Education Inequality in India

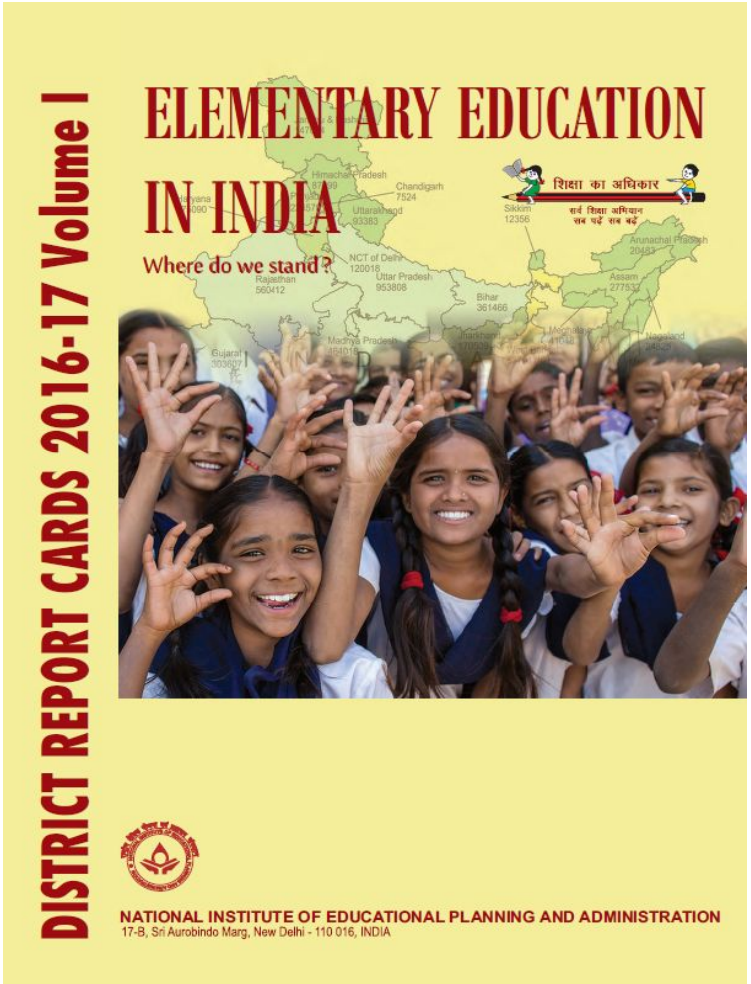
An exploration of school infrastructure and their impact on Enrolments

Mukul Kumar

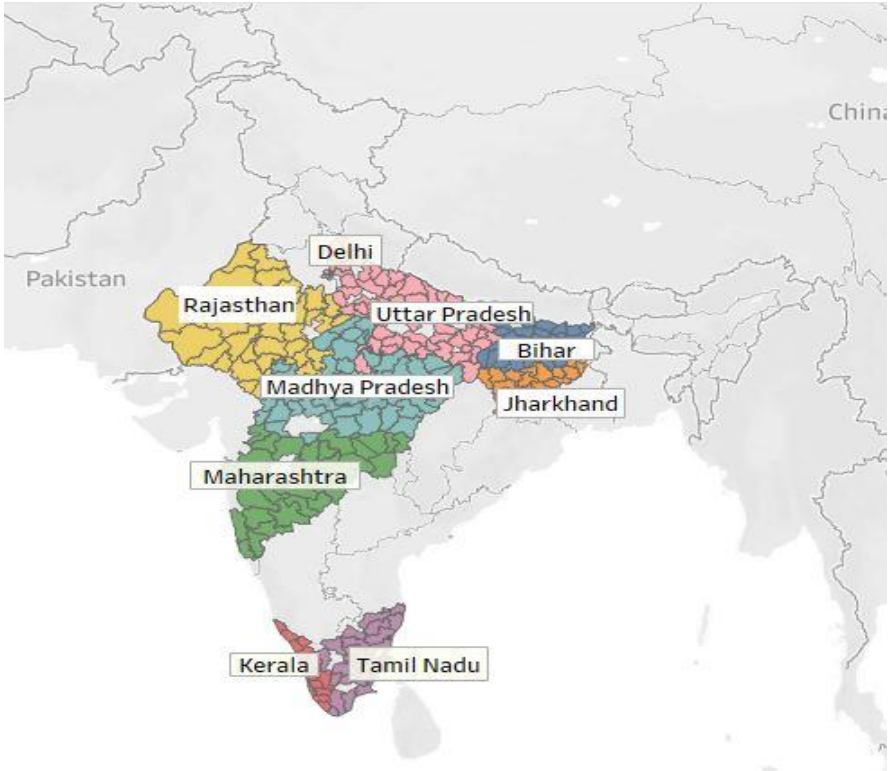
This study is based on the DRC data of DISE (District Report Card 2016-17)

It contains District wise data on education and school infrastructure indicators in India.

301 Districts | 9 States



States under study



Top States (Literacy > 80%)

Delhi	86.34%
Kerala	94.00%
Maharashtra	82.91%
Tamil Nadu	80.33%

Bottom States (Literacy <71%)

Bihar	63.82%
Jharkhand	67.63%
Madhya Pradesh	70.63%
Uttar Pradesh	67.63%
Rajasthan	67.06%

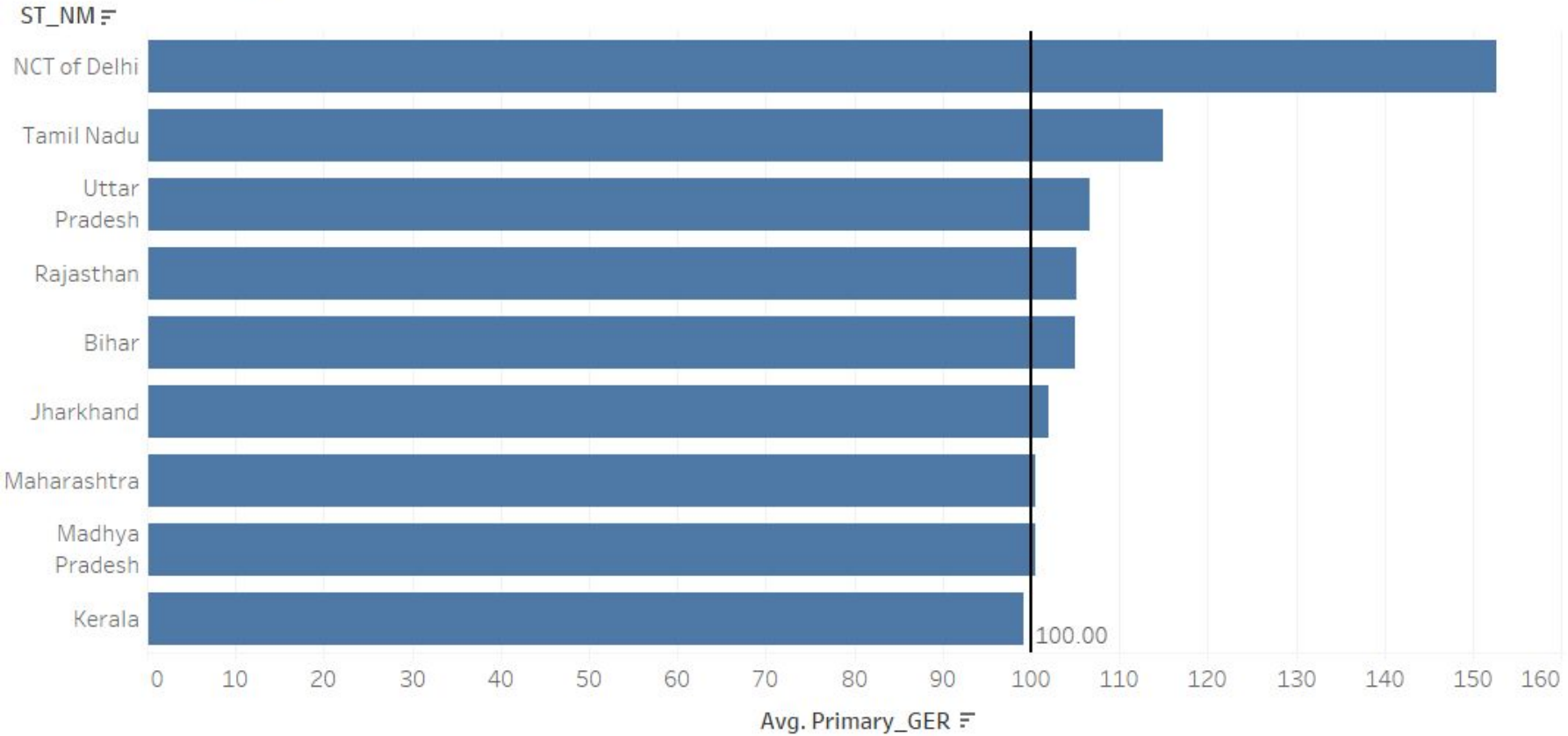
Gross Enrolment Ratio

Number of students enrolled in a given level of education, regardless of age, expressed as a percentage of the official school-age population corresponding to the same level of education.

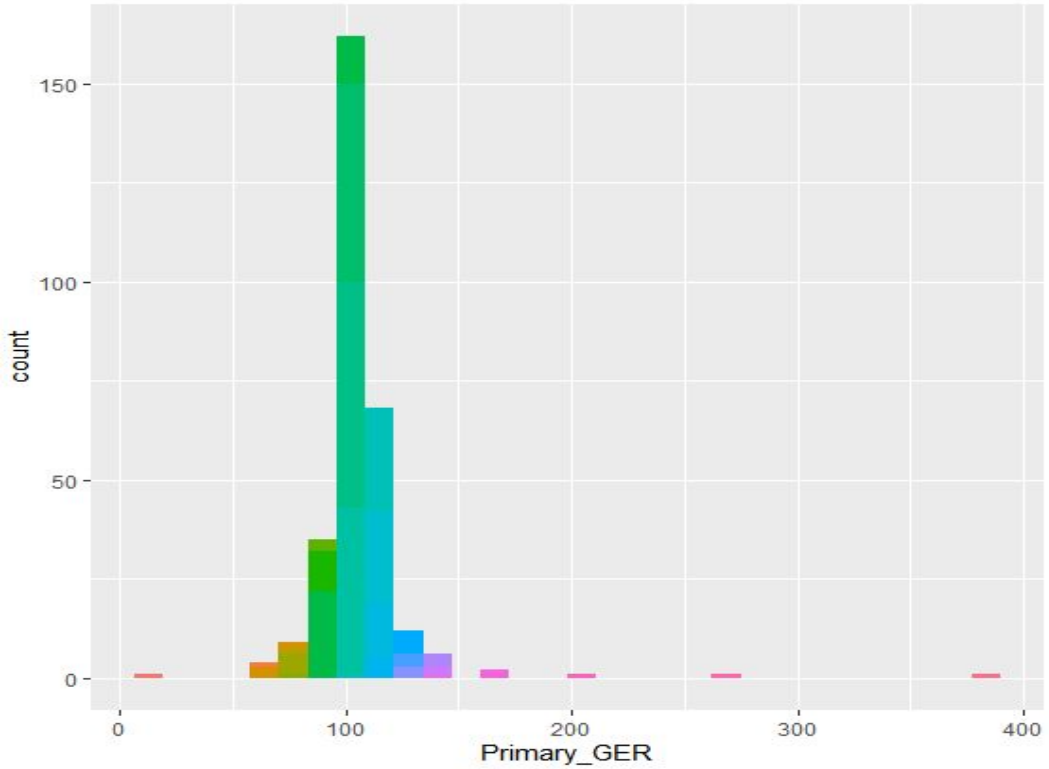
- Definition by UNESCO

Primary GER : GER for primary education (1st to 5th grade) Age group : 6 to 10 years
Upper GER : GER for upper primary education (6th to 8th) Age group : 11 to 13 years

