THE PROBLEM OF SKEWED SEX RATIO

An investigation using Data

"No country for women" is the phrase that comes to mind, when one has a glance at the state of sex ratio in some parts of the nation. Skewed sex ratio shows the state of deprivation women bear. They are deprived even of their right to life. A right, which is more fundamental than any other fundamental right.

Around 15 % of women are missing in some of the districts of India. While we leave the sociologists and anthropologists to study the implications of such skewed sex ratio , lets pause for a moment to realize the scale of this horrendous social crime being committed in some parts of the nation, with the help of data.

Globally in most of the countries where female foeticide/infanticide is absent females outnumber males. However even if we focus on the countries of South East Asia and South Asia it's 1043+-40. . The states of Haryana and Punjab which are the topic of study here has sex ratio of 879 and 895 respectively. So per 1000 men there are around 160 less women on an average , when compared to other nations. However , as we know aggregation hides more than it shows. There are districts with less than 800 women per thousand men in the state of Haryana. This is even more skewed once we get to smaller sub-divisions within districts.

While face within any sociological problem the usual policy approach is to hope that education and access to better quality of life would certainly help. This is specially true for a country like India. In fact education is the first go to step for any sociological problem. Education however itself is a very vague term.

While these two steps, mentioned above may work in many scenarios, here we are going to investigate their impact on sex ratio and sex ratio of children born in last 5 years ,which is one of the main focus of this project. This is because Child sex ratio is not only a good estimator for sex ratio in the future but also acts as an indicator for how prevalent this sociological evil is at the current moment. It also reflects the effect which interventionists steps such as 'BETI PADHAO BETI BACHAO' have on the sex ratio.

DATA

Whose

The states that I am focussing on this project is the state of Haryana and Punjab. The reasons for choosing these two states are:-

- They have a problem of acute skewed sex ratio(less than 900 women per 1000 males)
- They are neighbouring states so we can expect some overlap in their social set up.

- These states have a good mix of rural and urban districts providing a more representative data.
- These states are relatively well-off in economic terms. So they provide an opportunity to explore the relationship between gender justice and economic prosperity.

FROM WHERE

Coming to data for this project I have gathered the data from NFHS-4(Link). The reason for selecting this data is that it is:-

- Comprehensive
- Reliable
- Representative
- Recent

However, the data has one shortcoming. It is at District level, this may result in we missing the details at block levels and thus may prevent a more targeted action plan.

The data is available in the github repository for this project, whose link has been provided.

For analysis all the data has been scaled from 0-1(min-max) scaling. We have done scaling to ensure that any variable doesn't dominate the other. Also we have used relative scaling as we are more interested about relative difference between districts , so as to bring out the subtle difference between them.

## WHICH DATA IS COLLECTED AND WHY

The data that we have collected are :-

- Sex ratio(females/1000males)
- Child sex ratio(No of female to male births in last 5 years)
- Men's literacy rate(%)
- Women's literacy rate(%)
- Women who have had >10 years of schooling(%)
- Quality of life (combination of access to sanitation Drinking water, fuel as indicator)
- Institutional births(%) to measure access to maternal and child health services.

We also use mini-max scaling to scale these factors from (0,1) in order to bring out the relative difference.
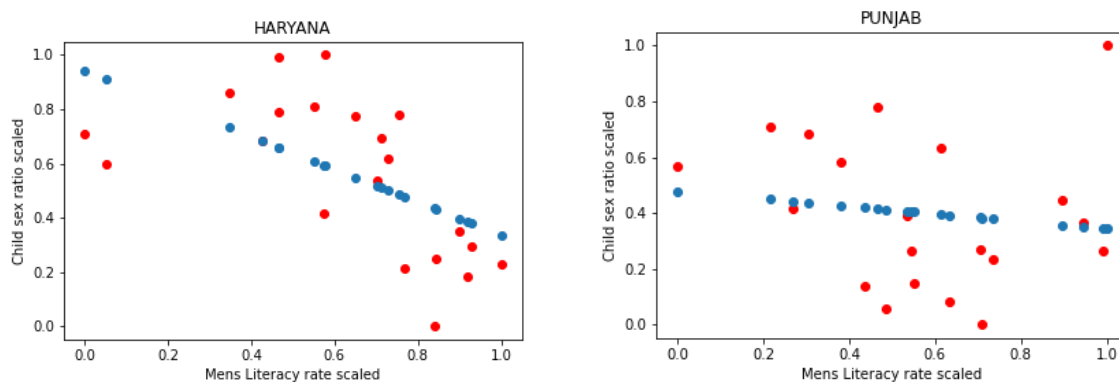
So lets explore the relationship of these variables with sex ratio and child sex ratio.

