Multivariate data analysis: Assigment 2

Disclaimer: this study is clearly not exhaustive, the original dataset contained more than 30GB of data.

In this Assignment, you will be working on the dataset called `DatasetAssignment2.csv` that you can find on GitHub. This dataset has been extracted from the (huge) collection of data from Julia Cagé & Thomas Piketty (2023): *Une histoire du conflit politique. Élections et inégalités sociales en France, 1789-2022*[1], Paris, Le Seuil. Here, I restricted the dataset to a smaller number of variables. All the data is given for each municipality of France. Note that because of some changes in geography I removed the municipalities that were not present for each variable.

You will find at the end of this document the description of all the variables.

Using this dataset and the techniques seen during the course, you will write a report (**maximum four pages** including graphs) by answering some of these questions in a quantified way, using graphs when you think it is useful. Note that these directions are only propositions, it is better that you treat three questions with care than rushing through all of them:

- Examine the relation between the variables, are there interesting relations, for example between the results of some of the candidates between 2022 and 2017?

- Are the results of the 1981 election a good predictor?

- Can you find the important sociological variables to predict the results of the 2022 election?

- Imagine a municipality belonging to an agglomeration of 50000 inhabitants, can you predict the hypothetical score of Macron in 2022, with which precision? Same question if we add that there are 25% of managers in the population? And if we add that the mean household revenue is 22000?

- Some of the candidates seems to be really close, do you see differences between the results for Mélenchon in 2022 and 2017? Between Arthaud and Poutou in 2022? Between Zemmour and Le Pen? All of these questions have been thoroughly answered by commentators, what can statistician bring to this analysis?

- The main conclusion of Piketty and Cagé was that Macron's vote was particularly linked with capital ownership and socioprofessional situation rather than for example the fact that people are living in urban areas.

- Feel free to add other interesting results you might find.

You can refer to the UK Statistic Authority *Standards for Statistical Reports* when preparing your report. Your report will be graded as follow:

- Presentation and writing: Your report is clearly written and comprehensible, it is structured, and the grammatical and typographical errors (if any) do not impedes understanding **[3 marks]**.

- Analysis and context: Appropriate and relevant use of graphics and statistical analyses. Your conclusion are supported by sound analysis and are accessible to a sociologist or even a political journalist. **[7 marks]**

---

[1]An history of political conflict. Elections and social inequalities in France, 1789-2022, the full dataset is available at `unehistoireduconflitpolitique.fr`

Description of the variables

First, we start with sociological information.

- `codecommune`: a unique code to designate the municipalities, it has the form $(X)XYYY$, where XX corresponds to the *département* and YYY a number inside the *département*. It is thus possible to split the data in geographic areas by using this code.

- `nomcommune` the name of the municipality.

- `capitalimmo2022` and `capitalimmo2017` the ratio of mean per habitant land value with respect to the national mean in 2022 and 2017.

- `revmoyfoy2022` and `revmoyfoy2017` the mean revenue per household in euros in 2022 and 2017.

- `revmoy2022` and `revmoy2017` the mean revenue per inhabitant in euros in 2022 and 2017.

- `ppropri2022` and `ppropri2022` the proportion of household that own their home in 2022 and 2017.

- `popagglo2022` and `popagglo2017` the population of the agglomeration the municipality belongs to in 2022 and 2017.

- `pop2022` and `pop2017` population of the municipality in 2022 and 2017.

- `perbac2022` and `perbac2017` percentage of the population that has a *baccalauréat* (equivalent to A-level) in 2022 and 2017.

- `persup2022` and `persup2017` percentage of the population that has a university diploma.

- `pagri2022` and `pagri2017` population that is employed as farmer in 2022 and 2017.

- `pcadr2022` and `pcadr2017` proportion of managers and leaders (employees of higher status) in 2022 and 2017.

- `pouvr2022` and `pouvr2017` proportion of physical workers in 2022 and 2017.

- `pempl2022` and `pempl2017` proportion of *employés* (that is employees that do intellectual work but are not part of the managers, equivalent to "clerk") in 2022 and 2017.

- `pchom2022` and `pchom2017` proportion of unemployed people in 2022 and 2017.

We are interested on the interaction of these variables with results of elections.

- `pvoixMITTERRAND` proportion of vote for François Mitterrand in 1981 (center left and left).

- `pvoixMARCHAIS` proportion of votes for Georges Marchais in 1981 (communist).

- `pvoixGISCARDDESTAING` proportion of votes for Valéry Giscard D'Estaing in 1981 (center right and right).

- `pvoixCHIRAC` proportion of votes for Jacques Chirac in 1981 (right).

- `pvoixMACRON2017` proportion of votes for Emmanuel Macron in 2017 (right).

- `pvoixMELENCHON2017` proportion of votes for Jean-Luc Mélenchon in 2017 (radical left).

- `pvoixMLEPEN2017` proportion of votes for Marine Le Pen in 2017 (far right).

- `pvoixHAMON2017` proportion of votes for Benoît Hamon (left and ecologists).

- `pvoixPOUTOU2017` proportion of votes for Philippe Poutou in 2017 (trotskyst).

- `pvoixARTHAUD2017` proportion of votes for Nathalie Arthaud in 2017 (trotskyst).

- `pvoixFILLON2017` proportion of votes for François Fillon in 2017 (conservative right).

- `pvoixMACRON2022` proportion of votes for Emmanuel Macron in 2022 (right).

- `pvoixMELENCHON2022` proportion of votes for Jean-Luc Mélenchon in 2022 (radical left).

- `pvoixMLEPEN2022` proportion of votes for Marine Le Pen in 2022 (far right).

- `pvoixZEMMOUR2022` proportion of votes for Eric Zemmour in 2022 (far right).

- `pvoixPOUTOU2022` proportion of votes for Philippe Poutou in 2022 (trotskyst).

- `pvoixARTHAUD2022` proportion of votes for Nathalie Arthaud in 2022 (trotskyst).

- `pvoixPECRESSE2022` proportion of votes for Valérie Pécresse in 2022 (right).

- `pvoixJADOT2022` proportion of votes for Yannick Jadot in 2022 (left and ecologists).