Primary GER by states

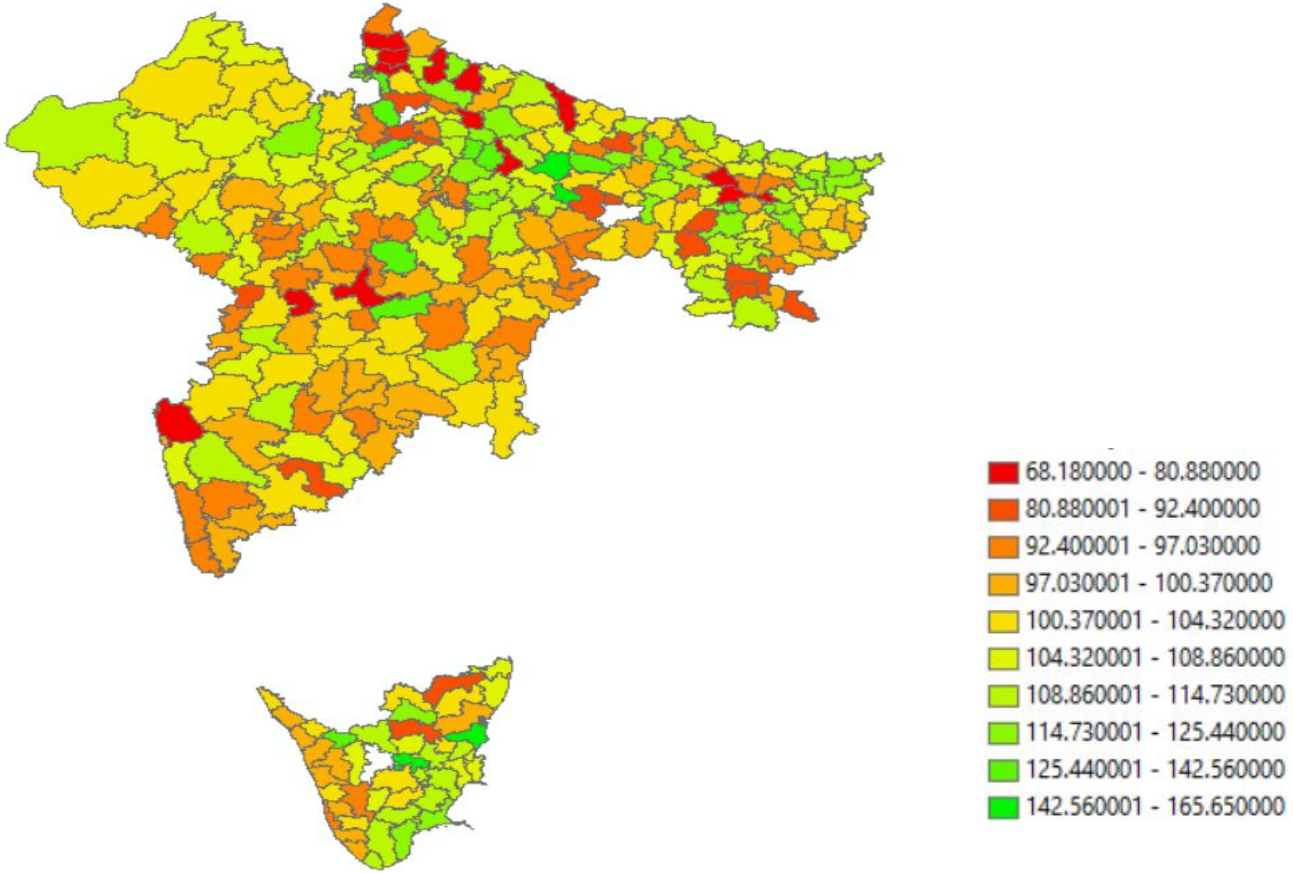


Primary_GER distribution



Mean	106.07
Standard Deviation	23.69
Median	102.98
1%	68.19
25%	98.33
75%	110.30
99%	165.62
MIN	11.76 (Saran)
MAX	382.61 (New Delhi)

Primary GER

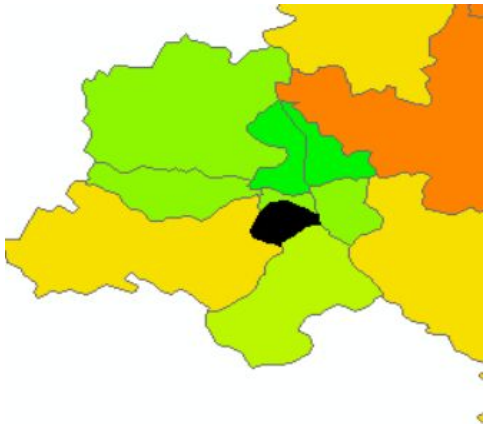

68.180000 - 80.880000
80.880001 - 92.400000
92.400001 - 97.030000
97.030001 - 100.370000
100.370001 - 104.320000
104.320001 - 108.860000
108.860001 - 114.730000
114.730001 - 125.440000
125.440001 - 142.560000
142.560001 - 165.650000

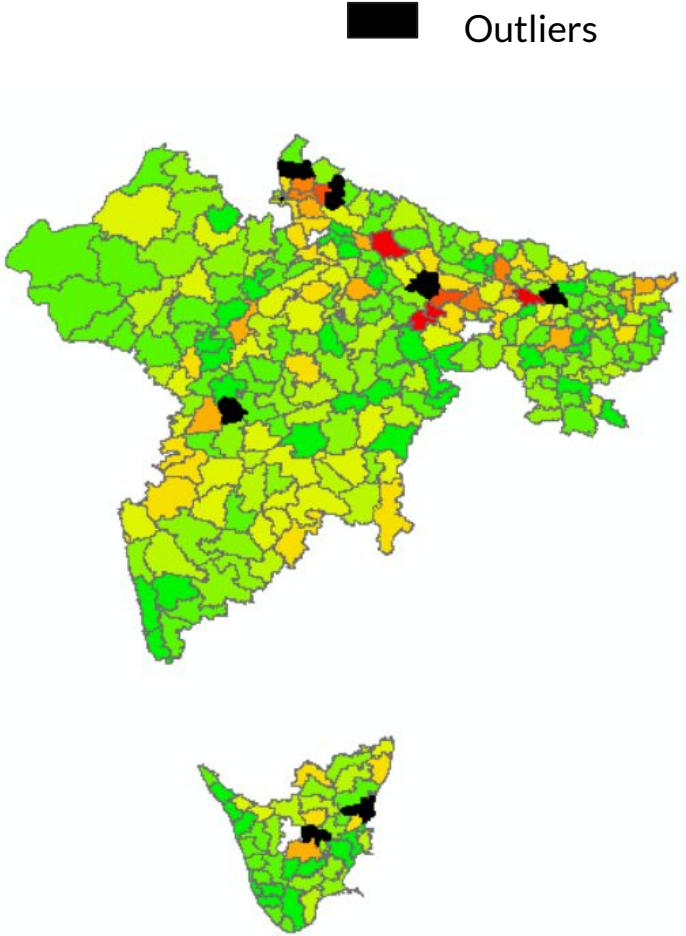
Outliers

8 Outliers using Winsorization

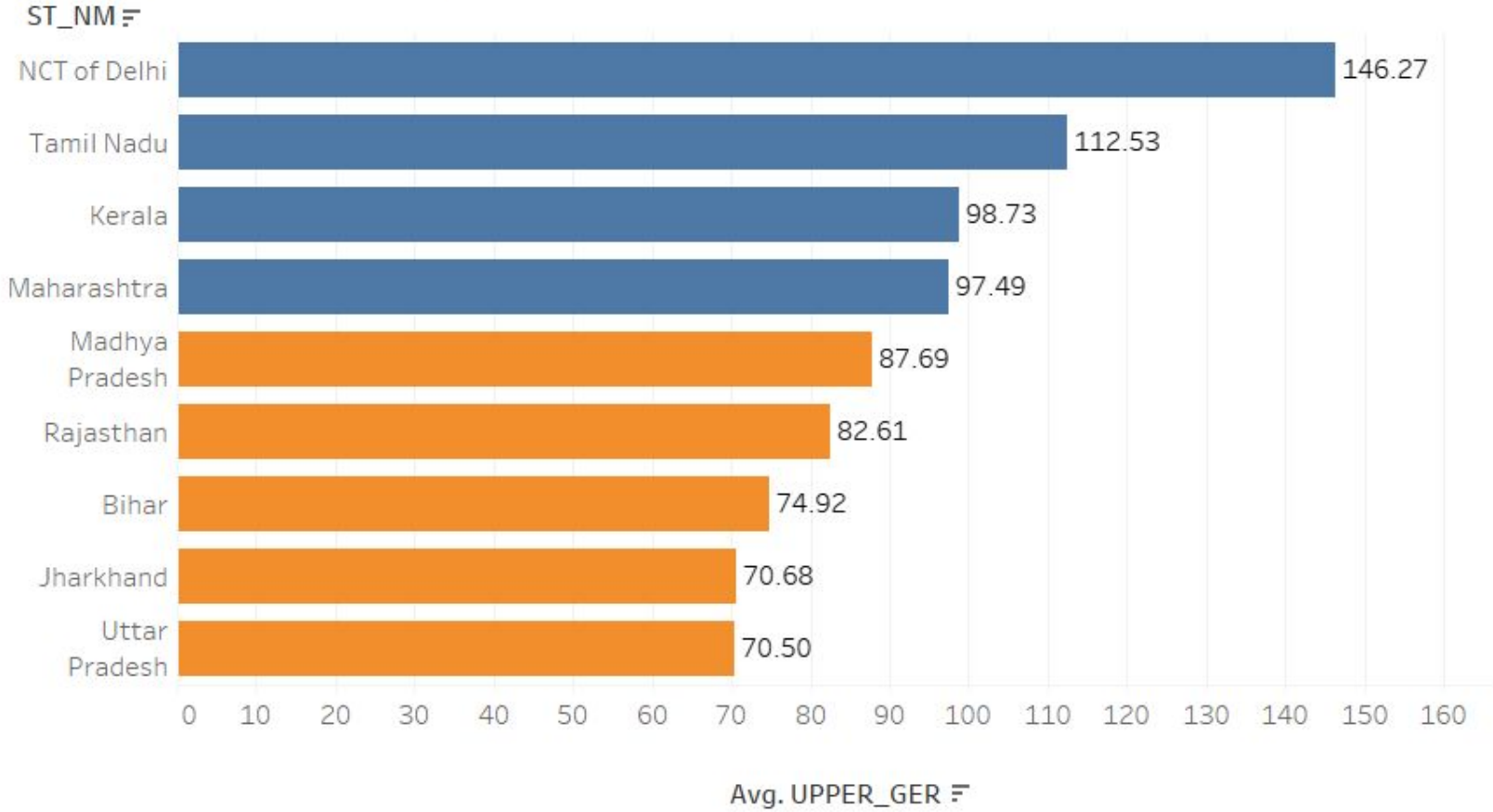
(>99 percentile and <1 percentile)