Men's Literacy Rate
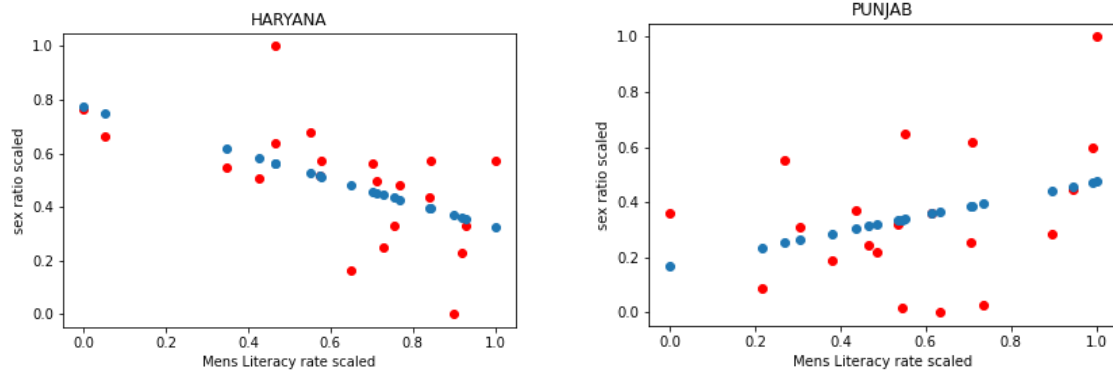
Comment

The relationship between men's literacy rate and sex ratio is inversely proportional in case of Haryana and directly proportional in-case of Punjab.
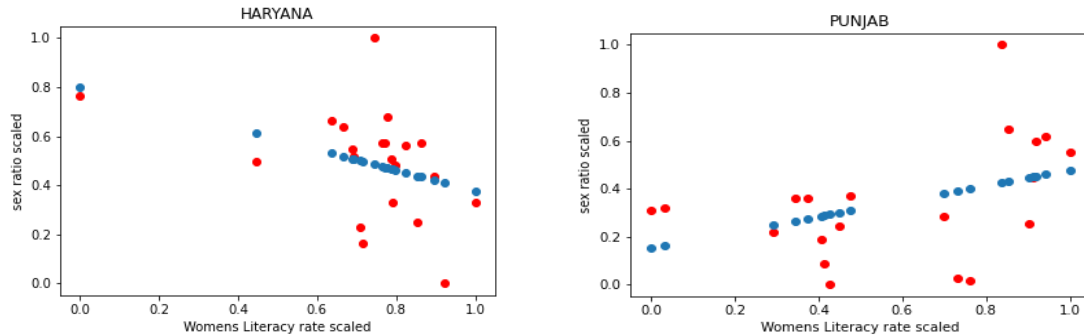

Comment

The relationship between Men's literacy rate and Child sex ratio is inversely proportional in-case of Haryana and flat in-case of Punjab.

Note:- Blue is the best fit-line.

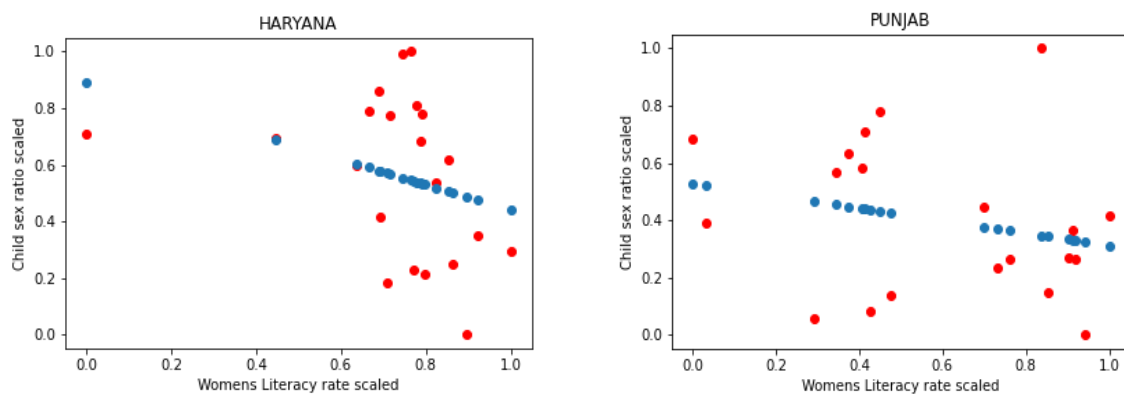Check appendix(1) for Correlation Matrix(

Check appendix(2) for, P value an R^2 value for slope between variables.

