

Project Report



Jharkhand Education

Submitted by:

Shrishti Singh

20070243031

Index

Sr. No.	Title	Page No.
1	Acknowledgment	3
2	Aim	4
3	Introduction	5
4	Detailed Description -Datasets -Analysis	7
5	Conclusion	24
6	Bibliography	25

Acknowledgment:

I have taken efforts for the project completion. It had been three – four months learning R programming under the subject Programming for Spatial Science, and I enjoyed working on the real – life dataset. I would like to express my gratitude towards professor Rajat Chopra for his guidance and supervision for the project.

Aim:

The principal goal of this project is to study the state Jharkhand in terms of educational field by importing the real-life data set, performing cleansing and tidying the datasets, creating some interactive maps based on datasets and performing basic exploratory data analysis to compare and analyze various patterns.

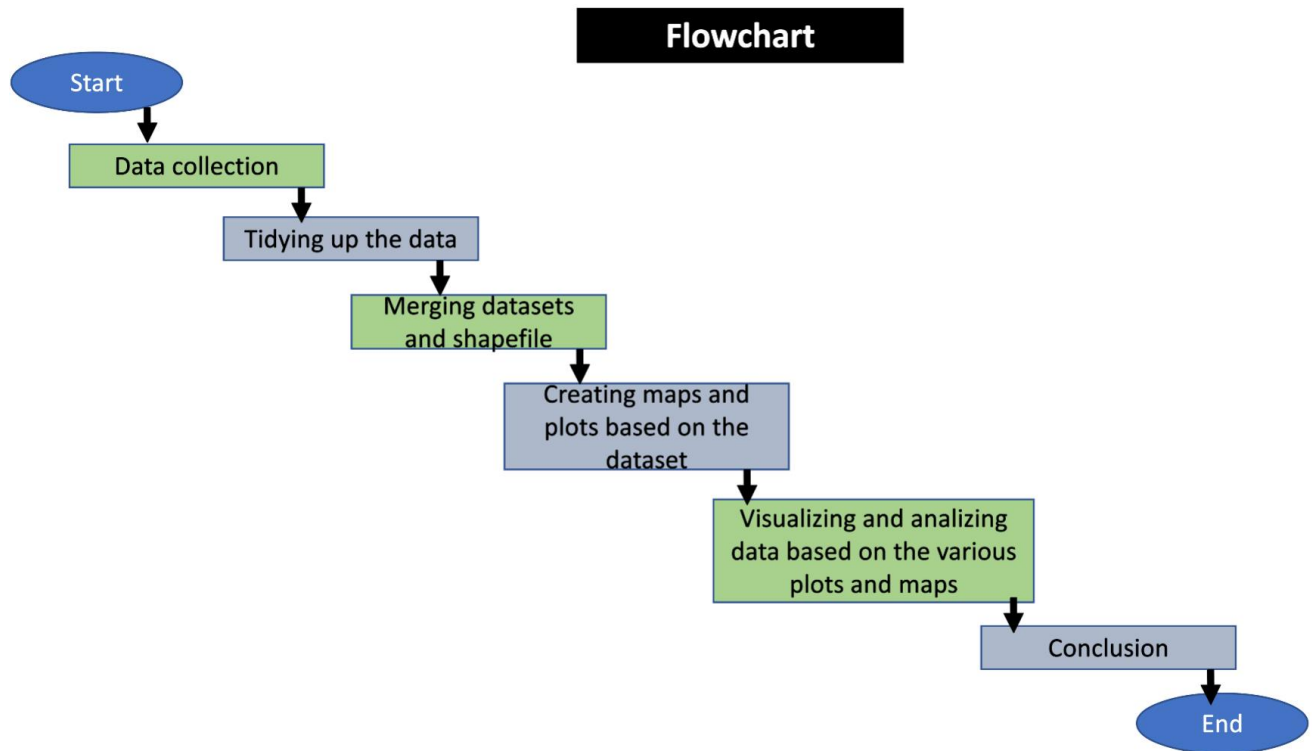
Introduction:

This project was done intending to study various patterns and analysis of the educational sector of the state Jharkhand. And to do so I have downloaded the real datasets from the government site and Arcgis. The data analyzed is of the year 2014. I have combined the available datasets and tried to compare and perform analysis on the available dataset. This project focuses on one of the Indian state Jharkhand. It is located in the northeastern part of the country and its capital is Ranchi.

The role of education is the most important lever for social, economic and political transformation. Accordingly, recognizing the importance of education in national development. And hence expansion of education, significantly improving the quality of education and on ensuring that educational opportunities are available to all segments of the society is of utmost important for the nation's growth and development.

My study on Jharkhand education is done in order to outlook how well is Jharkhand doing in the educational field. The study is done district wise. The state consists of in total 24 districts. I have applied my learning knowledge of R programming to explore and analyze the data and presented it in some innovative ways.

Flowchart Diagram of the project:



Detailed Description:

1.Datasets:

To perform the data analysis on education of Jharkhand I have used two datasets and Jharkhand shape file. To work upon the interactive maps I manually added the longitude and latitude to the data set district wise. And then have merged the obtained dataset with the available shapefile.

Dataset 1-

The dataset 1 named Jharkhand.data.csv consisted of the following attributes mentioned in the below table [Table 1]. This dataset was downloaded from Arcgis website along with the shape file. This dataset consists of the attributes, which can be used to analyze how well developed is the state of Jharkhand in the educational field. It consists district wise collected data which can be used to analyze like how many scheduled cast and scheduled tribe enrollments took place in the primary schools per district, or to look over the functioning of the government schools in each district by analyzing how many government schools provide mid-day meals or how many of them have school management committees, etc.

The original csv file did not contain valid Column names to perform data analysis and hence I had to tidy up the data by changing the column names to valid variable format.

[Table 1: attributes of Jharkhand.data.csv]

object_id	District_code
-----------	---------------

district_name	6_14_yrs_out_of_school
6_14_yrs_in_private_school	std_1_2_who_can_read_letters_word_more
std_1_2_who_can_recognize_no.s_or_more	std_3_5_who_can_do_subtraction_or_more
std_3_5_who_can_do_read_std1_texts_or_more	single_classroom_primary_schl
single_teacher_primary_schl	primary_school_with_girls_toilet
primary_school_with_boys_toilet	primary_school_with_Drinking_water_facility
gov_primary_school_providing_midDay_meal	gov_primary_school_with_school_managment_committee
TransitionRate_primary_to_UpperPrimary	Retention_Rate_primaryLevel
pupil_teacher_ratio_at_primary_level	pupil_teacher_ratio_at_upper_primary_level
sc_enrollments_in_primary_school	sc_girls_enrollments_in_primary_school
st_enrollment_in_primary_school	st_girls_enrollment_in_primary_school
year	globalid
st_areashape	st_lengthshape

Dataset 2-

The other dataset used was downloaded from the government website which consisted of the attributes mentioned in the below table[Table 2]. Originally this dataset did not contain longitude and latitude attributes. I manually added the longitude and latitude of each district in the dataset name Literacy_rank.csv. This dataset was useful to get the insights of literacy rate of total population of Jharkhand, excluding the age group of 0 to 6. It also consists of the literacy rank of each district in Jharkhand.