Delhi

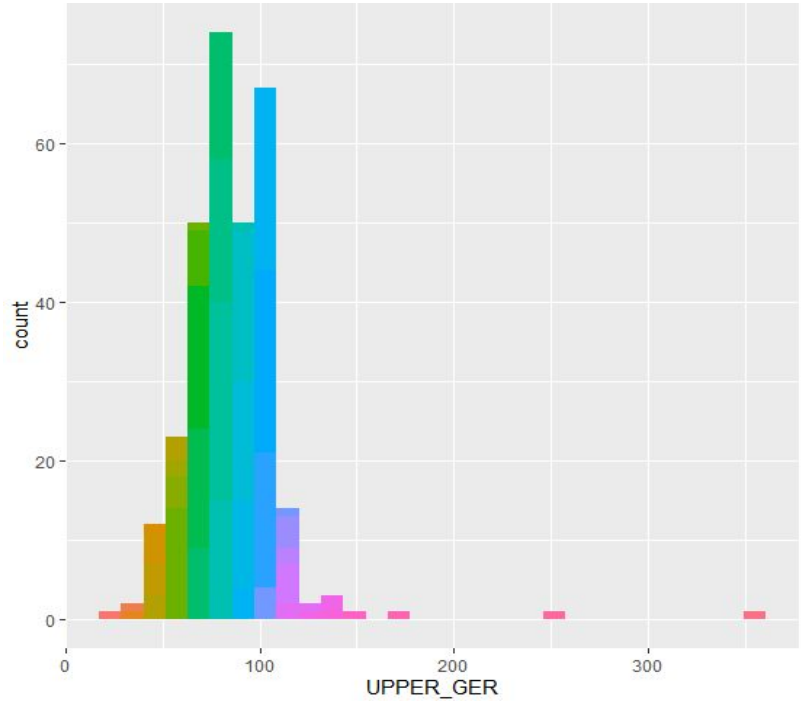




Upper primary GER by states

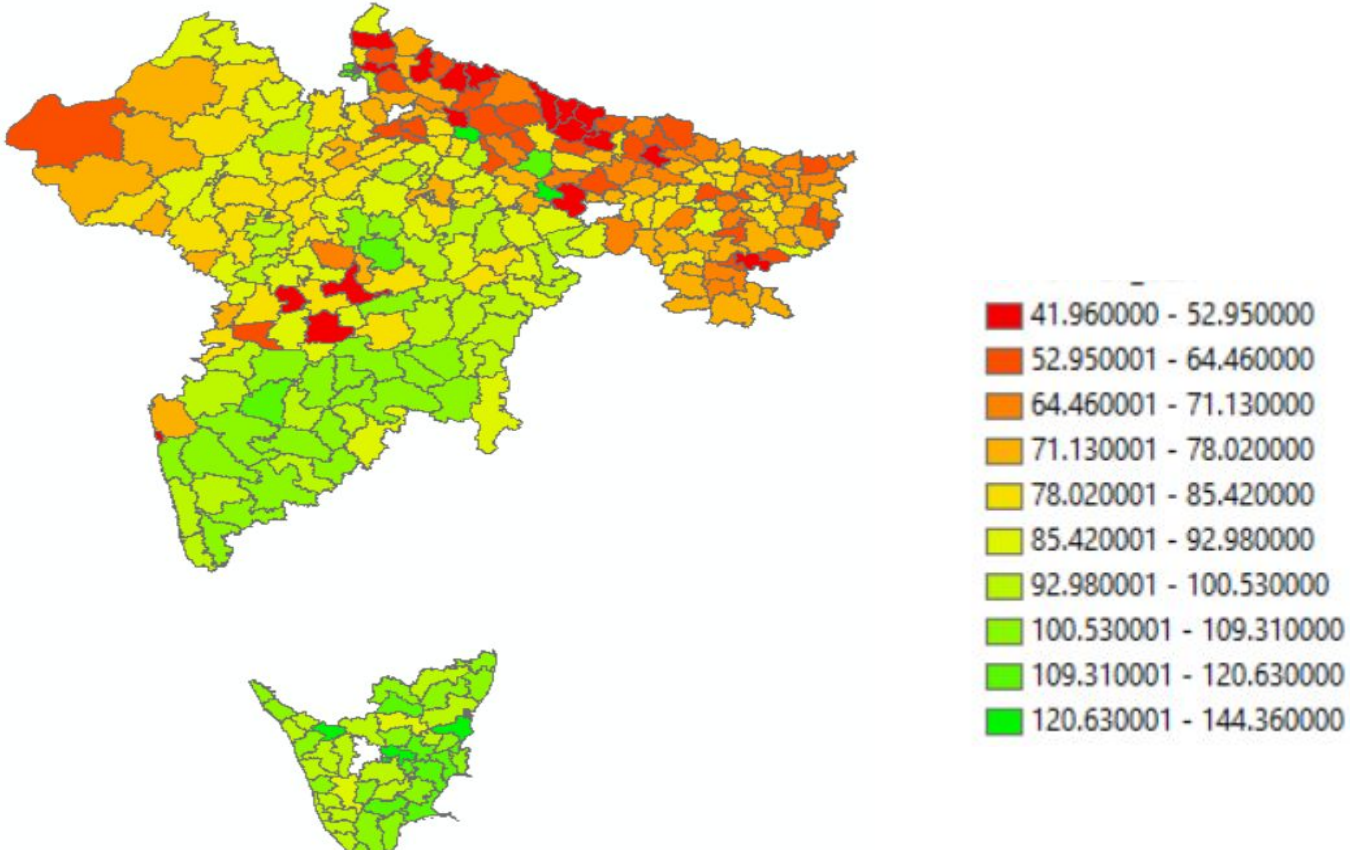


UPPER Primary GER distribution



Mean	86.02
Standard deviation	26.52
Median	83.48
1 %	41.99
25 %	72.54
75 %	98.89
99 %	144.28
MIN	34.52(Balrampur)
MAX	358.72(New Delhi)

Upper Primary GER

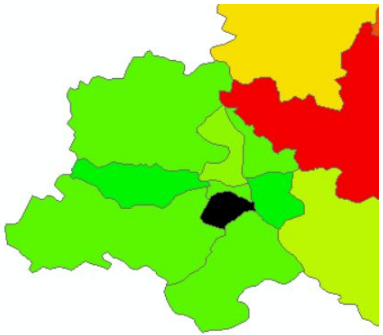


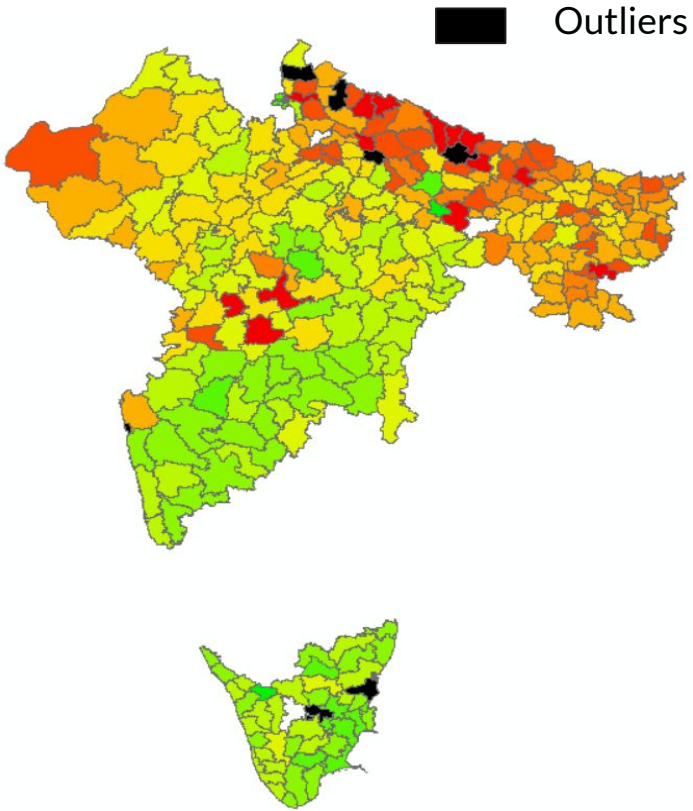
Outliers

8 Outliers using Winsorization

(>99 percentile and <1 percentile)

Delhi



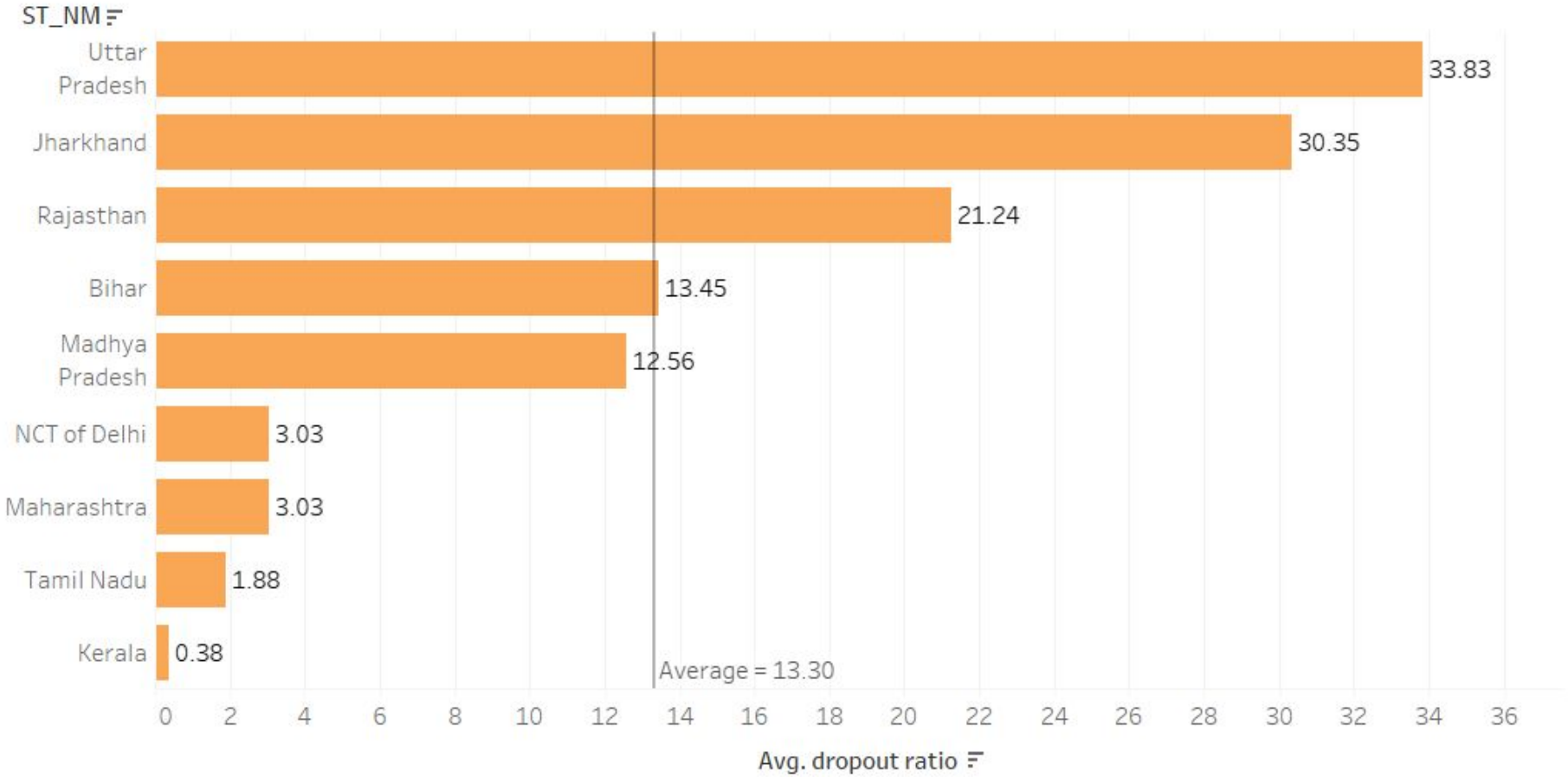


New Delhi, Coimbatore (T.N.), Perambalur (T.N.), Moradabad (U.P.),Muzaffarnagar (U.P.) are common outliers in both Primary and Upper Primary GER data.

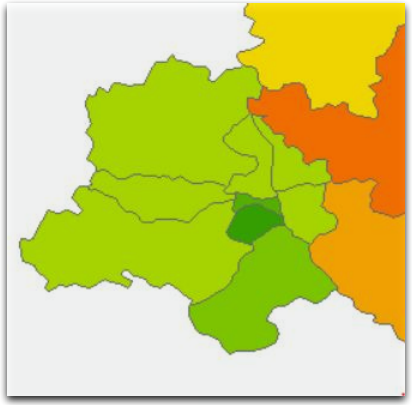
How many primary students make it to the upper primary education?