Women's Literacy Rate

Comment

The relationship between women's literacy rate and sex ratio also follows similar patterns as men's literacy rate and sex ratio, albeit less strongly in case of Haryana (low R^2 value)
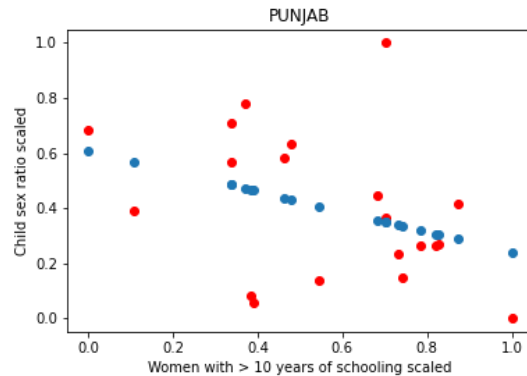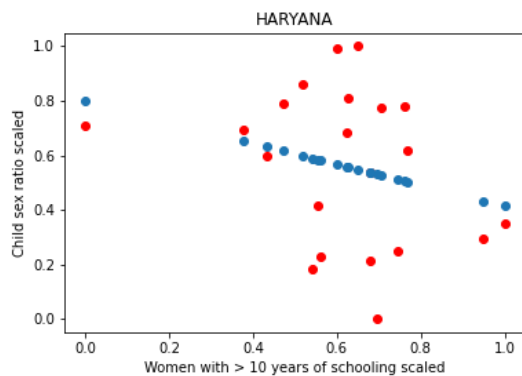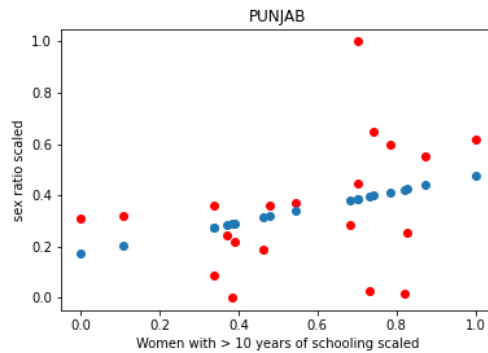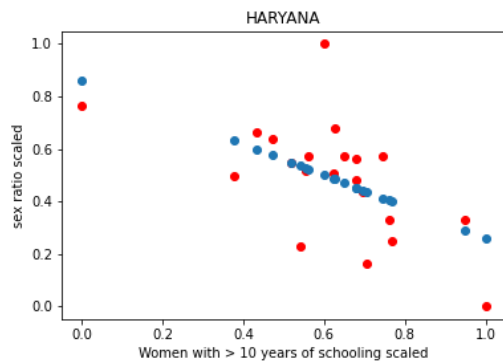

Comment

The relationship between women's literacy rate and Child sex ratio also follows similar patterns as men's literacy rate and sex ratio, albeit less strongly in case of Haryana.

Note:- since men's and women's literacy rate has high positive correlation in-case of Haryana(0.67) this is expected. (Refer correlation matrix).

Women With > 10 years of Schooling

Comment

The P-value for slope of Child Sex ratio vs women with greater than 10 years of schooling is not significant, and thus can be neglected in case of punjab.
Also the R^2 value is low in case of Haryana for child sex ratio.

For more details about significance of slope and R^2 values refer github link for the code.

Note:- Incase of social sciences an R^2 value of > 30 % is significant.

