



Hi ! Your feed is made up of these topics.



Follow



Follow



Follow



## Motivation

Women empowerment is the key to growth for any society is an established fact.

Through mining data , we get to find the deep rooted trends and issues faced by women and make corresponding policy decisions.

This is also a topic for research for many economic scholars that are employed by government to help in more informed decisions.

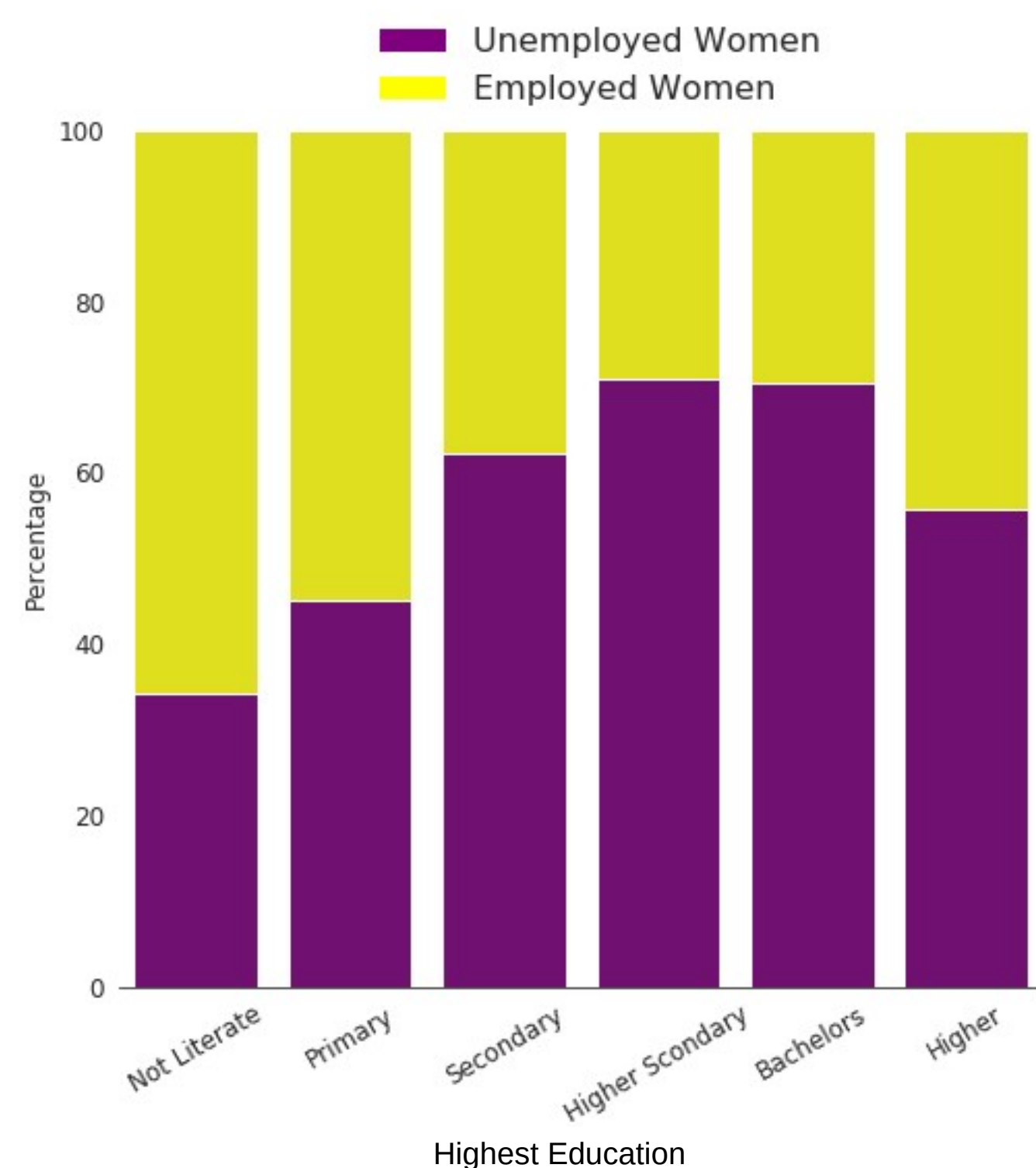
## Dataset

- IHDS Survey done by NCAER, New Delhi in collaboration with University of Maryland.
- Collected by interviewing 39,523 married women aged between 15-45 spread across 1,420 villages and 1,042 urban neighborhoods across India.
- Survey involved asking questions about health, education, fertility, family planning, marriage, and gender relations.
- Responses are categorized and only the labels of the categories are entered in the data files.

