

# CS595: DATA VISUALIZATION

## Education Infrastructure



**Sadem Vasant Kumar(224161010)**

**Department of Computer Science and Engineering**

**Indian Institute of Technology Guwahati**

**30 April, 2023**

# Contents:

**1.Data Sets**

**2.Introduction**

**3.Approach**

**4.Data Visualization**

**5.Results And Conclusions**

# how the education in India is progressing over the years ?

Given Data sets :

- Percentage\_of\_schools\_with\_boys\_toilet
- Percentage\_of\_schools\_with\_girls\_toilet
- Percentage\_of\_schools\_with\_Water\_Facility
- Percentage\_of\_schools\_with\_electricity
- Percentage\_of\_schools\_with\_Computers
- Gross\_Enrollment
- Dropout Ratio

Additional data set considered:

- Literacy\_Rate

## Approach

- ➡ After going through the data, I performed data cleaning to improve its usability.
- ➡ I attempted to analyze the correlation between water facilities and the availability of boys' and girls' toilets, as well as the correlation between computer facilities and electricity availability within the dataset comprising five facilities.
- ➡ Idea of analyzing to determine if there exists any correlation between different categories of schools for each of the facilities.
- ➡ Idea of comparing all 5 facilities for all india over the three reference years .
- ➡ Idea of exploring all the facilities data sets
- ➡ Idea of Introducing Facility score to states
- ➡ Idea of linking it to literacy rates

# Data Visualization And Results :

## "No Water No toilets!"

Scatter plot between percentage values of boys toilet facility and water facility of primary only category

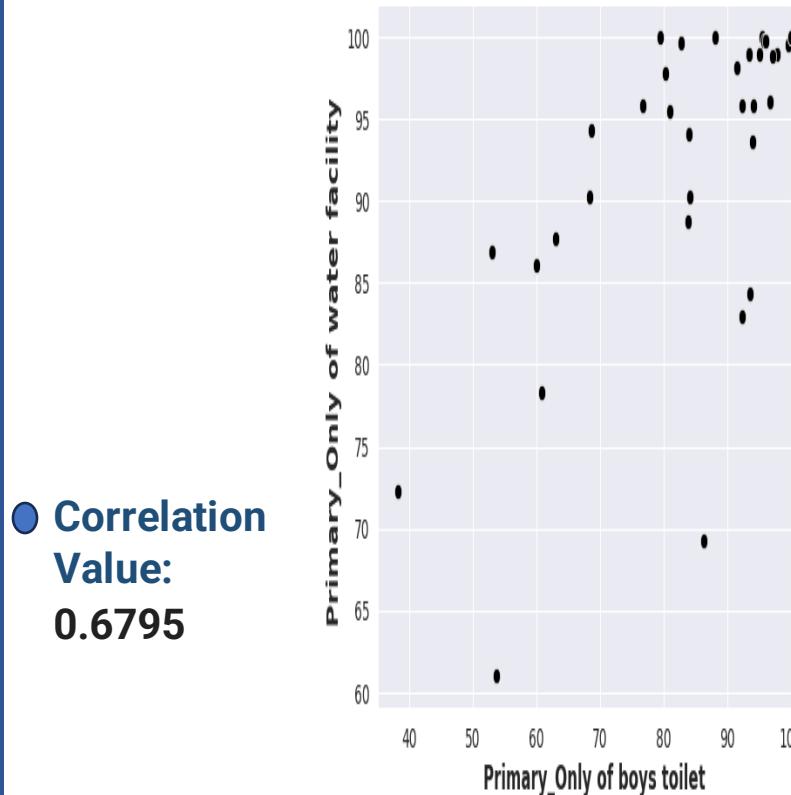


Figure .1

Scatter plot between percentage values of boys toilet facility and water facility of U\_Primary\_Only category

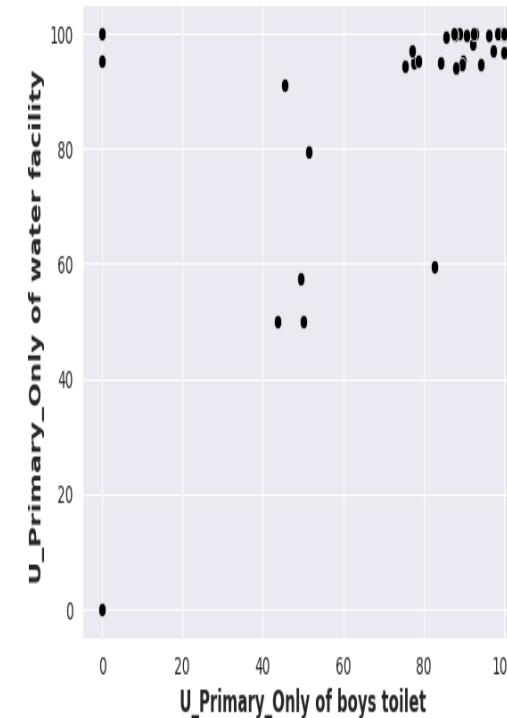


Figure.2

Correlation Value:  
0.6795

Correlation Value:  
0.727

# Data Visualization And Results :

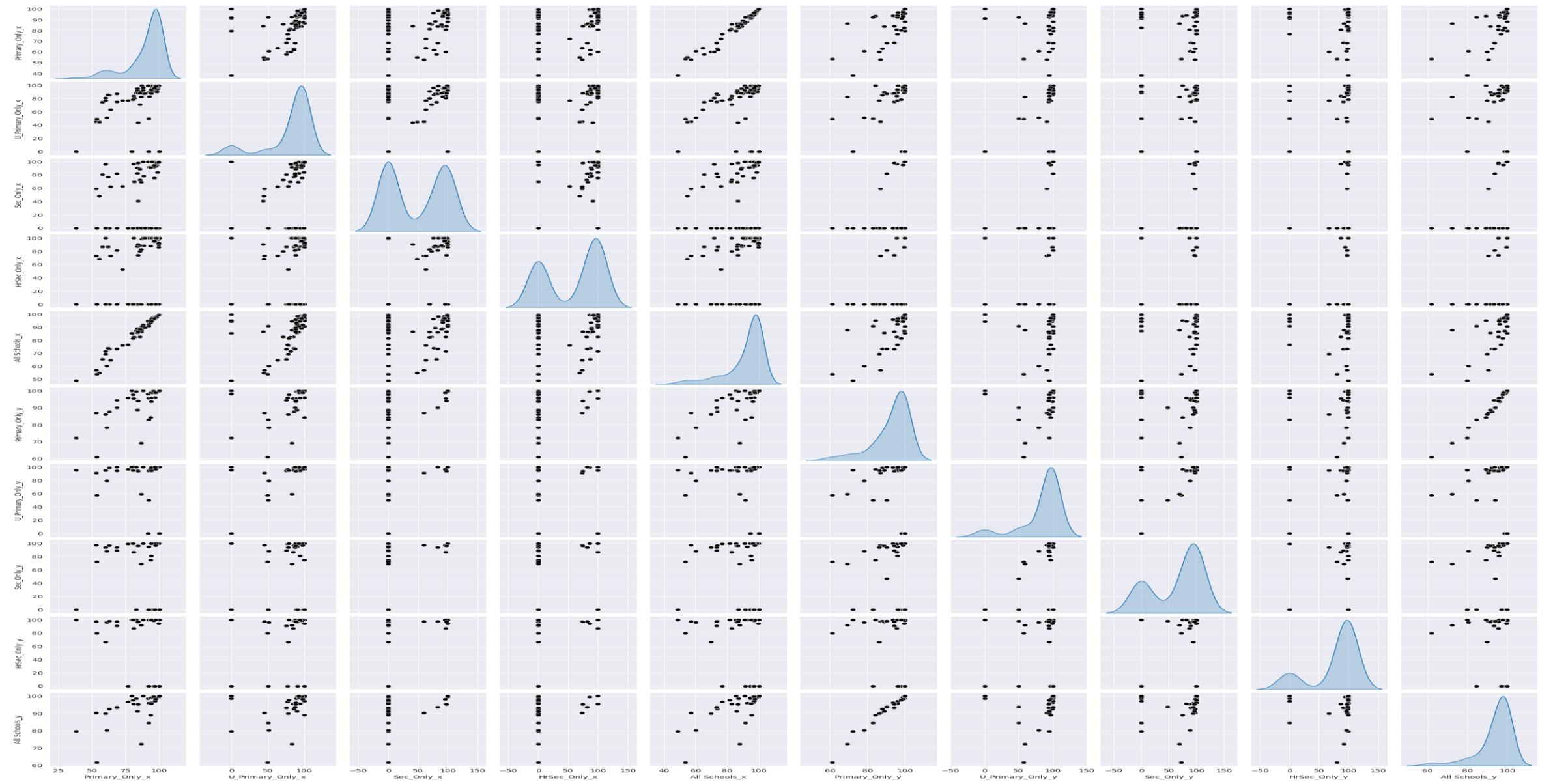


Figure .3

# Data Visualization And Results :

## "No Electricity No Computer Facility!"

Scatter plot between percentage values of electricity facility and computer facility of primary only category

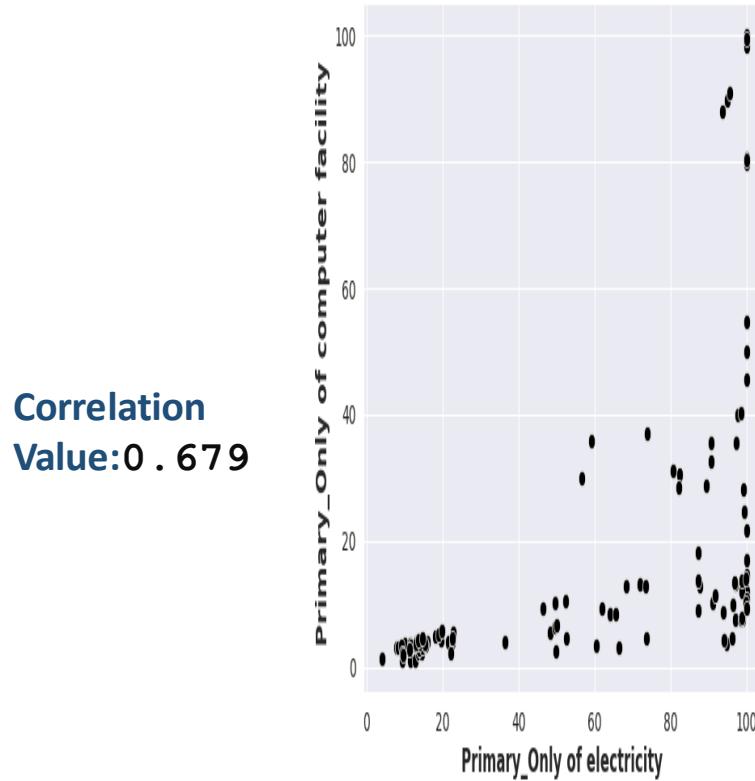


Fig.5

Scatter plot between percentage values of electricity facility and computer facility of U\_Primary\_Only category

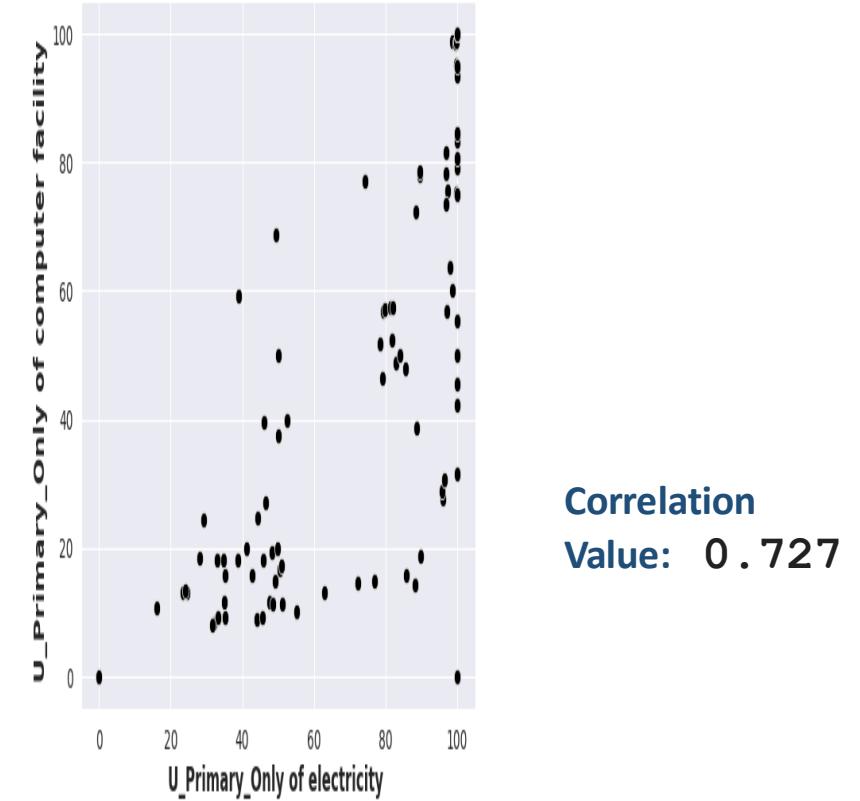


Fig.6

# Data Visualization And Results :

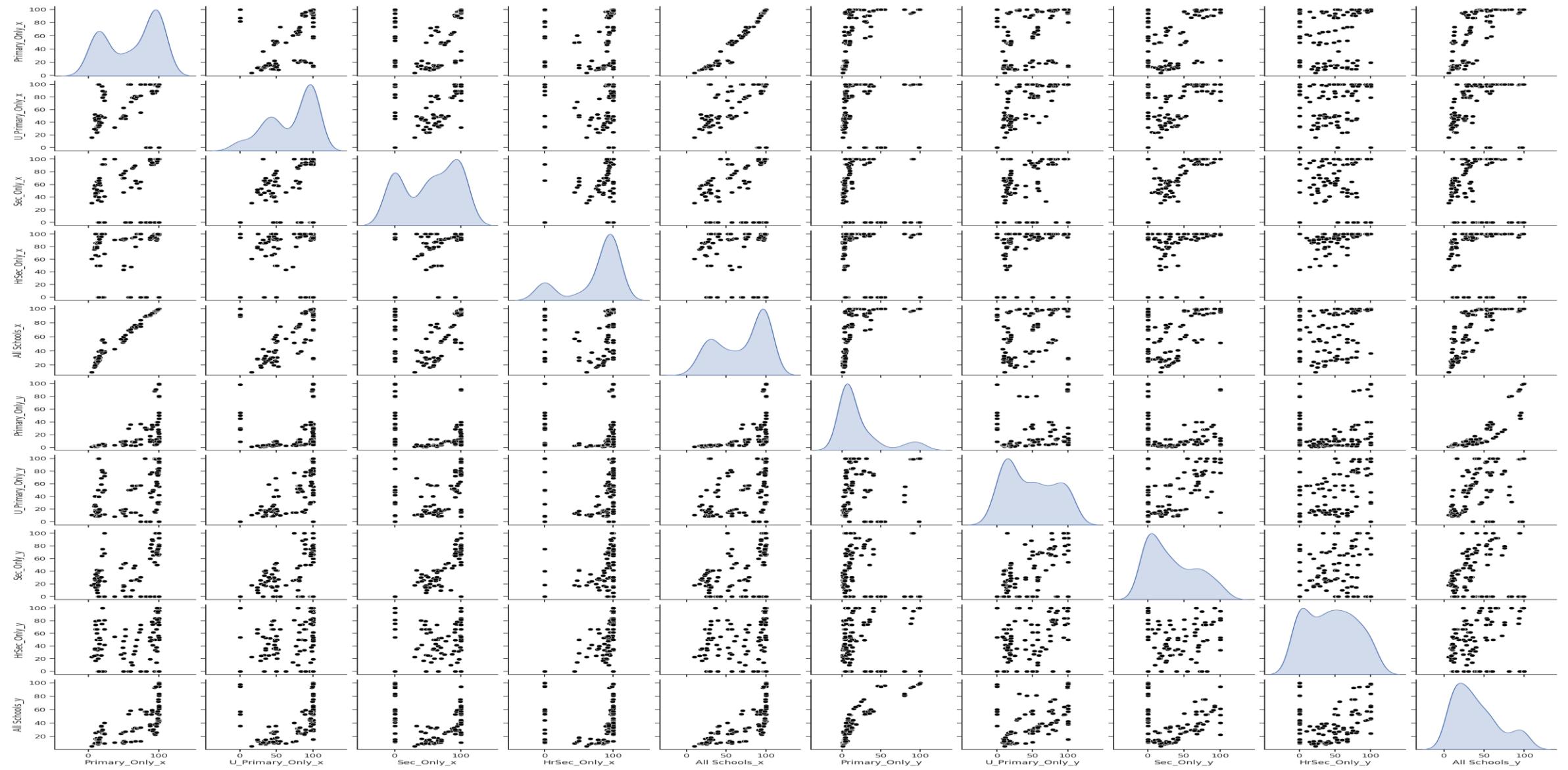
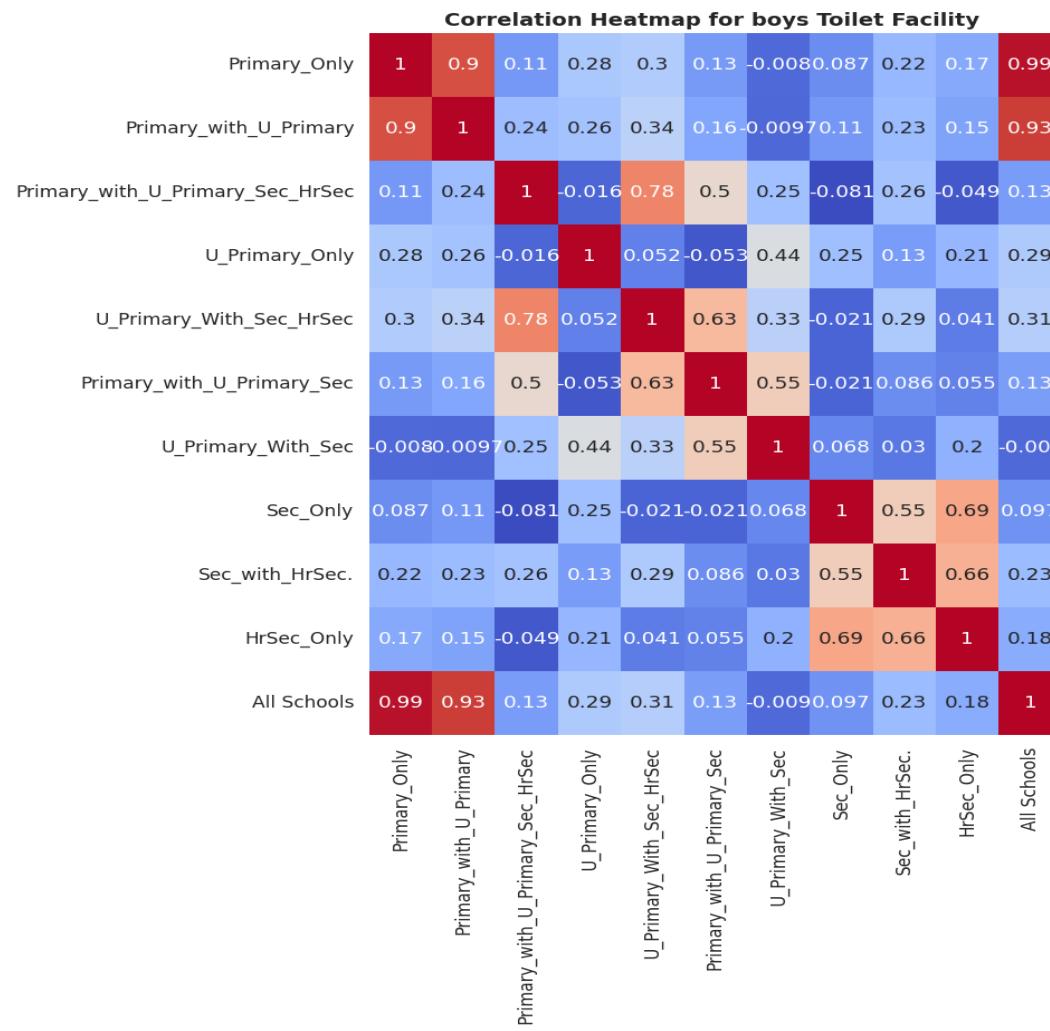