[Table 2 : attributes of Literacy_rank.csv]

Distict	Rank
Total_literacy_rate_excluding_0_to_6_age_grp	Male_literacy_rate_excluding_0_to_6_age_grp
Female_literacy_rate_excluding_0_to_6_age_grp	literacy_percent
longitude	latitude

Dataset 3-

This dataset was downloaded from the government website “www.embibe.com/indian-states/schools-in-jharkhand/” which consisted of the attributes mentioned in the below table[Table 3]. This dataset was useful to give counts of number of schools in each district of Jharkhand.

[table 3]

District_name	Number_of_schools
---------------	-------------------

Both the datasets contained district wise data so first I added the district_id from “Jharkhand.data” to “Literacy_rank” data where their district name matched. Later these two datasets were merged based on the unique district_id to form a single csv file named “jharkhand.data”. This “jharkhand.data” file was further used to perform merging of the csv and shapefile.

2.Data Analysis:

The data obtained from merging the csv file and the shape file was used to perform various data analysis by creating maps and statistical plots for the knowledge discovery.

a) Comparative analysis study by non-interactive maps:

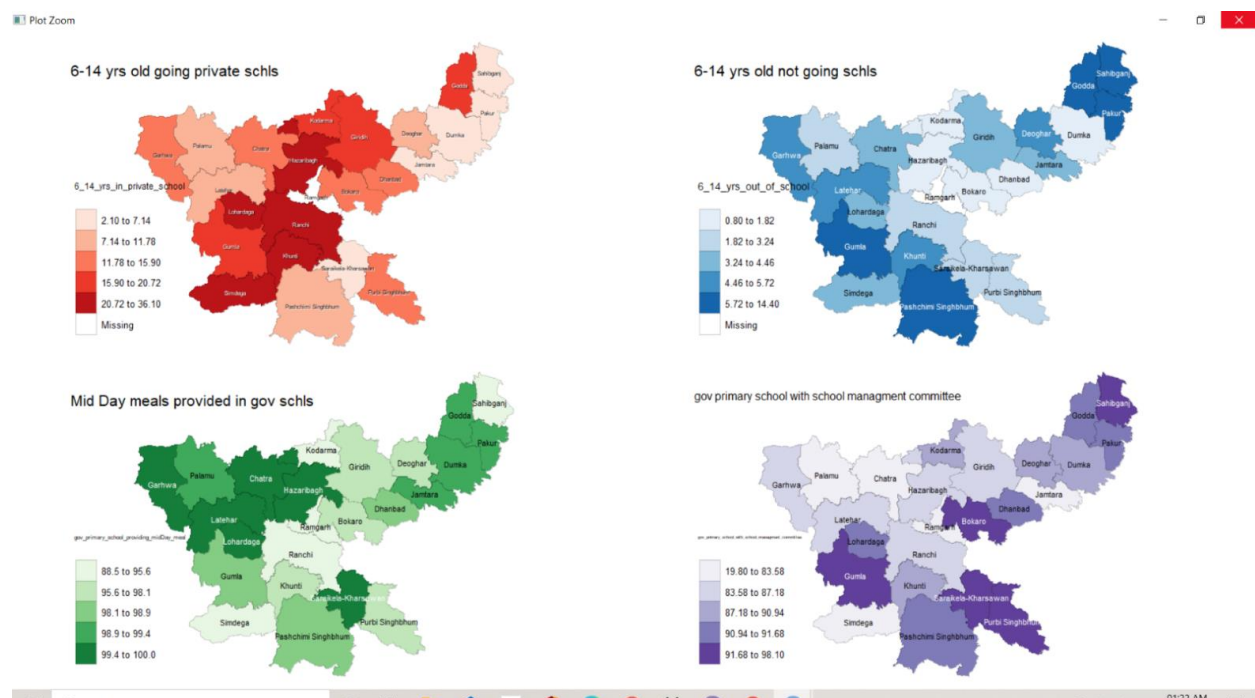
The maps have been created using “tmap” library functions. Where I used the functions like `tm_shape()`, `tm_fill()`, `tm_borders()`, `tm_layout()`, `tm_text()` and `tm_style()` to create the maps and have arranged the related maps in a single plot using the “`tmap_arrange()`”

Map1:

In the following map shown in the [img1], I created the individual maps of:

- 1] 6-14 yrs old going to private schools: -showcases the district wise count of the children between the age 6 to 14 who go to the private schools in percentage.
- 2] 6-14 yrs old not going to schools: - showcases the district wise count of the children between the age 6 to 14 who do not go to the any school in percentage.
- 3] Mid Day meals provided in government schools: - showcases the district wise count of mid day meals provided in percentage.
- 4] Government schools with school management committee: - showcases the district wise count of the government schools having school management committees in percentage.

[img1]



Map2:

In the following map shown in the [img1], I created the individual maps of:

