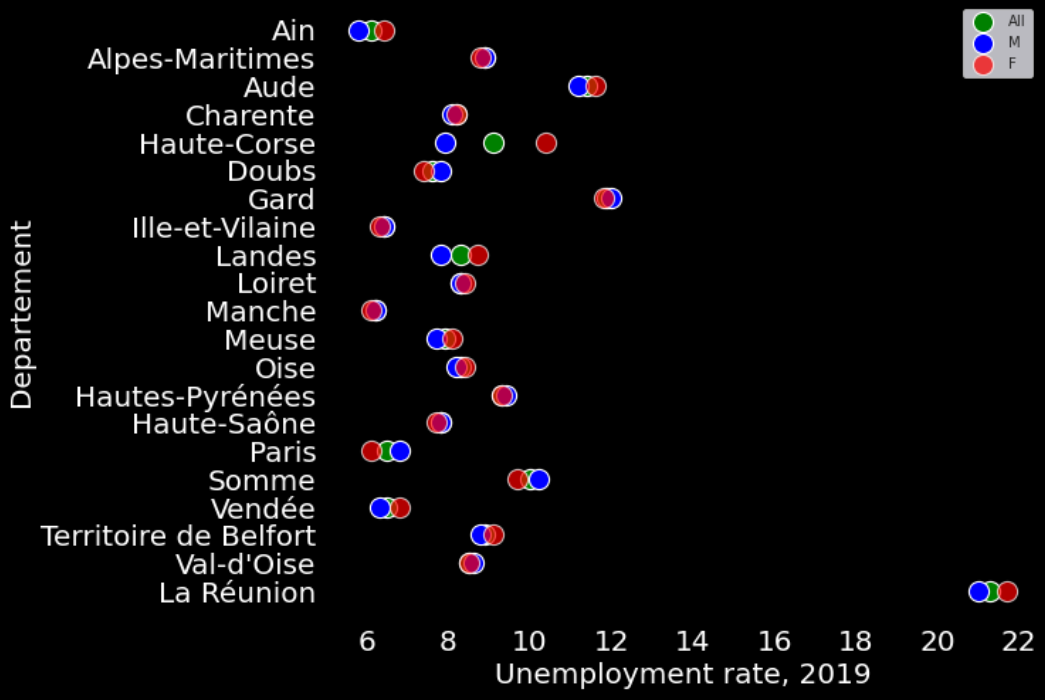
- POVERTY rate in 2018: people less than 30 years old live in more poverty.



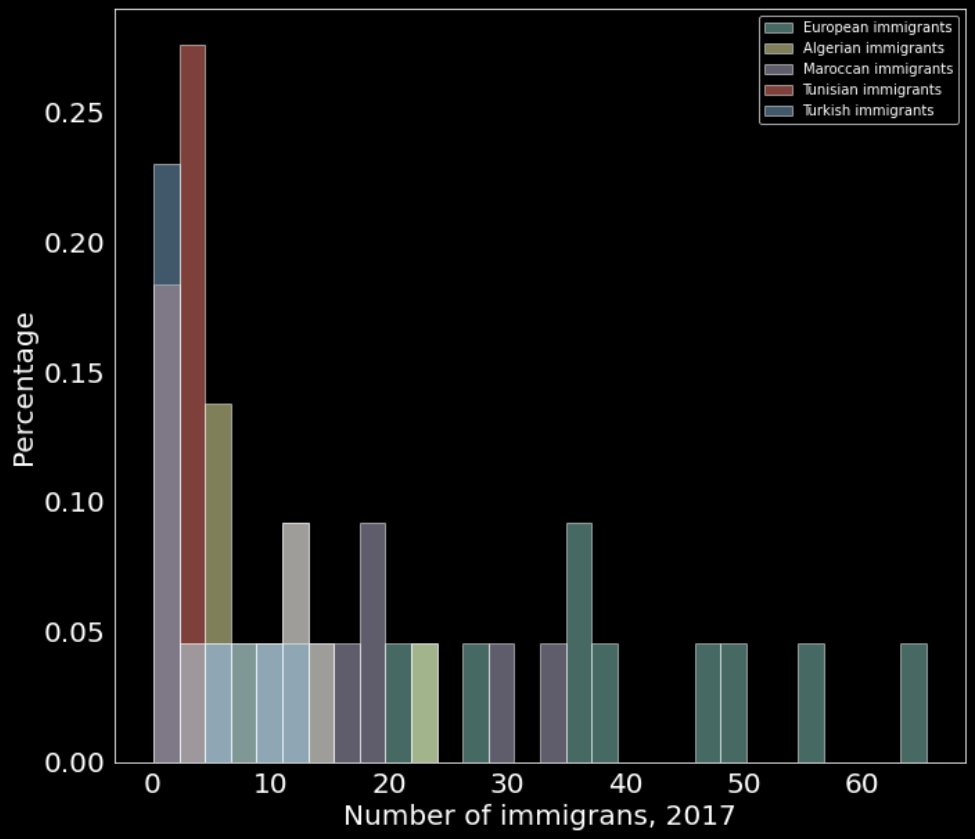
- EDUCATION, SUCCESS RATE in exams in 2019. These the results of BAC exams per each type of school (professional, technical, general). People who attend general BAC exams are more likely to perform better.