Quality of Life Indicator
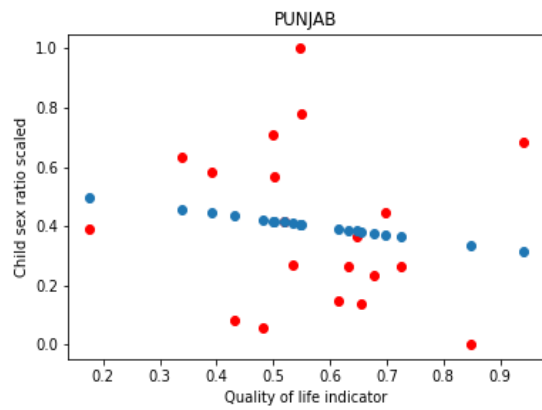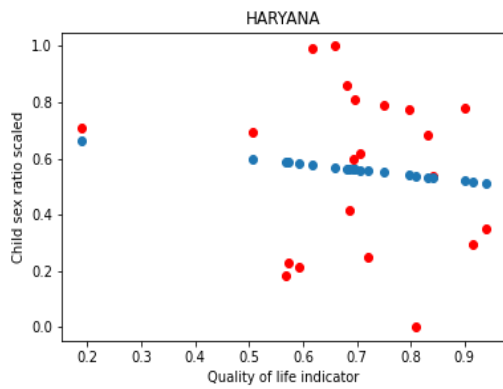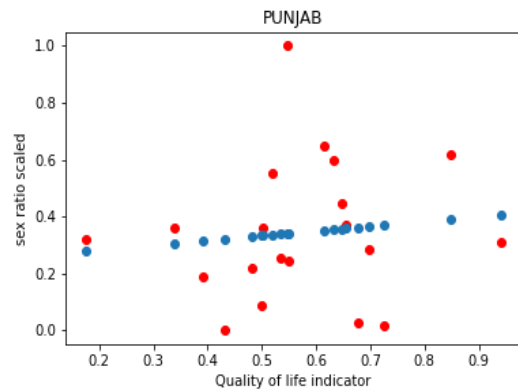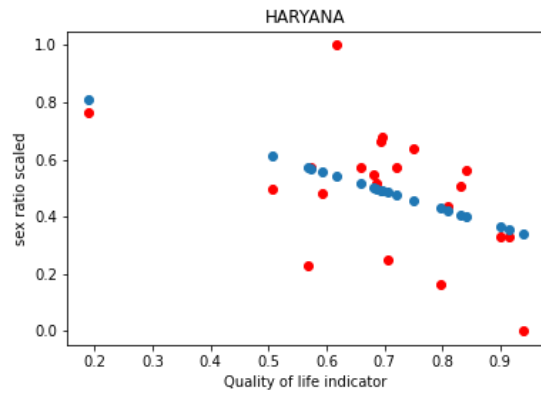
Comment:-

Since institutional births and Quality of life indicators are highly correlated. we restrict our analysis to Quality of life indicator.

ANALYSIS:-

When we evaluate the relationship of sex ratio and child sex ratio for the two states we can come to following conclusions:-

Sex ratio and child sex ratio is correlated negatively with most of the variables that we associate with indicators of development for the state of Haryana the trend is opposite for the state of Punjab. This shows that states that have historically shared geography, polity, culture and a common economic life can have very different directions when it comes to social progress. This also points towards flaws of aggregation based upon our biases.

This also shows that material progression may not necessarily lead to social progress, in fact if existing social order is not challenged, material progress can even lead to strengthening of existing social structure. This trend is visible in the state of Haryana where material progress leads to strengthening of patriarchal social order. This also highlights the absolute necessity of social campaigns such as 'Beti-Padhao, Beti- Bachao'. We need more such well directed campaigns.

This brings us to part two of the analysis which is to find districts which need greater attention.

CLUSTERING(PART 2)

Aggregation based upon our biases is flawed as it can mislead us and blind us towards subtilities in data. Nevertheless, aggregation is absolutely necessary for action, otherwise we would can never come to an actionable plan. Machine learning shows us the way towards aggregation based upon data. One widely used approach is K-means clustering.

K-means clustering is an unsupervised machine learning algorithm that categorizes data into K clusters. Data points are placed in clusters in-order to minimize the intra-cluster variability/distance from mean of cluster. To sum it up it groups data-points which are most similar. Similarity is judged based upon Euclidean distances from mean of cluster rather than based upon individual biases.

The optimum number of clusters are chosen based upon Elbow method. We need to find optimum number of clusters such that we are able to minimize both number of clusters and intra-cluster variability. The elbow point is that optimum point , where our twin objectives are met.

RESULTS:-

- The optimum number of clusters is 5.

- The districts with highest quality of life indicators have lowest sex ratio in case of Haryana. In fact, districts with lowest sex ratio have highest quality of life indicator. These are the districts of Ambala, Faridabad, Gurgaon, Panchkula. We need to keep in mind that are industrial urban centres and migration may further skew the sex ratio. As majority of migrants for employment are males.

- No such relationship was observed in case of Punjab, which was expected as sex ratio was weakly and positively correlated with variables in Punjab than in Haryana. So, clusters are more homogenous in case of state of Punjab.

- The other key finding is that in the state of Haryana, child sex ratio was lowest in cluster with second highest quality of life indicator. These are the districts of Jhajjar, Rewari, Mahendragarh , Yamunanagar. They are classified as rural as per NFHS-4. These districts are upcoming districts. The fact that they have the lowest child sex ratio means that as a districts transition to higher economic growth sex ratio declines. As low child sex ratio today points to low sex ratio in the future. These districts need maximum intervention currently.