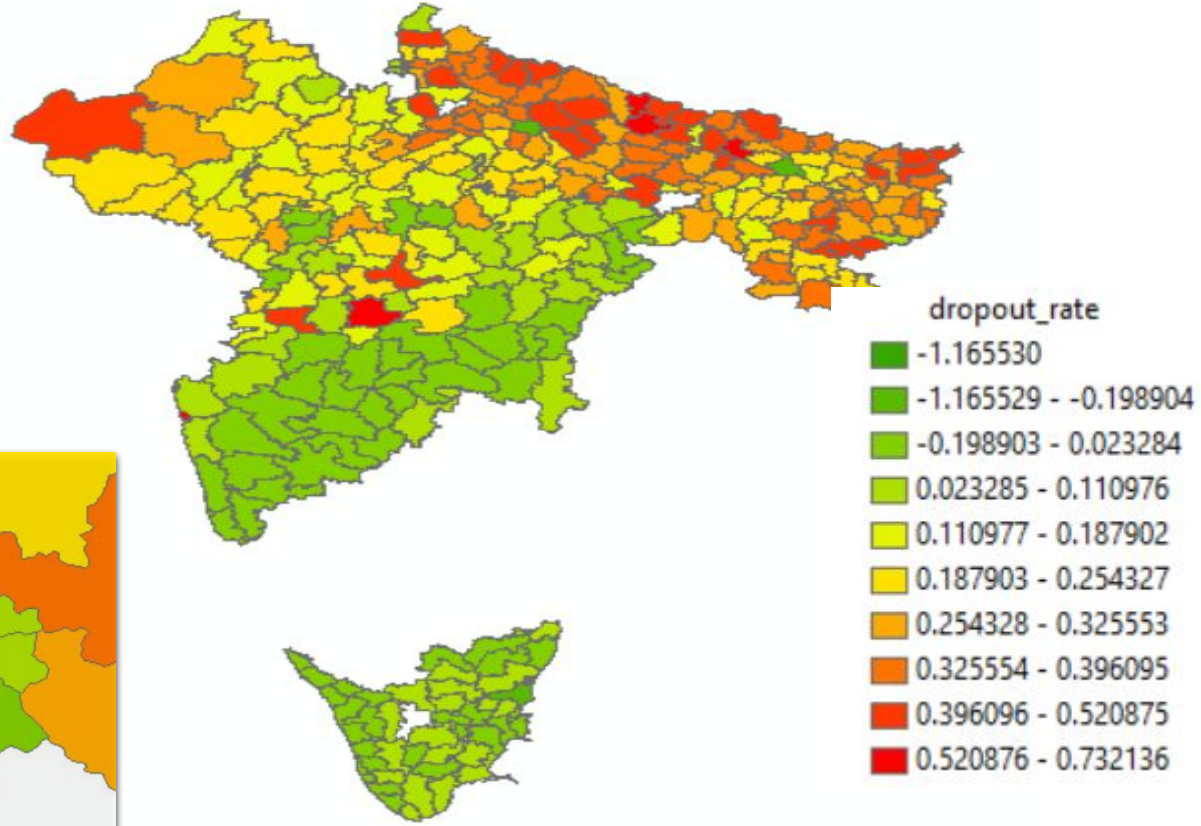
Dropout ratio = [(Primary GER) - (Upper Primary GER)]/(Primary GER)

