Fertility rates – worldwide

Mariana

Abstract

Dataset

World Development Indicators

Question

How have fertility rates evolved in the past decades and how can countries be organized based on fertility rates?

Methods

Multilinear regression
Clustering

Findings

1. The data collection methodology has probably changed overtime.
2. Fertility rates have been decreasing globally (between 1972 and 2012).
3. The indicators I choose to relate with fertility rates explain about XXX of the variance
4. The fertility rate in time seems to be bi-modal, applying clustering shows that there are multiple modes and those seem somewhat related to my ad-hoc understanding of fertility changes globally.

Motivation

I am exploring population growth, from the perspective of fertility rate.

Part 1, how does the proportion of rural population, GDP and energy use can be used to predict fertility rate. There are known relationships between all these indicators. Here, I am trying to understand what the most valuable contributions are within this subset of indicators.

Part 2, the data showed a bimodal distribution of fertility rates. Hence the second step involved clustering the evolution of fertility rates for the different countries. Though I didn’t expect it to cluster into two well isolated clusters, analyzing the clusters can help understand the effects of culture and policies.

Though it is not very easy to design policy to mostly positively influence these indicators, it is relevant to understand what can contribute to reducing overall fertility rates and hence population growth.

Dataset

- World Development Indicators Dataset

  “The World Development Indicators (WDI) is the primary World Bank collection of development indicators, compiled from officially-recognized international sources. It presents the most current and accurate global development data available, and includes national, regional and global estimates.” – From the World Bank website.

- Relevant indicators for this exercise:
  - Fertility rate, total (births per woman)
  - GDP per capita (current US$)
  - Rural population (% of total population)
  - Energy use (kg of oil equivalent per capita)

Data Preparation and Cleaning

The data was mostly organized.

I wanted to use information relative to countries only, I removed data relative to regions.

As I wanted to understand trends in time, I wanted to include as much data in time as possible. From looking at the data it became apparent that something had changed in the data collection, and I decided to focus on a subset of years.

Finally only a subset of countries had data for the selected years and indicators.

It was not a lot of data, but hopefully enough to illustrate the points I explored.

Research Question

Part 1. What is from the subset of indexes (GDP per capita (current US$), Rural population (% of total population) and Energy use (kg of oil equivalent per capita)) the best predictor for fertility rate? How much variance could such an approach explain?

Part 2. Given that the fertility rates across countries are not uniformly distributed, how many clusters would be appropriate to break the data into? In this case, looking at evolution of fertility rates in time.

Methods

Correlation: Explore relationships between GDP, Rural Population, Fertility Rate, Electricity Consumption, Energy Use, and Climate.

Data Regularization: Prepare data for following steps.

Multivariate Regression: Predict fertility rate using the selected indicators.

Dynamic Time Warping (DTW): Explore time evolution using DTW.

Cluster Analysis: Test different cluster sizes and apply Time Series KMeans clustering.

Findings (I)

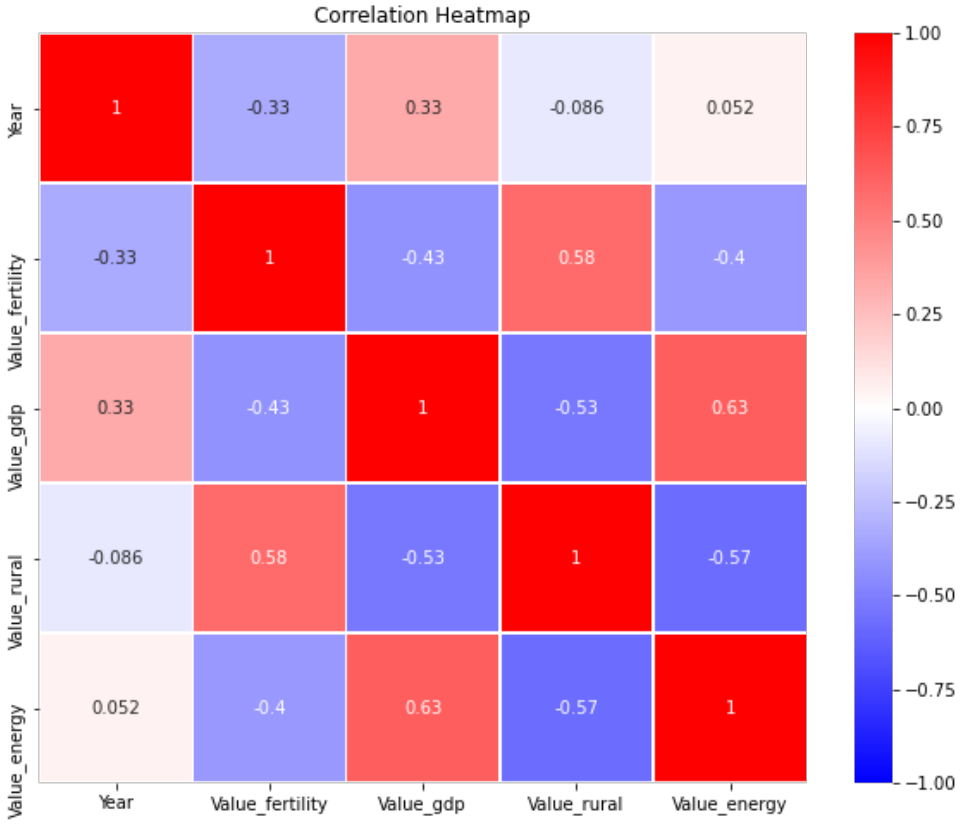
Observable correlations:

Fertility rate: decrease of fertility in time, decrease of fertility with increase in GDPs, increased in fertility for more rural populations, and decrease in fertility with increase is energy use.

GDP per capita: increase in GDP with increased energy use, increase of GDP in time (note: not inflation adjusted), decrease of GPD with increase of rural population

Rural: decrease in rural population with increase in energy use, and with time

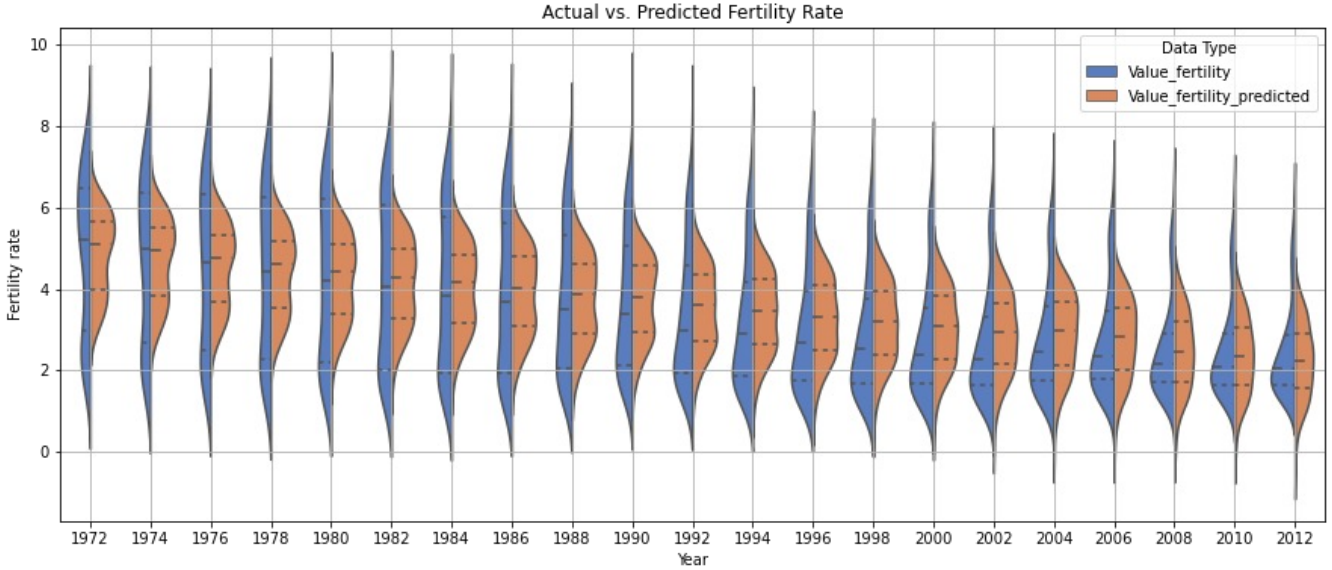
Energy: slight increase in energy use with time



Findings (II)

The first interesting observation is that in 1972 the world (in terms of fertility rates) seem to have two gears, a cluster around 2.5 and another around 6 and in 2012 there is mostly a single group around 2.