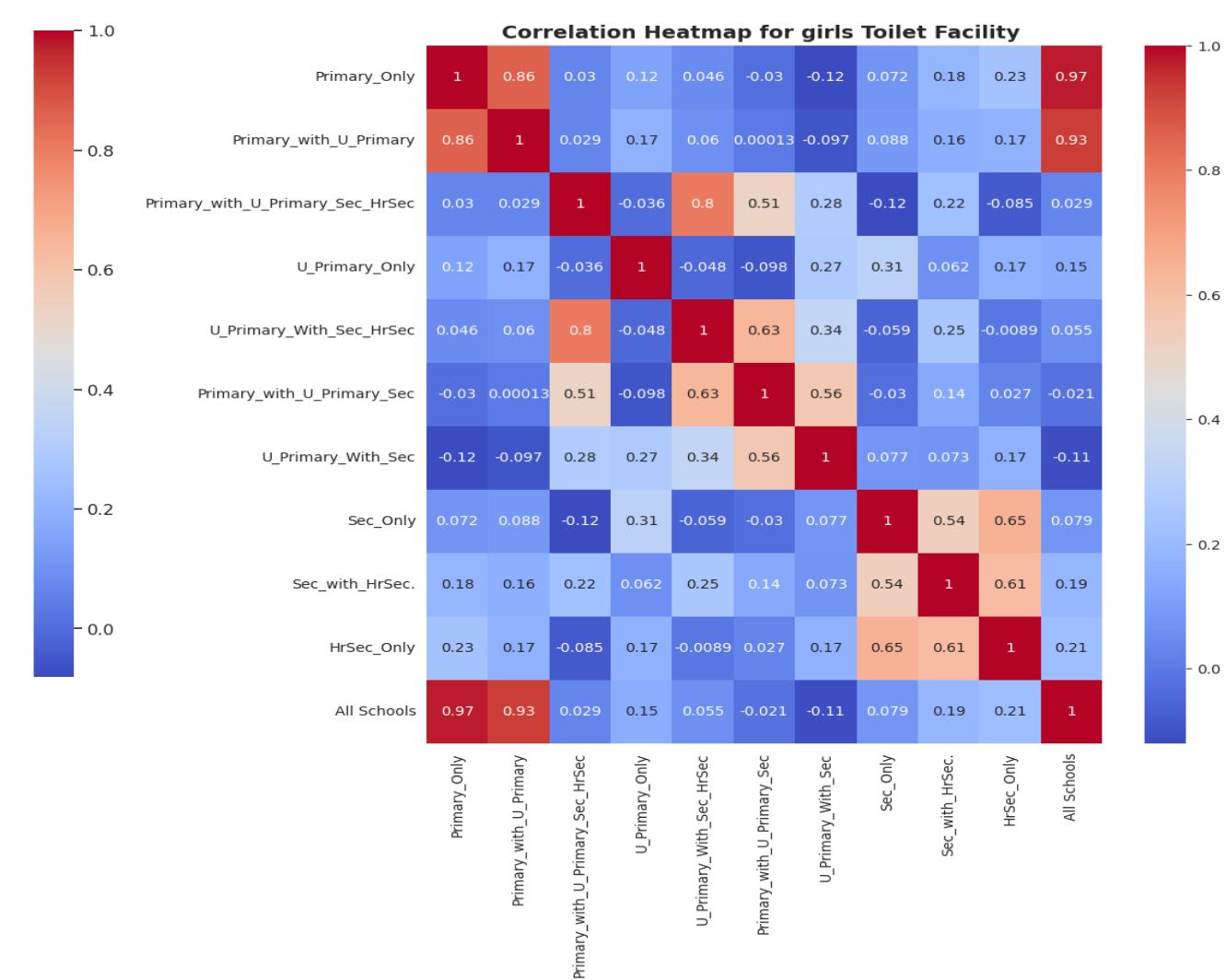
Fig.7

# Data Visualization And Results :

## Correlation B/w the categories corresponding to each Facilities::

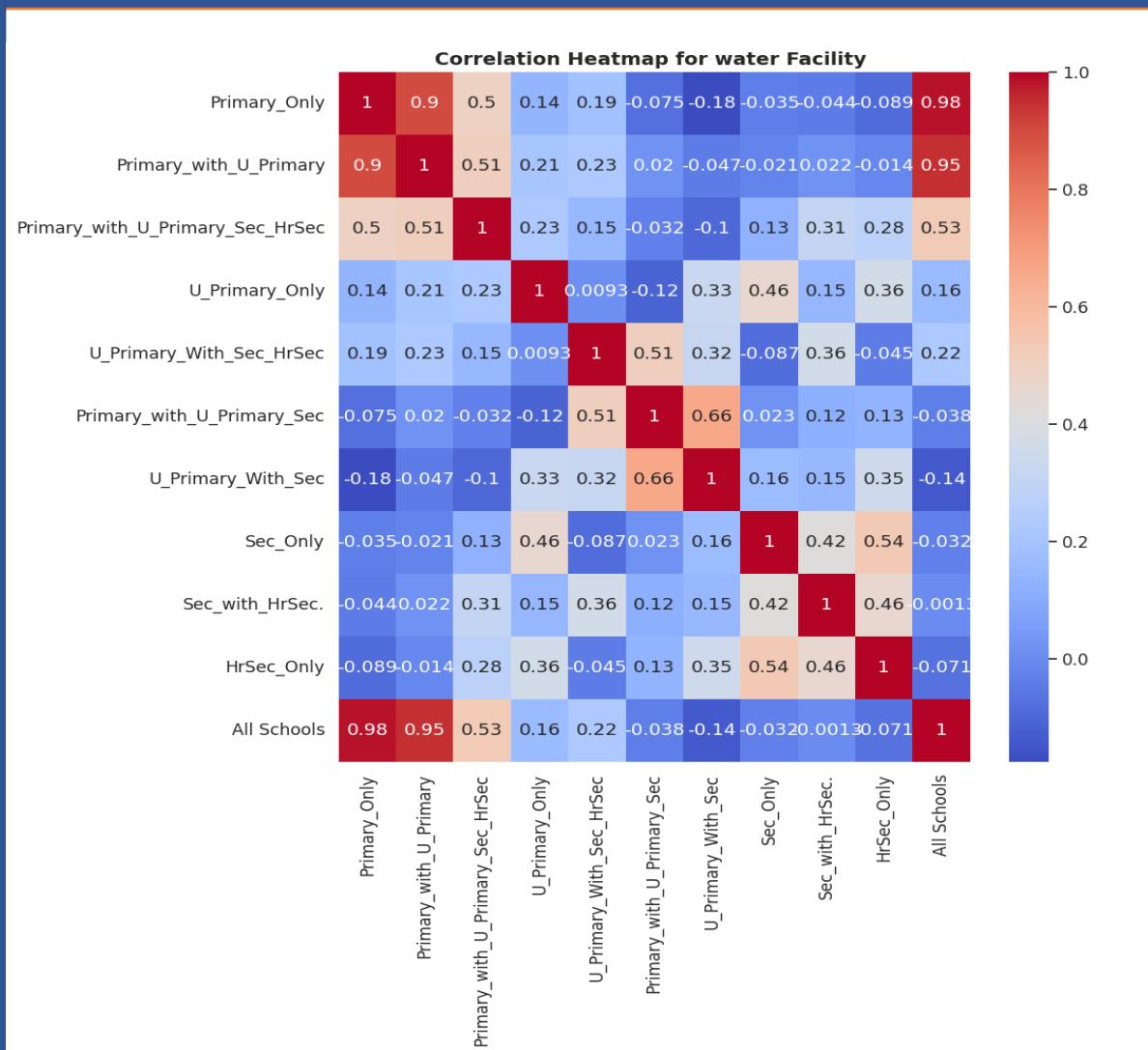


**Fig.8**

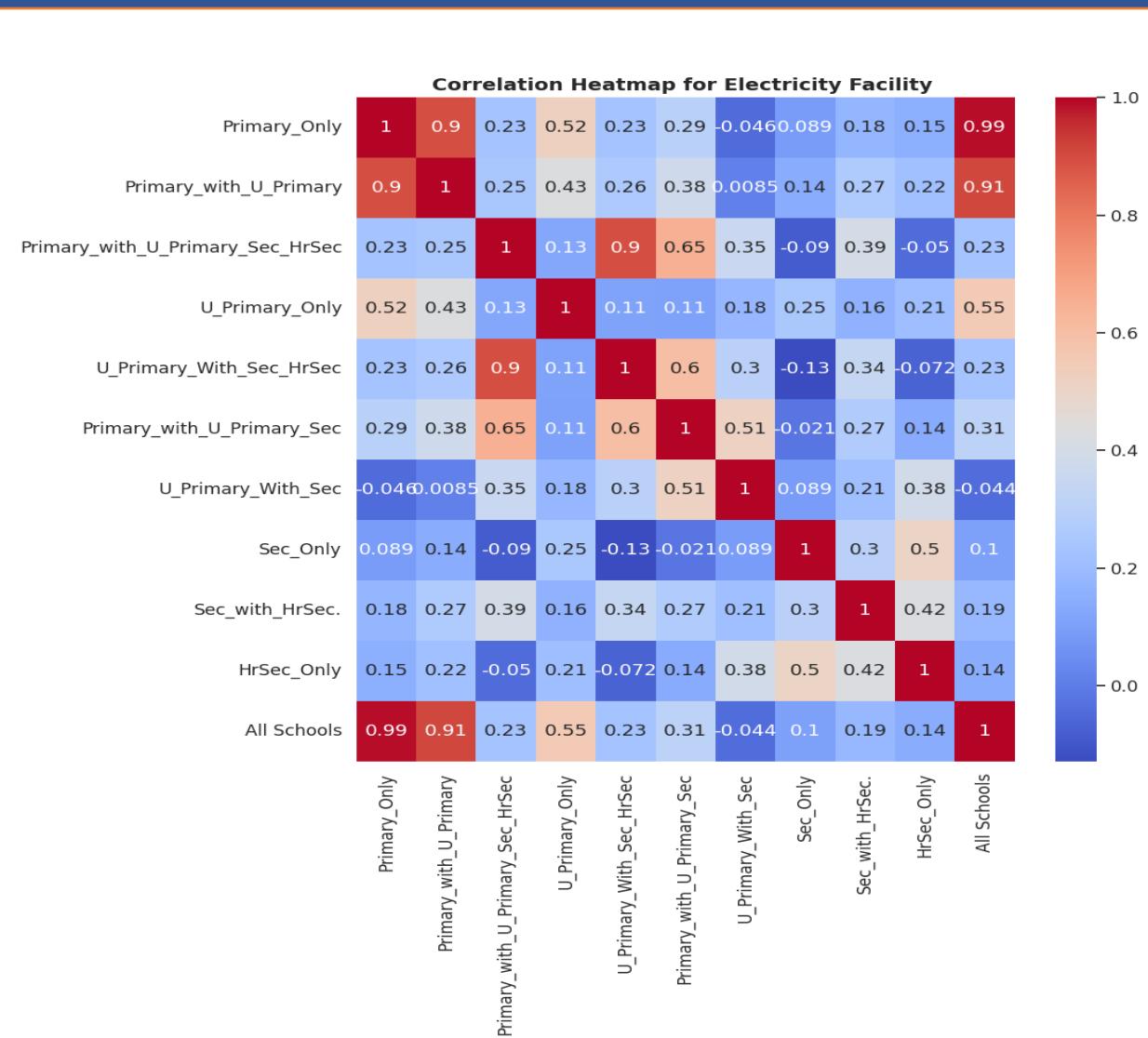


**Fig.9**

# Data Visualization And Results :

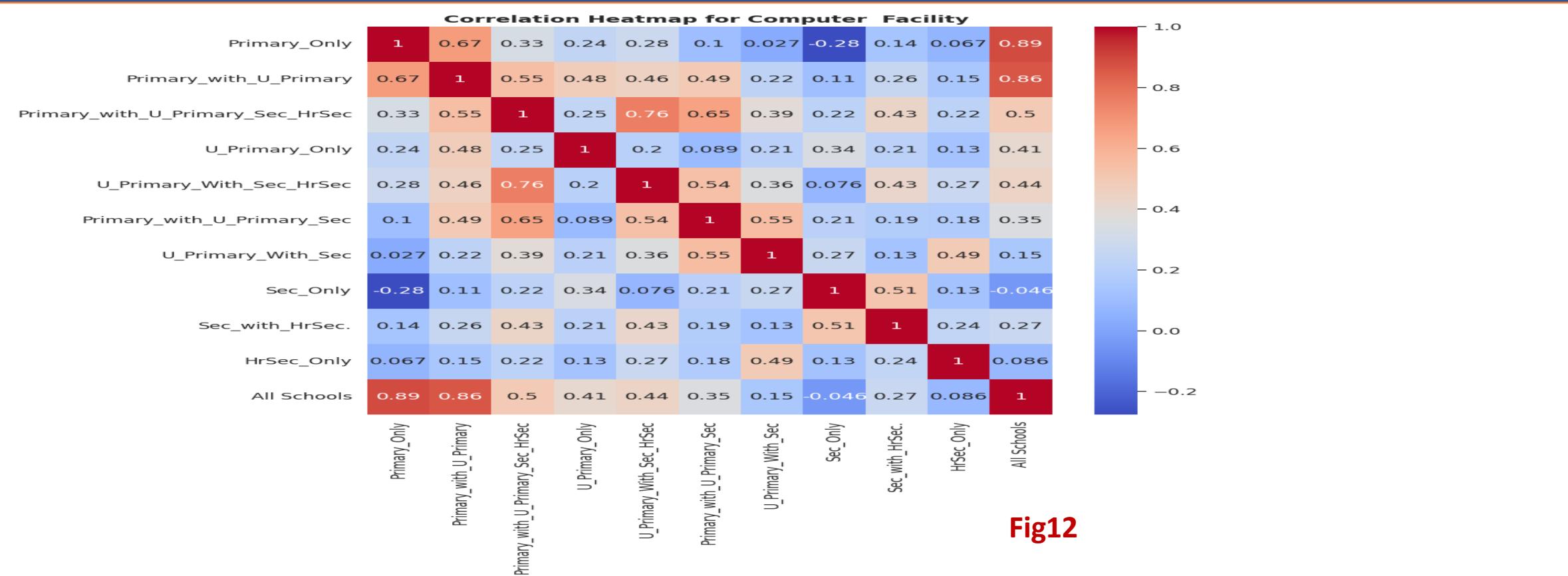


**Fig.10**



**Fig.11**

# Data Visualization And Results :



**Fig12**

►From all the heatmaps of the facilities we can clearly observe that the All\_schools is strongly correlated to primarily\_only category of Schools, which is an very important information .Later this is discussed in facility score concept

# Data Visualization And Results :

## Let's dive into the datasets!

### ►Comparision of all facilities of ALL India :

Education Infrastructure of India for the year 2013-14

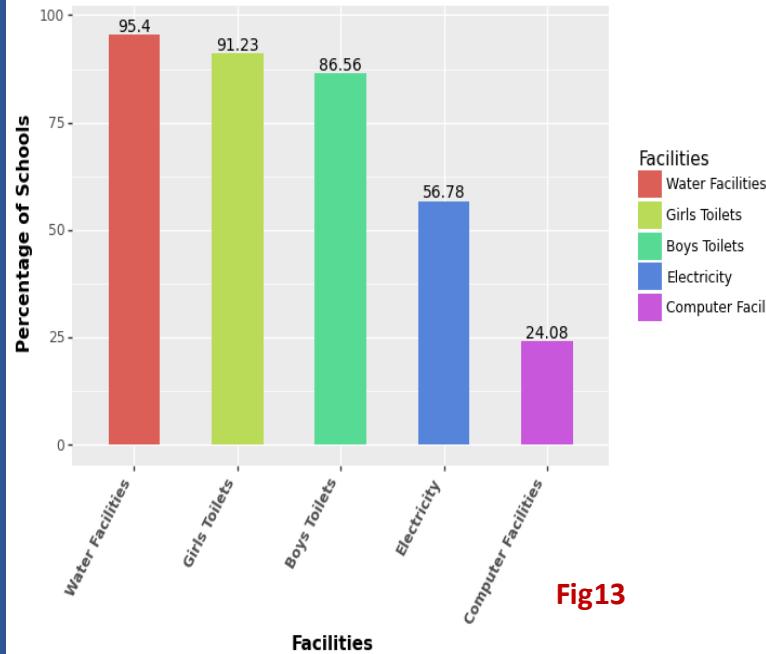


Fig13

Education Infrastructure of India for the year 2014-15

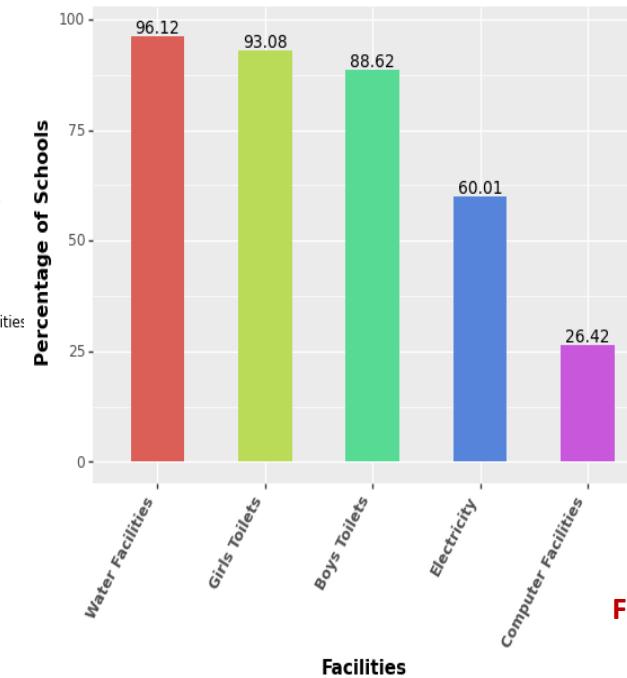


Fig14

Education Infrastructure of India for the year 2015-16

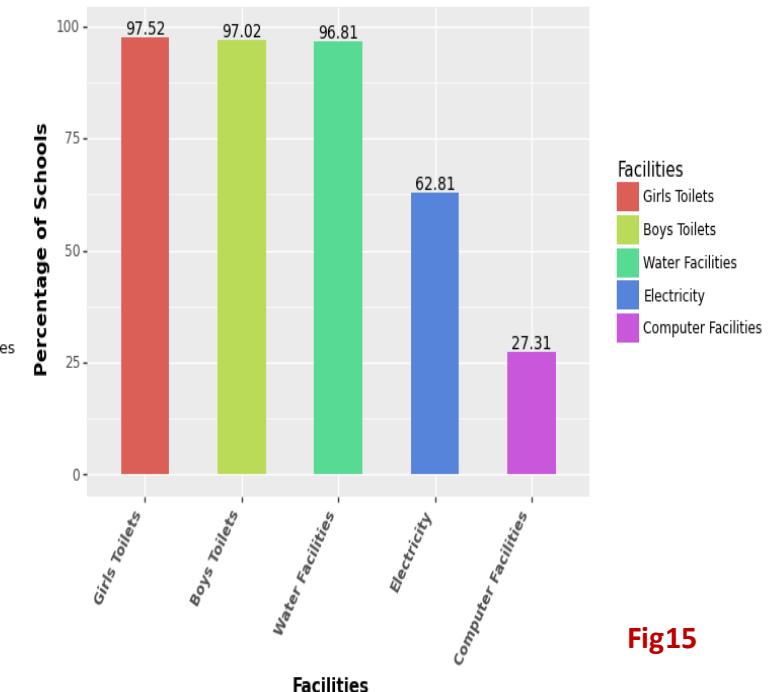
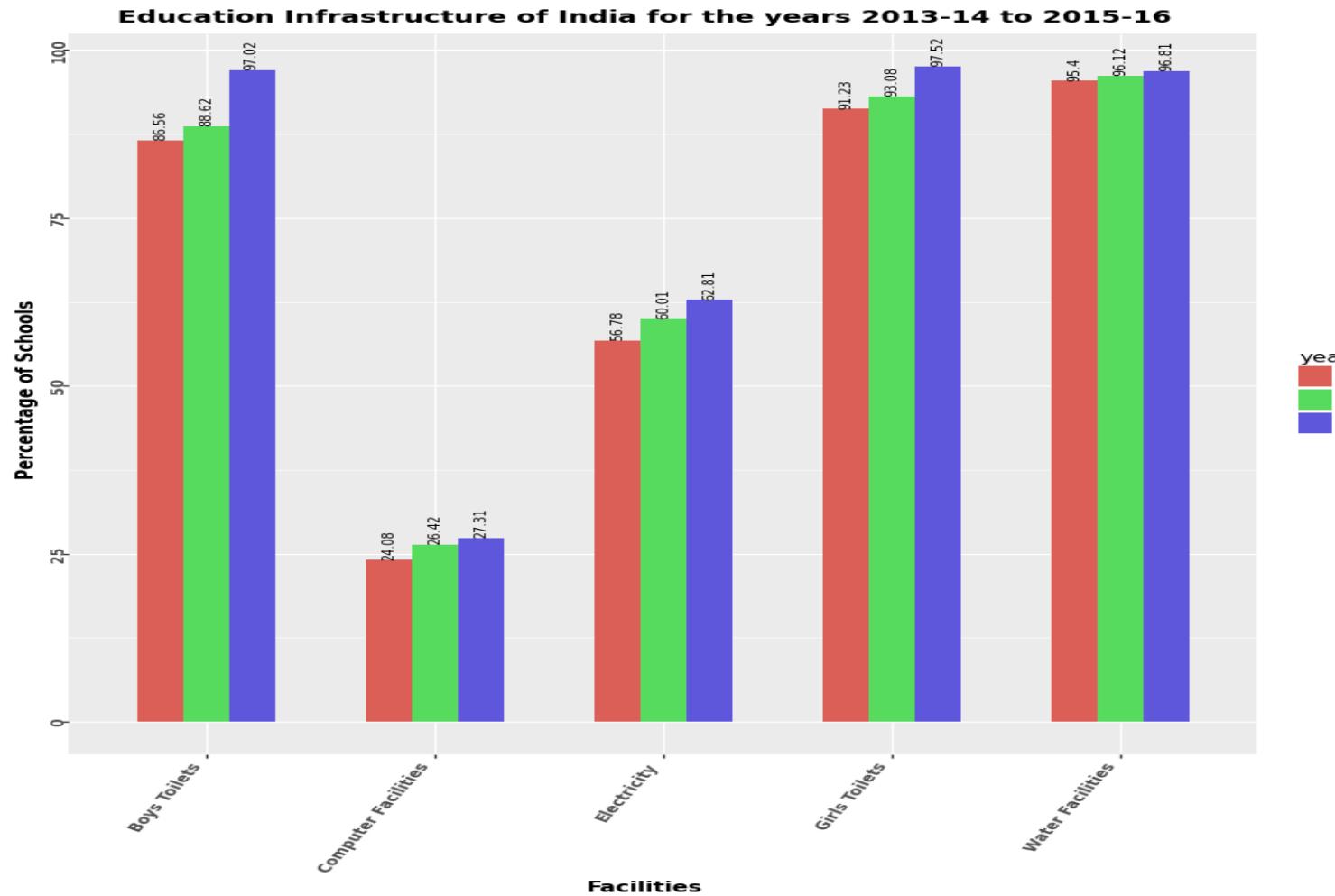


Fig15

►water facility remained same but we can see a good improvement in toilet facilities and the computer facility is remained unchanged in these three year poorly.

# Data Visualization And Results :



**Fig 16**

► From the above plot we can say The percentage of schools with boys' and girls' toilets facilities in India was almost equal, indicating an equal distribution of educational infrastructure for both genders.

# Data Visualization And Results :

## Let's explore more about Schools with Boys toilets !

Histogram of Percentage Value of All Schools with Boys Toilet Facilities in 2013-14

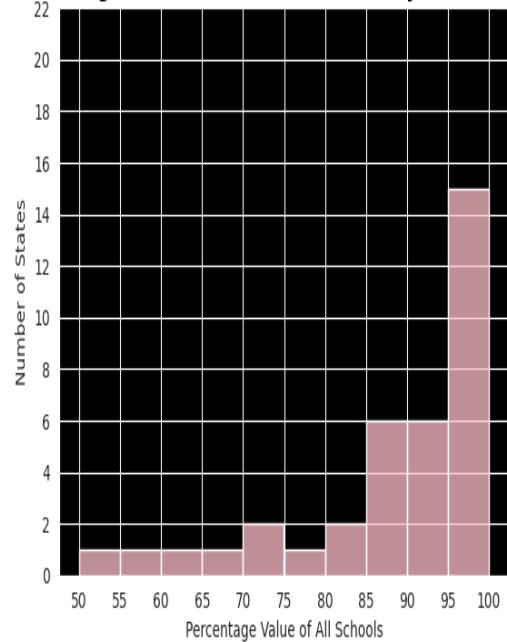


Fig17

Histogram of Percentage Value of All Schools with Boys Toilet Facilities in 2014-15

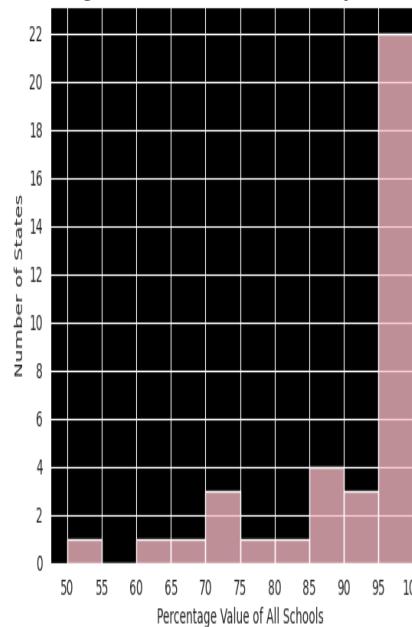


Fig18

Histogram of Percentage Value of All Schools with Boys Toilet Facilities in 2015-16

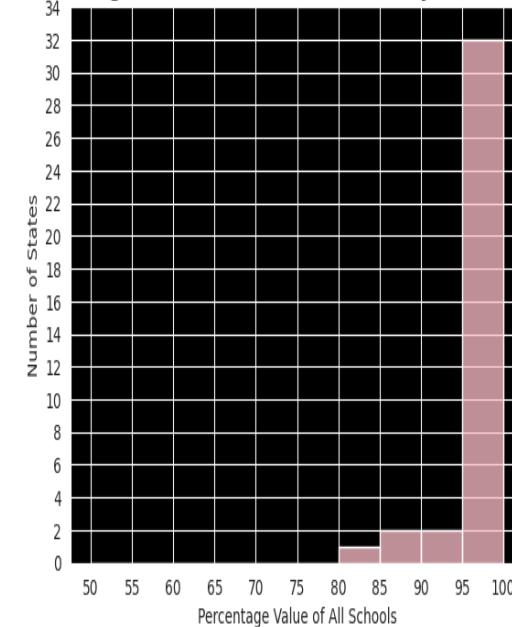


Fig.19

► By Comparing the two histogram plots for the years 2013-14 and 2015-16 , we can observe there huge improvement in the boys toilet facilities of states in India of all schools category.