- UNEMPLOYMENT rate in 2019 based on sex. Depending on the region, this parameter does not offer a general trend.



- IMMIGRATION rate in 2017 for each migrants' nationality. European immigrants compose a larger part of the migrant population, followed by Maroccan ones.



We realize, when starting to build the app with React, that having separate .csv files for each variable slows down our process. Moreover, the years are not consistent with each other.

INSEE complete, consistent, reliable database

We therefore start looking for other more integratable datasets and we find a complete dossier on the official INSEE website (https://www.insee.fr/en/accueil) which only includes the years 2007, 2012 and 2017. The .zip file is ~180 MB and it contains innumerable information for each town in these three years, from number of families per household to population, surface, and much more. We only take the relevant information:

- ☐ Surface area
- ☐ Population
- ☐ Number of empty/abandoned houses
- ☐ Number of people with a salary
- ☐ Number of unemployed people between 15 and 64 years of age
- ☐ Number of not educated people of 15+ years of age
- ☐ Number of men aged between 15 and 29
- ☐ Number of monoparental families

Since the data is per town, we group all the towns within the same department and we again only consider inland France, since the data for ex-colonies is incomplete.

We then integrate the **crime** (in years 2007, 2012, 2017) and **immigration** (unfortunately we only have the one for 2017, but it should be enough to show the correlations for that year) data from the crime dataset cited above.

In this way, we retrieve a fully working unique .csv file that we can import into Visual Studio Code and start working on it to create beautiful, powerful interactive visualisations.

VISUALISATION

We employ Visual Studio Code to develop the React website.

The website styling is determined by the .css file and most of the code development is done through the Typescript language. The function which projects each data point to the map is the most time consuming one.

We use linear models to show correlations between the crime data and the other parameters. On the app, we display the factor R2 every time the visitor chooses a parameter from the legend on the left of the map. This factor represents the proportion of the variance in the crime that is predictable from the chosen factor, hence symbolizing the strength of the correlation between them.

An interactive timeline is developed for both a geographical framework and a more statistical one, where we show the evolution in time of scatter plots (with one dot corresponding to each one of the 95 departments).

The scatter plots display the crime rate on the y axis, while the x axis represents the variables selected by the external user.

CHALLENGES

There were two main big challenges we encountered throughout the project:

- From a *statistical* point of view, making sense of the innumerable and disordered amount of data we collected was not trivial. We could have chosen an easier problem to investigate with an already cleaned dataset available, but our aim was to investigate a politically and socially relevant problem which had not yet been looked at in detail.
- From a *visualisation* point of view, correlating the data for each department with the geographical projection was the hardest part. Since we had innumerable data, the function mapping each data point to the map was taking a long time to compile.

PEER ASSESSMENT

Guillame was skilled at using Typescript and set up most of the skeleton of the React app. He developed the map projection and imported the data correctly to create the visualisation, implementing the linear regression model.

Elena focused on the data cleaning initial part and on making sense of the data through Jupyter notebook plots, while giving suggestions on the visualisations on a more theoretical level. She focused on the styling of the web page and process book.