Around 40% of the variability in the data can be explained by this model (R-squared ~0.4)



The proportion of rural population is the strongest predictor of fertility rate (not GDP)

Predictor	Coefficient
GDP	-0.045
Rural Pop%	0.847
Energy Use	-0.163
Time	-0.590

Findings (III)

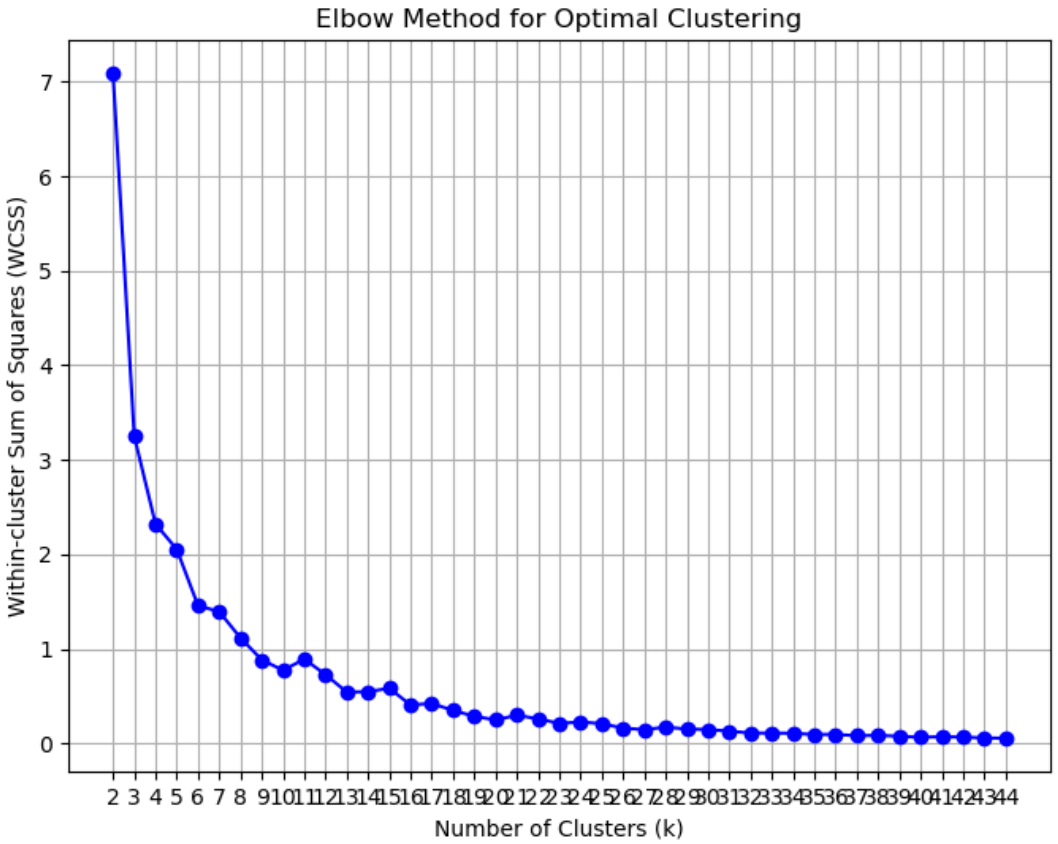
Though we tend to consider GDP a strong predictor of fertility rate, there are other factors, probably more related to livelihood, cultural habits, education, health care access that influence fertility rates.

In this toy exercise, I limited the question to rural populations (%) for the sake of having enough years/countries data.

Next, after noting that the fertility data seems to evolve from a bimodal to a unimodal distribution, I try to cluster the countries based on how fertility rates have evolved during this time (1972-2012).

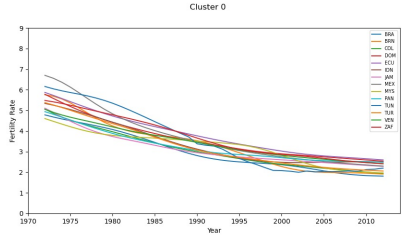
Findings (IV)

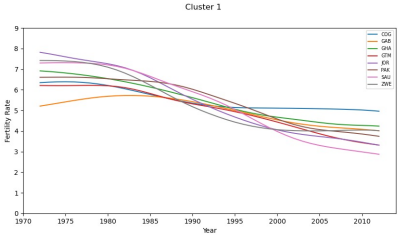
Using an elbow method for optimal clustering, given the data either the dynamic time warping or k-means clustering didn't result in a clear cluster cut off I decided on using 11.