## Limitations and Future Work

- Missing values in major decision-making variables that had to be removed after data cleaning.
- Misleading variables e.g. sister's highest education = NaN when respondent has no sister.
- More data exploration can be achieved by trying different scatter plots of other variables.
- Areas like health, finance and politics not covered.

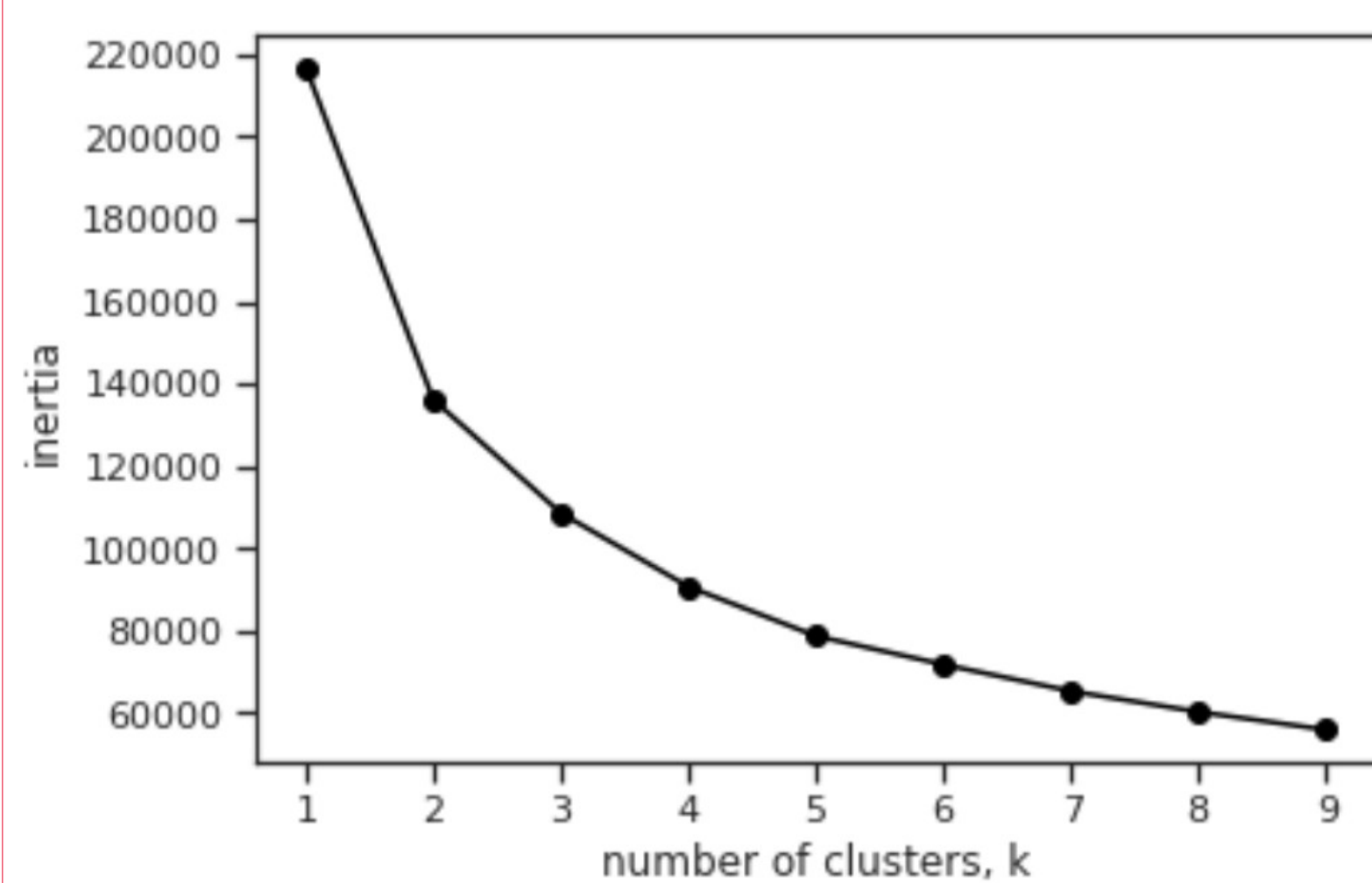
Does Education achieve Employability?



## Basic Data Analysis

- The first task was to map the column labels to their actual Labels (Questions asked) and data labels to the actual answers.
- Went through all the Questions and their explanations as given in the data book to find the interesting questions that can convey more information.
- Removed the questions that had very few responses.
- Filled the no responses with values that made sense for that particular column. Mean for a continuous variable and Mode for a categorical variable.
- Found two interesting categories Education and Employment. The adjacent figure shows the analysis done on these two categories.
- From the bar graph it is evident that the unemployment rate among women with education higher than secondary is greater than that of women without any education.