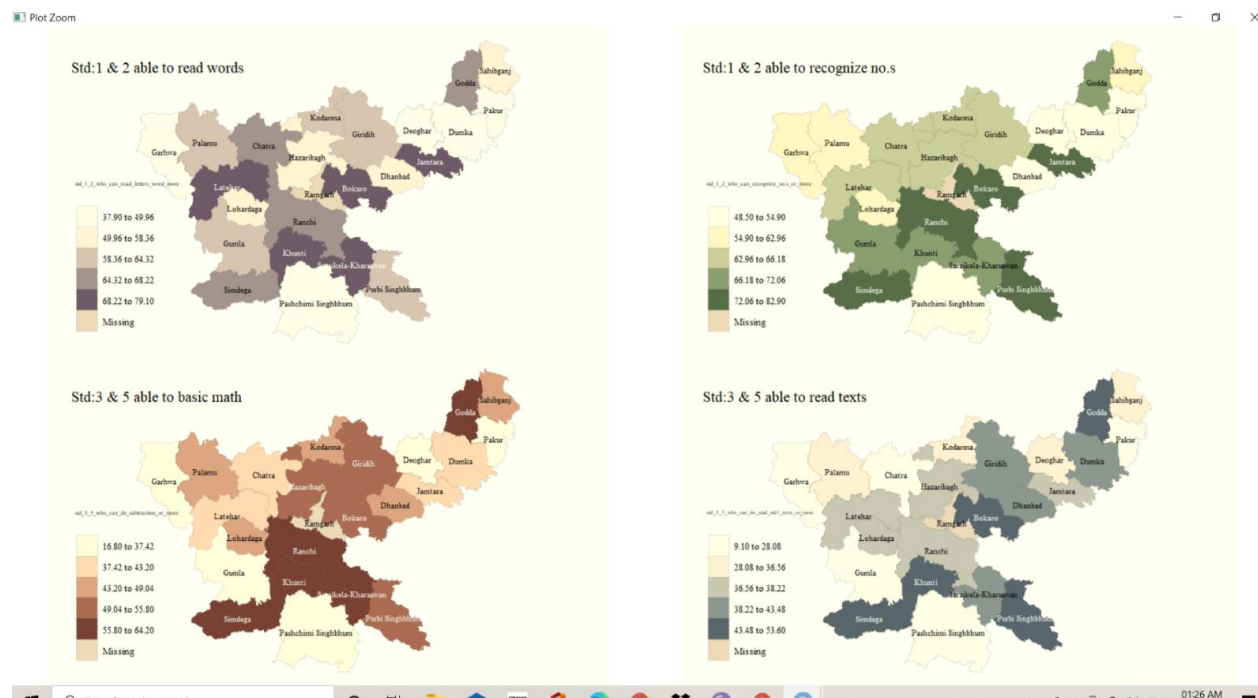
1] students of std 1 and 2 who are able to read words:- showcases the district wise count of the students of class 1 and 2 who are able to read some english and hindi words in percentage.

2] students of std 1 and 2 who are able to recognize numbers:- showcases the district wise count in percentage of the students of class 1 and 2 who are able to recognize numbers and read them.

3] students of std 3 to 5 who are able to do basic math:- showcases the district wise count in percentage of the students of class 3 to 5 who are able to perform some basic maths.

4] students of std 3 to 5 who are able to read texts:- showcases the district wise count in percentage of the students of class 3 to 5 who are able read english and hindi texts.

[img2]

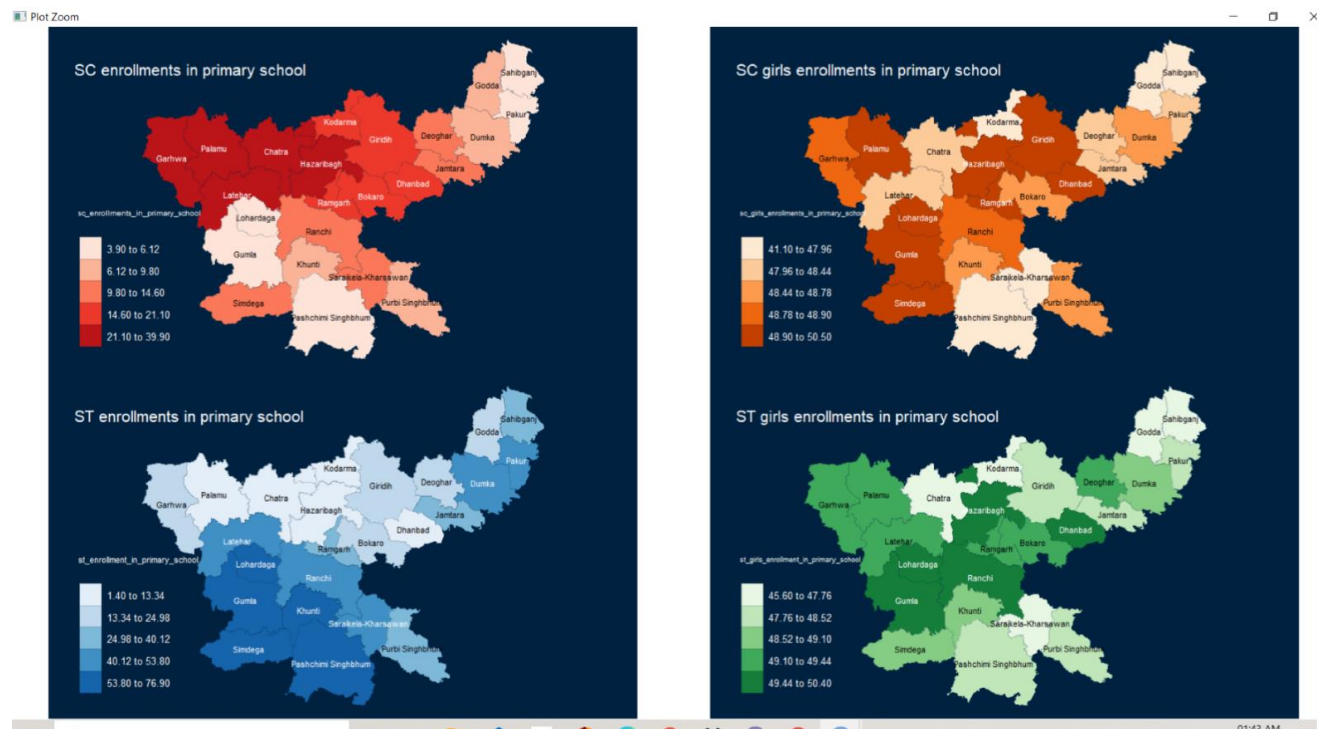


Map3:

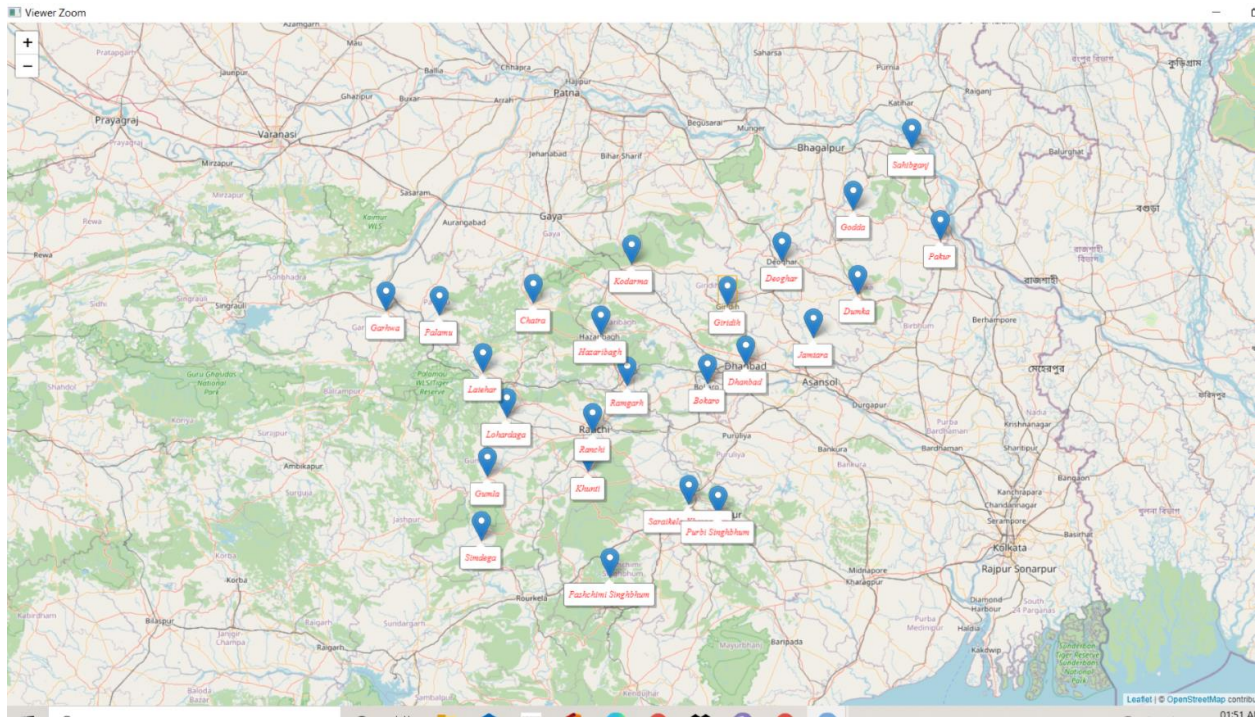
In the following map shown in the [img3], I created the individual maps of:

- 1] SC enrollments in primary schools:- showcases the district wise count of the scheduled cast enrollments in the primary schools in percentage
- 2] SC girls enrollments in primary schools:- showcases the district wise count of the scheduled cast girls enrollments in the primary schools in percentage
- 3] ST enrollments in primary schools:- showcases the district wise count of the scheduled Tribe enrollments in the primary schools in percentage
- 4] SC girls enrollments in primary schools:- showcases the district wise count of the scheduled Tribe girls enrollments in the primary schools in percentage

[img3]



b) Analysis study by interactive maps:

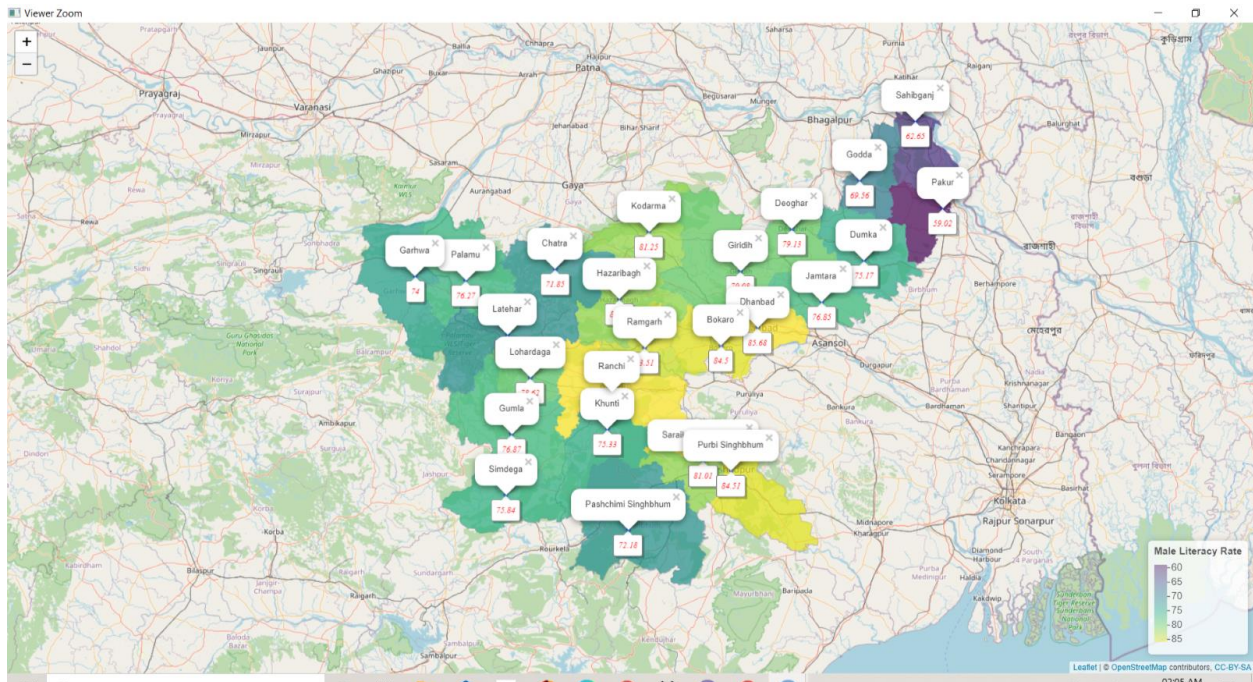


I have created the interactive maps using the functions of leaflet libraries. The functions I used for creating these maps are `addTiles()`, `addPolygons()`, `addCircles()`, and `addLegends()`.

I used the `colorNumeric()` for the maps to make them look good. As depicted in the following above image, I used the openstreet map as the base map and have plotted the data above base map with help of longitude and latitude of each district of Jharkhand. On moving the cursor above the map, the district tile rises up and darkens up the color of that district on which the cursor moves.