- One interesting observation which was observed is that districts which are clustered together by the algorithms also share geographical boundaries(generally). This was a bit both unexpected and expected. Unexpected because we have not included any geographical data. Expected because neighbouring districts generally have similar developmental, social and economic status.
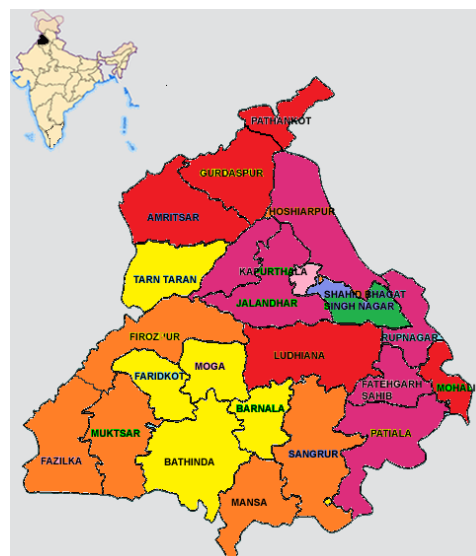
Haryana   **Punjab**

*different colours represents different clusters.

The possibility of using clustering algorithm for targeted policy making need more emphasis. Also with greater number of variables we can have better and more clear segregation at district or sub district level.

APPENDIX

1a.

Corr. Matrix	sex ratio scaled	Child sex ratio scaled	Men's Literacy rate scaled	Women's Literacy rate scaled	Women with > 10 years of schooling scaled	Quality of life indicator	Institutional births
sex ratio scaled	1.000	0.396	-0.575	-0.406	-0.589	-0.497	-0.136
Child sex ratio scaled	0.396	1.000	-0.594	-0.334	-0.289	-0.122	-0.257
Men's Literacy rate scaled	-0.575	-0.594	1.000	0.671	0.701	0.434	0.412
Women's Literacy rate scaled	-0.406	-0.334	0.671	1.000	0.930	0.841	0.881
Women with > 10 years of schooling scaled	-0.589	-0.289	0.701	0.930	1.000	0.855	0.740
Quality of life indicator	-0.497	-0.122	0.434	0.841	0.855	1.000	0.729
Institutional births	-0.136	-0.257	0.412	0.881	0.740	0.729	1.000

Correlation matrix for Haryana.

1b.

Correlation Matrix	sex ratio scaled	Child sex ratio scaled	Men's Literacy rate scaled	Women's Literacy rate scaled	Women with > 10 years of schooling scaled	Quality of life indicator	Institutional births
sex ratio scaled	1.000	0.214	0.347	0.421	0.343	0.124	0.180
Child sex ratio scaled	0.214	1.000	-0.142	-0.258	-0.379	-0.160	0.196
Mens Literacy rate scaled	0.347	-0.142	1.000	0.535	0.505	0.161	0.154
Womens Literacy rate scaled	0.421	-0.258	0.535	1.000	0.959	0.329	0.073

Women with > 10 years of schooling scaled	0.343	-0.379	0.505	0.959	1.000	0.325	0.026
Quality of life indicator	0.124	-0.160	0.161	0.329	0.325	1.000	-0.270
Institutional births	0.180	0.196	0.154	0.073	0.026	-0.270	1.000

Correlation matrix for Punjab.

2.a _Slope,R^2 and P Value of slope_

(Slope,R^2 and P Value of slope)	sex ratio scaled	Child sex ratio scaled
Mens Literacy rate scaled	(-.45,33%,≈0)	(-.6,35%,≈0)
Womens Literacy rate scaled	(-0.42,16%,≈0)	(-0.45,11%,≈0)
Women with > 10 years of schooling scaled	(-0.60,34%,≈0)	(-.38,.08,.01)
Quality of life indicator	(-0.62,24%,≈0)	(-0.2,1.4,≈.14)

Haryana

2.b

(Slope,R^2 and P Value of slope)	sex ratio scaled	Child sex ratio scaled
Mens Literacy rate scaled	(.31,12%,≈0)	(-.13,2%,0.02)
Womens Literacy rate scaled	(.32,18%,≈0)	(-0.21,6%,0.04)
Women with > 10 years of schooling scaled	(.30,11%,≈0)	(-.37, 14%,.08)
Quality of life indicator	(.17,1.5%,55)	(-0.23,2%,≈0)

Punjab

GitHub Link For the Code and data set :-https://github.com/ML-NOOB/Project-on-sex-ratio