# Data Visualization And Results :

lets having a quick comparision for all the categories of schools in brief

Histogram of boys toilets Percentage Value of All categories in 2013-14

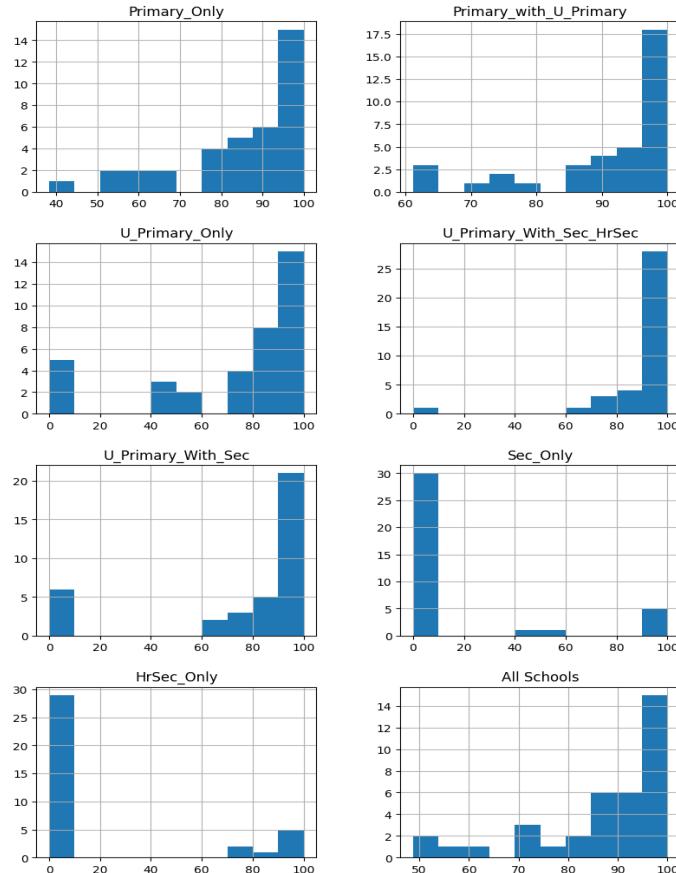


Fig.20

Histogram of boys toilets Percentage Value of All categories in 2015-16

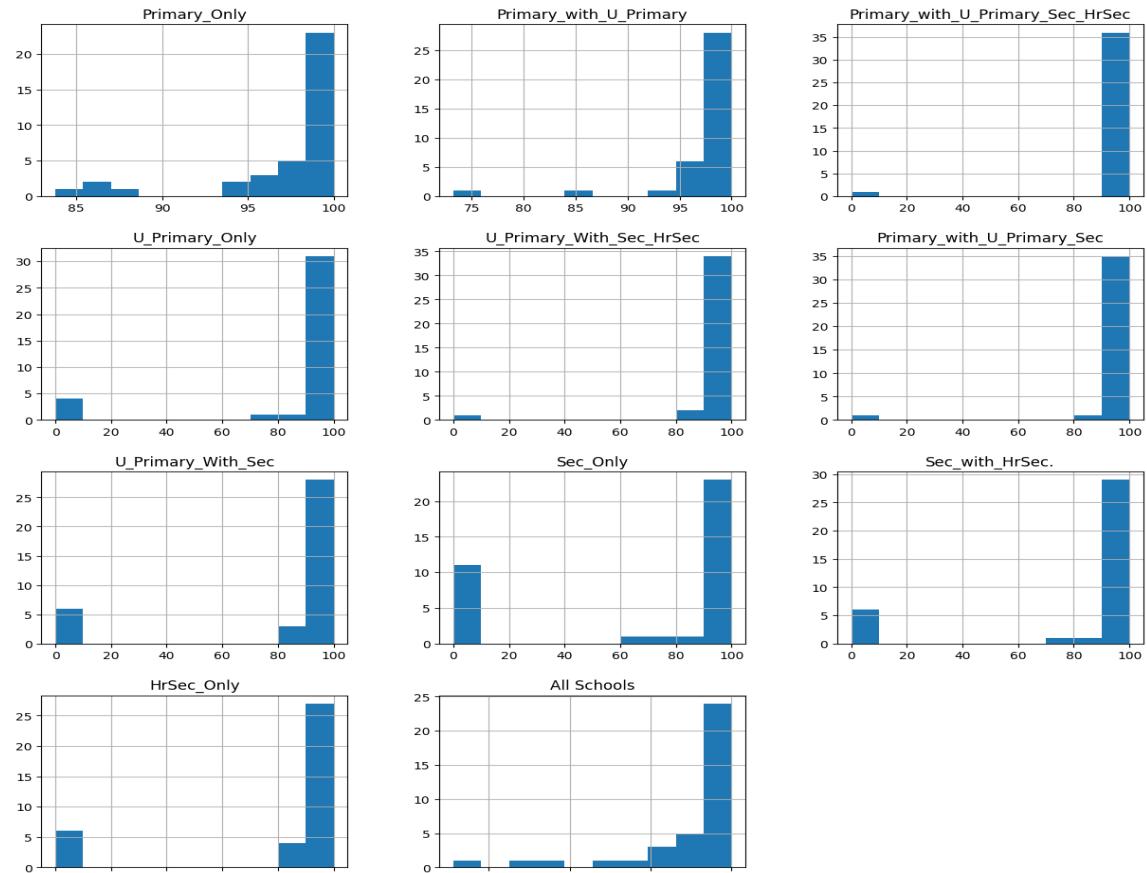


Fig.21

# Data Visualization And Results :

Percentage of Schools with Boys Toilets of all schools by State for the year 2013-14

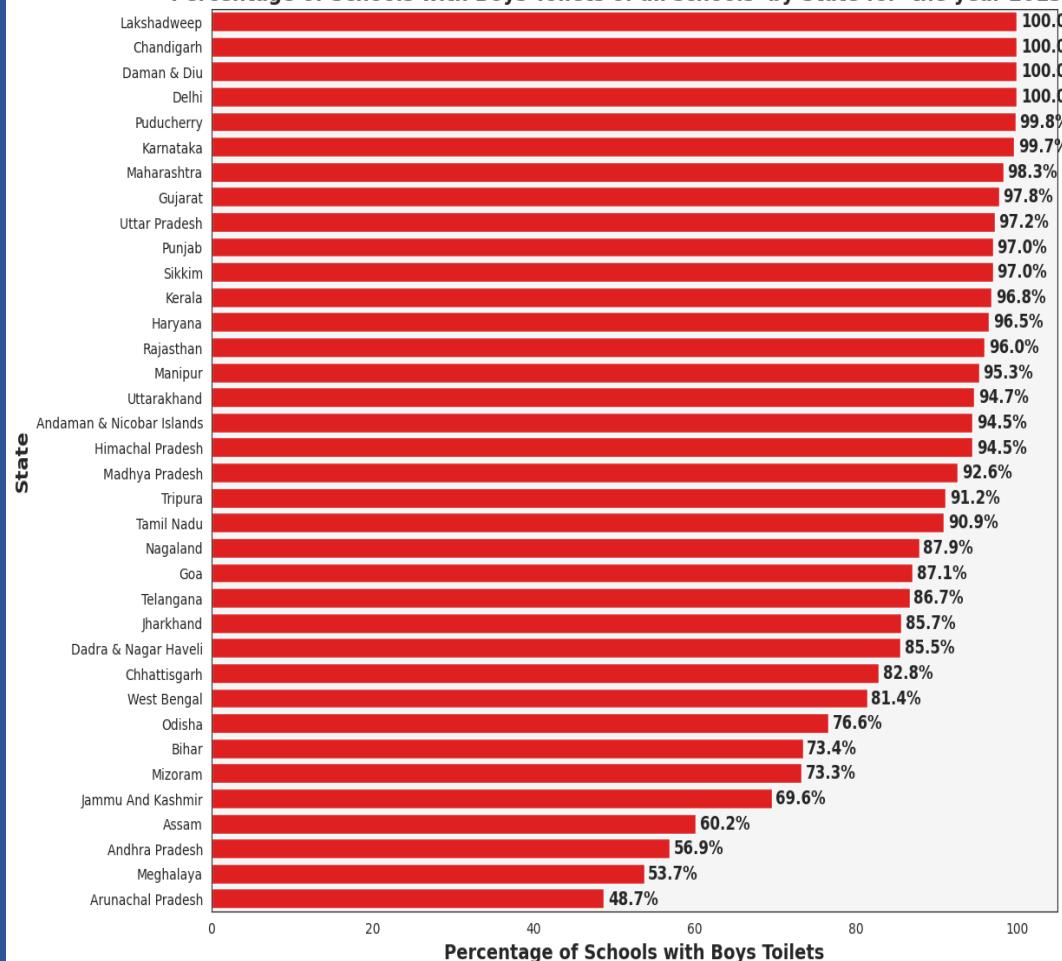


Fig.22

Percentage of Schools with Boys Toilets of all schools by State for the year 2015-16

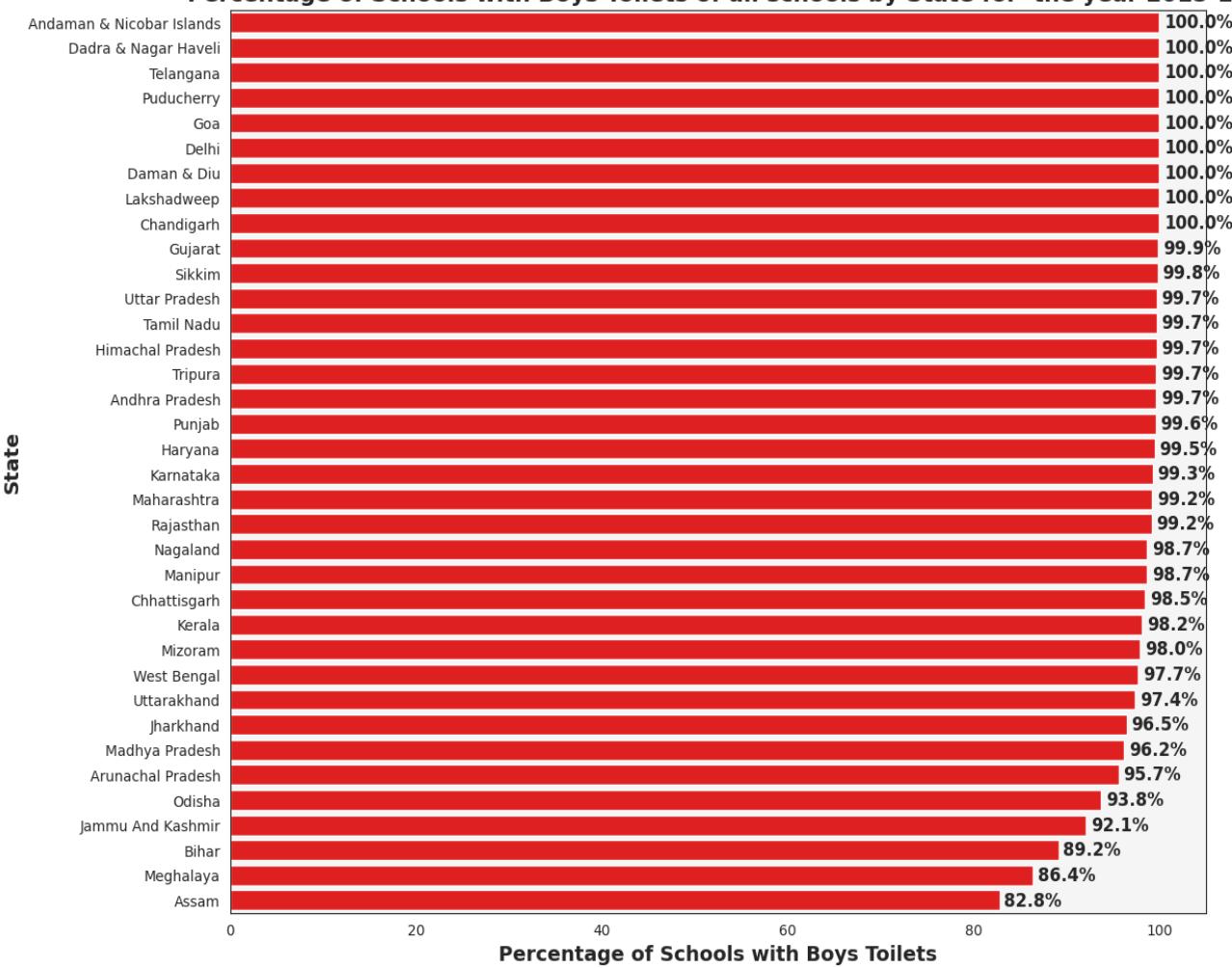
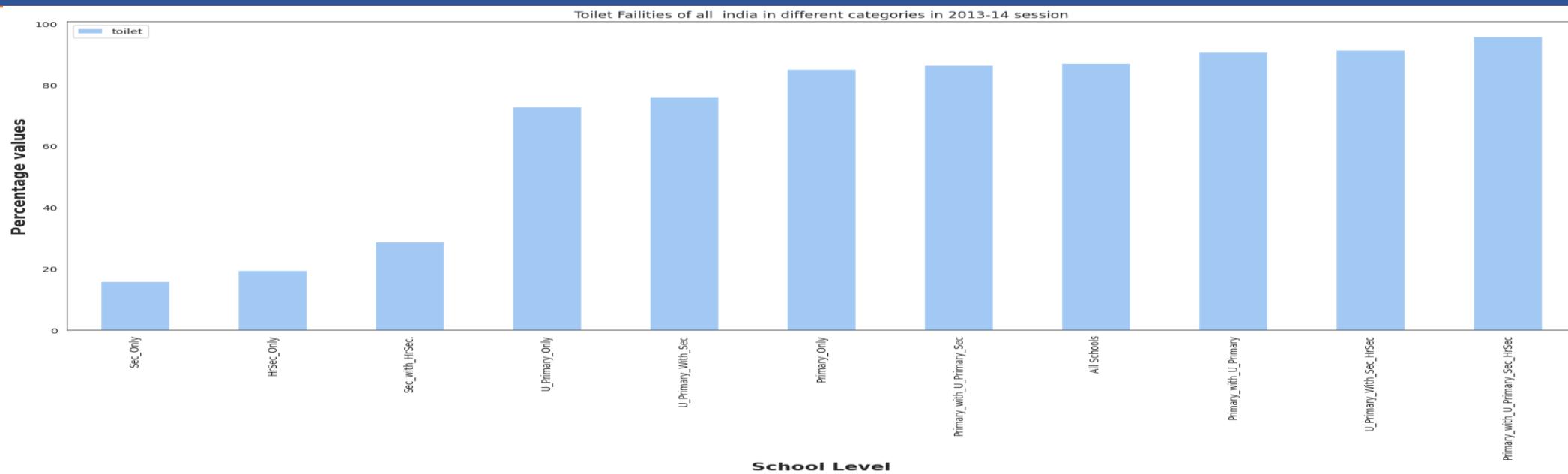
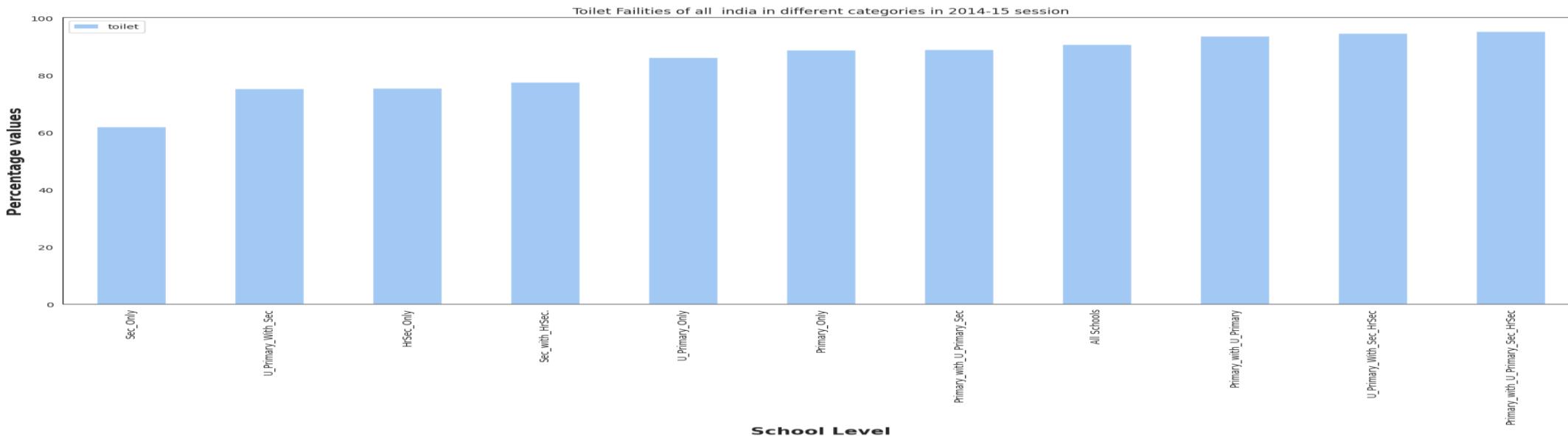


Fig.23

# Data Visualization And Results :



► From the bar chart above we can clearly see that toilets facilities weren't much in secondary schools in 2013-14 session

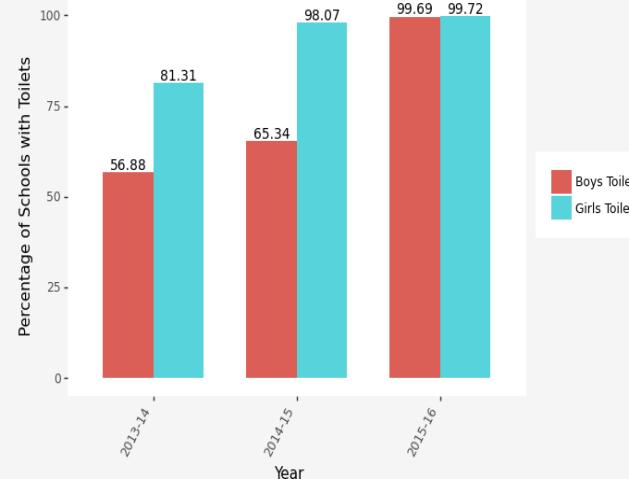


► From the bar chart above we can see rapid rise in the toilets facilities of in secondary schools in 2014-15 session, compared to last session, this is a big improvement

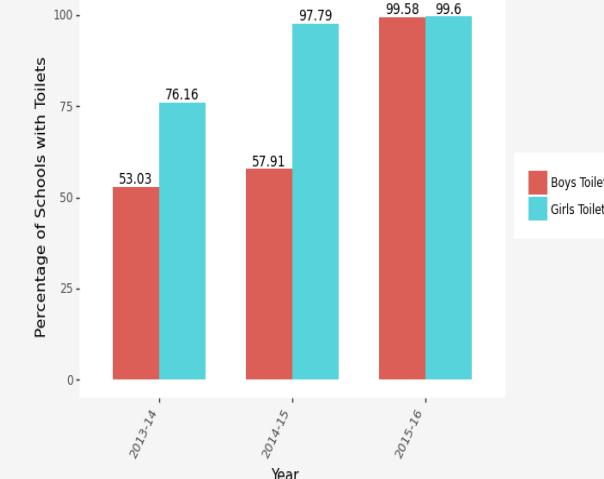
# Data Visualization And Results :

Let's explore for a particular state , for e.g. Andhra Pradesh

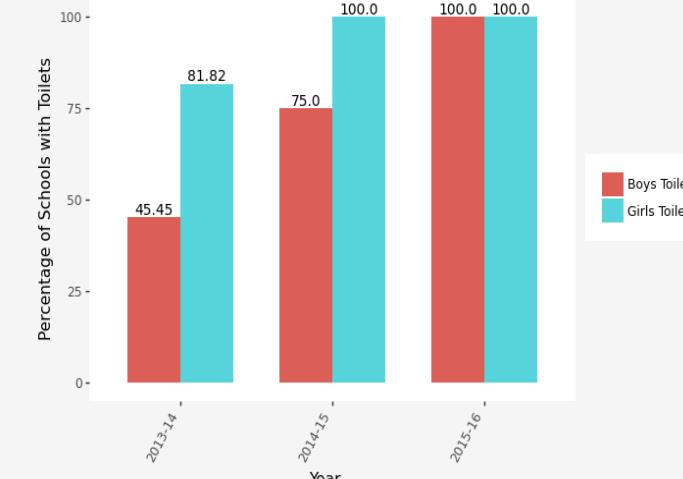
Year-wise Toilet Facilities Analysis for the state Andhra Pradesh for all schools category



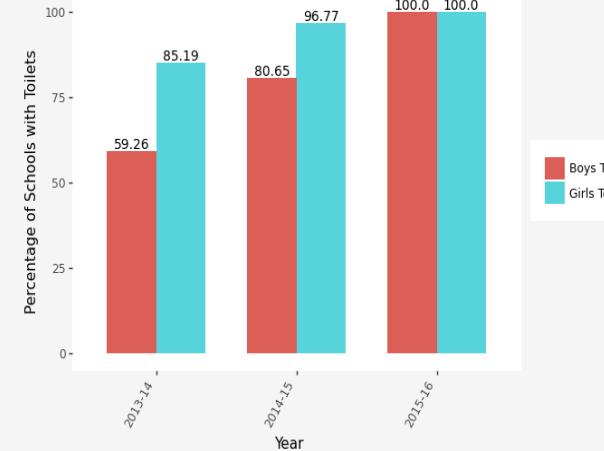
Year-wise Toilet Facilities Analysis for the state Andhra Pradesh for primary only category



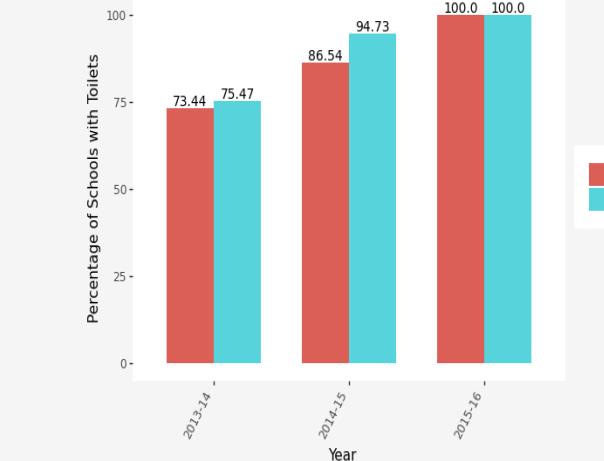
Year-wise Toilet Facilities Analysis for the state Andhra Pradesh for U\_Primary\_Only category



Year-wise Toilet Facilities Analysis for the state Andhra Pradesh for Sec\_Only category



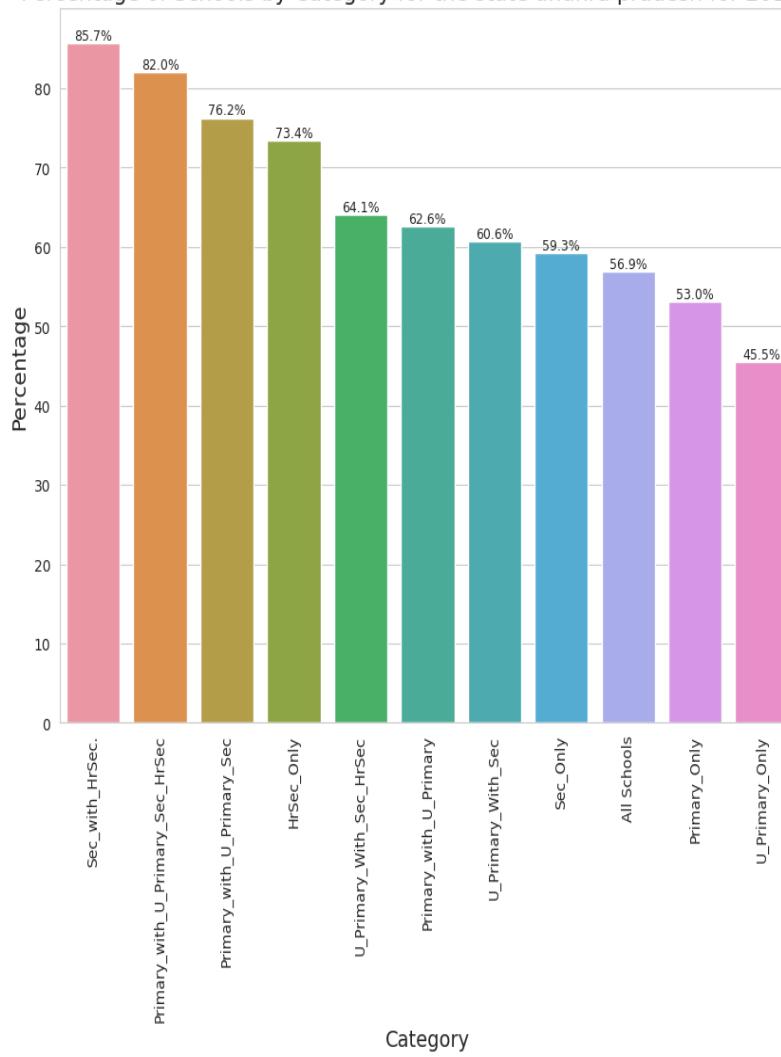
Year-wise Toilet Facilities Analysis for the state Andhra Pradesh for HrSec\_Only category



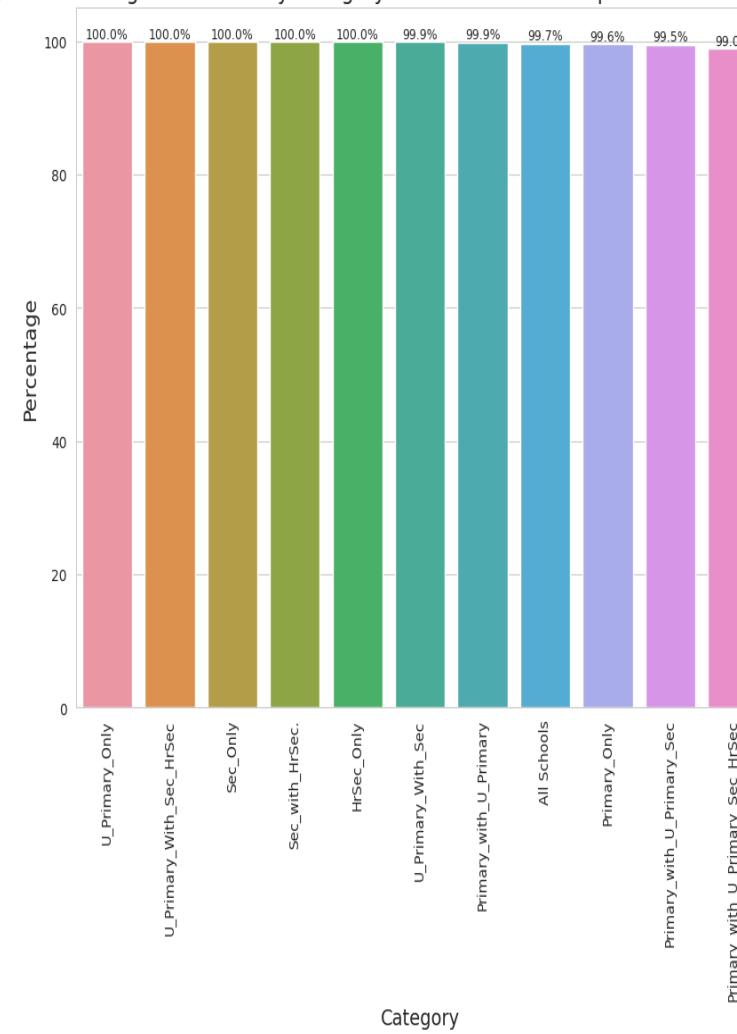
► We can clearly see that there is a good improvement in toilet facilities with year in both genders for the categories of schools .