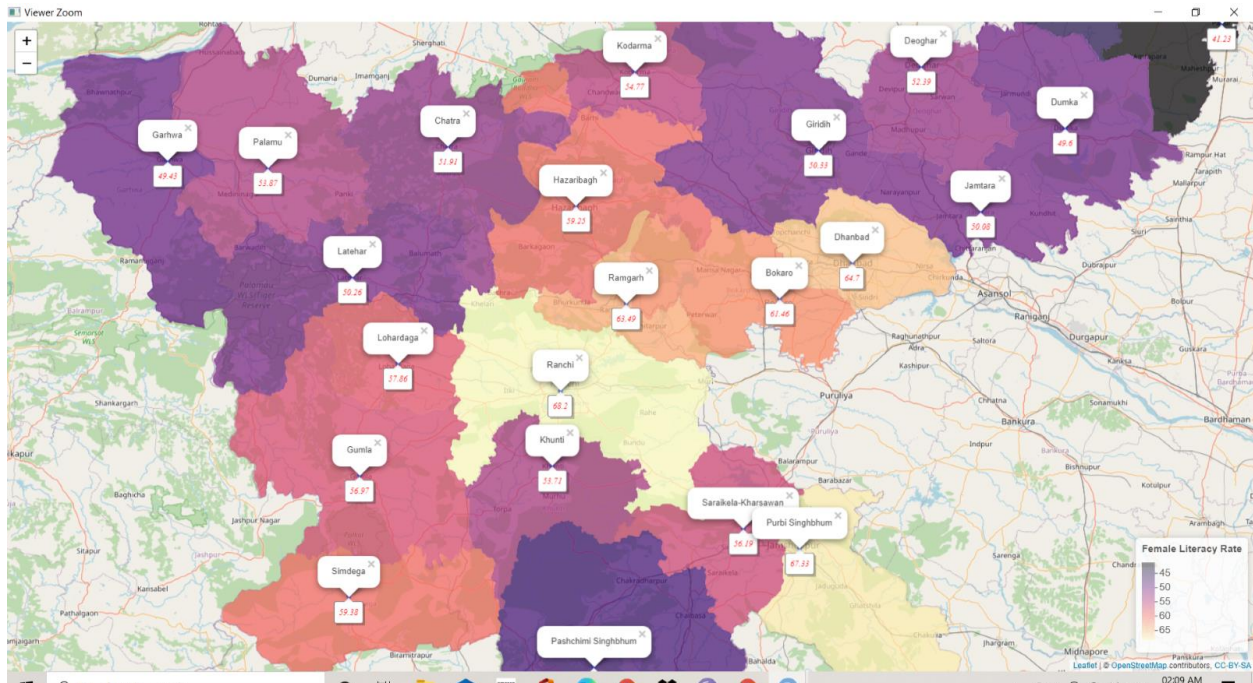
Map1:

The following map showcases the male literacy rate of each district along with the name of the district on the map. And the legend is shown at the bottom- right to understand the map.



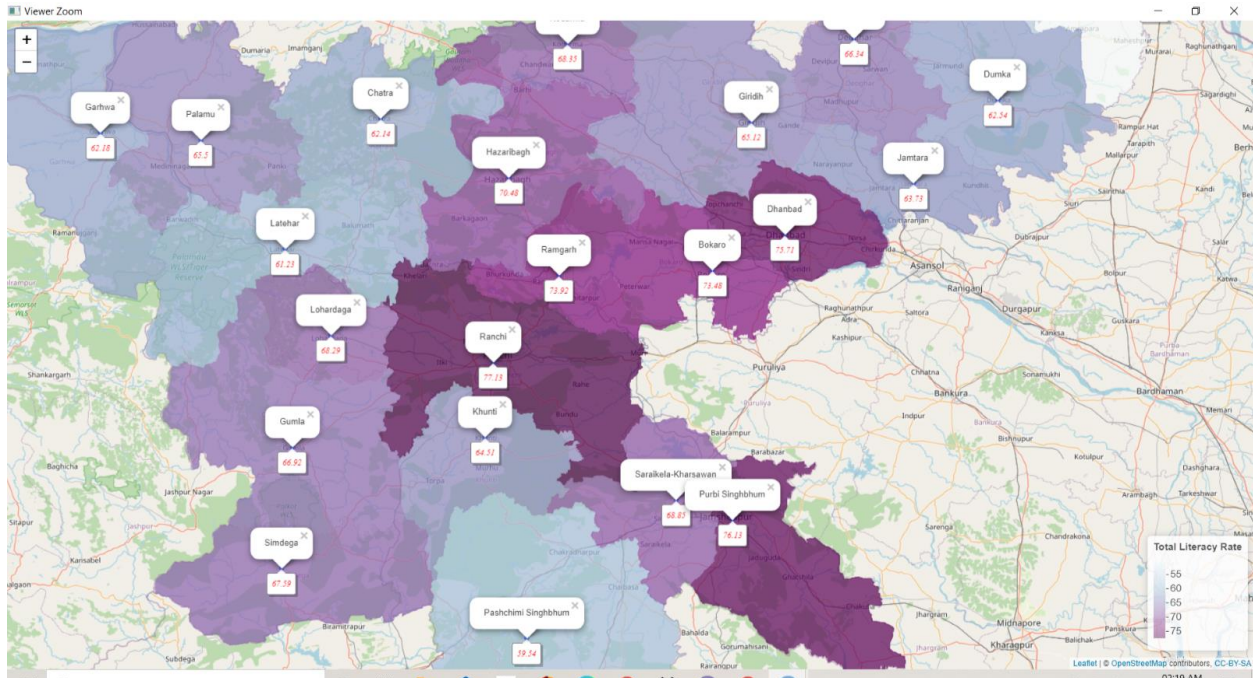
Map2:

The following map showcases the female literacy rate of each district along with the name of the district on the map. And the legend is shown at the bottom- right to understand the map.



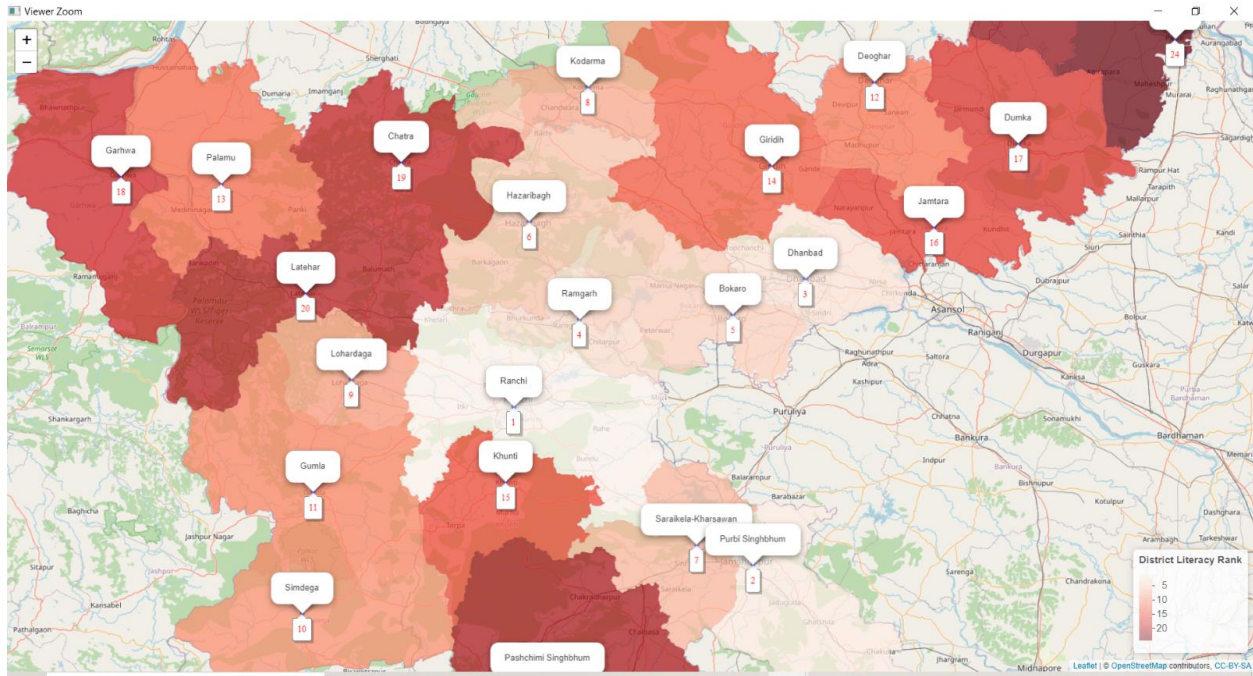
Map 3:

The following map showcases the total literacy rate of each district along with the name of the district on the map. And the legend is shown at the bottom- right to understand the map.



Map4:

The following map showcases the literacy Rank of each district along with the name of the district on the map. And the legend is shown at the bottom- right to understand the map.



C)Statistical Data Analysis:

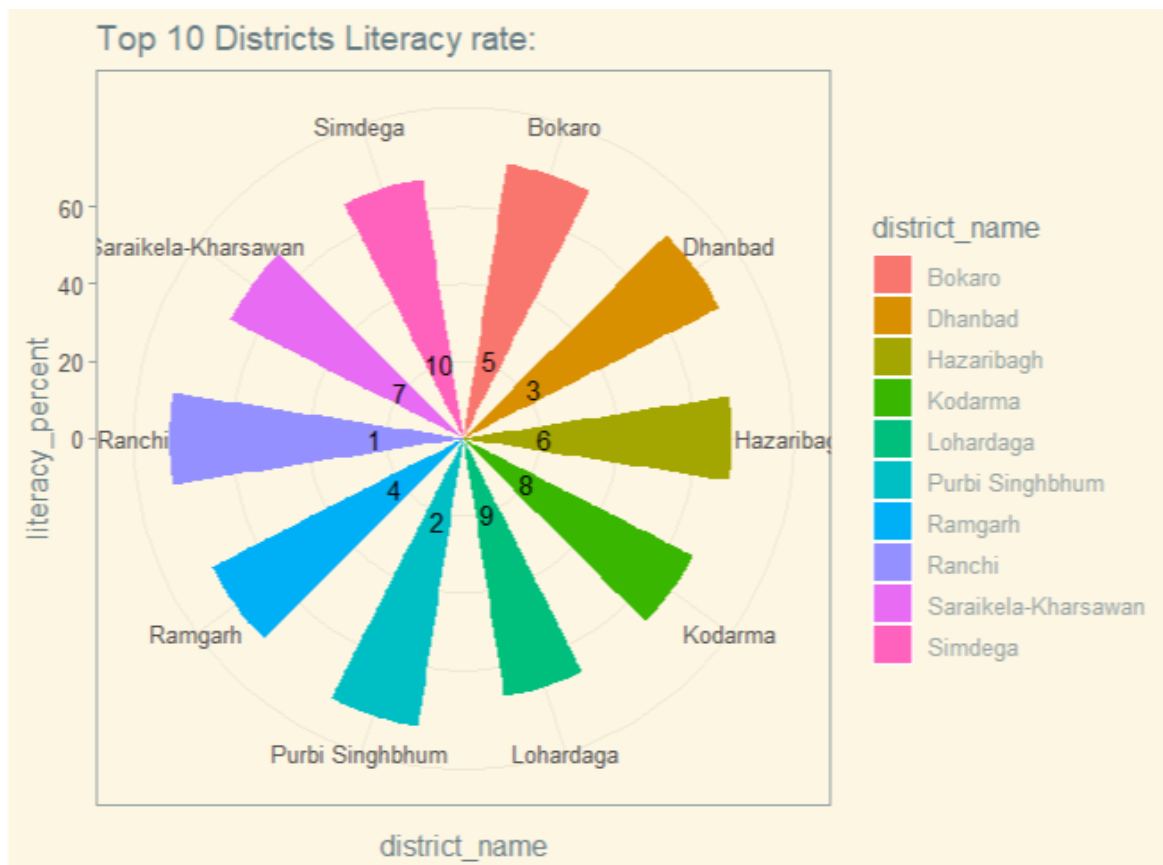
Plot1:

The given plot is created using functions of “ggplot2” libraries, “ggthemes” library is used to add theme to the plot.

The plot shows the Top 10 districts literacy rate in percentage.

The top 10 ranked districts of Jharkhand are shown in the pie chart along with the districts name and their rank with each district showing different color.

