Université Jean Monnet

Course: Data Mining

Analysis of Illiteracy of different states of India.

Project Repot

Submitted by:
Prem Prakash

Date: March 24, 2017

Objective:

Analyze the literacy of Indian states across the spectrum of rural, urban, male, female of different age-groups. And explore how it is related with different metric characteristics of state like GDP, state expenditure on health or education, woman parity in higher education, and unemployment rate. As all these factors directly or indirectly impact the literacy of the state and sometimes are in feedback loop.

Problem Understanding:

India is a rich country but with a lot of poor people. Approximately 21% of people live under poverty, that's a lot people around 200 million. What is even worse is that despite having high GDP poverty at some places/remote areas more horrid then Africa. A lot people have lived in complete neglect either because of lack of education or government reachability to those locations. People are so poor that at some time parent has to pull their children from school and put them to work which create a feedback loop of keeping people illiterate and thus having diminished chances of coming out from social pit. Here an exploratory analysis is done to see how is state illiteracy related to GDP, education, health, woman parity in higher education, unemployment rate. Does all these parameters have any impact on literacy or literacy has nothing to do with these metrics.

Data Understanding:

To analyses the problem and find some pattern which could throw light on the problem with the help of data, which has been collected and accumulated from different sources available online

Following are the different metric data collected:

- Literacy/Illiteracy at all the ages across the states of India: Government literacy census data: Literacy of people across all ages(7-100 years) for men, woman and all for all of states
- Indian states GDP data: GDP of Indian states in US dollars
- Expenditure of states on public health:
- Public expenditure on education by states:
- Woman parity in higher education: Woman in higher education compared to man.
- Unemployment rates of Indian states:

Data Preparation:

Data has been cleaned and NA type has been removed using R code.

Name of Data Set	Number of Rows	Number of Attributes
Literacy Data	7481	15
GDP of states	36	5
Health Expenditure	36	8
Expenses on education	36	7
Woman higher education parity	36	4
Unemployment rates	36	4

Main preparation of data has been done literacy where data has been accumulated for different age groups for literacy and illiteracy across different states.
Age groups are: {"7-25", "26-50", "51-75", "76-100"}. A new data frame is created for all these age groups using R function which generates the rows for each state across different age groups
All the other data is just for comparative analysis of factors which could have influence on the literacy of state.

Modeling and Evaluation:

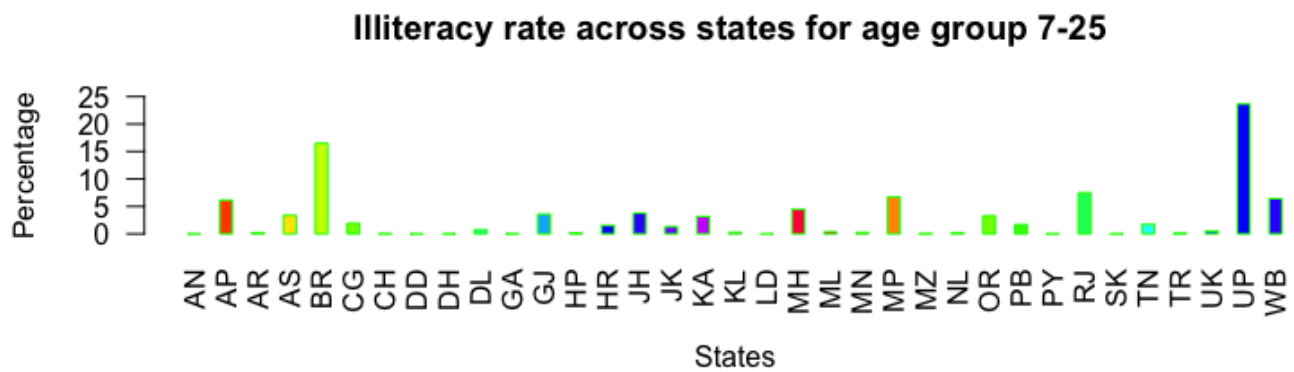Comparison on illiteracy of different age group across all the states.

Illiteracy rate across states for age group 7-25

Fig 1: Illiteracy rate for age group: 7-25

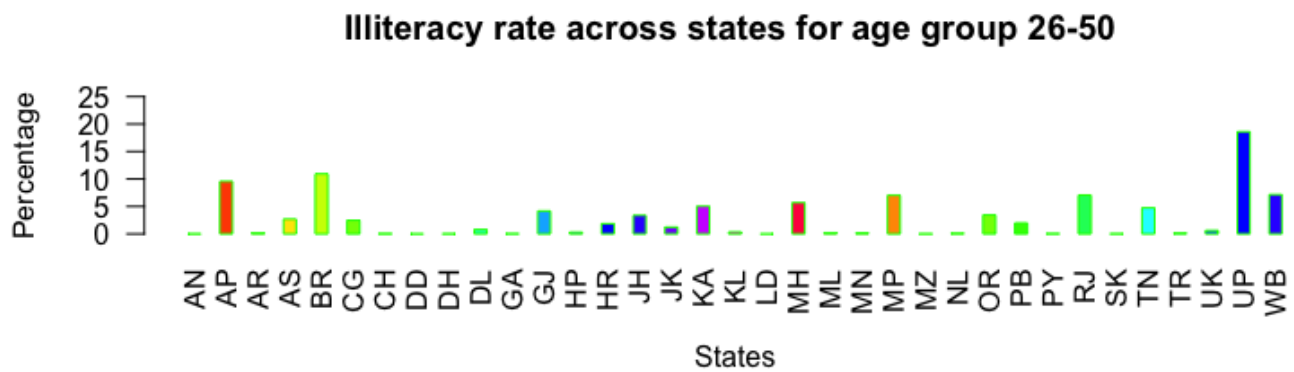**Illiteracy rate across states for age group 26-50**