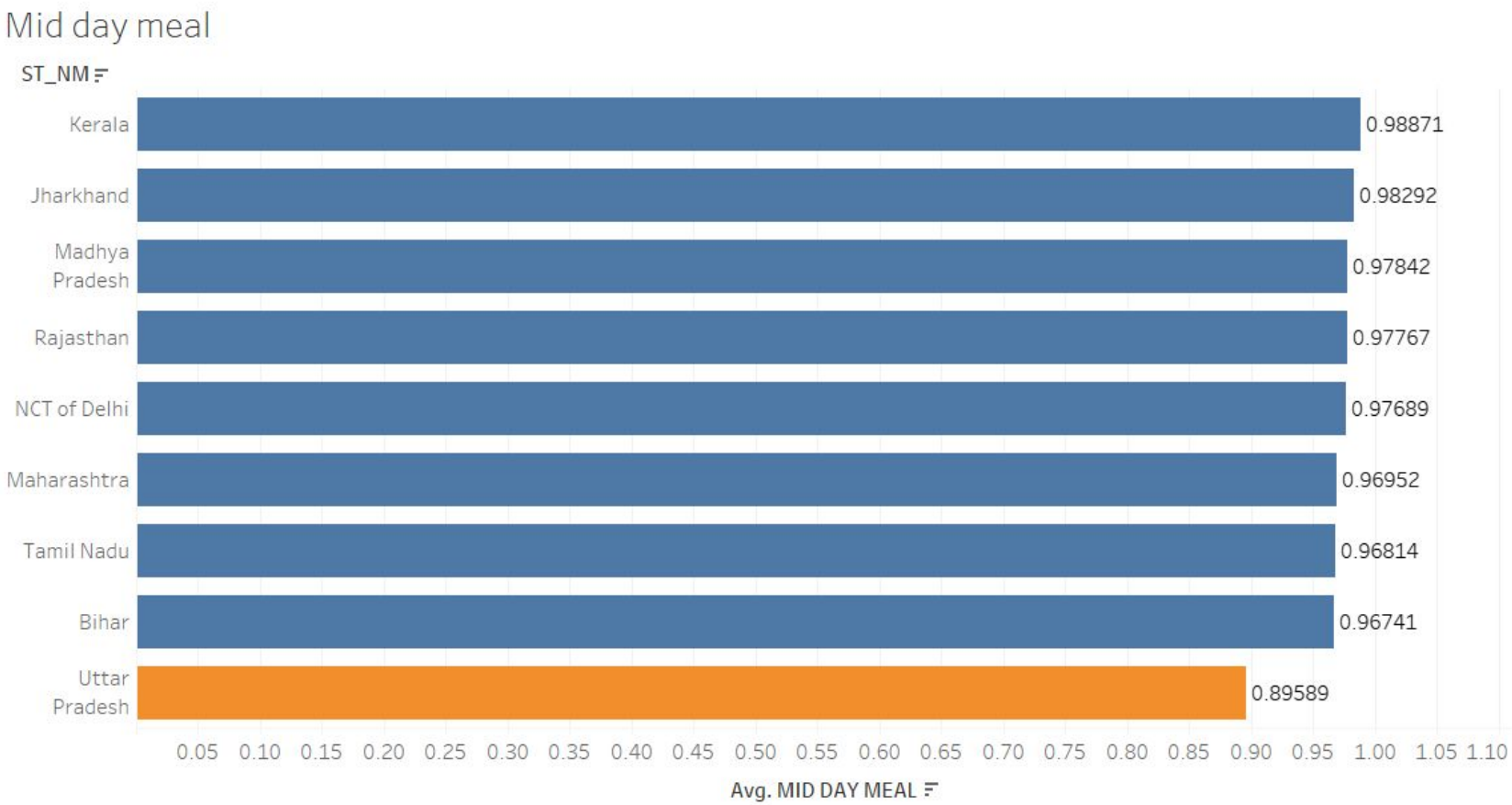
Drop out ratio



Dropout rates





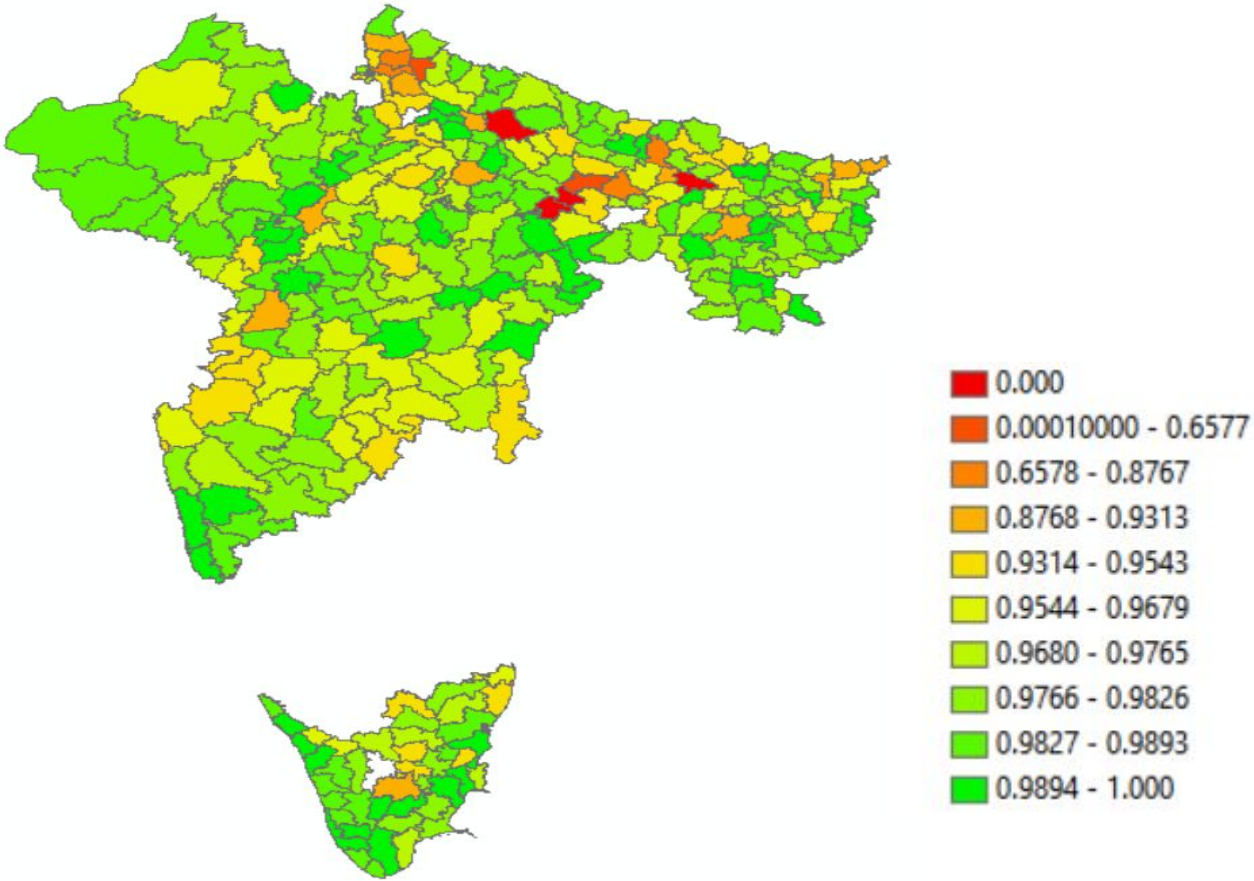


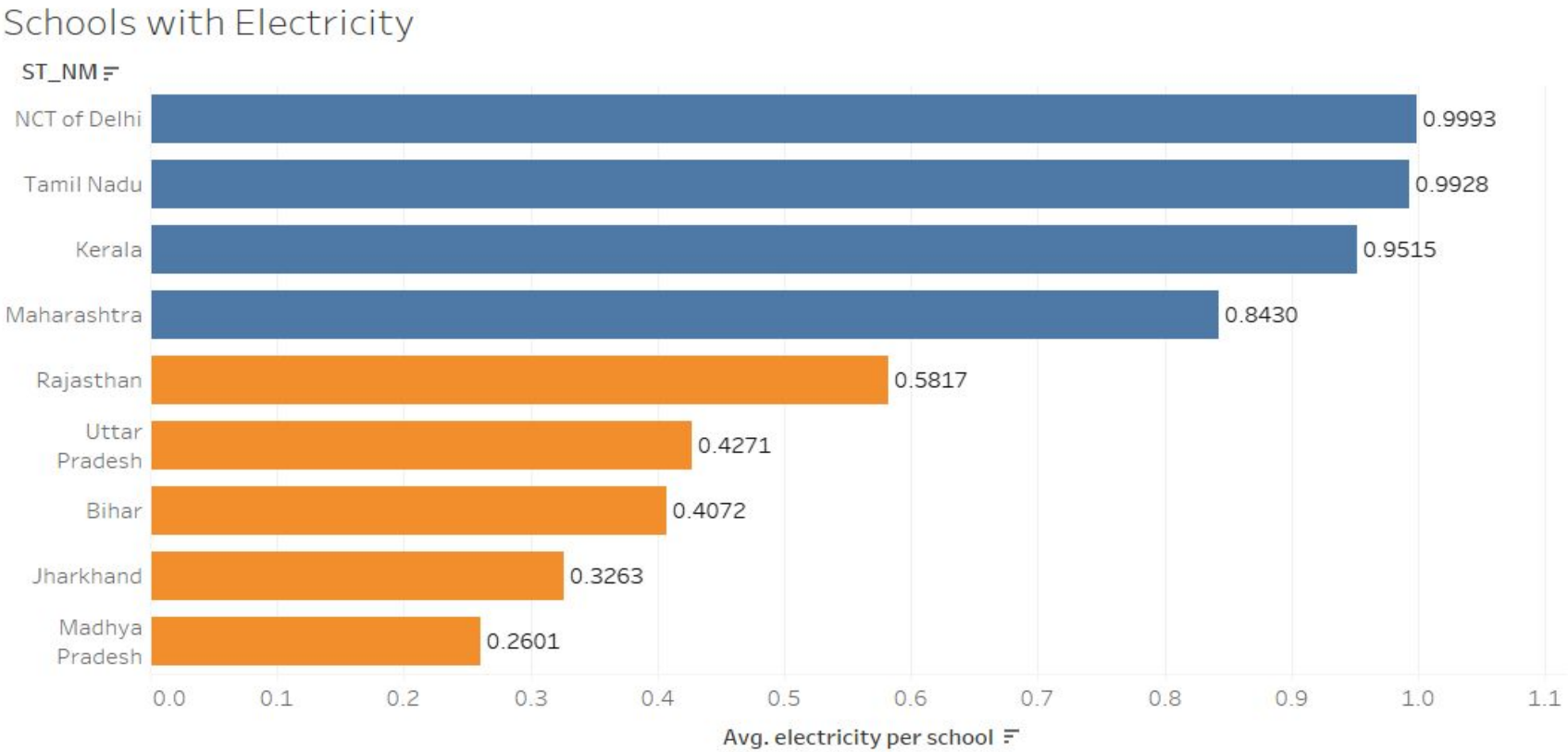
Clustered Using K-means clustering

Cluster 1

Cluster 2

Mid day meal


0.000
0.00010000 - 0.6577
0.6578 - 0.8767
0.8768 - 0.9313
0.9314 - 0.9543
0.9544 - 0.9679
0.9680 - 0.9765
0.9766 - 0.9826
0.9827 - 0.9893
0.9894 - 1.000