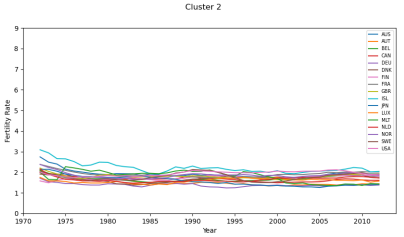


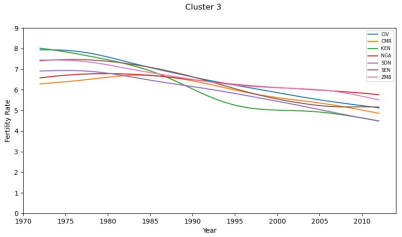
Findings (V)

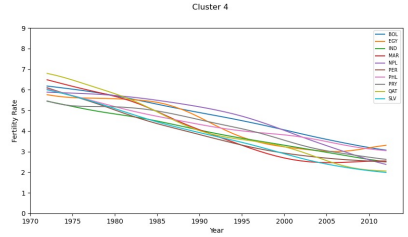
Apologies for the dense visuals. Overview of all the clusters.

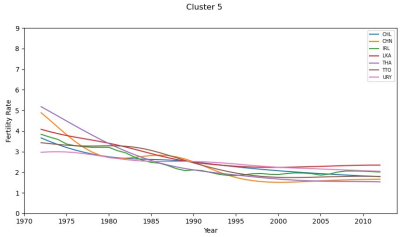


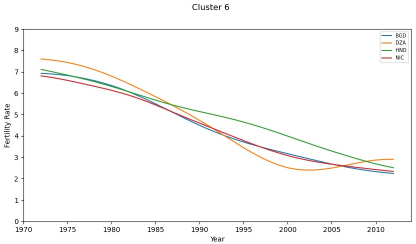


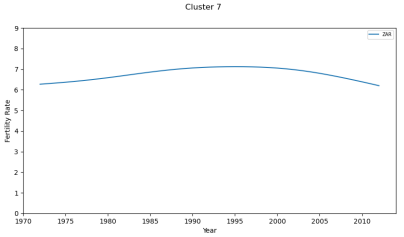


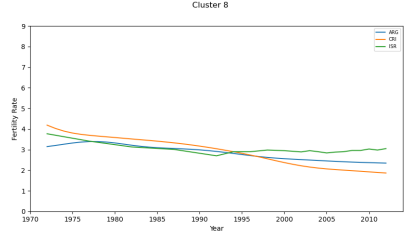


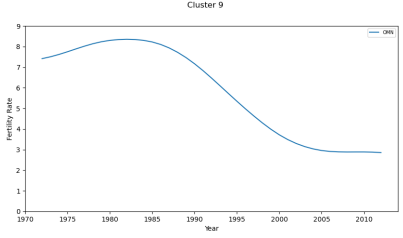


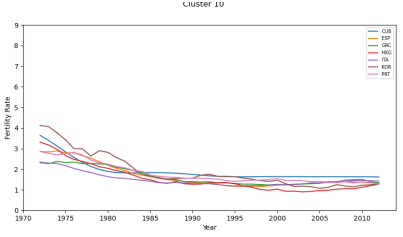






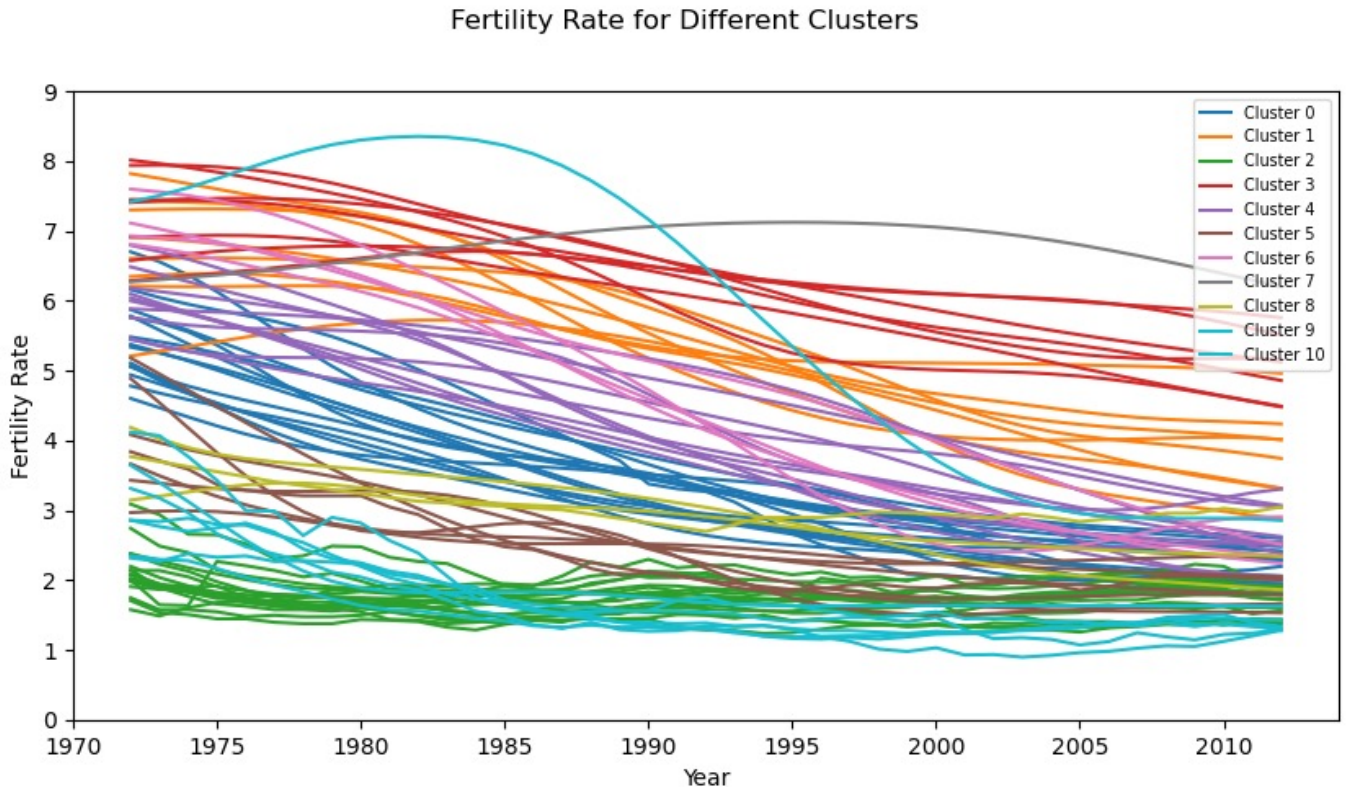






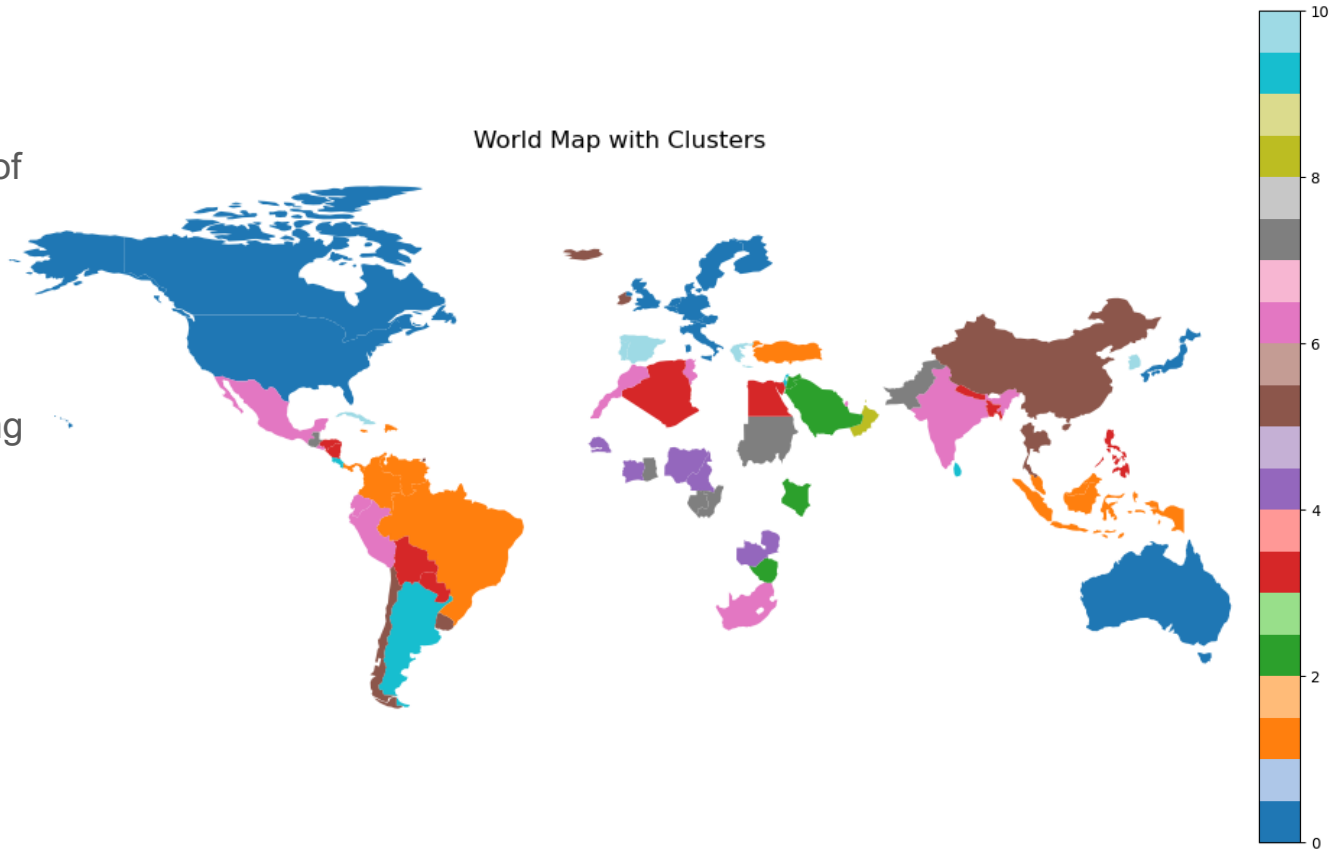
Findings (VI)

For the most part the different clusters tend to capture somewhat different modes of how the fertility rates have been trending. Mostly differentiating the rate of change in fertility rates across different countries.



Findings (IV)

The map, though with a lot of countries missing, is still shows some geographical relationships that seem to make sense. For example, “Western countries” including parts of northern Europe, North America, Australia belong to the same cluster. Sub-Saharan countries also organize in their own clusters.