Fig 2: Illiteracy rate for age group: 26-50

Fig 3: Illiteracy rate for age group: 51-75

Fig 4: Illiteracy rate for age group: 76-100

Figure 1 to 4 tells that illiteracy has been uniform across all the age group for each states except slight fluctuation in some of the states for the younger age. Now let us analyze how is this related to the metrics selected for learning it's impact literacy of the states

1.) How is literacy related to GDP of the states.
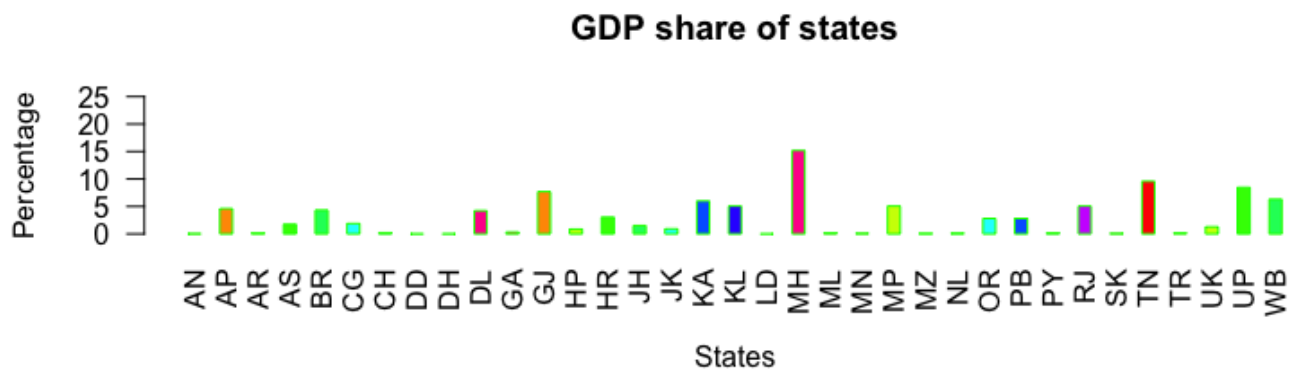

GDP share of states

Fig 5: GDP of the states.

Before even we compare the GDP of state with illiteracy. We could say that more the illiteracy higher the chances that GDP will be low because people will be involved more in unskilled work/services. So, if a state wants to improve it's GDP it must work at the grass root level of educating it's mass.

2.) Does public expenditure on health have any influence on the illiteracy rate. Because to each state should invest fair share of it's revenue for public health as health is important for educating children.

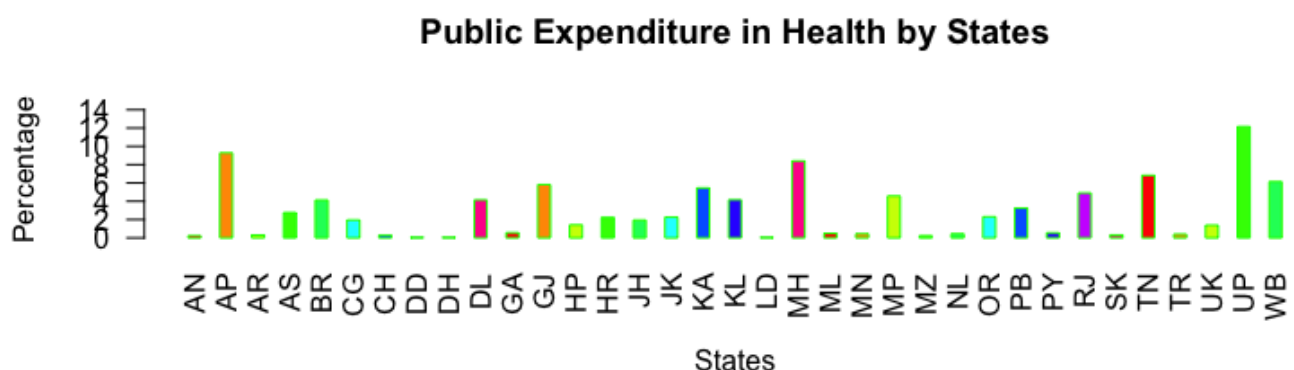**Public Expenditure in Health by States**

Fig 6: Public expenditure on health.

