Data Mining Project

On

Comparative Analysis of the Population of Disabled Workers between the age of 15-59 in INDIA according to Population Census 2011

By

Navdeep Kumar

Master 1 MLDM Saint-Étienne, France

1. Objective :

The Objective of this comparative analysis is to analyse the data set of the Disabled population of India according to the Census 2011 and compare the working conditions different types of disabled population between the age of 15-59. In this analysis we will use five different types of disabilities and categorically compare them as per their percentage of working or not working.

Dataset chosen:

I have selected a dataset from the official website of the Indian Government: http://www.censusindia.gov.in/2011census/population_enumeration.html

In India, census is conducted once in ten years. So I have selected the most recently conducted census (Census 2011) for the analysis.

Software Requirement:

For the analysis I have worked on R software which is one of the most efficient and popular software for data analysis.

GoogleVis Package: GoogleVis is an interactive library package for working and drawing maps

and other graphical data. The main advantage of this package is that it interactively plot the data on the map which is very effective to visualize. For the map, it generates an html file with javascript which is then viewed in any browser. For this one must require an internet connection in order to work with Googlevis package.

2. Problem Understanding:

In India, there is a huge population of disabled people. It is very difficult for them to earn the living due to their disability. Even if some of them may get the work, it is only for a temporary period of time. Government of India has launched many programmes for their benefits but a very few get the employment under these schemes as there is less information available about the density of the disabled persons.

My main motivation for this problem is to analyse the data of disabled people and their disabilities and compare the results in order to know in which state, the disability is more affected by unemployment and which gender suffers more as compare to other gender. These results are useful to make assessment about the affected disable persons and Government can make efforts towards a particular disability in which most of the persons are unemployed.

3. Data Understanding:

I have a good dataset of the statewise population of the disabled persons taken from the official website of Government of India. This data comes from the Indian Population Census conducted in 2011 and it is the most recent census of the population of India.

This dataset has around 5000 observations for statewise different categories of the of disabled workers. It also has information about male and female disable workers. All in all, around 18 out of 22 attributes are useful for the analysis. Since actual population is in millions, so I have scaled the data in the form of percentage (per 100).

4. Data preparation:

Data preparation is the first step for analysis. In this phase data is going to be cleaned so that our analysis gives good results.

In my problem we are concerned with only those disable persons who are having age between 15-59 because only these persons can be employed. Further I have selected only five kind of disabilities which contributes the most towards the total disables population. These disabilities are:

- In Seeing: Those that can not see.
- In Hearing: Those that can not hear.
- In Speech: Those that can not speak.
- In Movement: Those who have difficulties in moving.
- Mentally Retard: Those who have mental disorders.

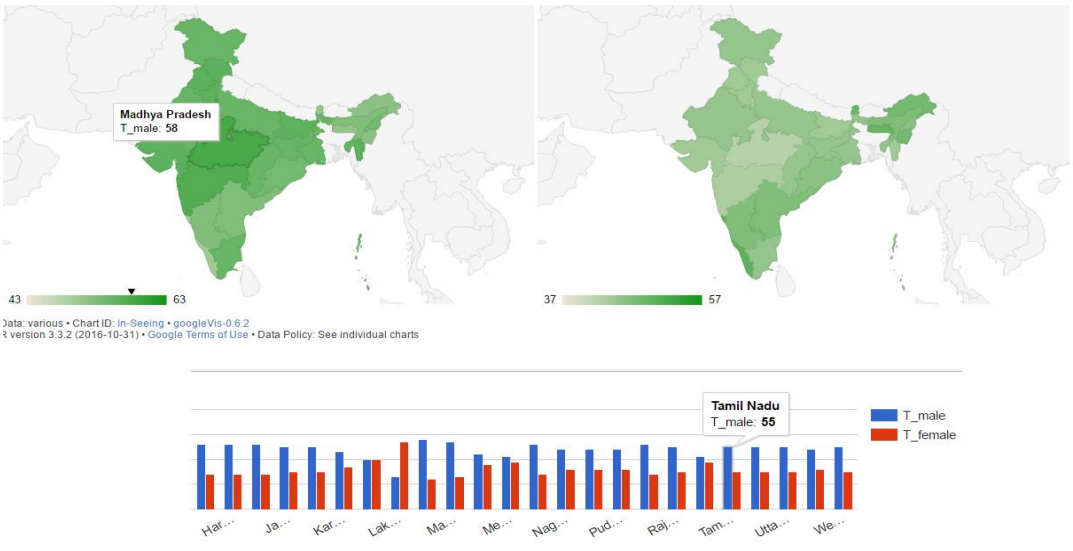
5. Data Modelling:

This is the main phase in the exploratory analysis. For this purpose, I have worked in 'R' software. In this phase I draw the state wise data map of India using GoogleVis package. I also draw the column charts of various states for comparing male and female disabled population for each of five selected disabilities. Following plots show about how data is distributed in various states.