(Apologies, I couldn’t find a way to make the cluster legend better.)

Limitations

To start with I used only a subset of the data, both in time and included countries. This was somewhat out to my control; it was related to the dataset I used.

The multilinear regression, assumes a linear relationship between the considered indicators, and that does not have to the correct. Have a better understanding of the indicators and how the data was collected could have helped in using more complex relationships and a better explanation of the data.

Regarding the clustering, I tried a few things that didn’t generate stable clusters. My solution is stable, but probably not very generalizable. I tied it to ‘culture’ or ‘policies’ that would allow to ‘validate’ the clustering results.

Conclusions

Fertility rate trends: from 1972 to 2012 fertility rates have been decreasing. The earlier data showed more of a bimodal distribution that with time has been diluting and converging to a unimodal distribution. This might be related to a more global transition from rural and agricultural societies to more industrialized and service-oriented societies where families decide to have less children. This is supported by the fact that rural population is a better predictor of fertility rates and GDP.

When trying to understand how fertility rates compare across different countries one notices that there are adjacent countries that fall into similar clusters, or countries that have similar ‘societies’ (Western societies).

Looking that this rich data and a few indicators, for a few years in a few countries one can already notice patterns that seem to differentiate so much of our global landscape. Moving forward it would be relevant to tie this work to policy changes and expand it to explore more indicators.

Acknowledgements

No real acknowledgement to do here, other to the web browsing I did to fix issues with my script and learn about different packages.

References

I don’t have references to cite.