Comparing the public expenditure on health and illiteracy does seem to be inversely proportional. It seems that public health expenditure does not tell much if the health of a state has any direct impact on the illiteracy.

3.) Studying the expenditure on education by states and literacy:
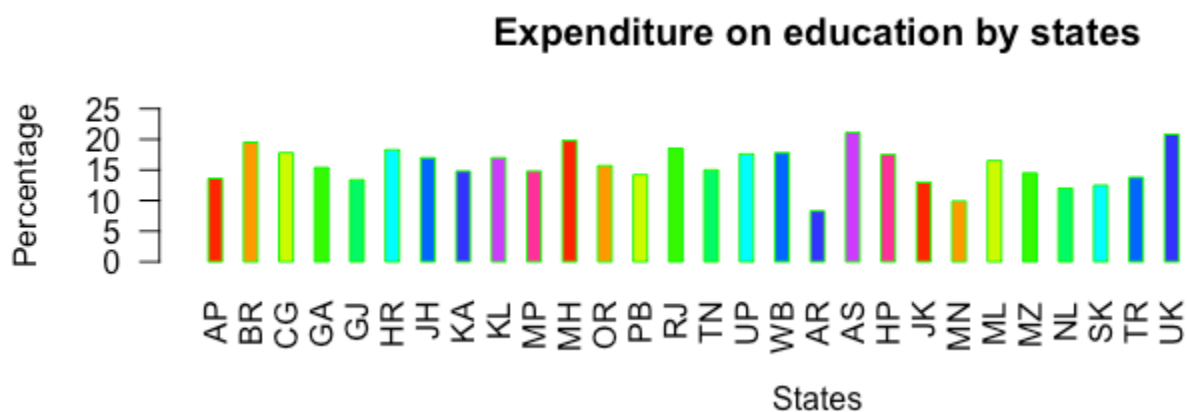

Fig 7: Public expenditure on education.

Each state seems to spend uniformly on education of it's total expenditure, which is good sign of that things are changing irrespective to GDP of the state. Although the above plot is just the recent expenditure so, nothing much can be said about the past that and how it was related to literacy of the state.

4.) Inspecting the impact of woman in higher education on literacy of the state. Because it is believed and observed for a country which a patriarchal if the woman is educated in the family then it has low chances of it's family member to be uneducated irrespective of economic condition.
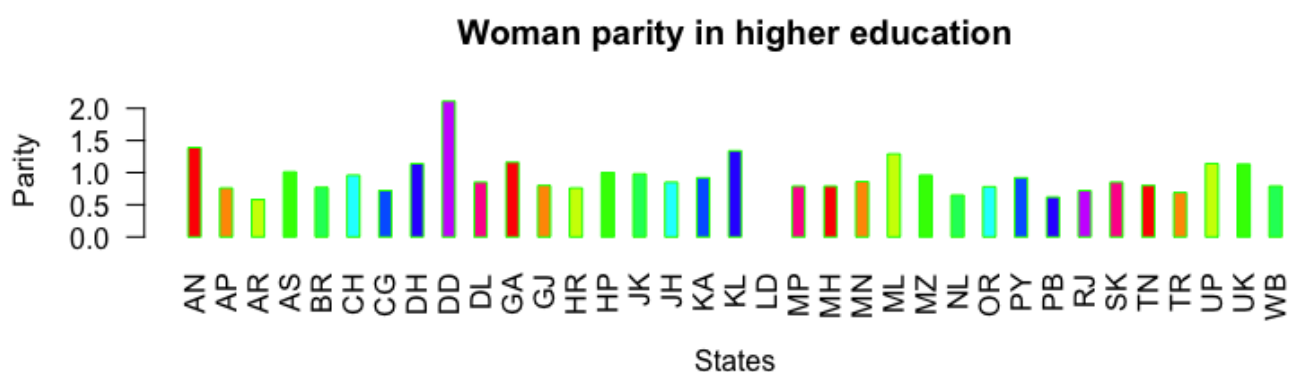

Fig 8: Woman parity in higher education.

Woman parity in higher education. How many women go into higher education compared to man? We can see that it has direct relation to literacy of the state. If the parity is low the chances are higher for the state to have higher illiteracy. So it can be concluded that state should encourage and put more emphasis on woman education which a path to have higher literacy rate.

5.) Examining unemployment rate influence on the literacy rate.

Fig 8: Unemployment rate per 1000 people

As expected higher the unemployment rate higher the illiteracy of the state. Education take a back seat when stomach is empty. No one read empty stomach. And to tackle such problem Indian government take steps to curb unemployment by providing income through various social schemes which requires beneficiary to be available for a fixed number of days to called on work if required, and in return government provide minimum survival emolument, whose impact has been positively seen in reducing the illiteracy of the state.

Conclusion:

Things are changing. Corruption going down, government is trying to change the mentality of people and encouraging people to educate their children and to promote such thing many social scheme have introduced for example mid-day meal and more importantly there is further encouragement on girls education. Because if girl is educated in family then entire family is educated.