# Data Visualization And Results :

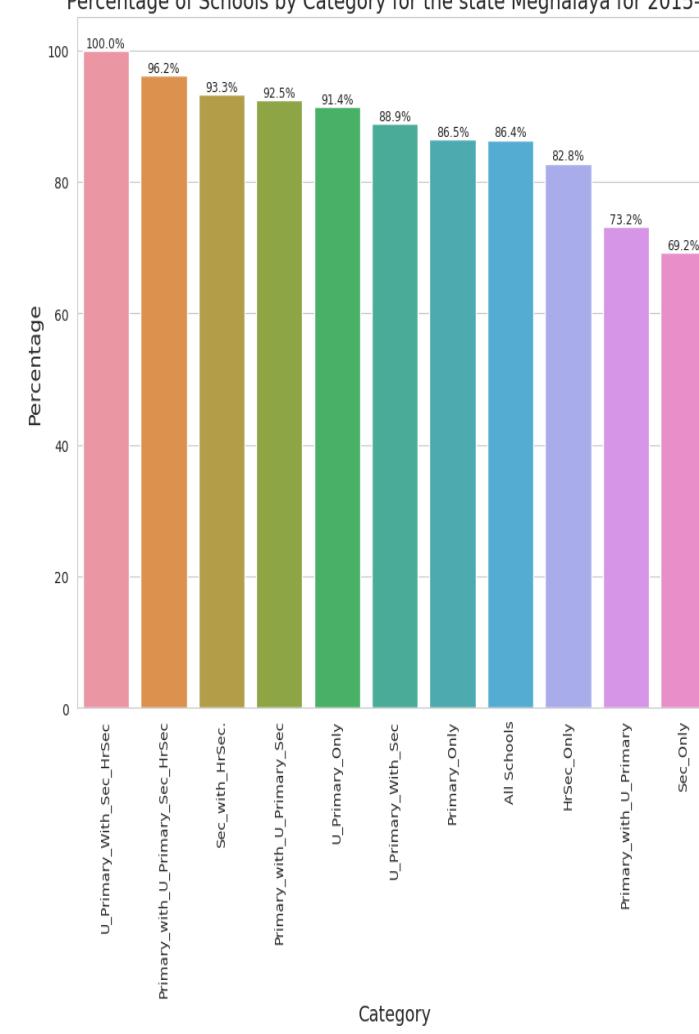
Percentage of Schools by Category for the state andhra pradesh for 2013-14



Percentage of Schools by Category for the state andhra pradesh for 2015-16



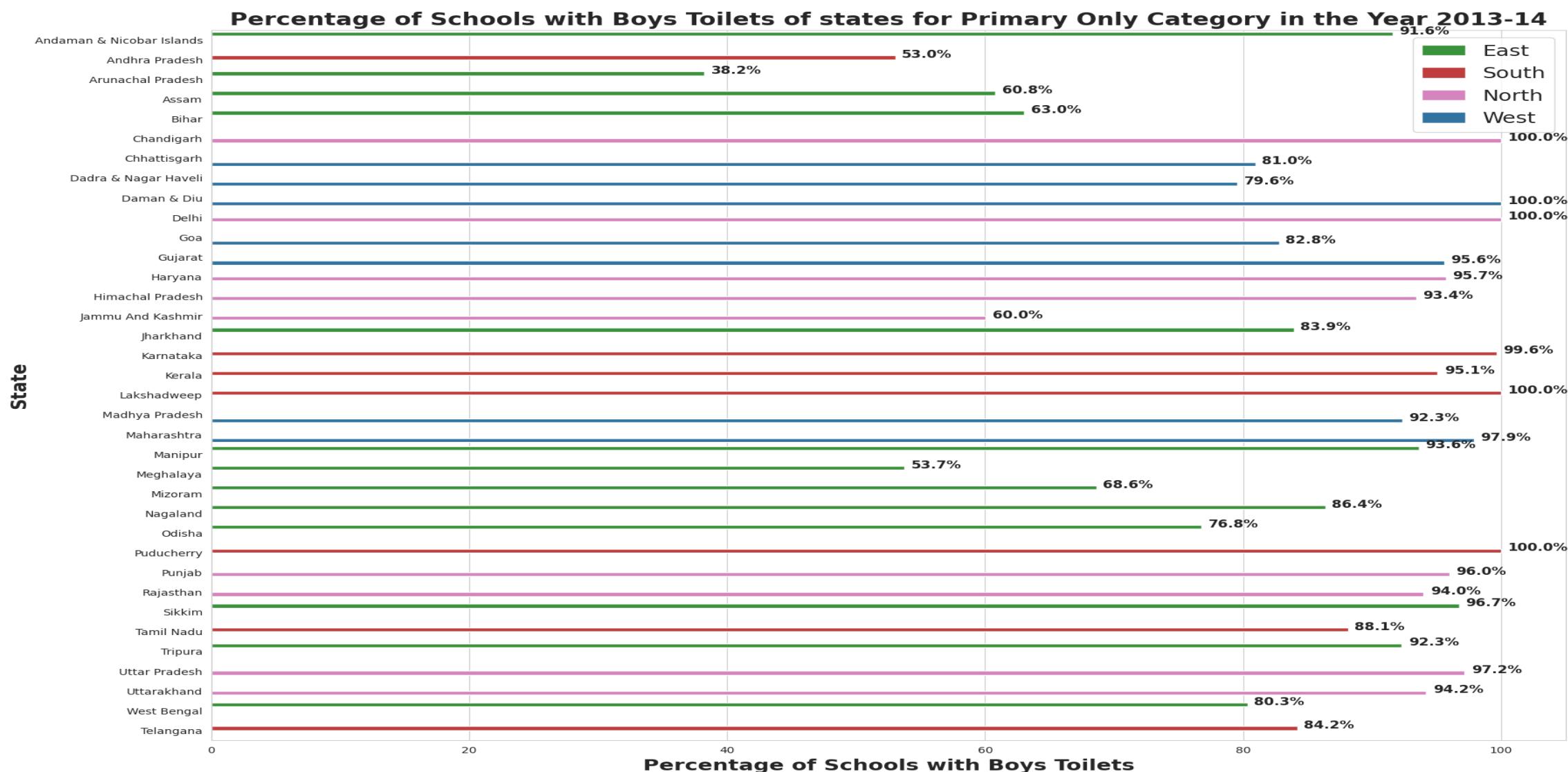
Percentage of Schools by Category for the state Meghalaya for 2015-16



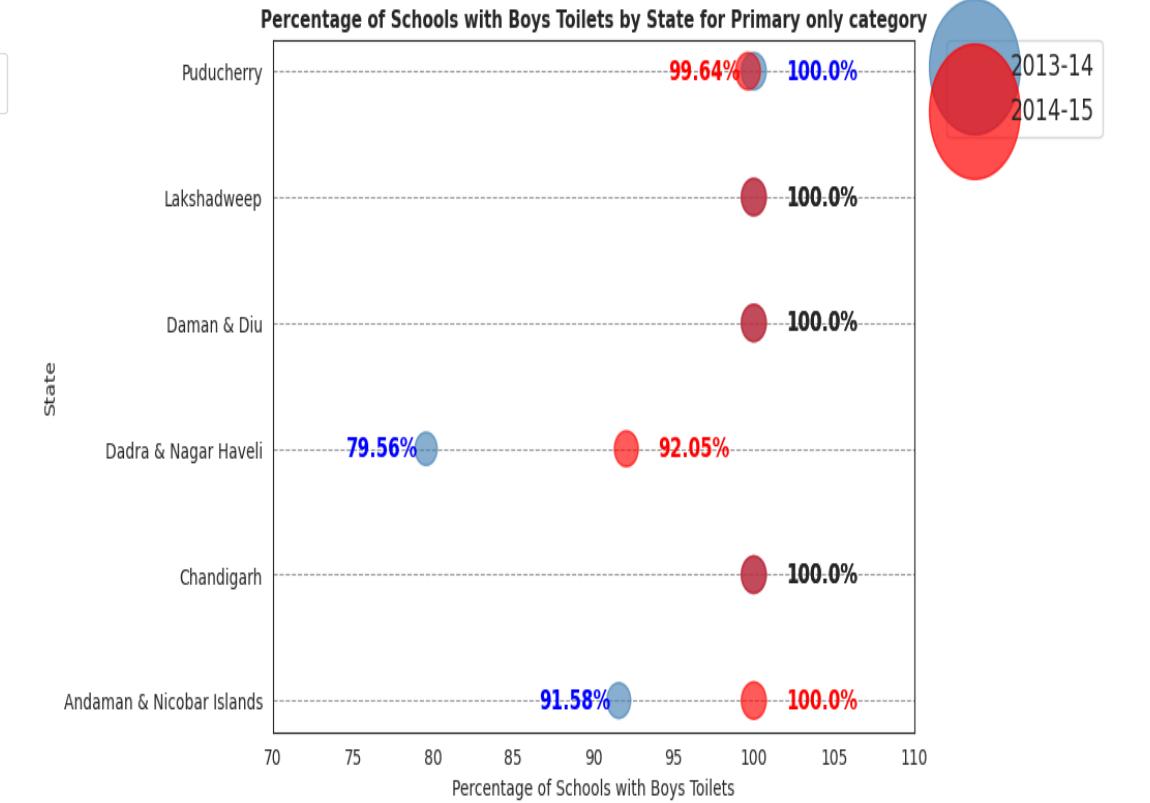
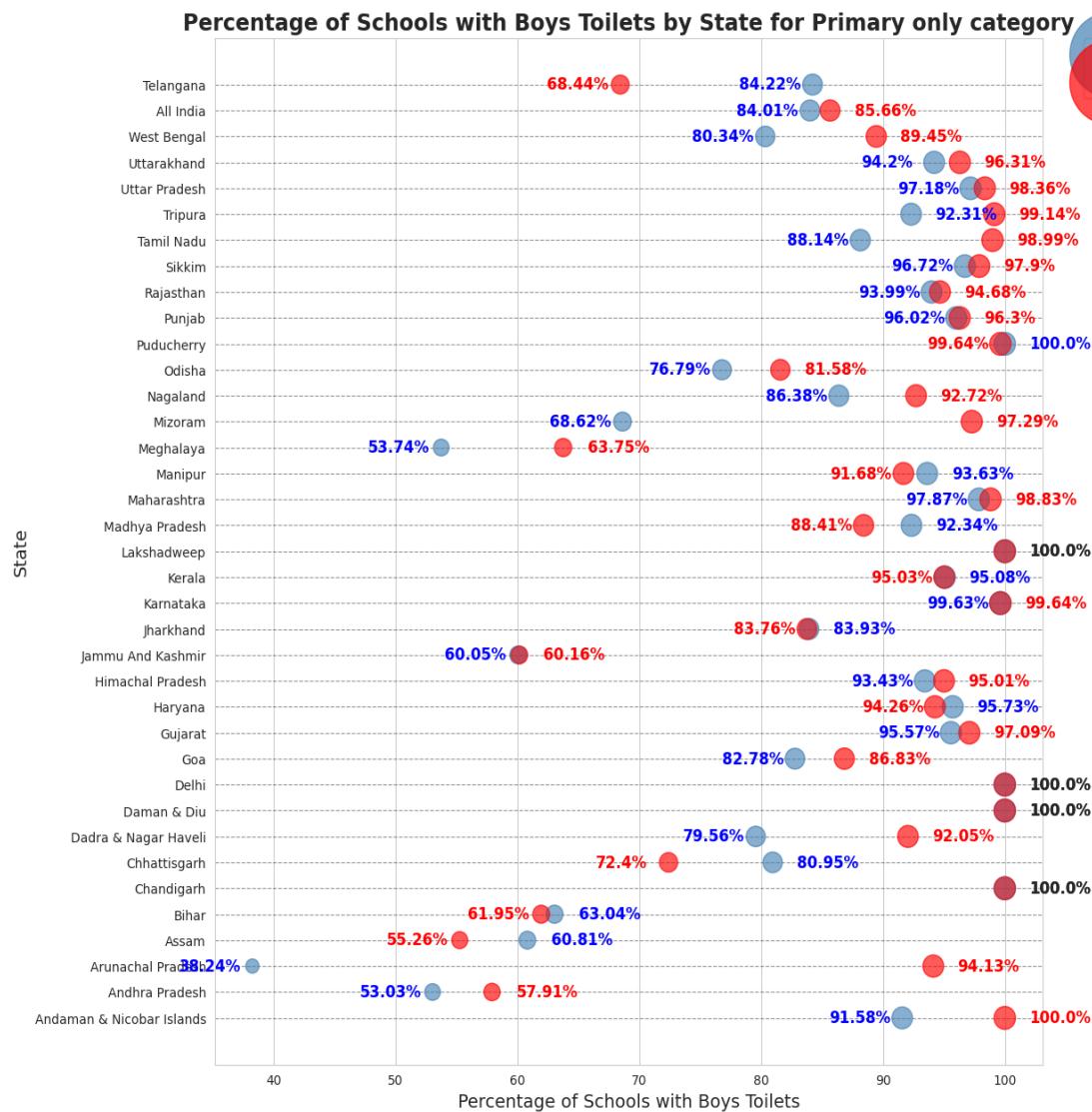
- Here for andhra pradesh we can see there great improvement in all categories and we can compare its facilities with other state meghalaya for the year 2015-16.
- For meghalaya secondary category schools have less toilet facilities .

# Data Visualization And Results :

## Region wise Comparision b/w the states



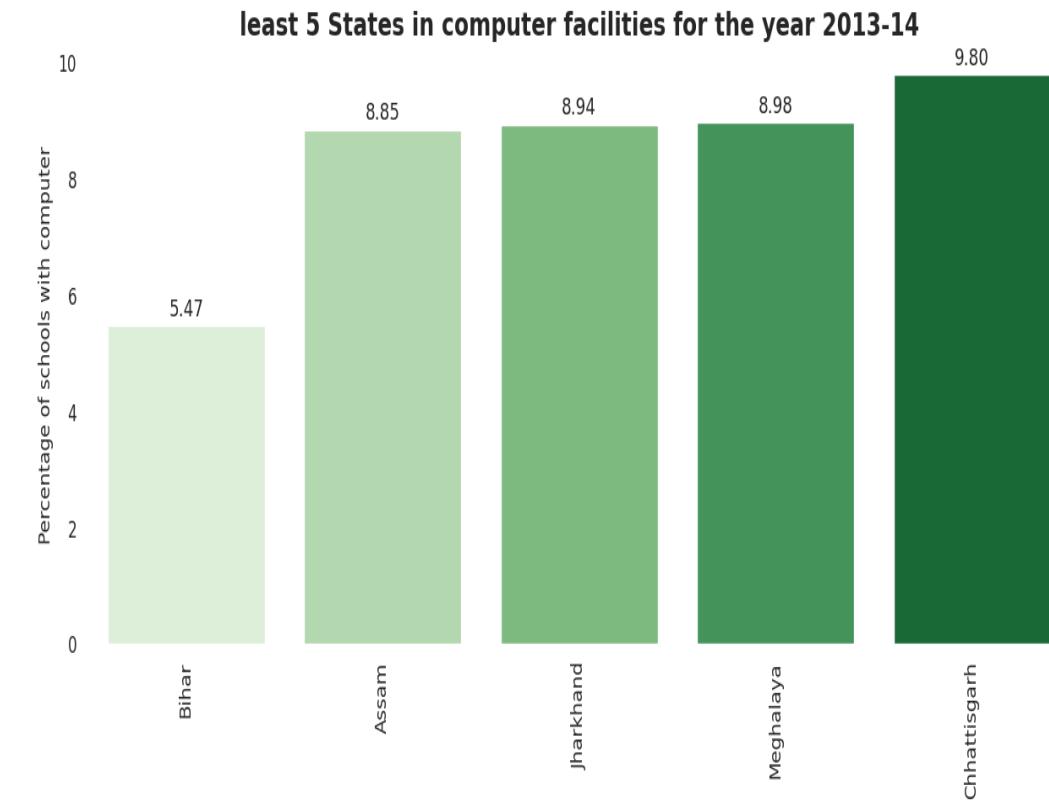
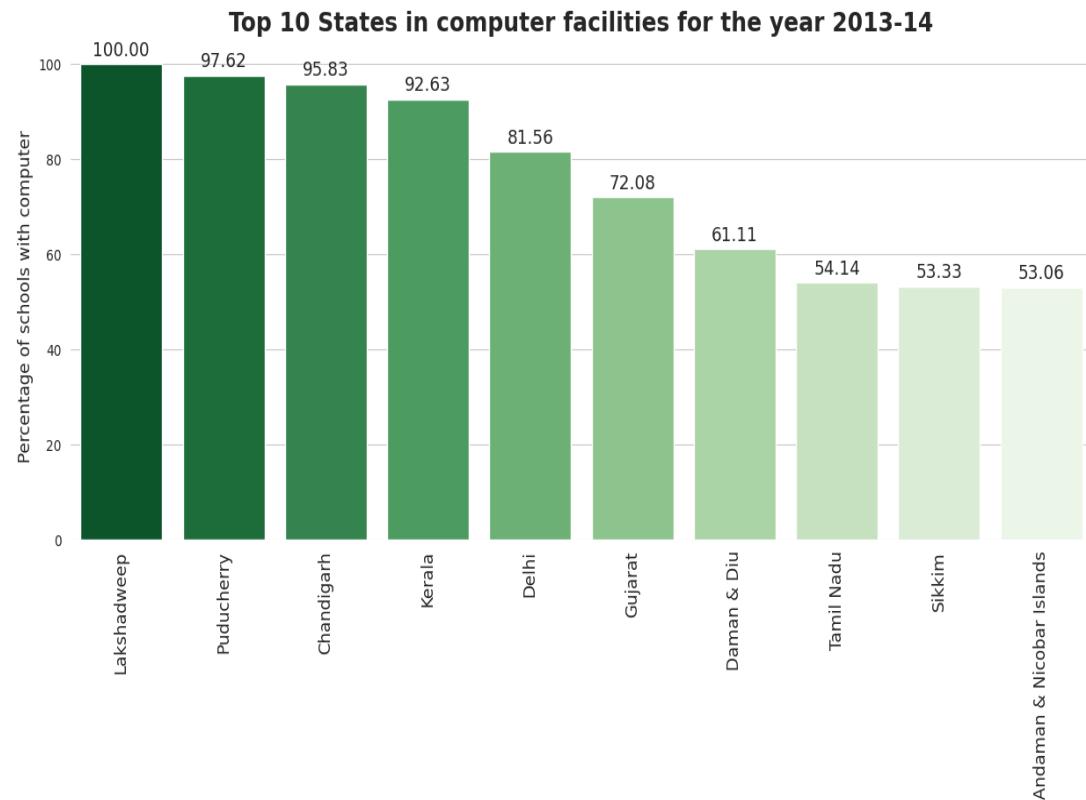
# Data Visualization And Results :



From these plots we can see the change in facility values in both the years 2013 and 2014

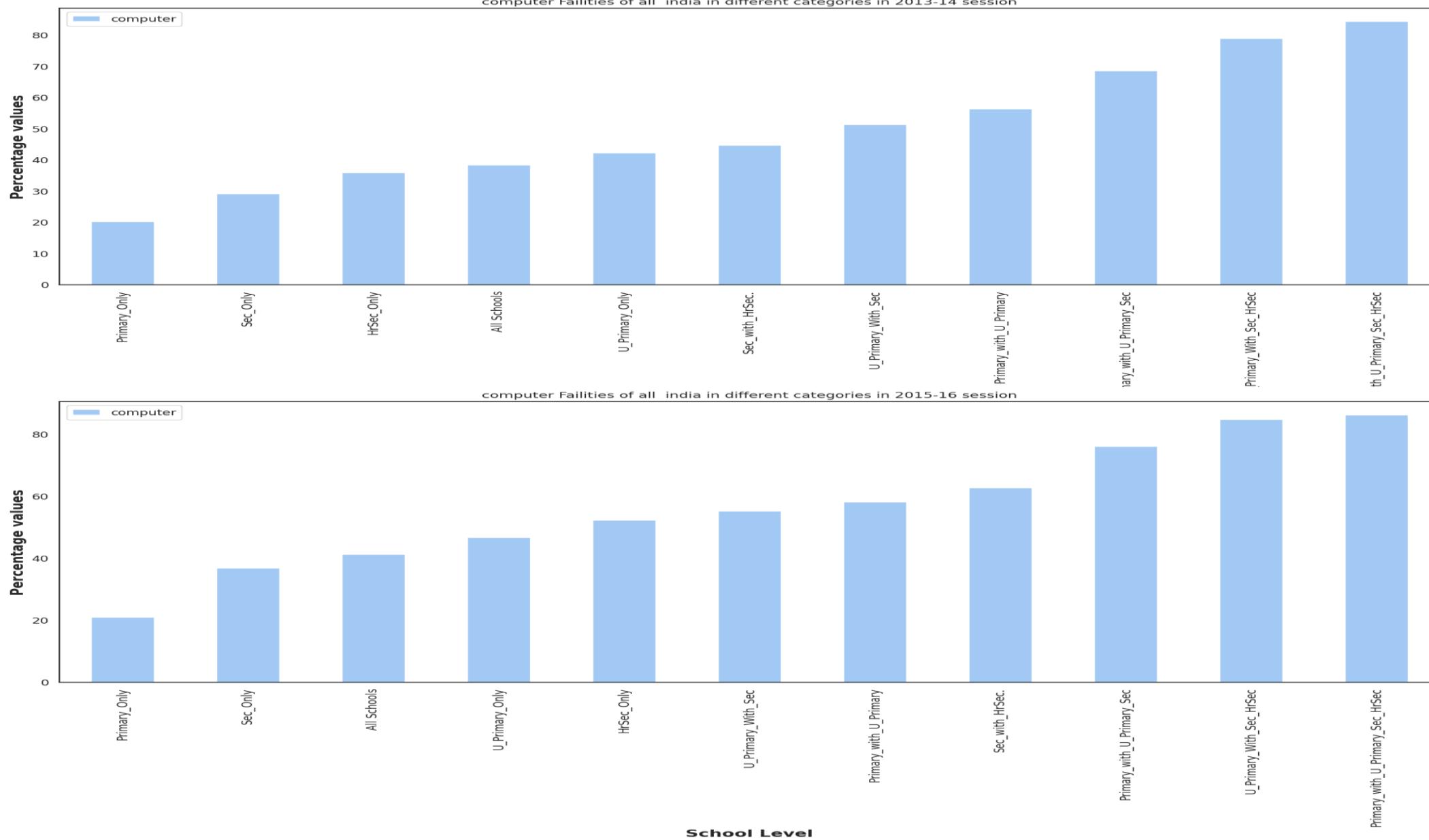
# Data Visualization And Results :

## Let's explore the computer facilities!



► Very poor! computer facilities for above 5 states as compared to top 10 states, we see the effect of having these less values in further analysis of facility score and grouping into Tiers.