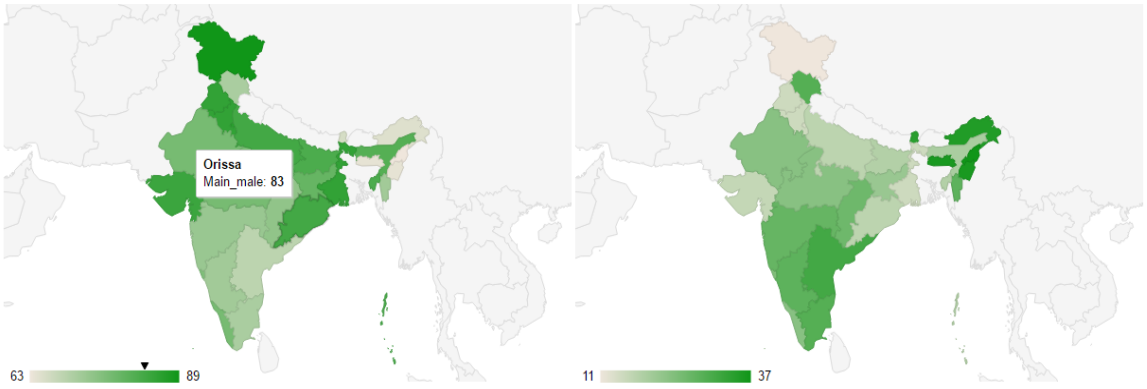
Here the intensity of the colour shows the density of data i.e. the more the colour intensity the more the percentage of corresponding population of the state.

One of the good visualization feature of GoogleVis is that we can interactively visualize our plots in the browsers. For example if we place mouse over a particular area, it dynamically shows the data density of that area.

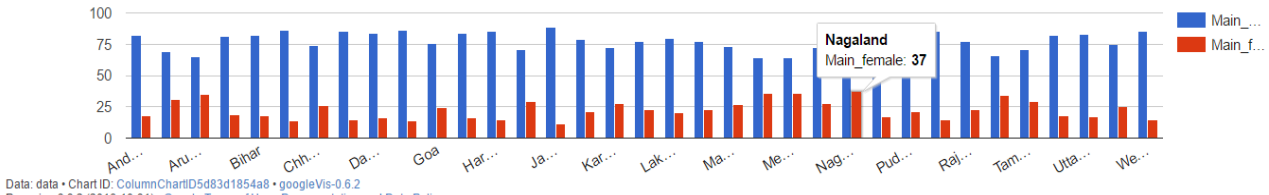
- **Plot.1: Statewise Total population of In-Seeing disability between the age 15-59 (Male(Left Side Map), Females(Right Side Map))**



- **Plot.2: Statewise population of Main Workers of In-Seering disability between the age 15-59 (Male(Left Side Map), Females(Right Side Map))**

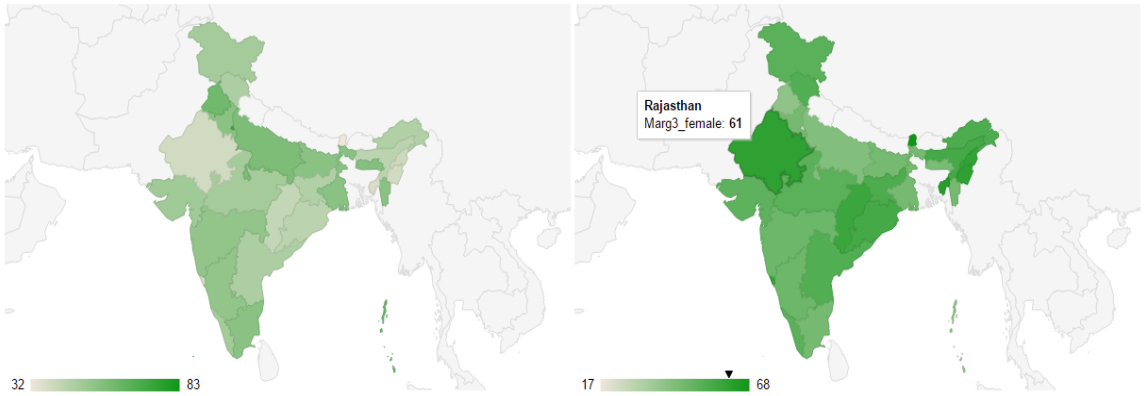


Data: various • Chart ID: In-Seeing • googleVis-0.6.2
R version 3.3.2 (2016-10-31) • Google Terms of Use • Data Policy: See individual charts

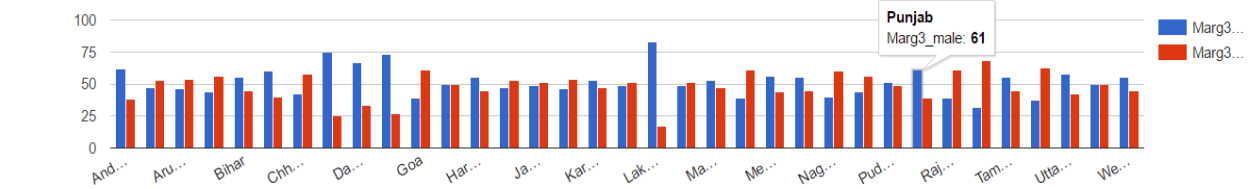


Data: data • Chart ID: ColumnChartID5d83d1854a8 • googleVis-0.6.2
R version 3.3.2 (2016-10-31) • Google Terms of Use • Documentation and Data Policy

