* My name…
* Masci
* Dataset coming from OECD pisa which every 3 years (starting from 2000) coducts a series of surveys for students of 15 years old in order to detect information about the level of education of different countries.
* We are focused on the dataset of 2018 which contains….
* The questions in the survey are not only to test learning capabilities but also about the general background and habits of the students
* Goals (dataset ampio, molte possibilità di analisi)
* Welth:indicatos: esempi numero di macchine etc, immigration:tre livelli
* We started obv exploring our dataset and for simplicity we began with italy. In our dataset there are some missing values (as we can see by this plot) and therefore we have the issue of how to handle that
* Sebbene siano pochi (i maggiori sono al massimo 20%) ma magari su righe diverse quindi not negligible. We are currently deciding how to handle this issue. Se ridurre il numero di features (comunque ne abbiamo molte, quindi potremmo semplicemente fare una variable selection migliore), data filling (anche se con categoriche non sappiamo se ha molto senso), or obviusly decide to discard the rows w/ missing values
* Pca