# Data Visualization And Results :

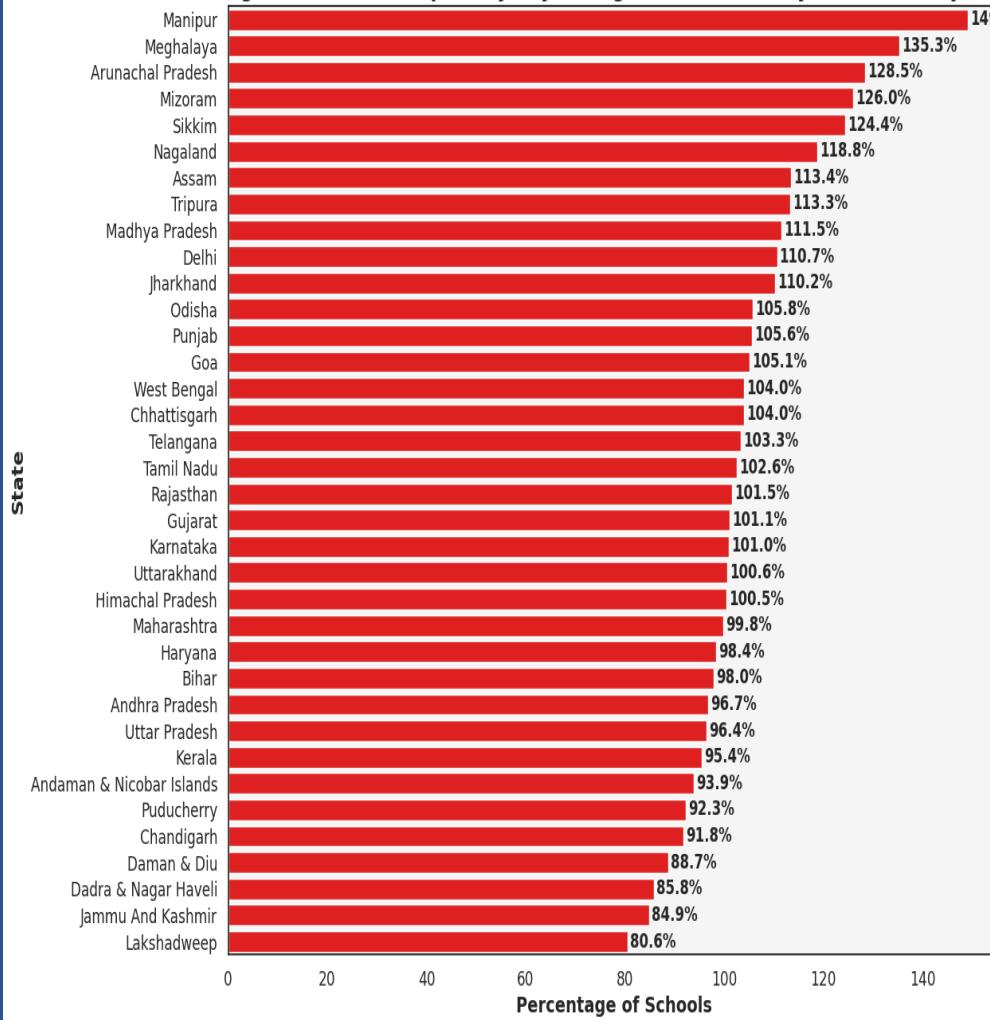


► Here we can notice the facilities increased for Higher secondary category

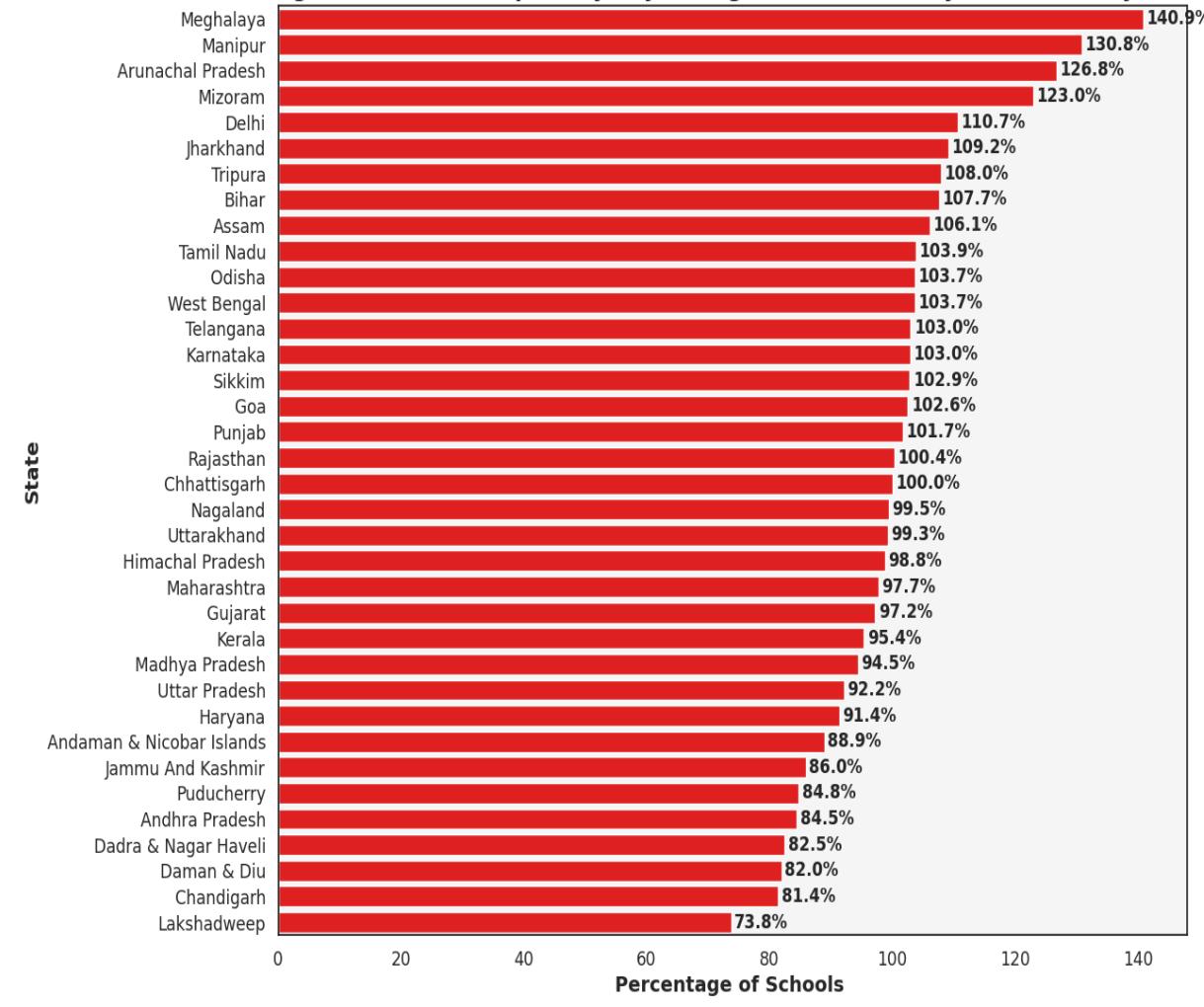
# Data Visualization And Results :

## Let's explore Gross Enrollment data set!

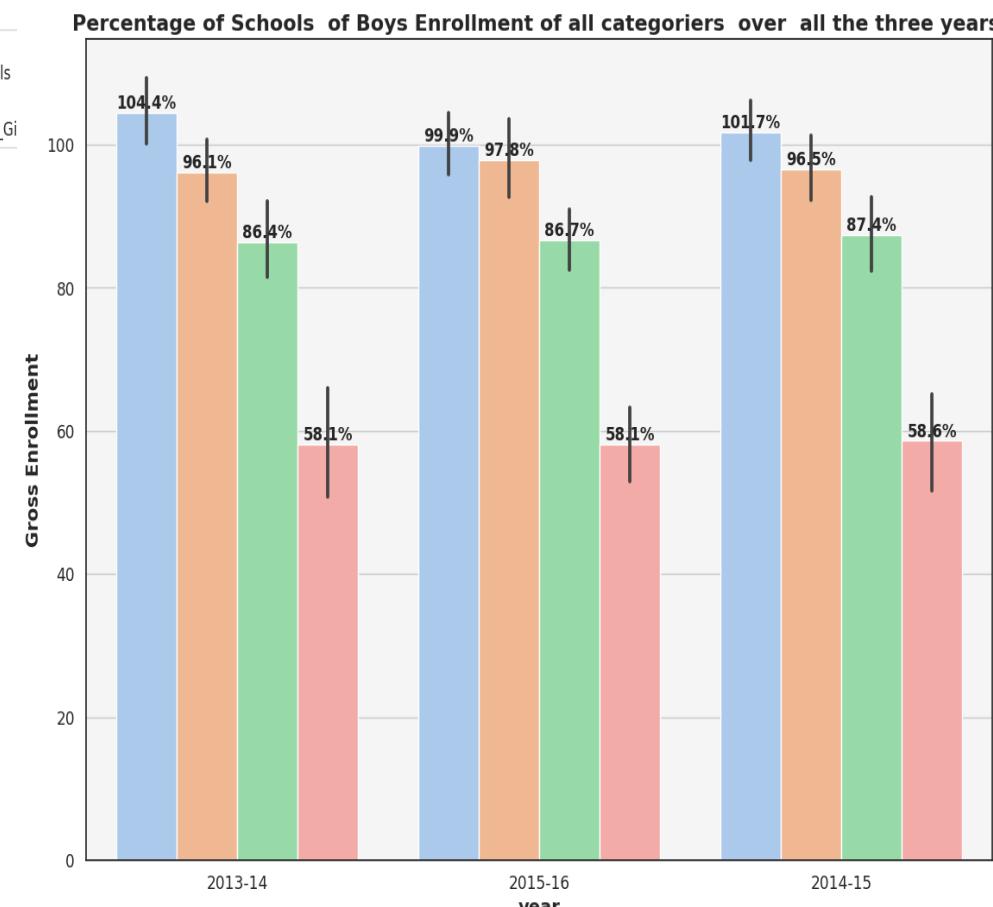
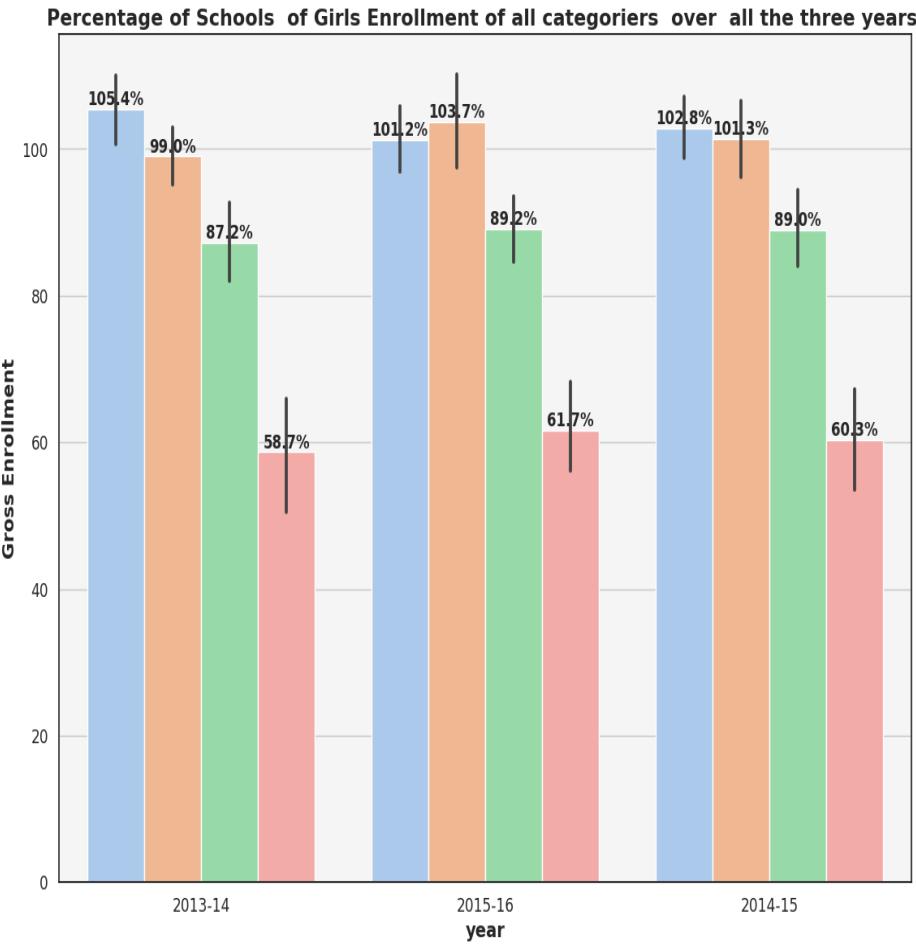
Percentage of Schools with primary Boys and girls enrollment by States in the year 2013-14



Percentage of Schools with primary Boys and girls enrollment by States in the year 2015-16



# Data Visualization And Results :

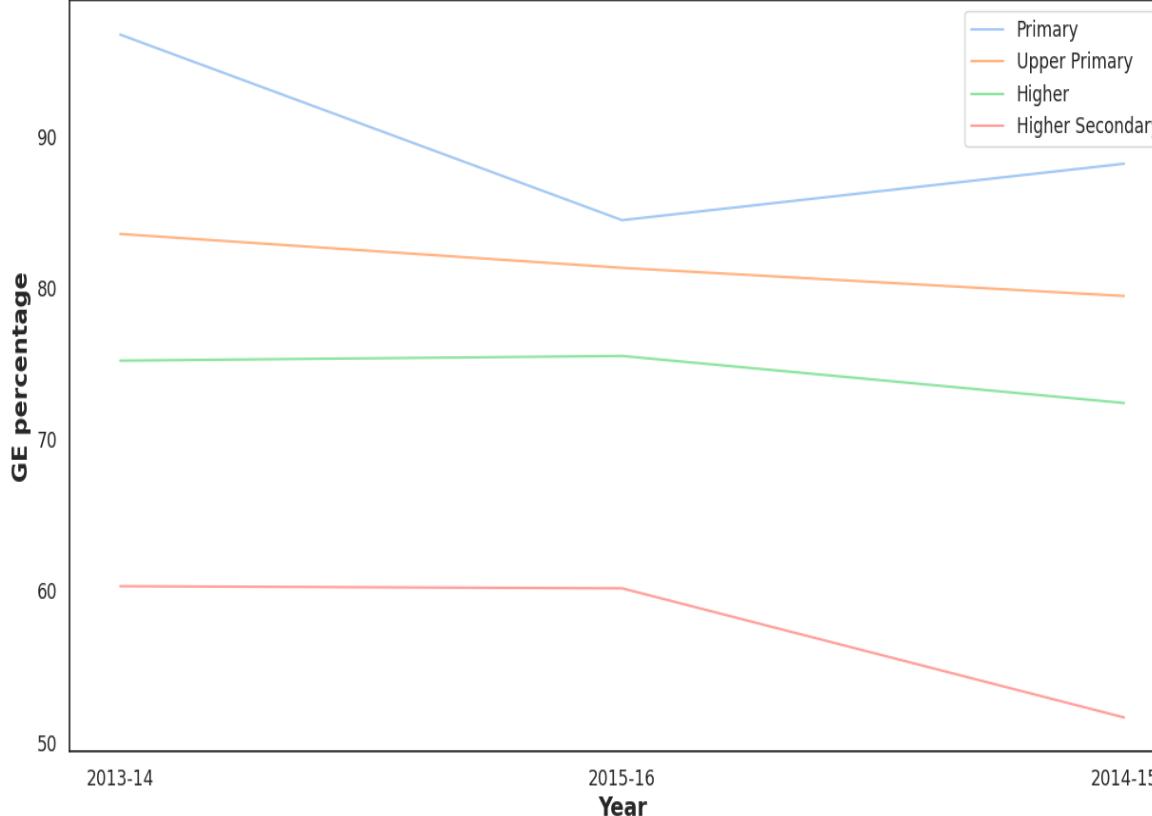


- Here Girls who join Primary School or upper primary have exceeded 100 GER which means the students also include who are repeating a grade, or those who have enrolled late and are older than their classmates, or those who have advanced quickly and are younger than their classmates.
- We can also observe that there's a big difference between Secondary and Higher Secondary Enrollments for both boys and girls . There could be many reasons like Students might go to better colleges in cities/towns for JEE/NEET preparations.
- There's a considerable increase in GEP from 2013,2016 for both girls and boys who are joining Upper Primary, Secondary. This is a very good positive sign.

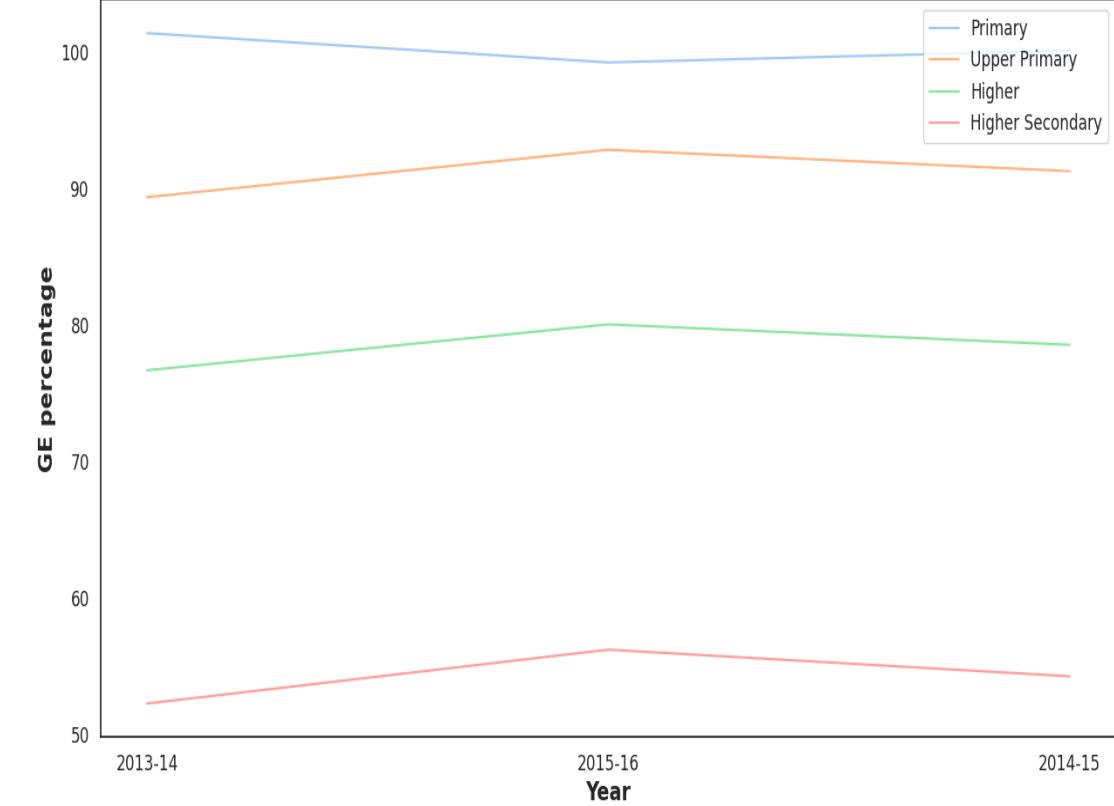
# Data Visualization And Results :

## Let's explore Gross Enrollment data set!

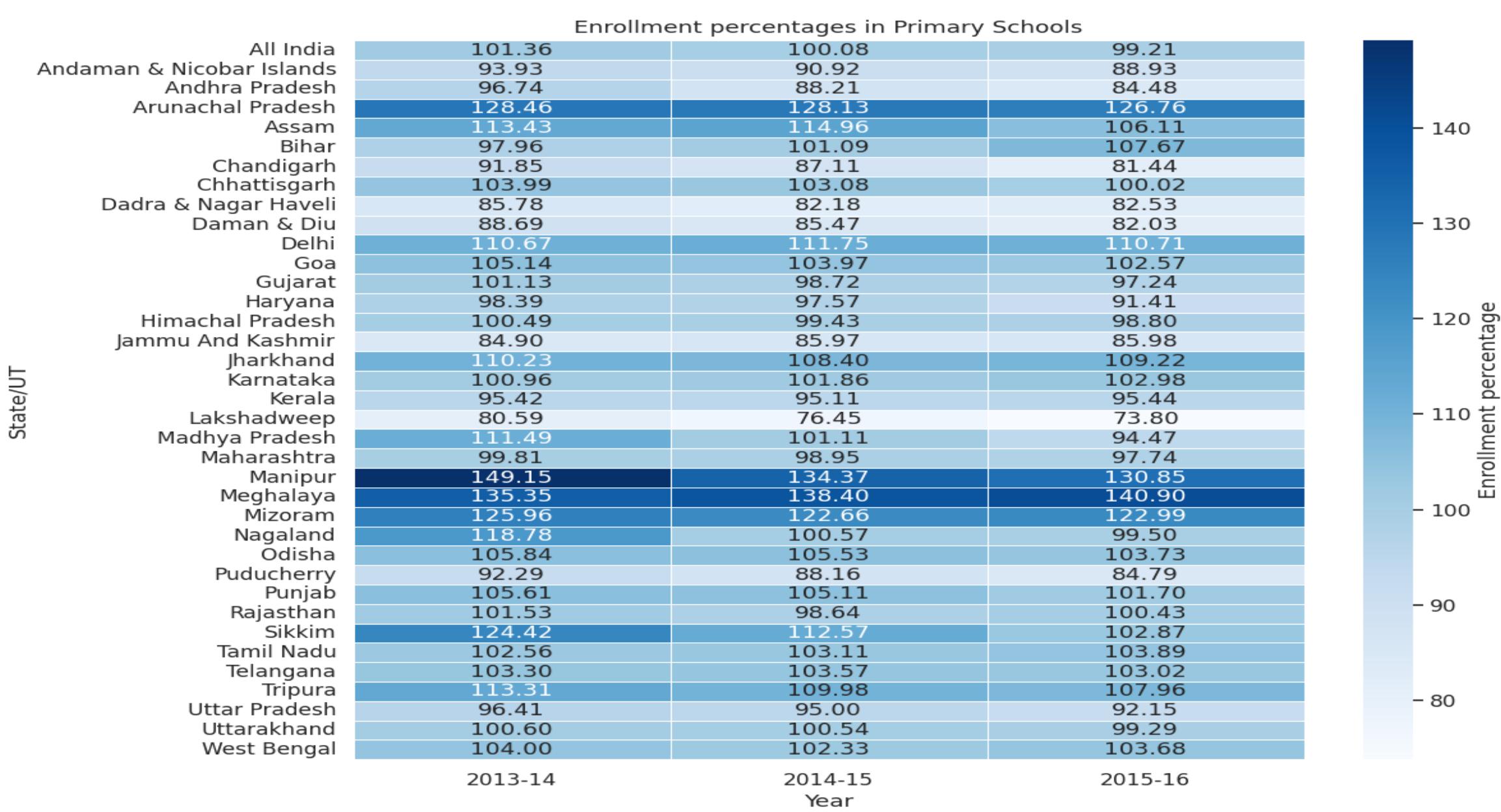
Line plot of Gross Enrollment of Andhra pradesh over three years