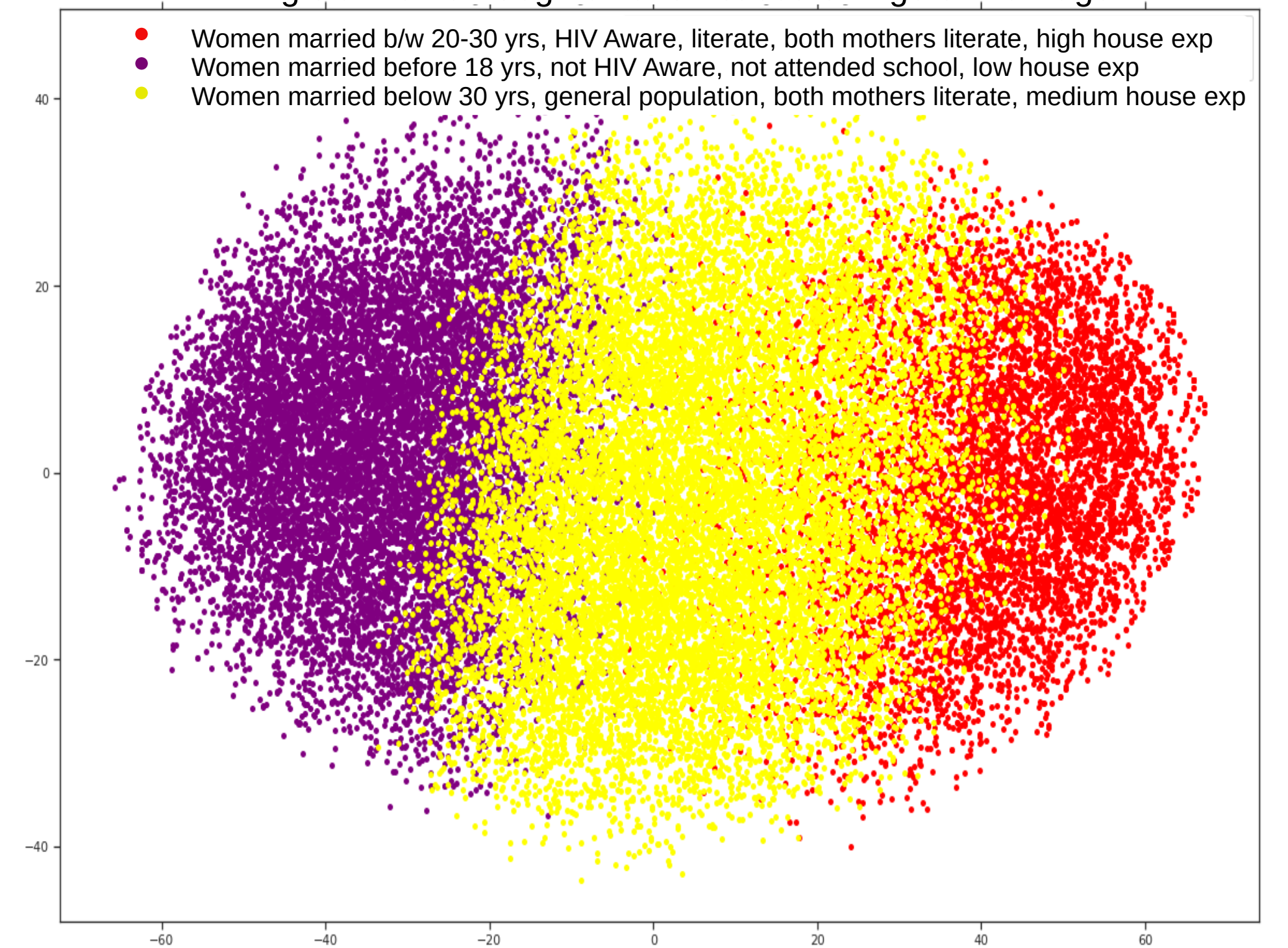
How many clusters to look for?