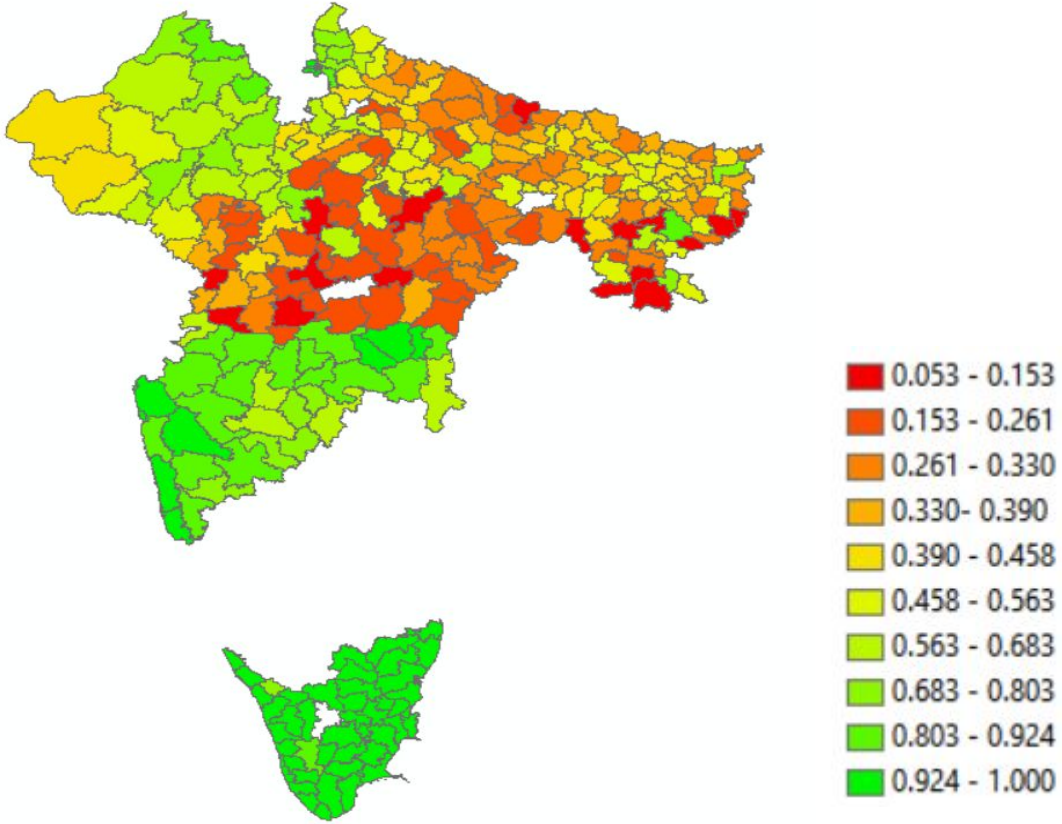


Clustered Using K-means clustering

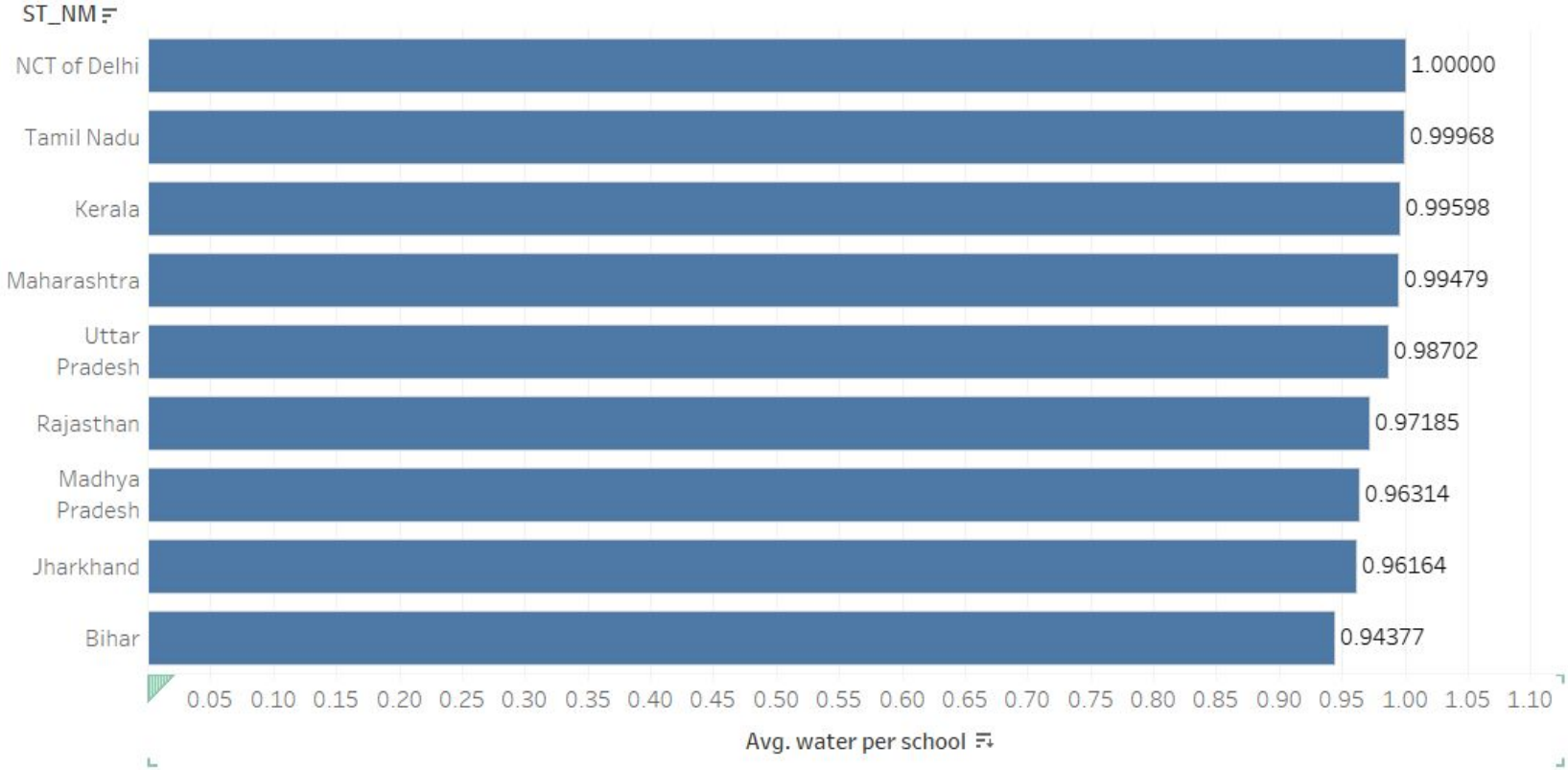
Cluster 1

Cluster 2

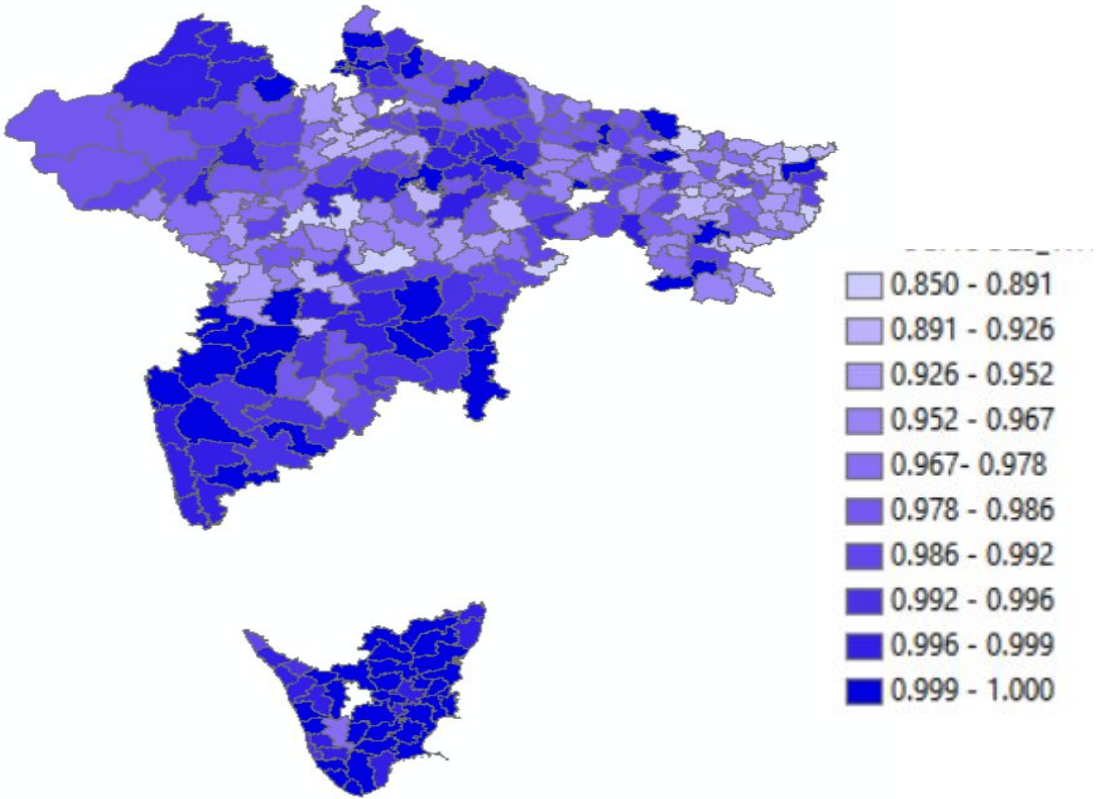
Availability of Electricity

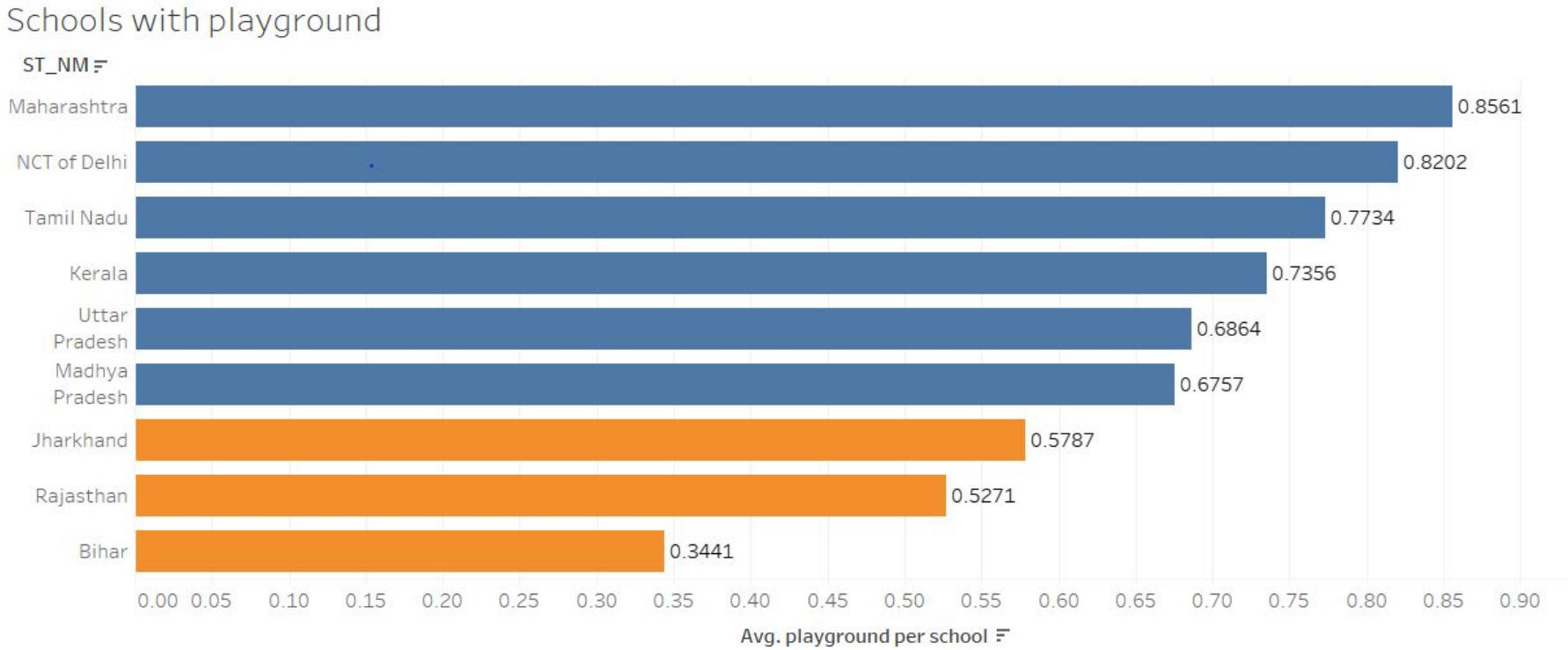


Availability of Water



Water availability

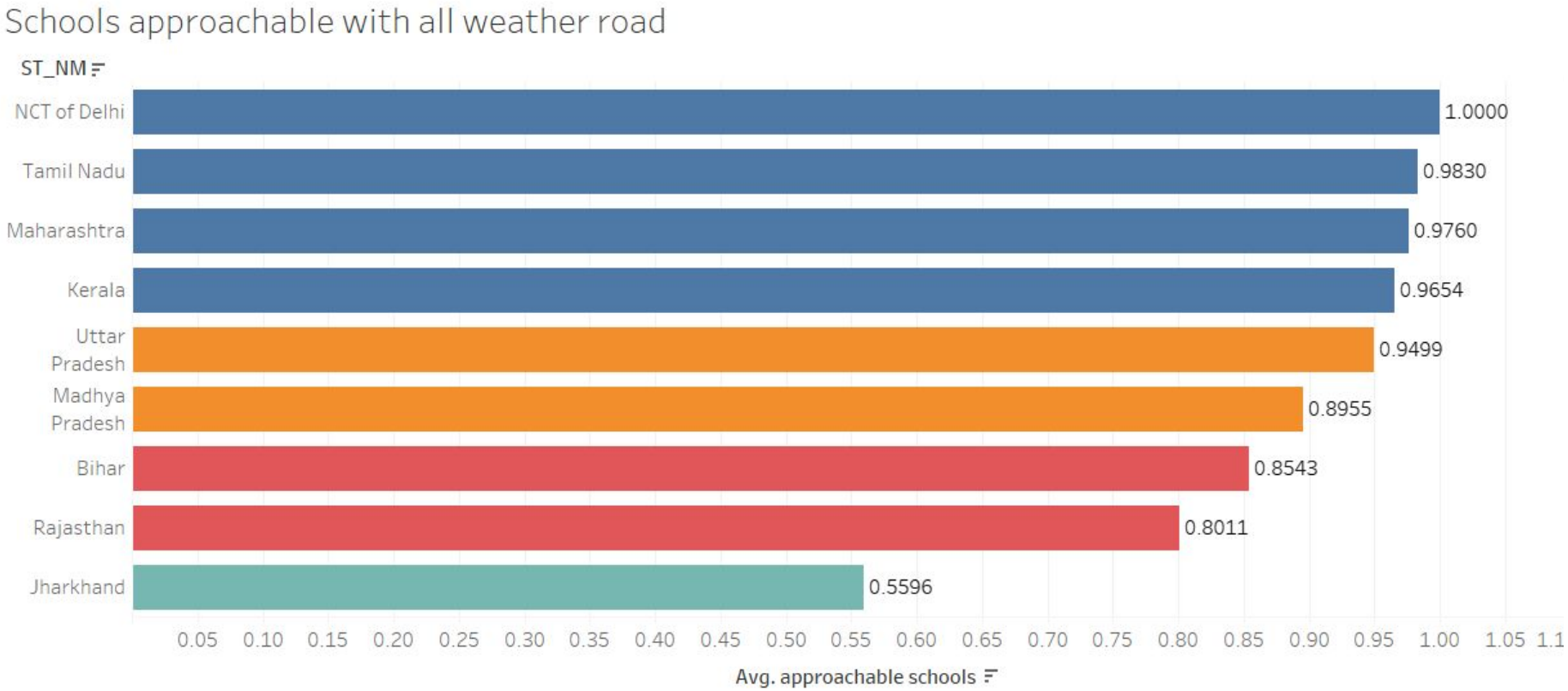




Clustered Using K-means clustering

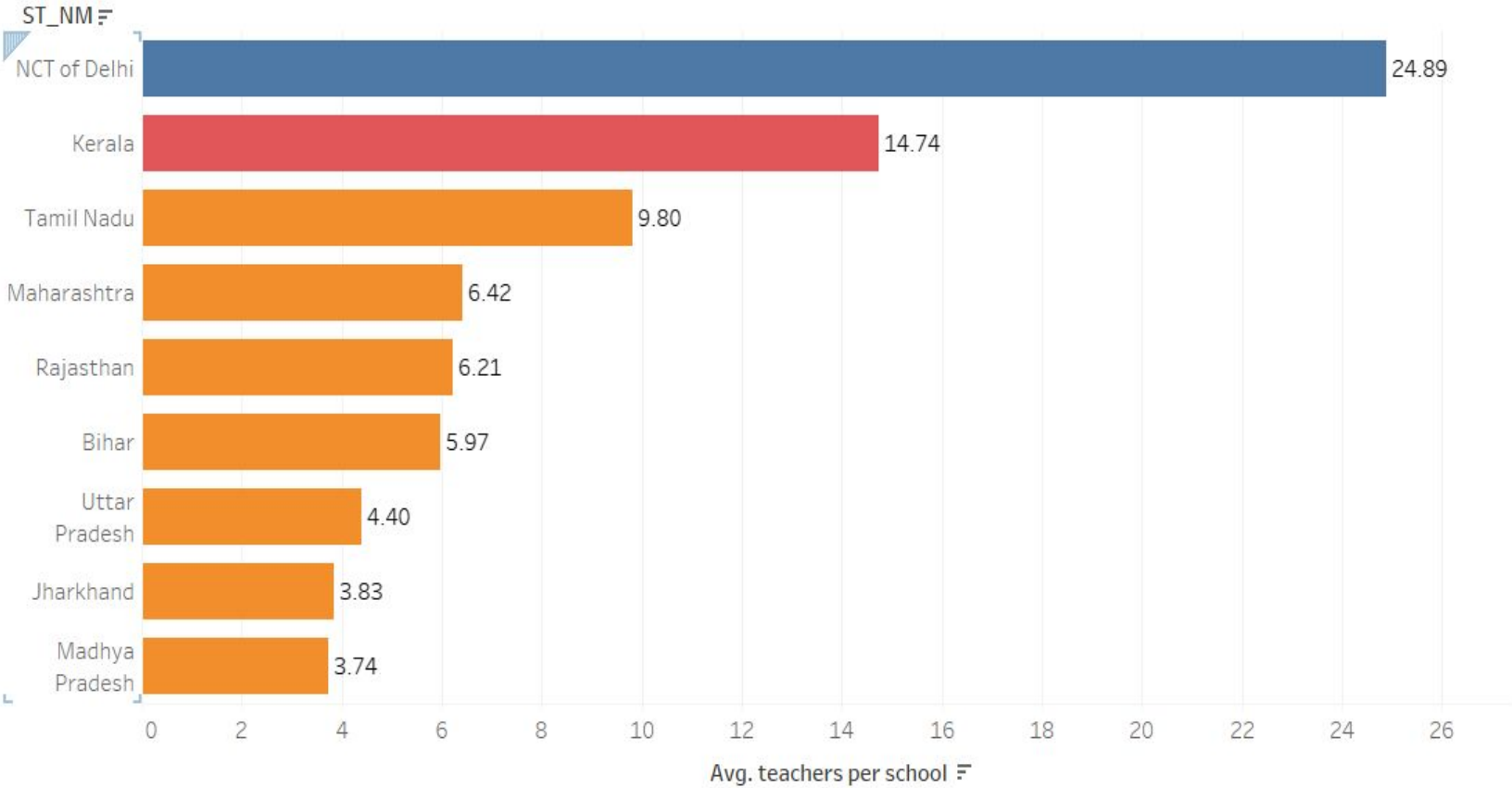
Cluster 1

Cluster 2



Clustered Using K-means clustering

Teachers per school



Where the hotspots are?