Line plot of Gross Enrollment of all india over three years

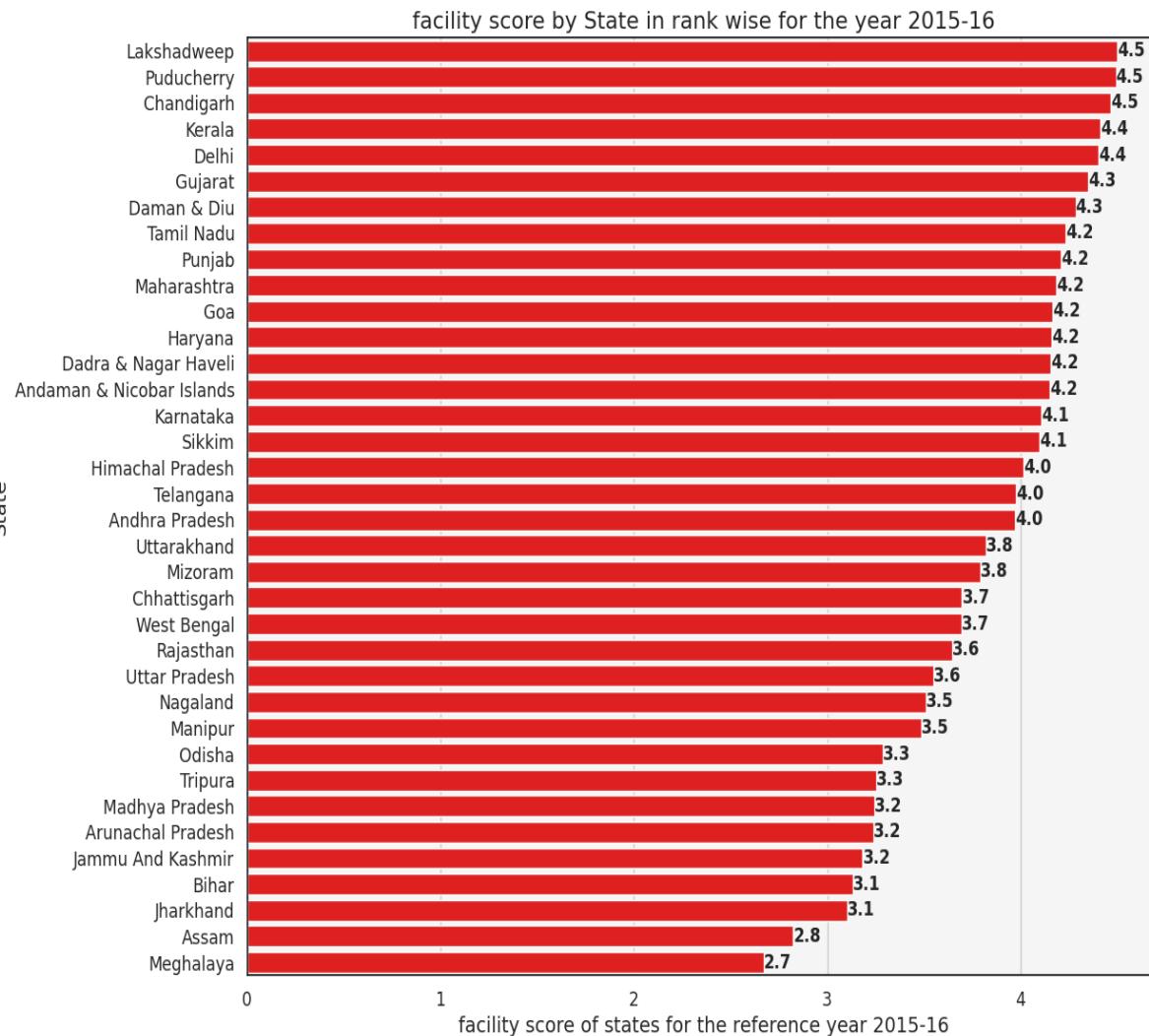
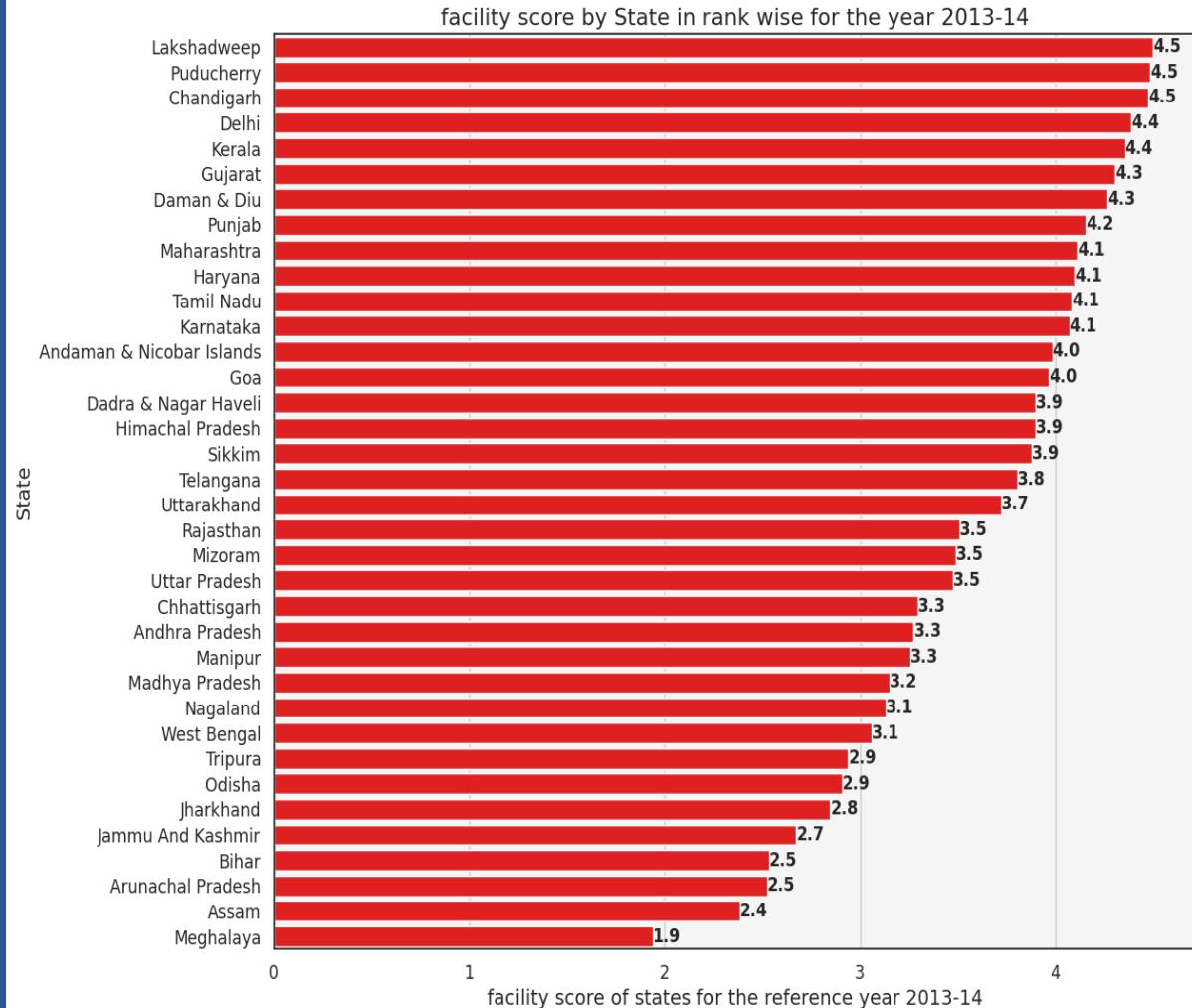


# Data Visualization And Results :



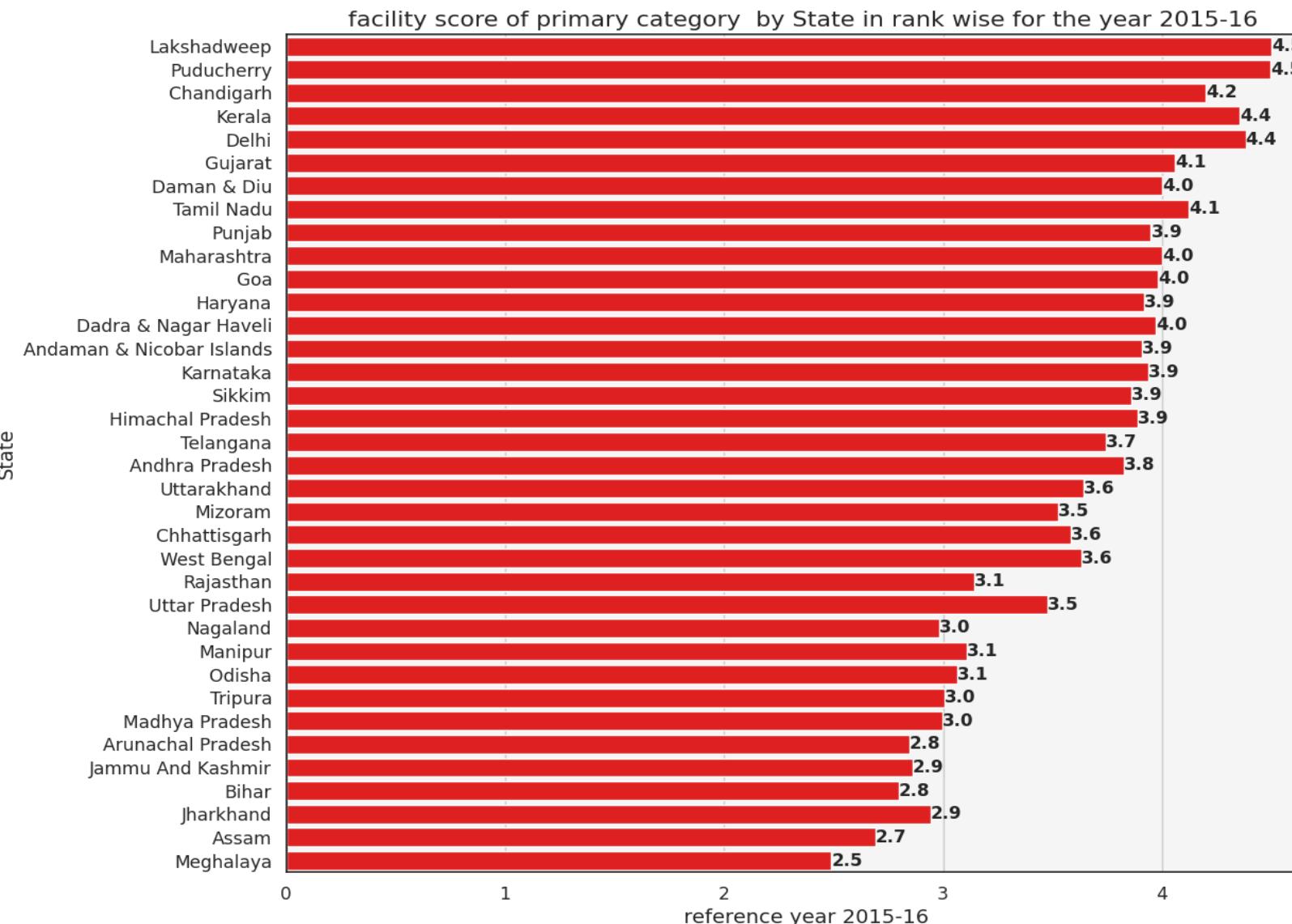
# Data Visualization And Results :

Facility score\_all\_school=  $(\text{All Schools\_girls\_toilets}/100)*1 + (\text{All Schools\_boys\_toilets}/100)*1 + (\text{All Schools\_water}/100)*0.9 + (\text{All Schools\_computer}/100)*0.6 + (\text{All Schools\_electricity}/100)*0.8$



# Data Visualization And Results :

►Similarly for facility\_score\_primary , we take primary category values in calculation.



►we can see that they are almost similar to that of facility score of All schools , we come across about this thing at starting that the primary\_only and allschools category are strongly positive correlated.

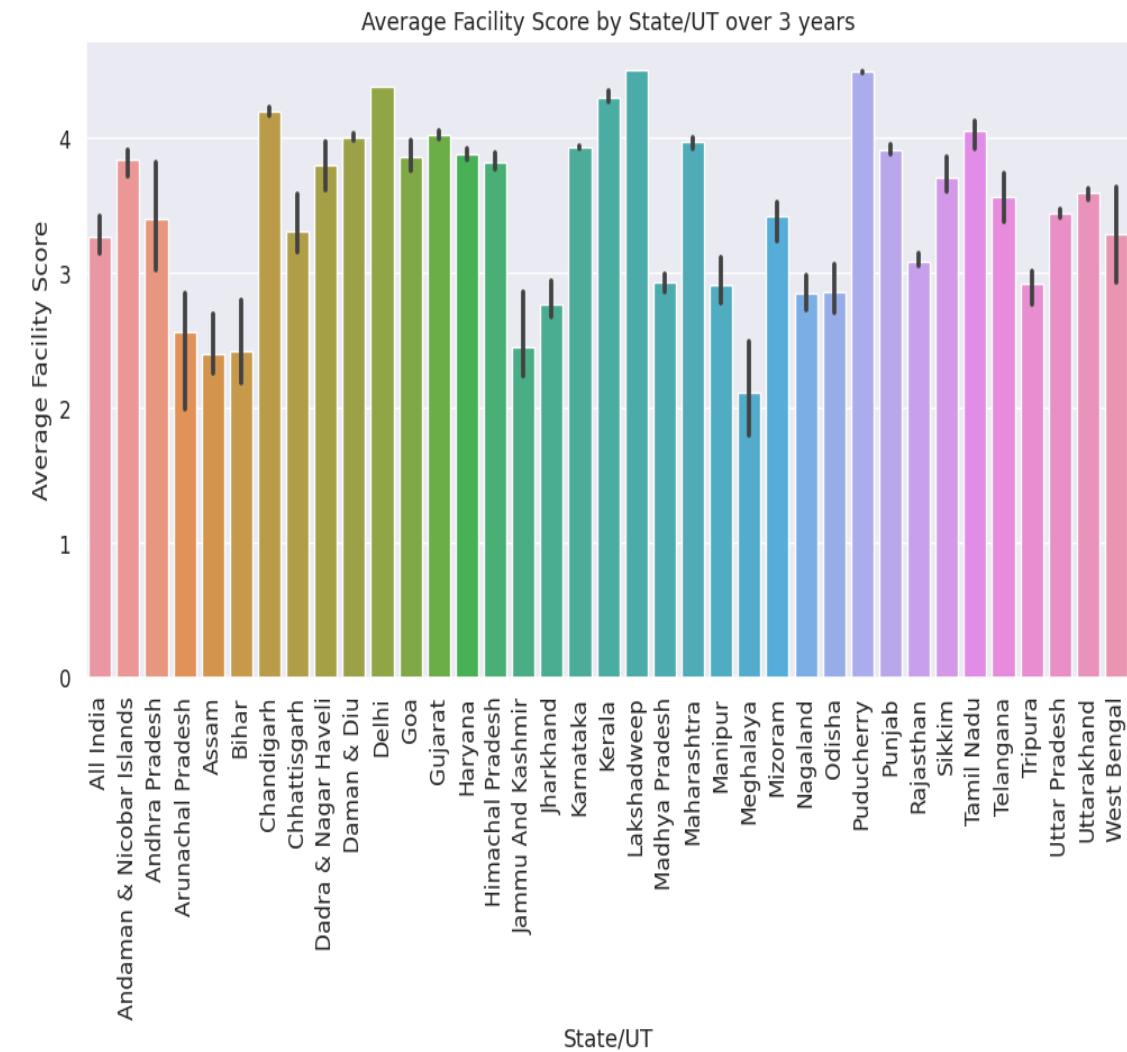
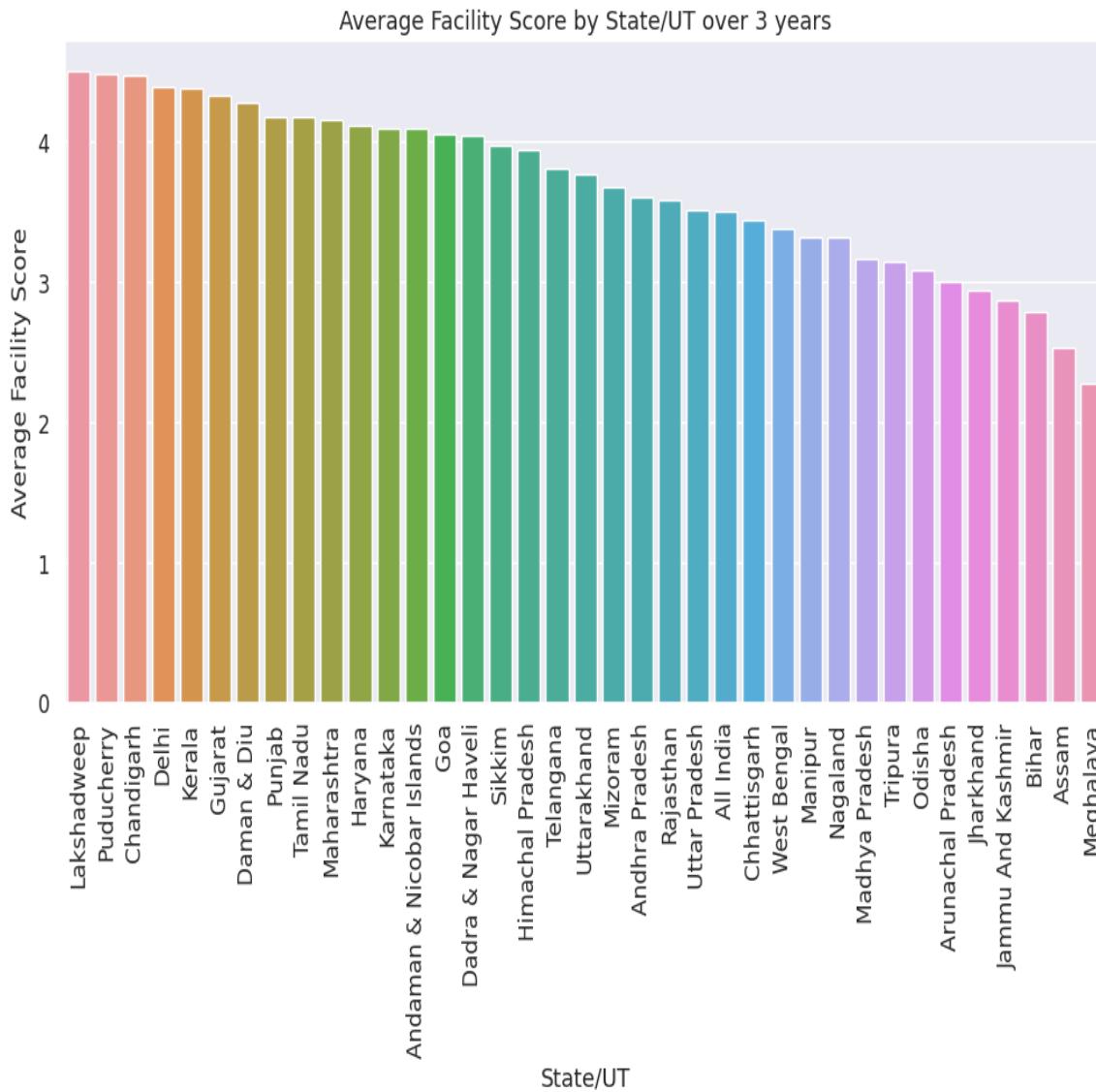
# Data Visualization And Results :

► we compare facilityscore values of states for three years to see the performance of states



# Data Visualization And Results :

calculate the avg of three year facility score of states and sort and plot them

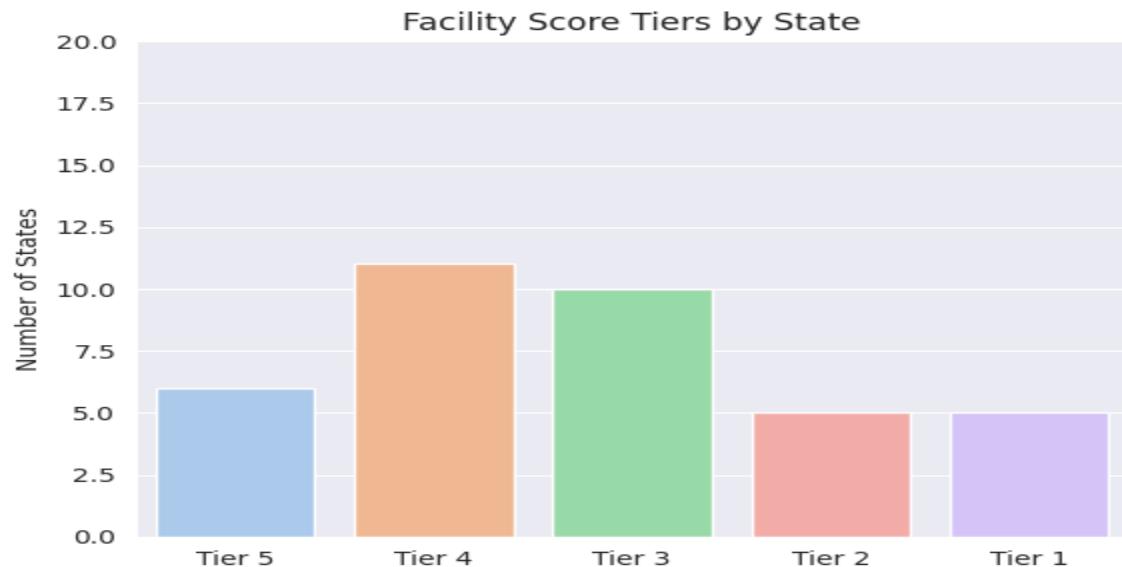


# Data Visualization And Results :

Sorted according to avg\_facility\_score(i.e for All school catg.)

