Mowy's talk

Mowy

May 2017

1 Our goals

Before running through the data-set itself, I recall what is our main purpose.
Quoting our abstract we want to individuate the best countries where to invest; so first of all, what does best mean for as?
Best country means the one for which we will forecast the largest development. But there are a lot of theoretical indicators of growth, we will have to match our idea of development with the indicators that the dataset will allow us to analyze.
Finally we will leave as the last step witch investment we are going to suggest, taking into consideration that could be financial or entrepreneurial or both.

2 Raw dataset

As you can see, our dataset give us a lot of information; and we can link our tables with the framed variables.
What really surprised us of this dataset is not only the huge number of indicators (more then 1300), but also the possibility to have a really detailed information about what the indicators represent and how the measurement is done (unit of measure, Periodicity...)

3 Examples of indicators

Here we can see how it is concretely done the table with the values and on what, instead, we want to perform our analysis:
we wish to create a 3D matrix that has on one dimension the years, on the second the countries and on the third the indicators.

4 3D matrix problems

But this desire gets stuck for some intristic problem of our data.
First, the numbers of indicators per year are different.
Second, there are countries with indicators not theoretically comparable: let's

think about agriculture indicators on countries as San Marino, Andorra, etc; some with too few data, some in which data are a lot but concentrate in a few years, or in general not well distributed.

Third, the indicators cover really a wide range of topics, some not useful for our analysis, some with too incomplete.

5 Graphs

In the first graph we can see the inhomogeneous number of indicators per year (counting them multiple times if present for multiple countries).

In the second we show the problems with the countries that we ran through before: USA are a good country to be analysed, Afghanistan present a distribution of indicators too inhomogeneous along the years, finally San Marino has too few indicators overall.

In the end, from this histogram we can observe the wide range of indicators; each category in reality contains a large number of different topics that group our indicators. We can also observe the predominance of the economic indicators: reason for which this dataset gives us the possibility to reach our goals. Finally, taking a close look at our variables we can notice that some indicators are really less significant than other for our purpose.