- **Plot.3: Statewise population of Marginal Workers of In-Seeing disability between the age 15-59 (Male(Left Side Map), Females(Right Side Map))**

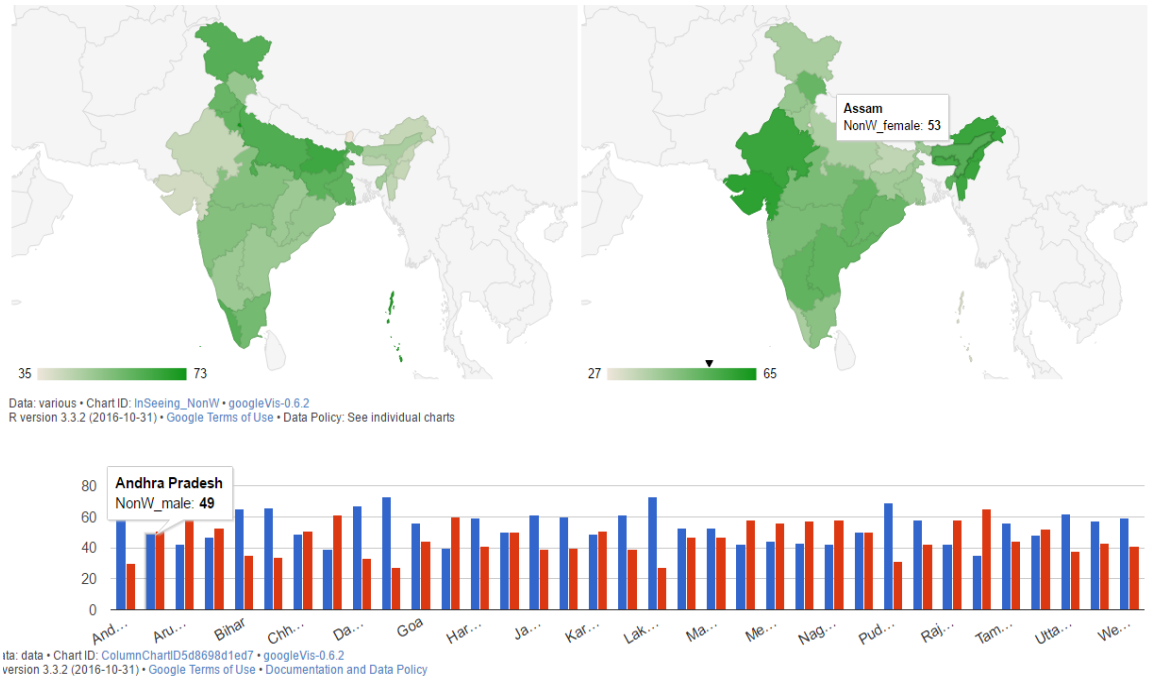


Data: various • Chart ID: InSeeing_Marginal • googleVis-0.6.2
R version 3.3.2 (2016-10-31) • Google Terms of Use • Data Policy: See individual charts



ata: data • Chart ID: ColumnChartID5d84ad012a2 • googleVis-0.6.2
version 3.3.2 (2016-10-31) • Google Terms of Use • Documentation and Data Policy

- **Plot.4: Statewise population of Non Workers of In-Seeing disability between the age 15-59 (Male(Left Side Map), Females(Right Side Map)).**



Similarly we have 4 plots of each **of 'In-Hearing Disability', 'In-Speech Disability', In-Movement Disability**' and **'Mentally-Retard Disability'** total of 20 plots.

(Due to page constraints I am not showing other plots)

6. Comparative Analysis:

I have done an exploratory analysis of the Categorically disabled population of India.

In this analysis we can see some interesting colour intensity patterns in the map.

In the northern areas like Punjab, Haryana, Uttar Pradesh, majority of the main workers are males whereas in the southern part (Kerala, Andhra, Tamilnadu) of India, female disabled populations is the dominant main working class. In the eastern ,westerns and middle areas, disable populations is moderately distributes between the two genders. This might be due to the fact that male to female sex ratio is more in the northern parts of India whereas in the south of India population of females are more. Overall we can infer that those areas where the male to female sex ratios are more, male gender is the dominant working class and where this ratio is less females comprises the main working class.

For disabled non workers, north-eastern part of the India has more percentage of male non workers as compared to female non workers. Western part of India (Gujarat, Rajsthan) has more number of disabled female non workers. In the middle and south eastern part of India(Madhya Pradesh, Orissa, Chattisgarh), population of non workers are almost equally distributed between two genders.

7. Conclusion:

From the above analysis we can conclude that, disable population is not evenly distributed across all parts of India. Disable population is getting more employment in male dominant areas whereas in female dominant areas, non workers are more. Further the most dominant category of disability which have more non workers are 'Mental-Retardation' and 'In-Movement' category is getting more employment than other disability classes. In this analysis we have got an idea that employment generations programmes should be implemented where there disable population is dominated by females.