	State_UT	facility_score
19	Lakshadweep	4.500000
27	Puducherry	4.486533
6	Chandigarh	4.471580
10	Delhi	4.392900
18	Kerala	4.383383
12	Gujarat	4.328410
9	Daman & Diu	4.282920
28	Punjab	4.181700
31	Tamil Nadu	4.177963
21	Maharashtra	4.155913
13	Haryana	4.120547
17	Karnataka	4.093347
1	Andaman & Nicobar Islands	4.092070
11	Goa	4.059633
8	Dadra & Nagar Haveli	4.042473
30	Sikkim	3.969437
14	Himachal Pradesh	3.942827
32	Telangana	3.807515
35	Uttarakhand	3.770027
24	Mizoram	3.680570
2	Andhra Pradesh	3.607807
29	Rajasthan	3.582853
34	Uttar Pradesh	3.514887
0	All India	3.502287
7	Chhattisgarh	3.439987
36	West Bengal	3.381407
22	Manipur	3.320720
25	Nagaland	3.317107
20	Madhya Pradesh	3.163340
33	Tripura	3.142357
26	Odisha	3.078847
3	Arunachal Pradesh	2.998777
16	Jharkhand	2.942333
15	Jammu And Kashmir	2.867643
5	Bihar	2.791473
4	Assam	2.533300
23	Meghalaya	2.278837

Histogram of Division of states into tiers



**Tier 1 States:** ['Chandigarh' 'Delhi' 'Kerala' 'Lakshadweep' 'Puducherry']

**Tier 2 States:** ['Daman & Diu' 'Gujarat' 'Maharashtra' 'Punjab' 'Tamil Nadu']

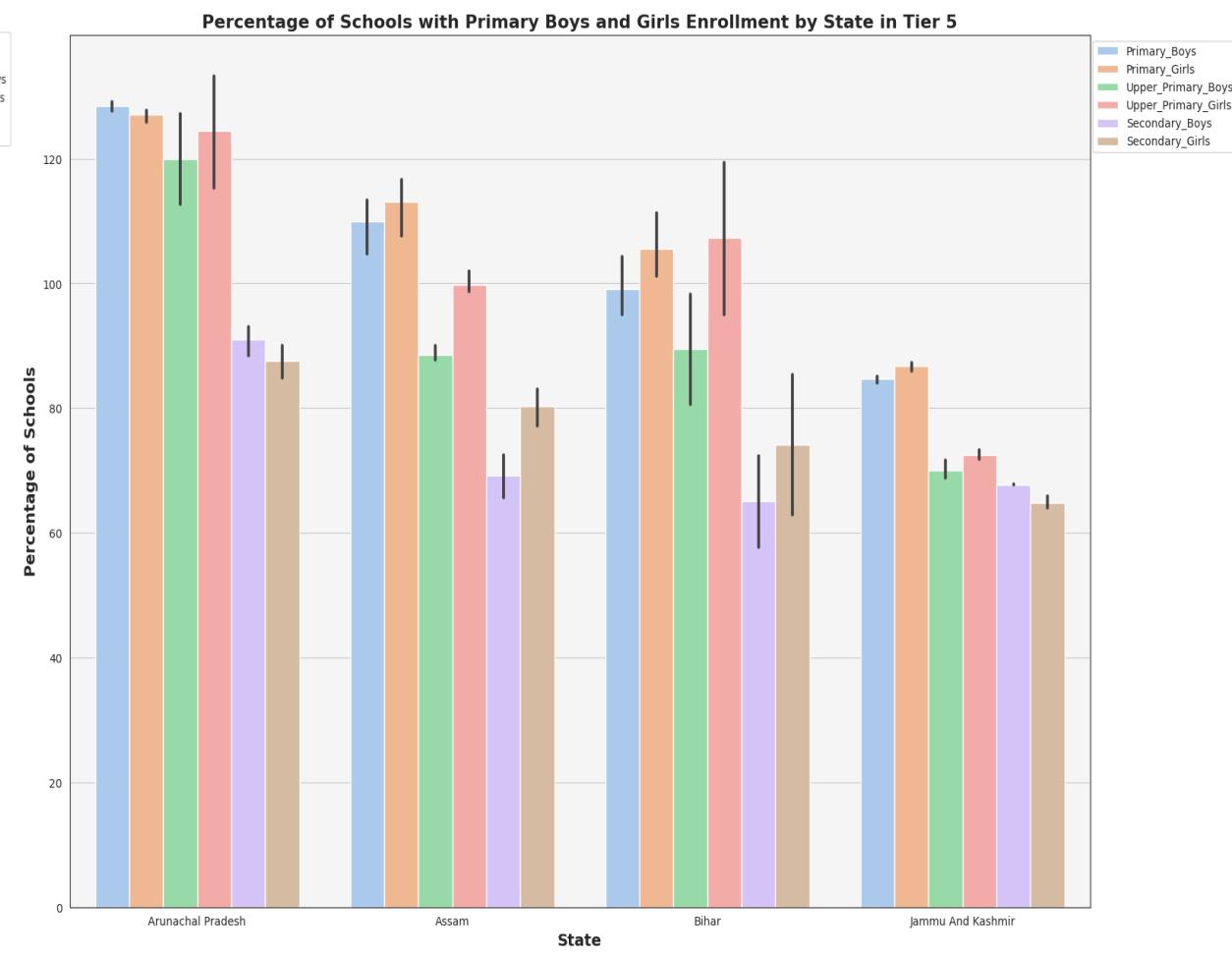
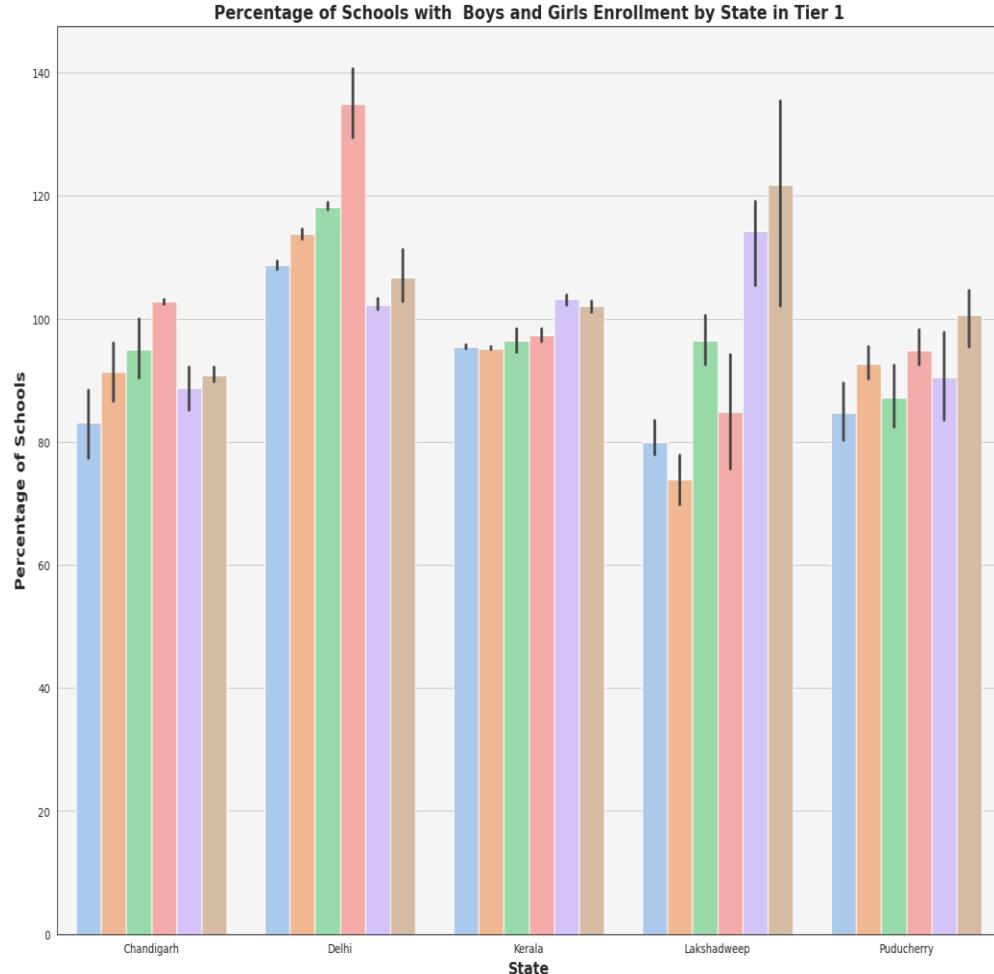
**Tier 3 States:** ['Andaman & Nicobar Islands' 'Dadra & Nagar Haveli' 'Goa' 'Haryana' 'Himachal Pradesh' 'Karnataka' 'Mizoram' 'Sikkim' 'Telangana' 'Uttarakhand']

**Tier 4 States:** ['All India' 'Andhra Pradesh' 'Chhattisgarh' 'Madhya Pradesh' 'Manipur' 'Nagaland' 'Odisha' 'Rajasthan' 'Tripura' 'Uttar Pradesh' 'West Bengal']

**Tier 5 States:** ['Arunachal Pradesh' 'Assam' 'Bihar' 'Jammu And Kashmir' 'Jharkhand' 'Meghalaya']

# Data Visualization And Results :

►Now we can do plots for each tier separately for having a better comparision between states, How do you stand out compared to those at your level ?

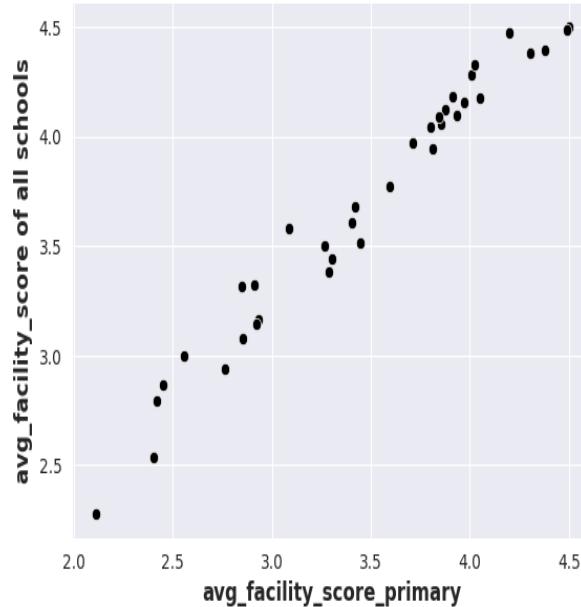


# Data Visualization And Results :

► now we add enrollment and check its relation with fac score of states. As enrollment file have for each category so we do average of all category percentage and get a mean percentage value for each state in a year.

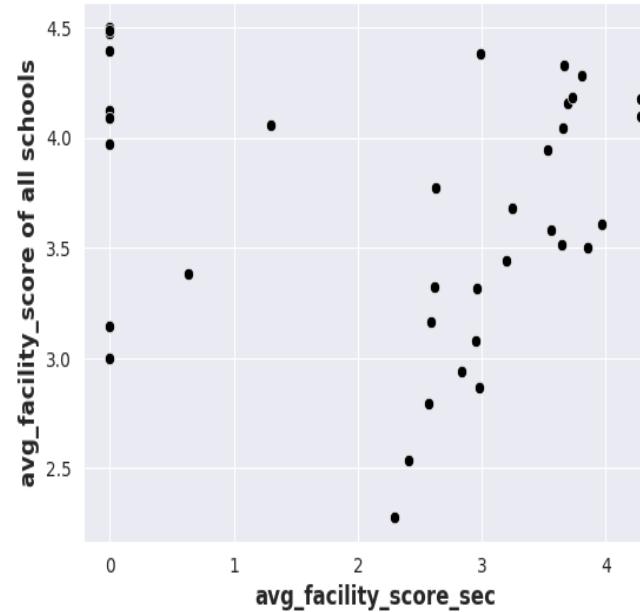
Column names of my latest data frame are: 'State\_UT', 'Primary\_Boys\_mean', 'Primary\_Girls\_mean', 'Primary\_Total\_mean', 'Upper\_Primary\_Boys\_mean', 'Upper\_Primary\_Girls\_mean', 'Upper\_Primary\_Total\_mean', 'Secondary\_Boys\_mean', 'Secondary\_Girls\_mean', 'Secondary\_Total\_mean', 'Higher\_Secondary\_Boys\_mean', 'Higher\_Secondary\_Girls\_mean', 'Higher\_Secondary\_Total\_mean', 'All\_Categories\_Mean', 'avg\_facility\_score', 'avg\_facility\_score\_primary', 'avg\_facility\_score\_upper\_primary', 'avg\_facility\_score\_sec', 'avg\_facility\_score\_Hrsec', 'tier'

Scatter plot between average facility score of primary and all schools



Correlation Value:  
0.983

Scatter plot between average facility score of secondary and all schools



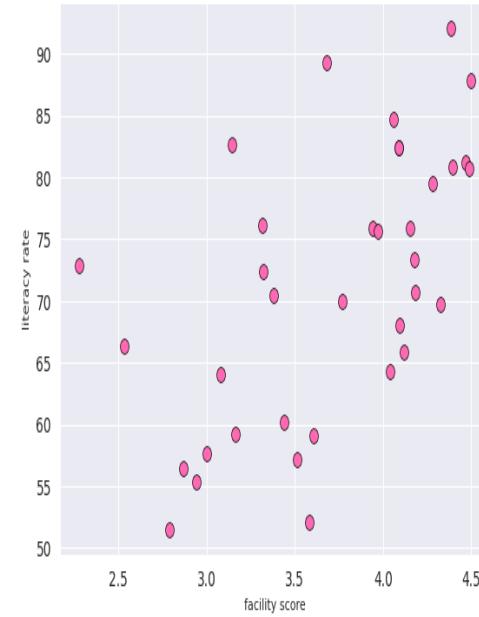
Correlation Value:  
-0.133

# Data Visualization And Results :

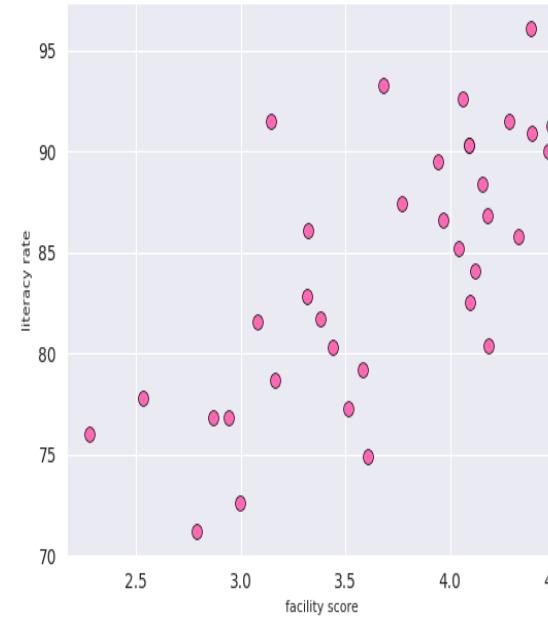
**"link our latest dataframe with literacy rate!"**

Column names of latest data set: 'State\_UT', 'Primary\_Boys\_mean', 'Primary\_Girls\_mean', 'Primary\_Total\_mean', 'Upper\_Primary\_Boys\_mean', 'Upper\_Primary\_Girls\_mean', 'Upper\_Primary\_Total\_mean', 'Secondary\_Boys\_mean', 'Secondary\_Girls\_mean', 'Secondary\_Total\_mean', 'Higher\_Secondary\_Boys\_mean', 'Higher\_Secondary\_Girls\_mean', 'Higher\_Secondary\_Total\_mean', 'All\_Categories\_Mean', 'avg\_facility\_score', 'avg\_facility\_score\_primary', 'avg\_facility\_score\_upper\_primary', 'avg\_facility\_score\_sec', 'avg\_facility\_score\_Hrsec', 'tier', 'Literacy Rate in % (Person)', 'Literacy Rate in % (Male)', 'Literacy Rate in % (Female)'

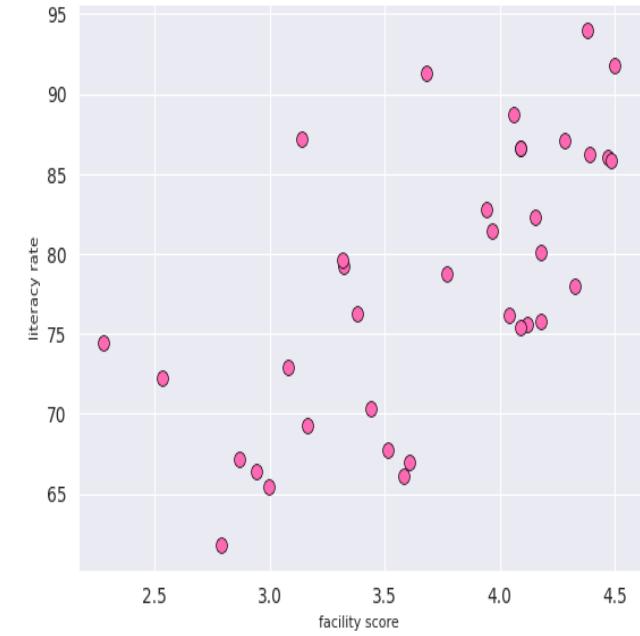
scatter plot of literacy rate of female vs facility score of the states



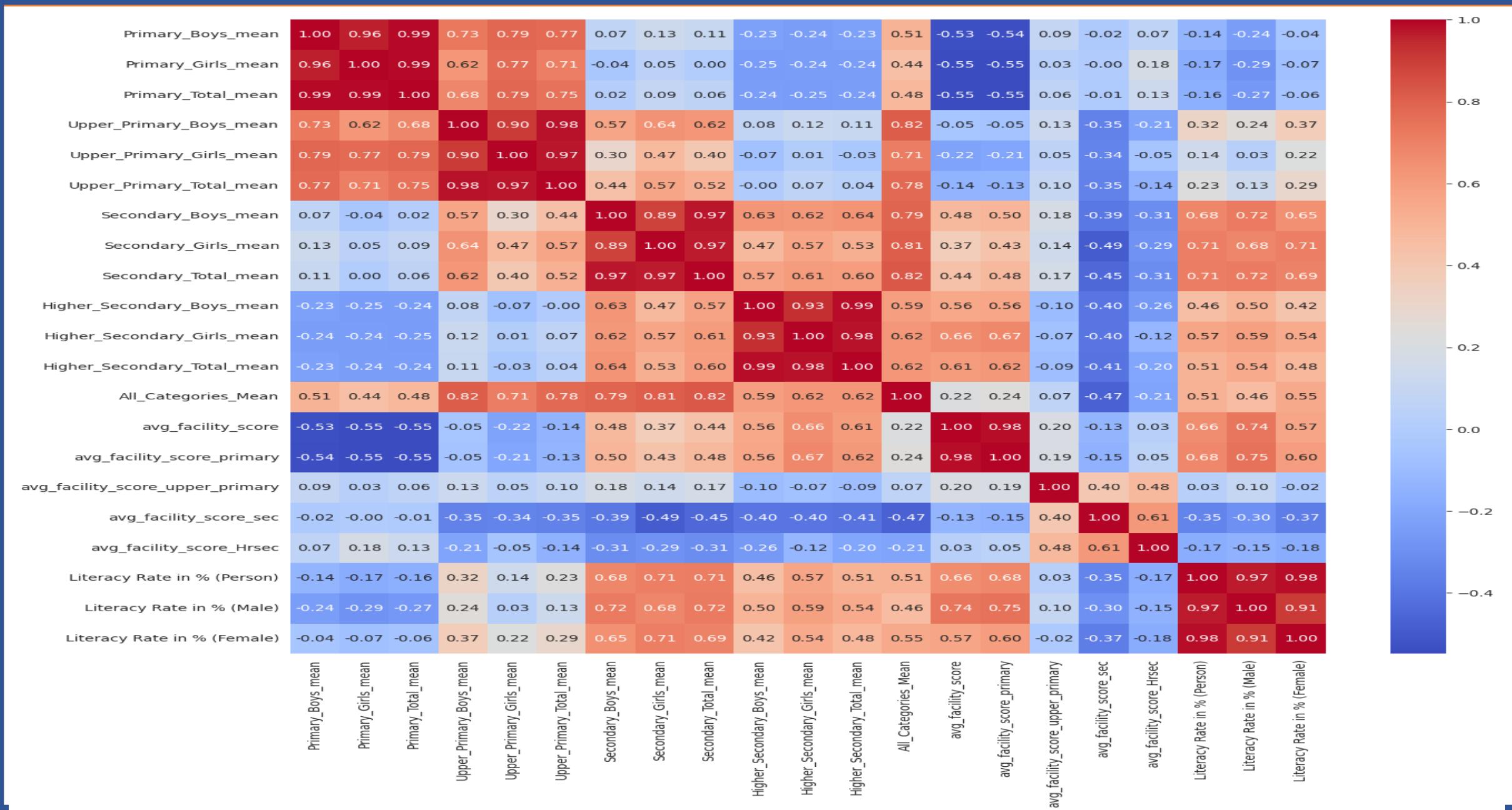
scatter plot of literacy rate of male vs facility score of the states



scatter plot of literacy rate vs facility score of the states

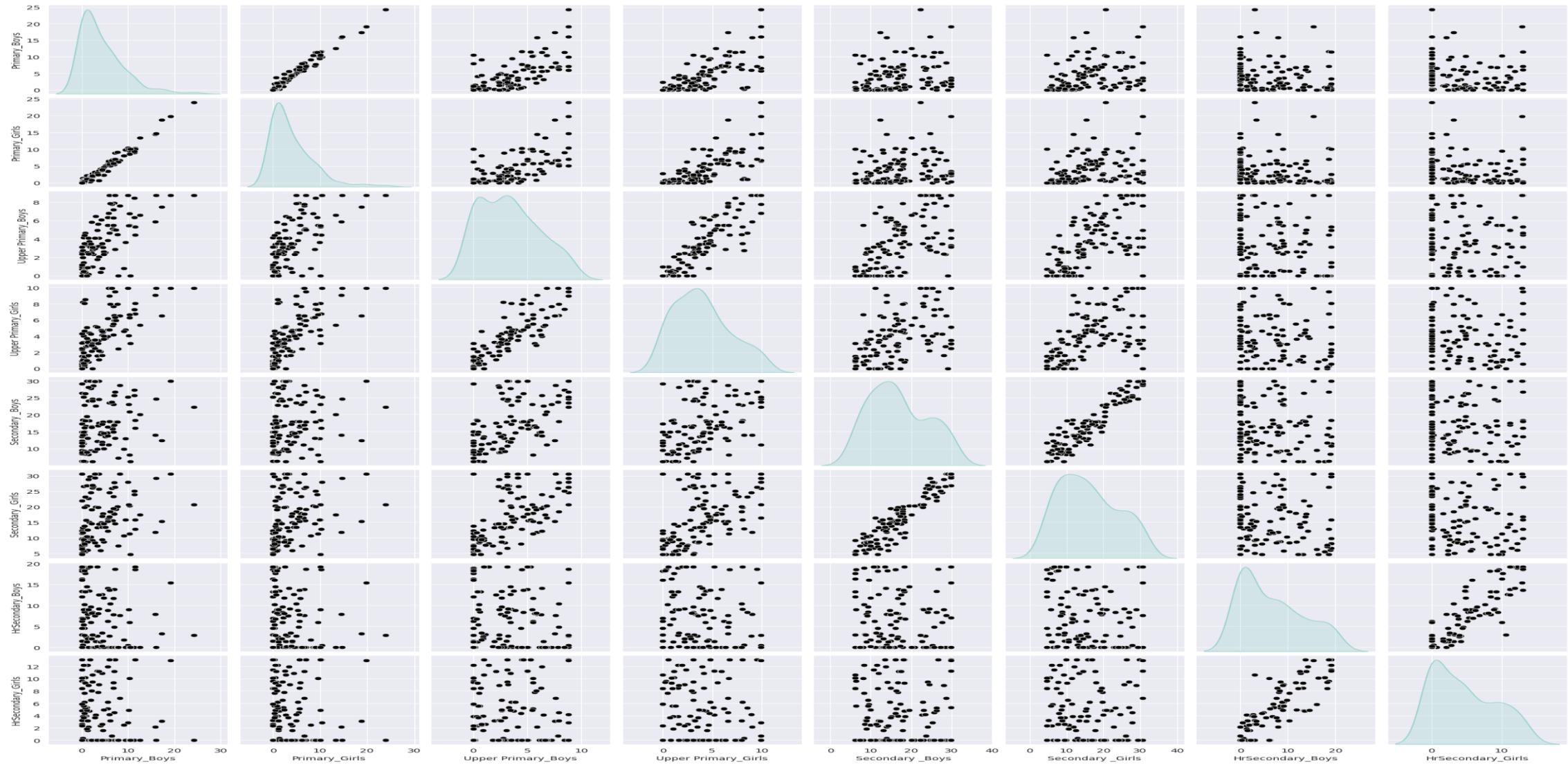


# Data Visualization And Results :

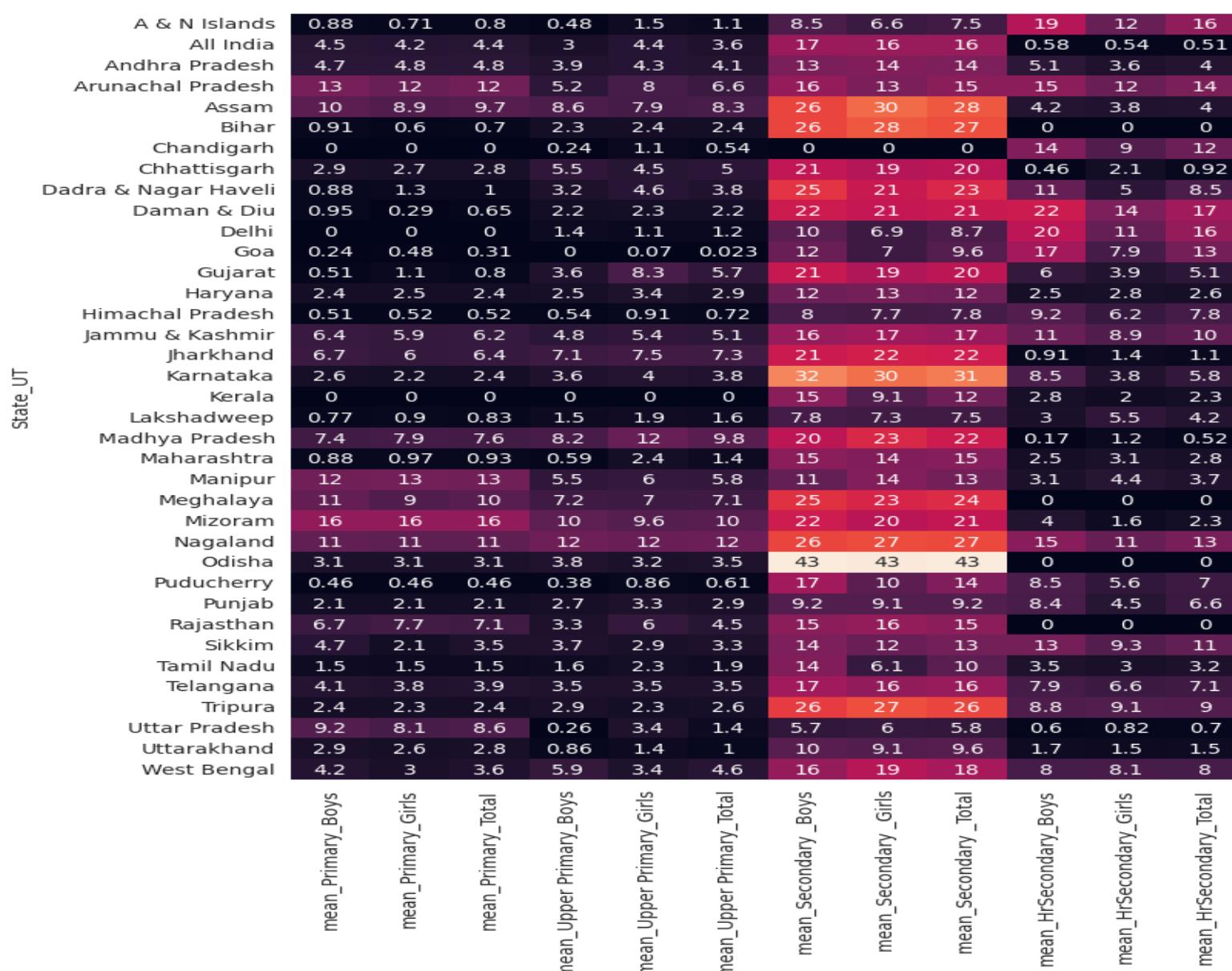


# Data Visualization And Results :

Let's explore Dropout ratio data set!