A hotspot is a cluster of polygons with significantly high or low values.

Hotspot Detection using **Hotspot Analysis (Gi * statistic)**

Two kinds of hotspots :

One with High GER Value

One with low GER Value

The Getis-Ord local statistic is given as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}} \quad (1)$$

where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \quad (2)$$

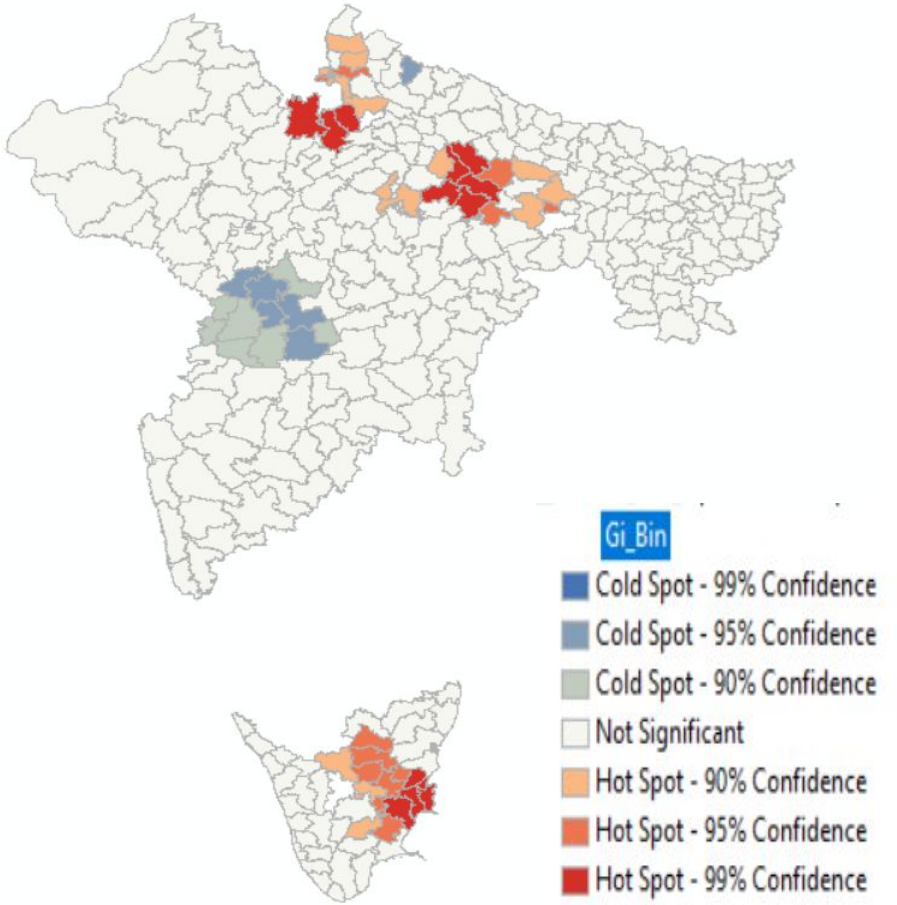
$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \left(\bar{X}\right)^2} \quad (3)$$

The G_i^* statistic is a z-score so no further calculations are required.