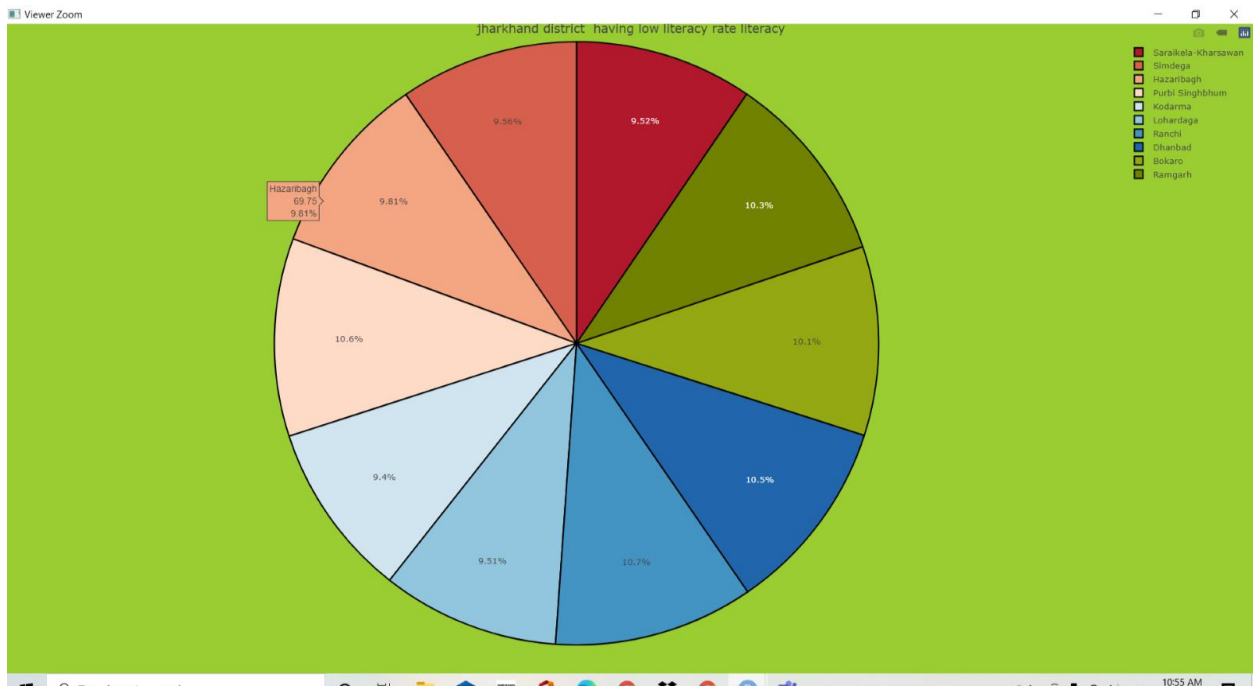
To do so, first I created a dataset named q1 by filtering out only those data which falls under literacy rank of 1 to 10 and then I created this pie chart.



Plot2:

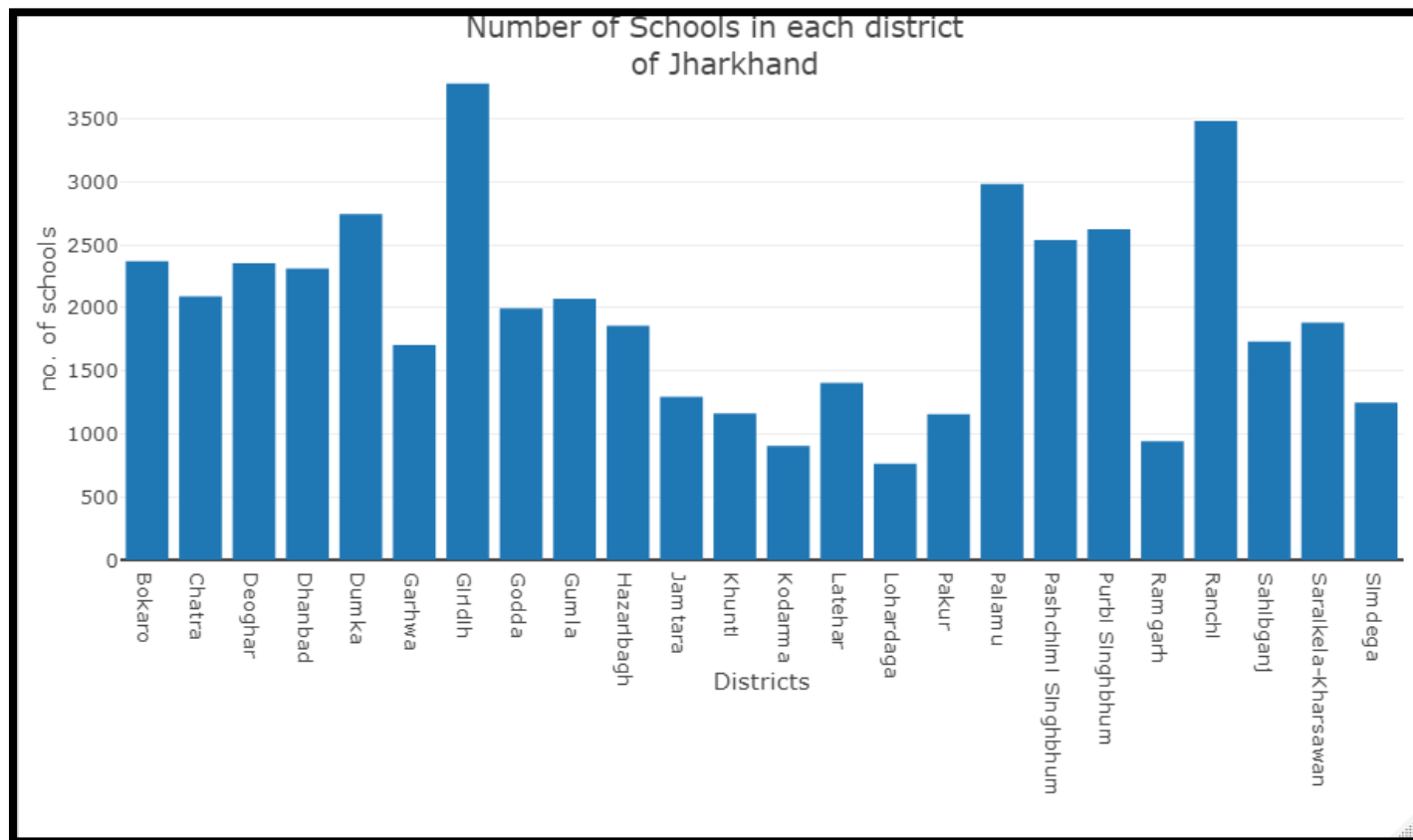
The following pie chart has been created using the function `plot_ly()` from the “plotly” libraries and added the layout using the `layout()`.

The pie chart shows the districts having low literacy rates. To do so, I have created a dataset named `q2` in which I have included only those data having literacy rank greater than 15 that is, I excluded the top 15 ranked districts data. On moving the cursor above the pie chart, the district name and its literacy rate is shown. And the legend is added at the top-right corner for understanding which block belongs to which district.



Plot 3:

The following bar plot presents each district as a individual bars and for each bar I.e. for each district the count of number of schools is shown.