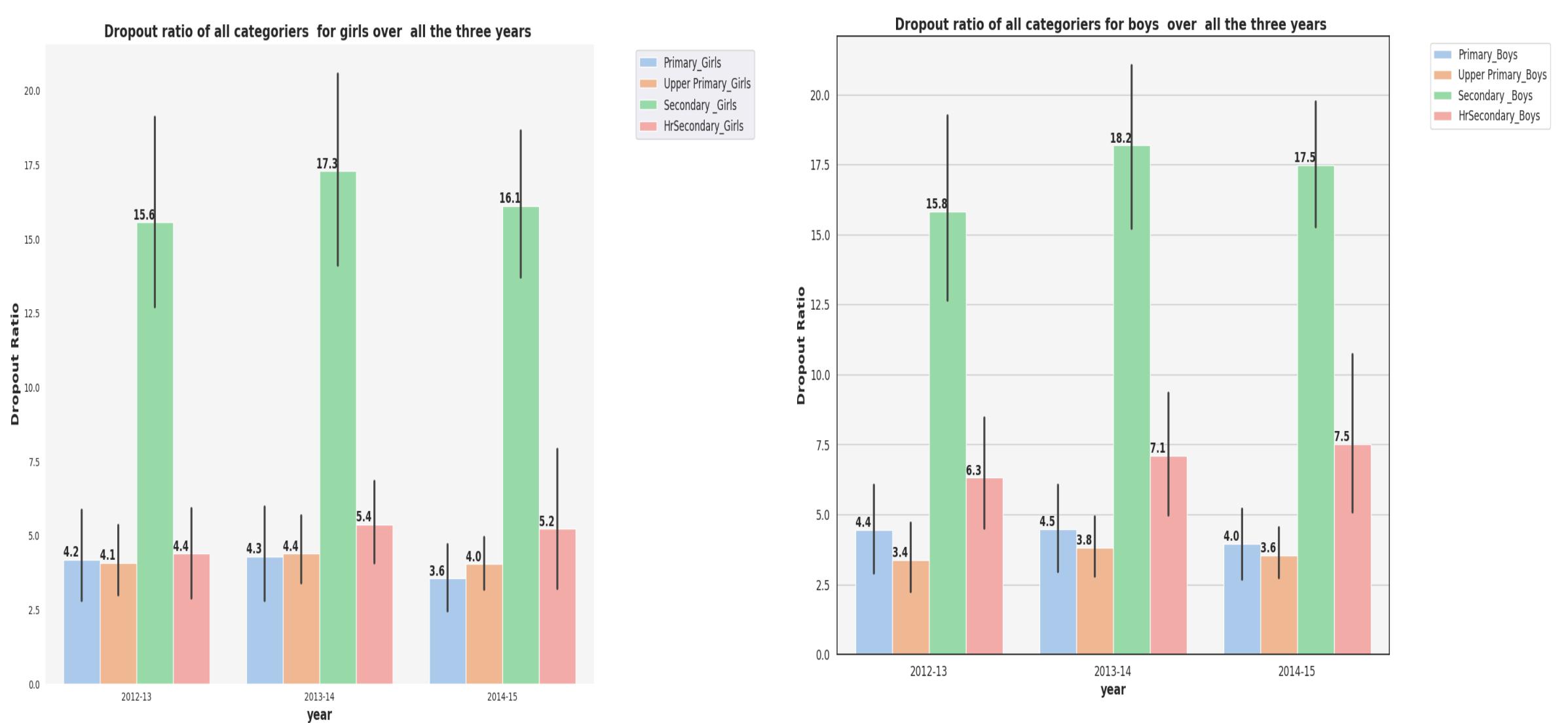
# Data Visualization And Results :



► The above heatmap is the mean of all three year values  
 ► From the above heat map we can draw the conclusion that "Odisha" has the top most school drop out  
 ► The point to be considered on high alert is that dropout ratio is high for the states: Odisha, Karnataka, Assam, Bihar, Nagaland, Tripura, and yes we can relate it with the tier that they belong to we decide upon the facility score, YES! 5 out of the 6 belong to tier 4 and tier 5.

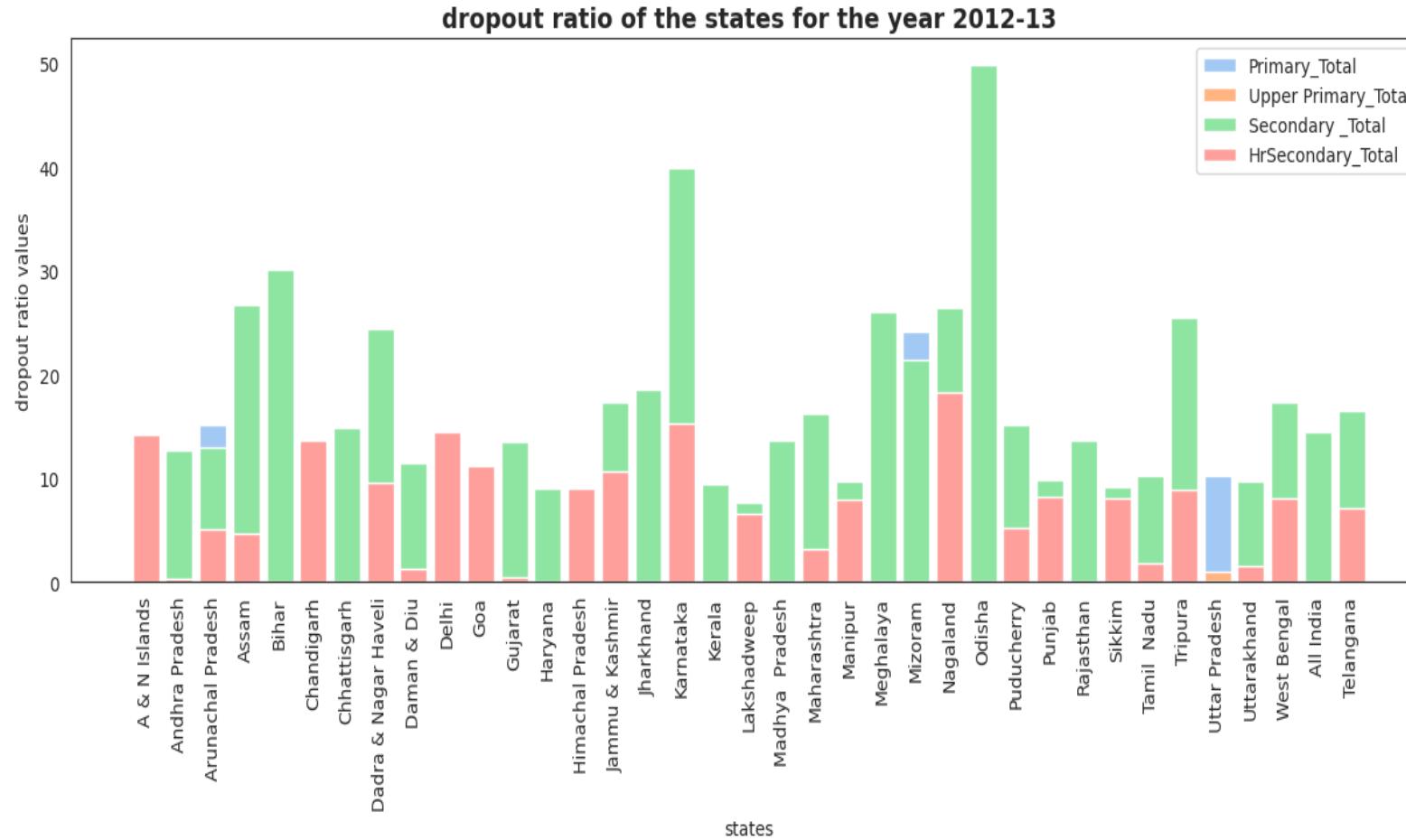
► In terms of school category, we can see that secondary category schools have the most drop out ratio, we can observe this in below plots

# Data Visualization And Results :



# Data Visualization And Results :

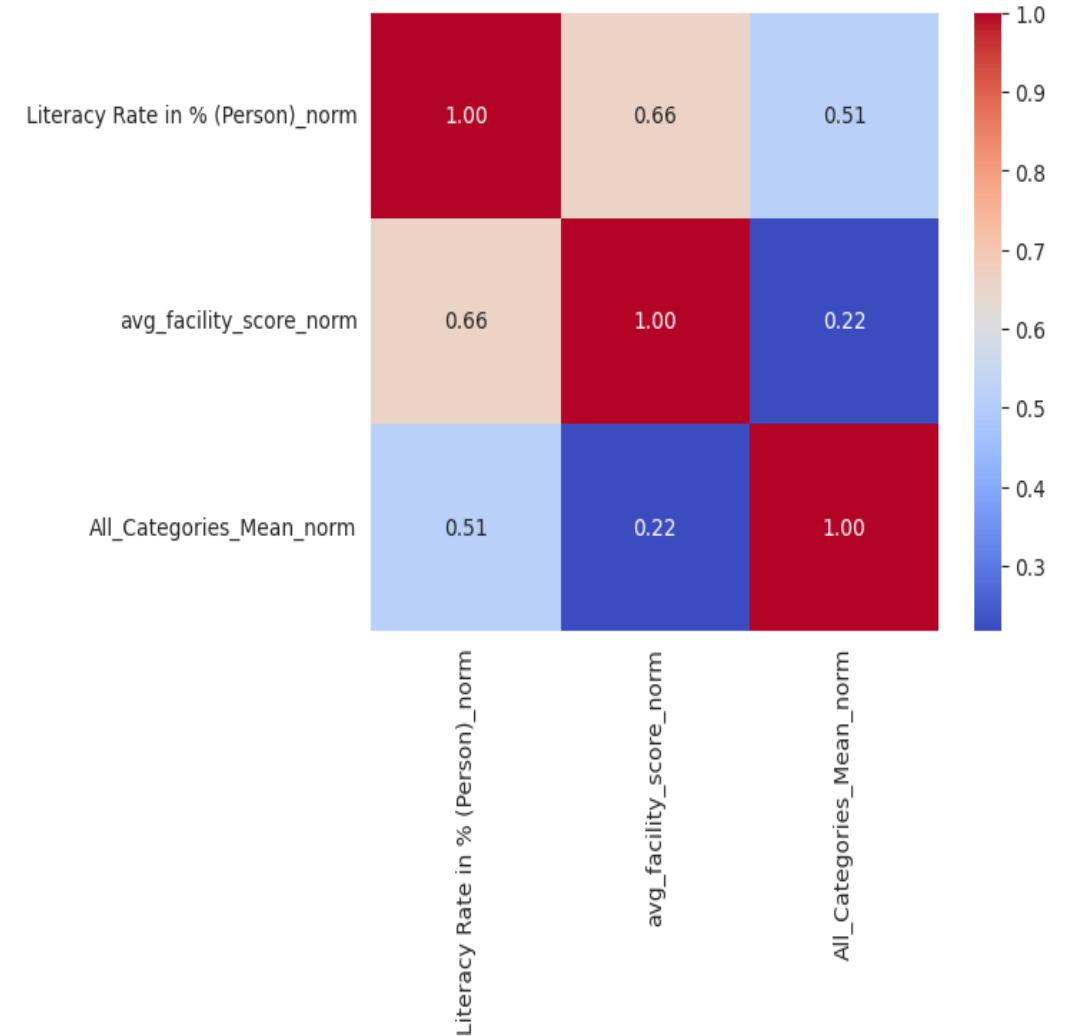
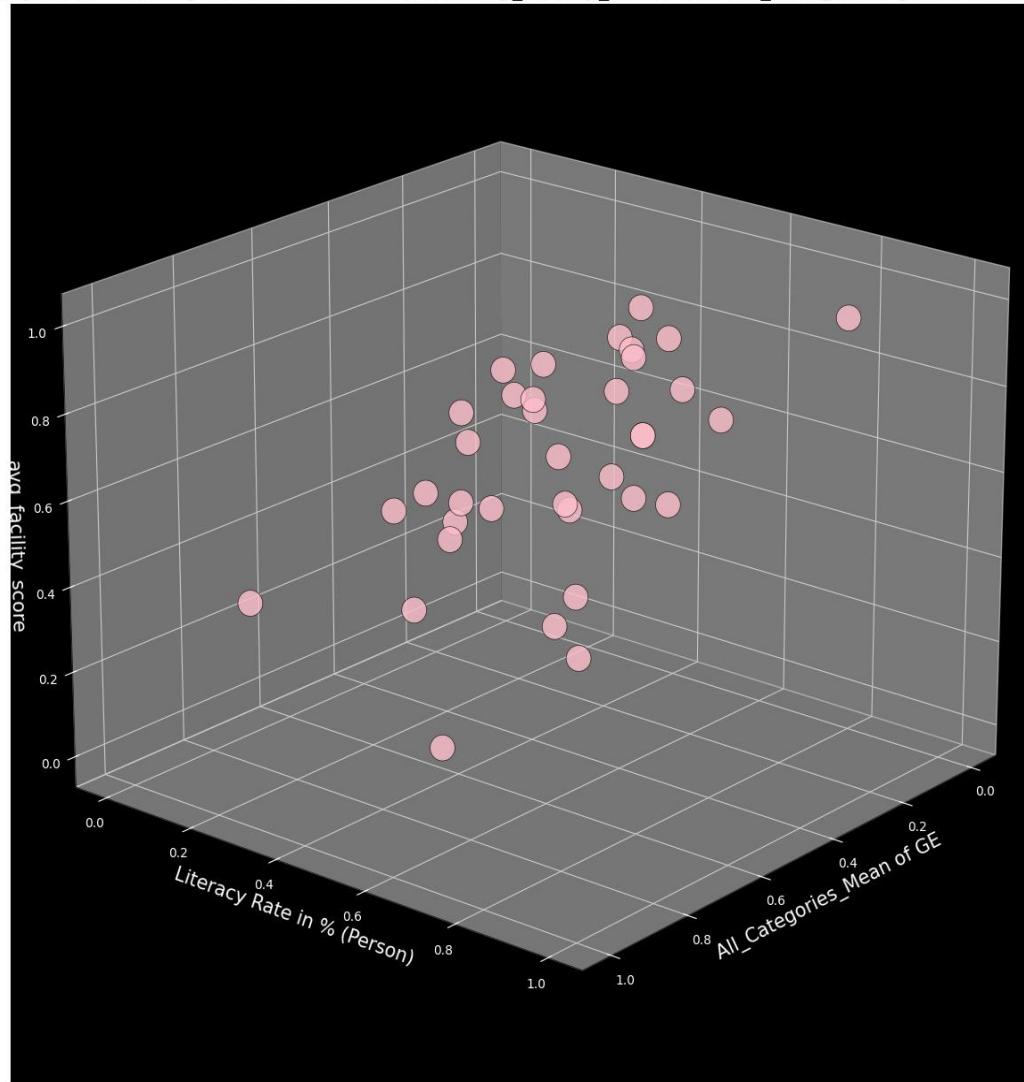
► Now for a particular state which category have max dropout ratio ? If we know this that particular state will take action accordingly . Now this can be seen in above heatmap ...but now for vizualization it is shown in below plot



► so now for odisha secondary school category dropout is alarming for the year 2012-13..simimlary primary category for uttar pradesh

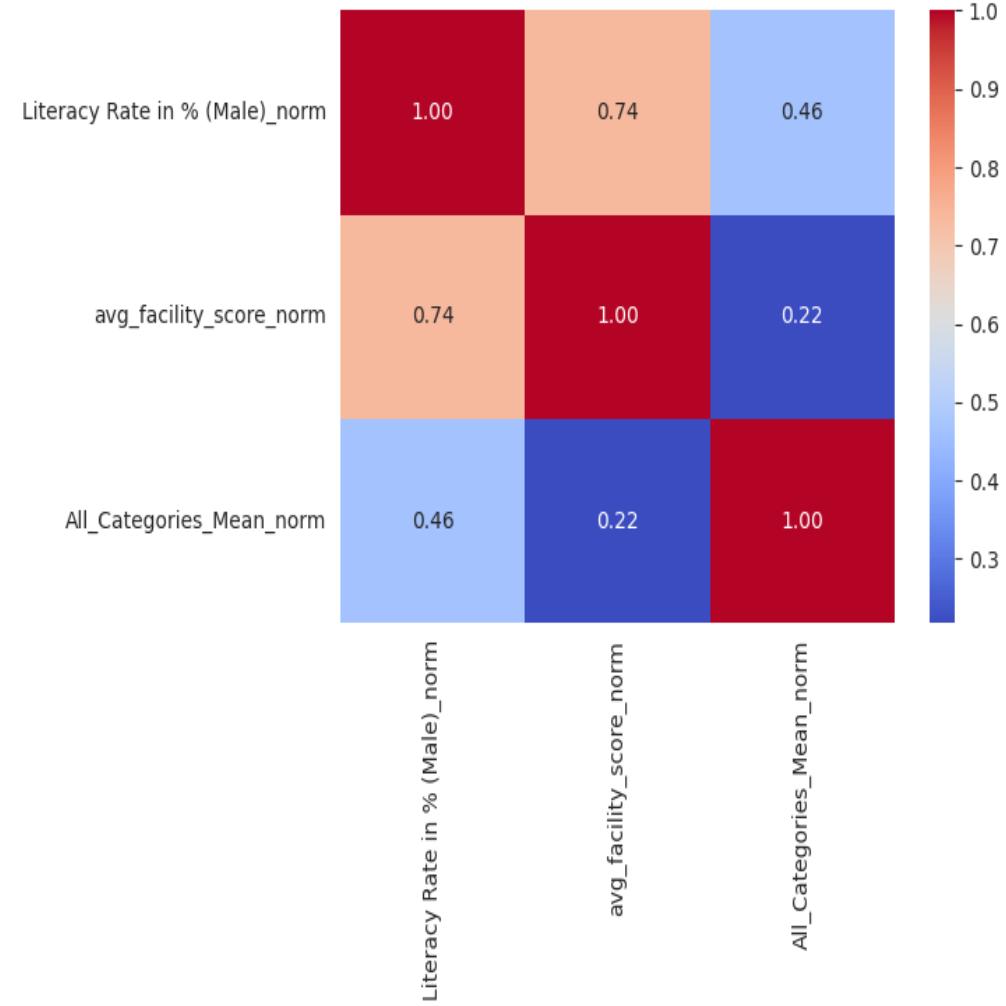
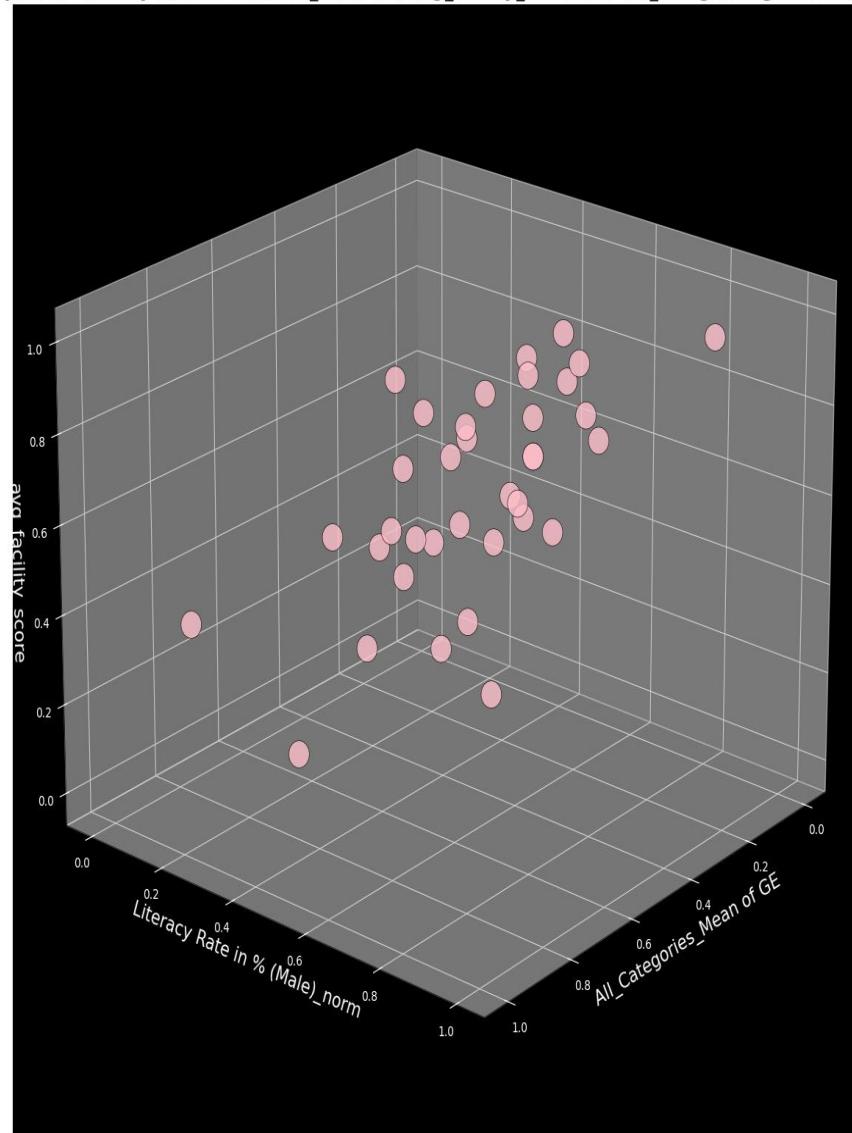
# Data Visualization And Results :

3-D scatter plot of Literacy Rate in % (Person) and avg\_facility\_score and All\_Categories gross enrollment mean



# Data Visualization And Results :

3-D scatter plot of Literacy Rate in % (Male)\_norm and avg\_facility\_score and All\_Categories gross enrollment mean



**Thank You!**