Hotspots for Primary GER

More Red means high Upper GER

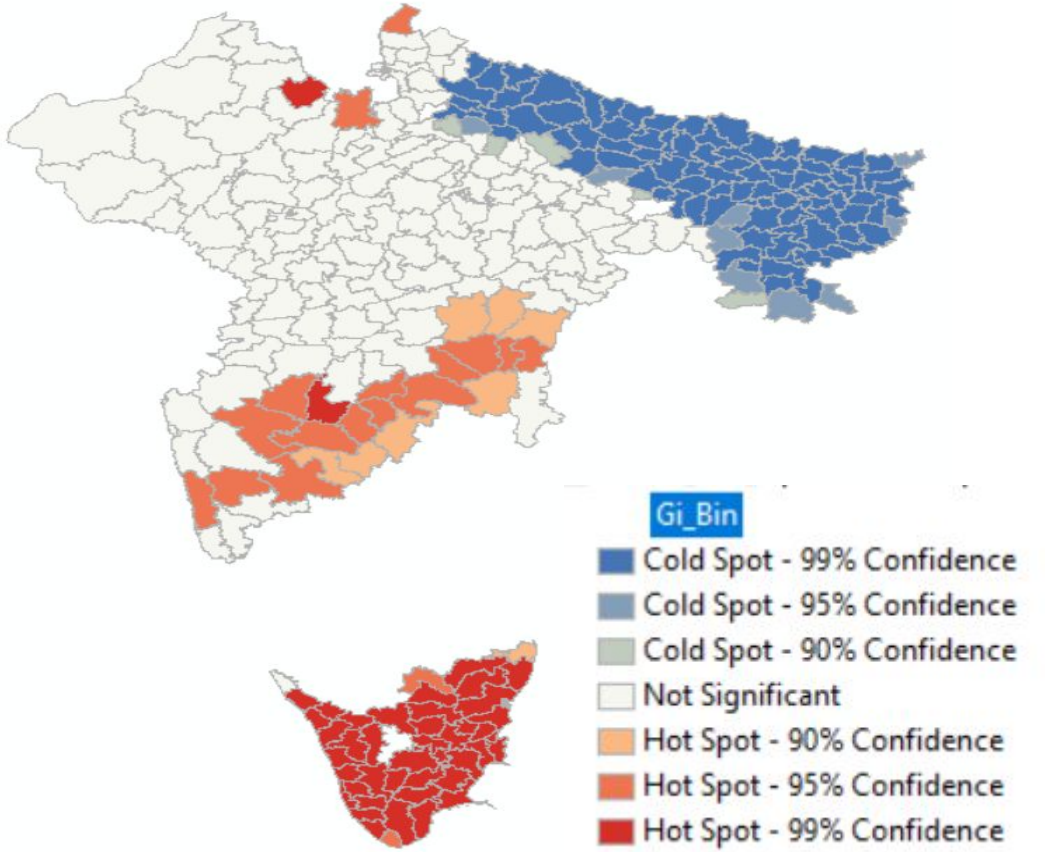
More Blue means low Upper GER



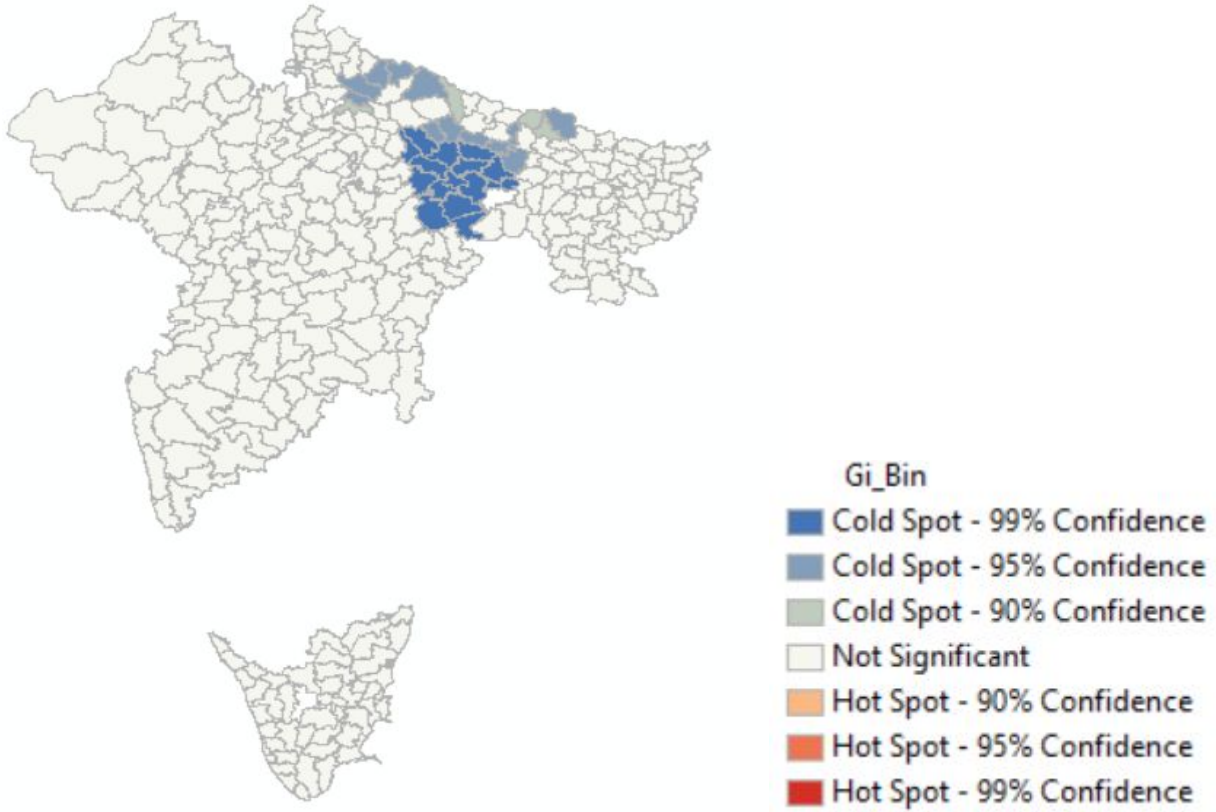
Hotspots for Upper Primary GER

More Red means high Upper GER

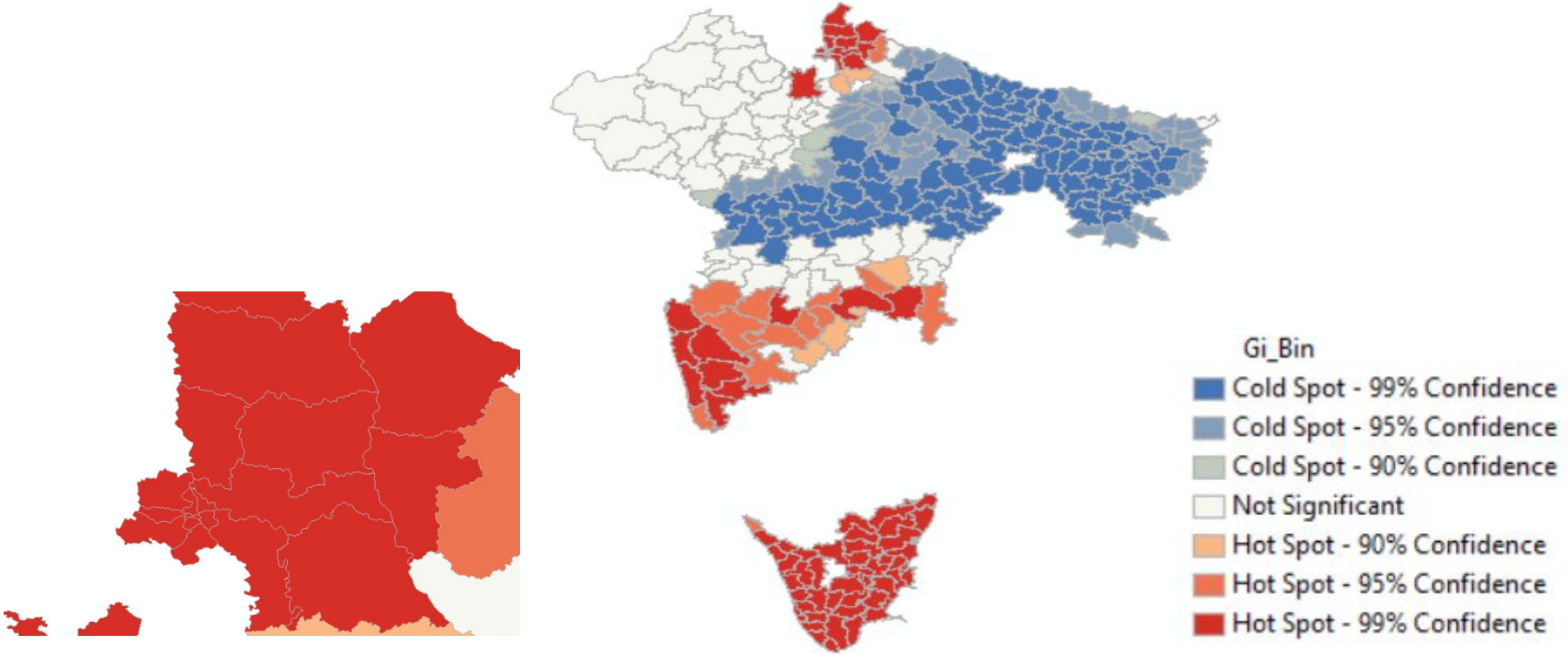
More Blue means low Upper GER



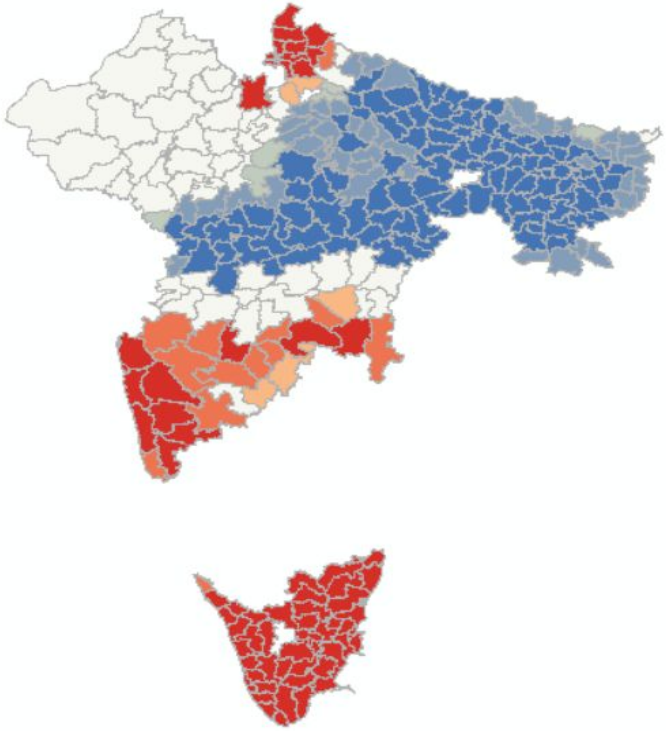
Hotspots for mid day meal



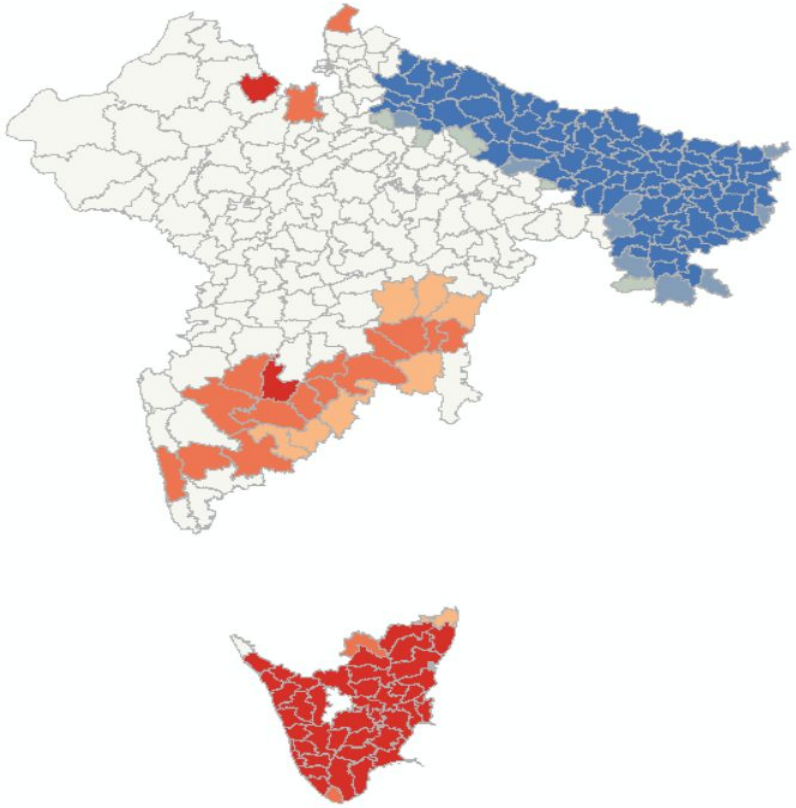
Hotspots for Electricity availability



The overlap between hotspots



Electricity hotspots



Upper Primary GER hotspots

Some Regression Analysis

Dependent variable -> Total Enrolment
Independent variables -> Schools with electricity facility, Schools with water facility, Schools with Mid day meal facility, Schools with Boys toilets, Schools with Girls toilet, Schools with Playgrounds

For top states :

Coeff	Value
Electricity	201.11
Water	1446.06
Boys toilet	-1975.22
Girls toilet	411.13
Playgrounds	65.99

For bottom states :

Coeff	Value
Electricity	76.03
Water	427.56
Boys toilet	-116.40
Girls toilet	-151.78
Playgrounds	-107.94

t-tests

Top states

T-value for Electricity (1.68) > t-critical at p-value 0.05 (1.66)

T-value for water (10.81) > t-critical at p-value 0.05 (1.66)

Similar with Girls and Boys toilet

For playground t-value < t-critical

Bottom States

T-value for electricity (1.46) < t-critical at p-value 0.05

For water t-value > t-critical

For boys toilet t< t-critical

For girls toilet, t-value> t-critical

For playgrounds, t-value <t-critical

The effect of different facilities have different effect on the enrolment in schools

Top States

With More electrified schools, enrolments increases, same with water facility and girls toilets.

Boys toilets surprisingly are negatively correlated.

However, the presence of playground in the school doesn't affect the decision to enrol in a school

Bottom States

Availability of water in schools increases enrolment.

Electricity and boys toilets have no effect.

Girls toilets surprisingly are negatively correlated.

Playgrounds again have no effect on enrolment.

But Midday is only available in Govt. schools !!

For top states:

Midday meal = 44.81

T-value (0.40) < t-critical

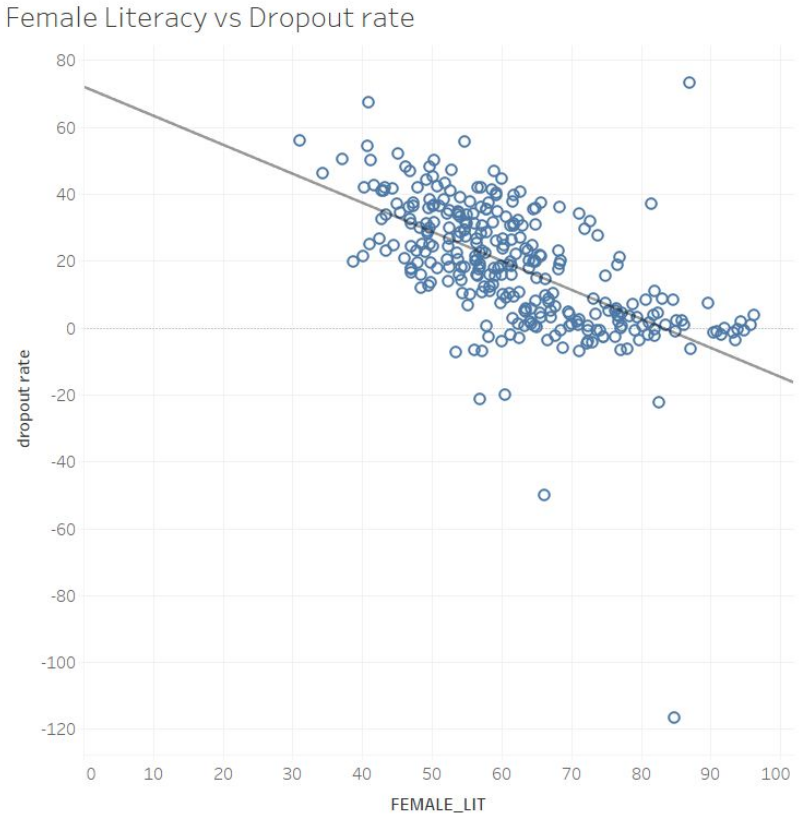
For bottom States :

Midday meal = 82.53

T-value (1.40) < t-critical

No effect of mid day meal on Govt School Enrolments

How does existing literate people affects Dropouts?

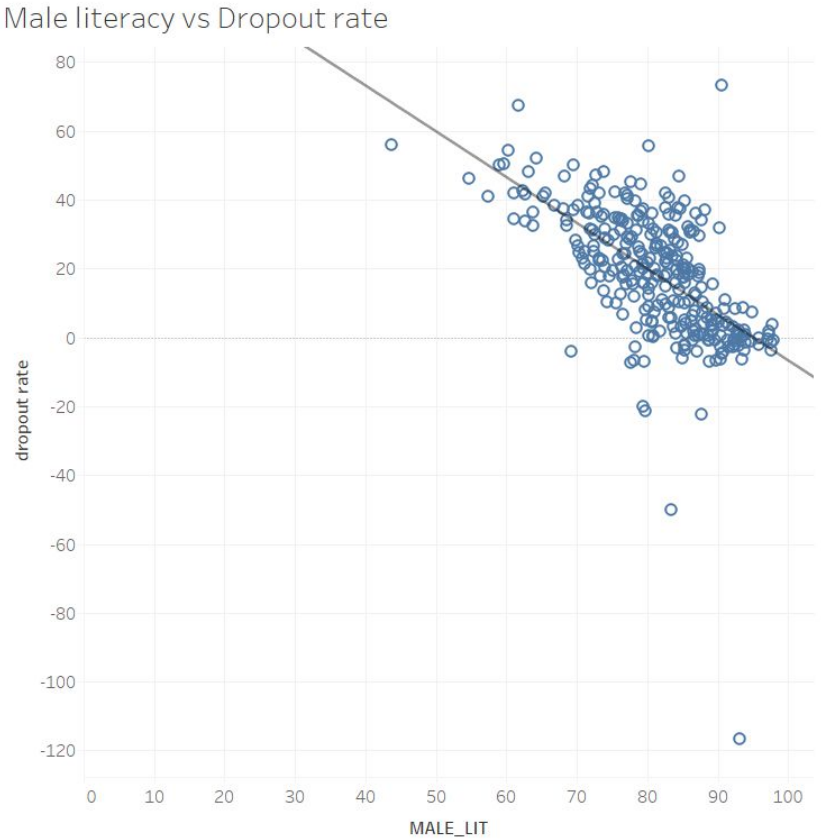


dropout rate = -0.867476*FEMALE_LIT + 72.0391

Coefficients

Term	Value	StdErr	t-value	p-value
FEMALE_LIT	-0.867476	0.0672432	-12.9006	< 0.0001
intercept	72.0391	4.25516	16.9298	< 0.0001

dropout rate = -1.33097*MALE_LIT + 126.371

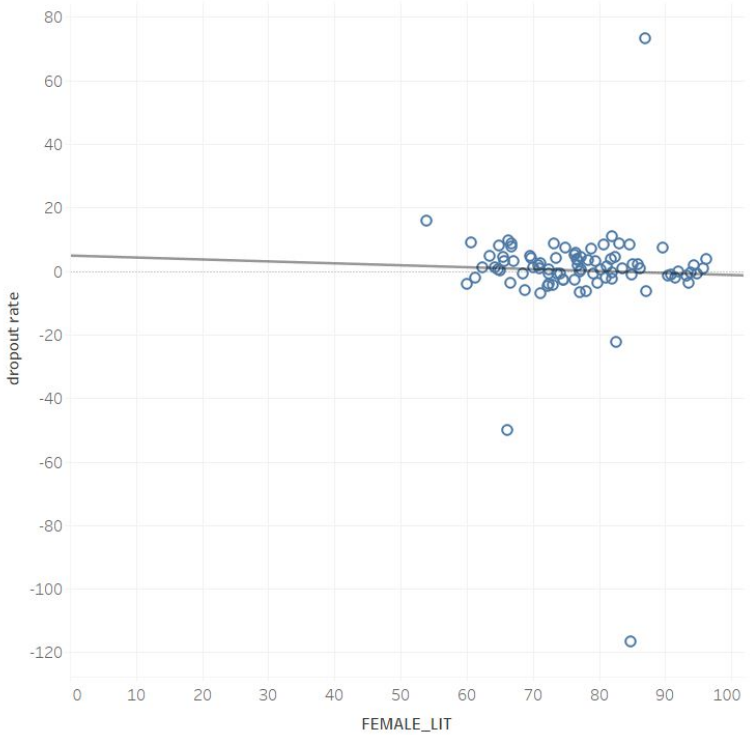


Coefficients

Term	Value	StdErr	t-value	p-value
MALE_LIT	-1.33097	0.0995307	-13.3724	< 0.0001
intercept	126.371	8.12714	15.5493	< 0.0001

Top states vs Bottom States

Female Literacy vs Dropout rate (Top)



Female Literacy vs Dropout rate (Bottom)

Educated parents, Educated children

Make 1% males of the district literate, and the dropout will significantly get reduced by 1.33%

Make 1% of females of the district literate, and the dropout will significantly get reduced by 0.87%.

However, in Top states, this effect of female literacy on dropout rates is lower than in the bottom states. Same pattern with male literacy.

Conclusions

- The difference between top states and bottom states in terms of facilities in schools is clearly visible and spatial differences can be observed whether in the availability of water or electricity.
- Availability of water increases the enrolment of students, so every state should ensure that their schools should have water facility.
- Electricity is also an important factor behind enrolments, so all the schools should be electrified.
- Contrary to belief, midday meal does not have any effect on the enrolments.
- Increasing Enrolments can lead to better literacy rate which further leads to less dropouts hence this reinforcement effect can help achieving the goal of full literacy.


Thank you.