Plot 4:

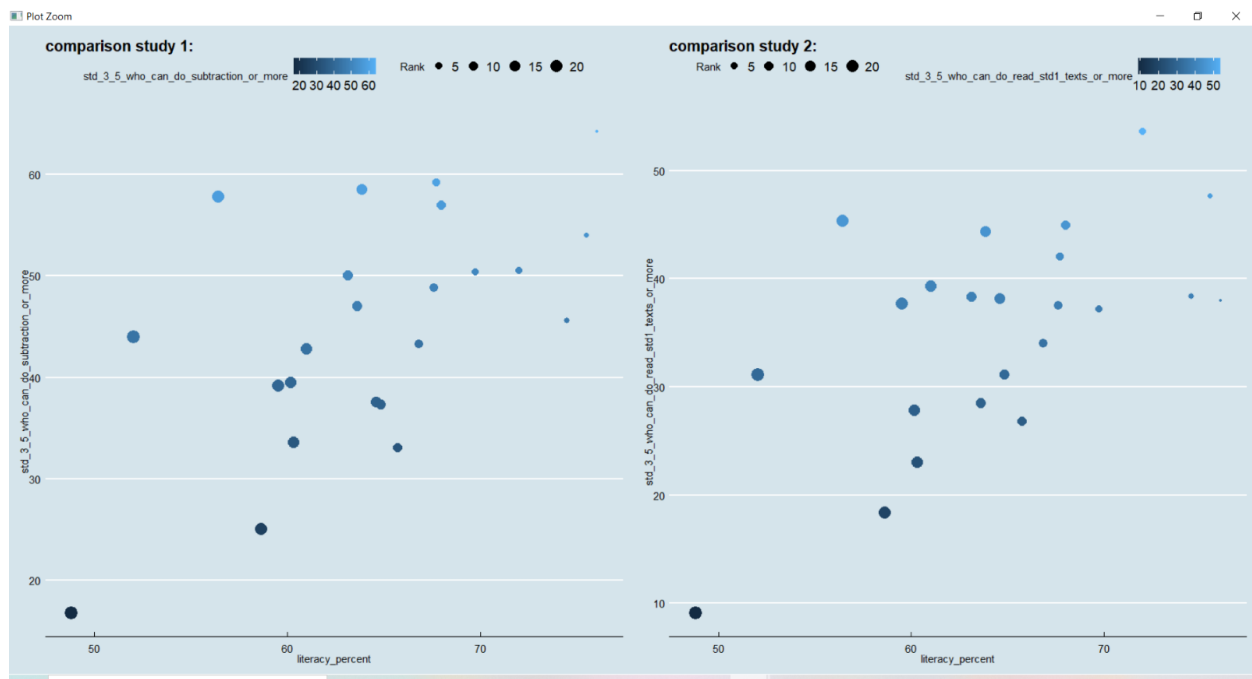
The following plot is created using the functions of the library “ggplot2”. I have used functions `ggplot()` to enter the dataset, `geom_point()` for creating scatter plot, `ggtitle()` for giving the title head to the plot and also have used `theme_economit()` to give a background theme to the plot.

I have plotted the plots together using the “`ggarrange()`” function from “`ggpubr`” library.

It consists of two plots:

- 1] literacy percent plotted against std_3_5_who_can_do_subtraction_or_more
- 2] literacy percent plotted against std_3_5_who_can_do_read_texts_or_more

The plot depicts that the districts where the literacy percent is more the count of the students who are able to do the basic math and those are able to read the texts is also more.



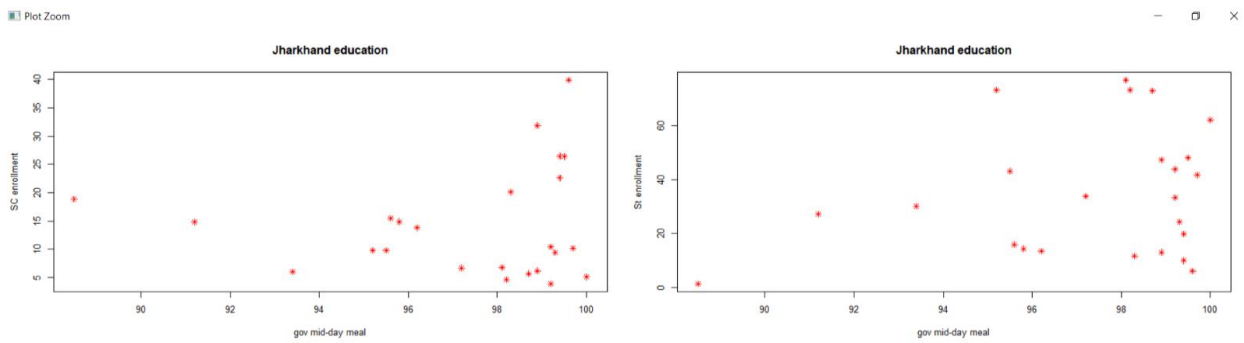
Plot 5:

This is a simple plot created using the plot() function.

It consists of the following two plots:

- 1] gov mid-day meal plotted against SC enrollments
- 2] 1] gov mid-day meal plotted against ST enrollments

On observing the following plots it can be concluded that SC and ST enrollments in government schools is greater in the districts where the government schools provide mid-meals.



Conclusion:

Based on my Data analysis following can be concluded:

- The capital of Jharkhand has ranked top in the literacy rate followed by purvi singhbhum and Dhanbad, ranking 2 and 3 respectively.
- The districts Pakur, shaibganj, and Godda have the least literacy rate.
- SC and ST enrollments in the government schools are more where the government schools provide mid-day meals.
- The primary class students are performing well in the districts where the literacy rate is higher than in those where literacy rate is poor.

What I learned during the project completion:

It hadn't been long since I started learning R programming for spatial science but I'm still happy about the fact that I was able to do this project where I worked on a real-life dataset.

In the initial stage of finding the datasets for the project, it is convenient that several open source websites including the government websites provide the real datasets which helps us to enhance our study. After obtaining the datasets, I found that the datasets are not cleaned datasets, they contain either some missing values or some dummy/noisy data. And hence we have to tidy up the data before using it to get results. Tidying and transforming the data is important because the consistent structure will help focusing on our analysis. My original dataset had the column names in way that it

always resulted in error and so I had to change the column names in a valid variable name format before using dataset.

During the data analysis I learned about several more libraries and their functions and how and where they work. I specially enjoyed while I was working with the interactive maps because that's where I was able to explore more. Personally, I thought working on interactive maps would be a bit difficult and time consuming but once I started working on it, I learned it easy to code for them. It's just that one should know which functions is to be used where. R studio is a very interactive software to work with. The help option of the R studio helped me to analyze my errors and to correct it.

Overall, I had great experience working on R studio with real life dataset. And hence I extend my gratitude towards my professor Rajat Chopra for assigning such projects to us and for his guidance.

Bibliography:

The class notes shared by professor Rajat Chopra.

Jharkhand Economic Survey 2016-17

<https://hub.arcgis.com/datasets/esriindia1::jharkhand-education/data>

<https://geocompr.robinlovelace.net/adv-map.html#interactive-maps>

https://slcladal.github.io/maps.html#1_Getting_started_with_maps

<https://www.embibe.com/indian-states/schools-in-jharkhand/>