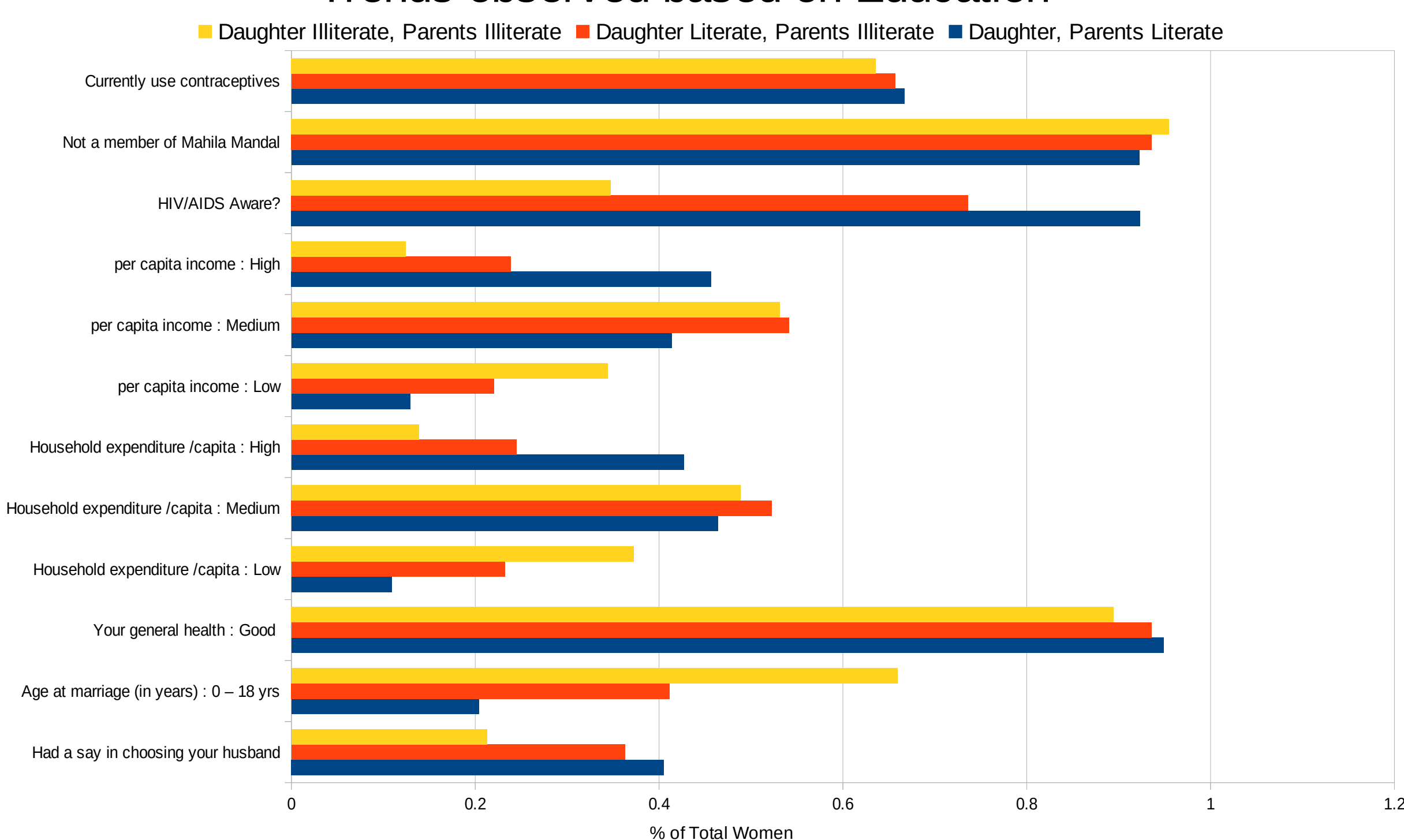
## Clustering

- Manually cherry picked 20 features after data cleaning.
- Applied Affinity Propagation to estimate number of clusters - obtained 777 clusters.
- Switched to Principal Component Analysis (PCA) – obtained top 3 components with highest inertia.
- Applied K-Means clustering with 3 clusters on normalized data and analyzed graph for each cluster for each feature to check for trends.

Eligible Women Segmentation based on Age of Marriage



## Trends observed based on Education



## Association Rule Mining

- Manually cherry picked 18 features after data cleaning.
- We cannot run the Association Rule Mining on all 18 variables at the same time, because the support values for each feature can be in a different range based on how many categories they have.
- To avoid this issue, we divided the data into 3 categories, as mentioned in the graph shown to the left.
- We then tried the different values of support values to obtain meaningful relations between the features.
- It is interesting to note that relation between education, women who get married before 18 and whether or not they got a say in who their husband would be.
- One other interesting observation was that we were expecting that women who are educated would be more inclined to join mahila-mandal and self-help groups but that doesn't seem to be the case, and education has no effect on health of women.