

# Analysis of District Level Data to Identify Reasons for Student Dropouts

Initial motivation for this project was to correlate the data with the dropout rate using regression analysis. The data is huge enough to make it successful. But before directly jumping into it, checking the data carefully first was found to be a good idea. The Indian Education system in my opinion isn't so bad. But it isn't so good either. This assumption was mostly based on my own experience and it lacked the understanding of ground truth. Is the situation worse than what I thought? or is it better? We will see in a few moments.

The data that is used for this work is from <http://udise.in/drc.htm> website. The data is district level and it has total 661 districts. Another data that was used was state level data <http://udise.in/src.htm>. This state level data was used as a support for the district level data. The state level data for secondary schools is available from year 2012 and thus for a fair comparison, 2012 to 2015 data was used from both of these websites.

## Data Corrections:

- 1 Telangana State did not exist in some data and thus it was eliminated entirely from the data in the data cleaning process.
- 2 Some data fields were left blank for some reason and thus the data in those fields is assumed 0.
- 3 Arunachal Pradesh and Meghalaya had 4 extra districts in some data. For fair comparison, those districts were not taken into consideration as well.
- 4 Districtwise 2013-14 data has wrong state Id's, used it from 2014-15 data.

## Coordinates of DataPoints:

A State and District name combination was obtained from the data. The combination is fed into the geocoder library to retrieve coordinates of those states from Google API for plotting purposes.

Coordinates were saved to avoid calling the API again and again to avoid the excess API calls  
Locating the districts with the dropouts

The dropout rate for each district was obtained using the following formula

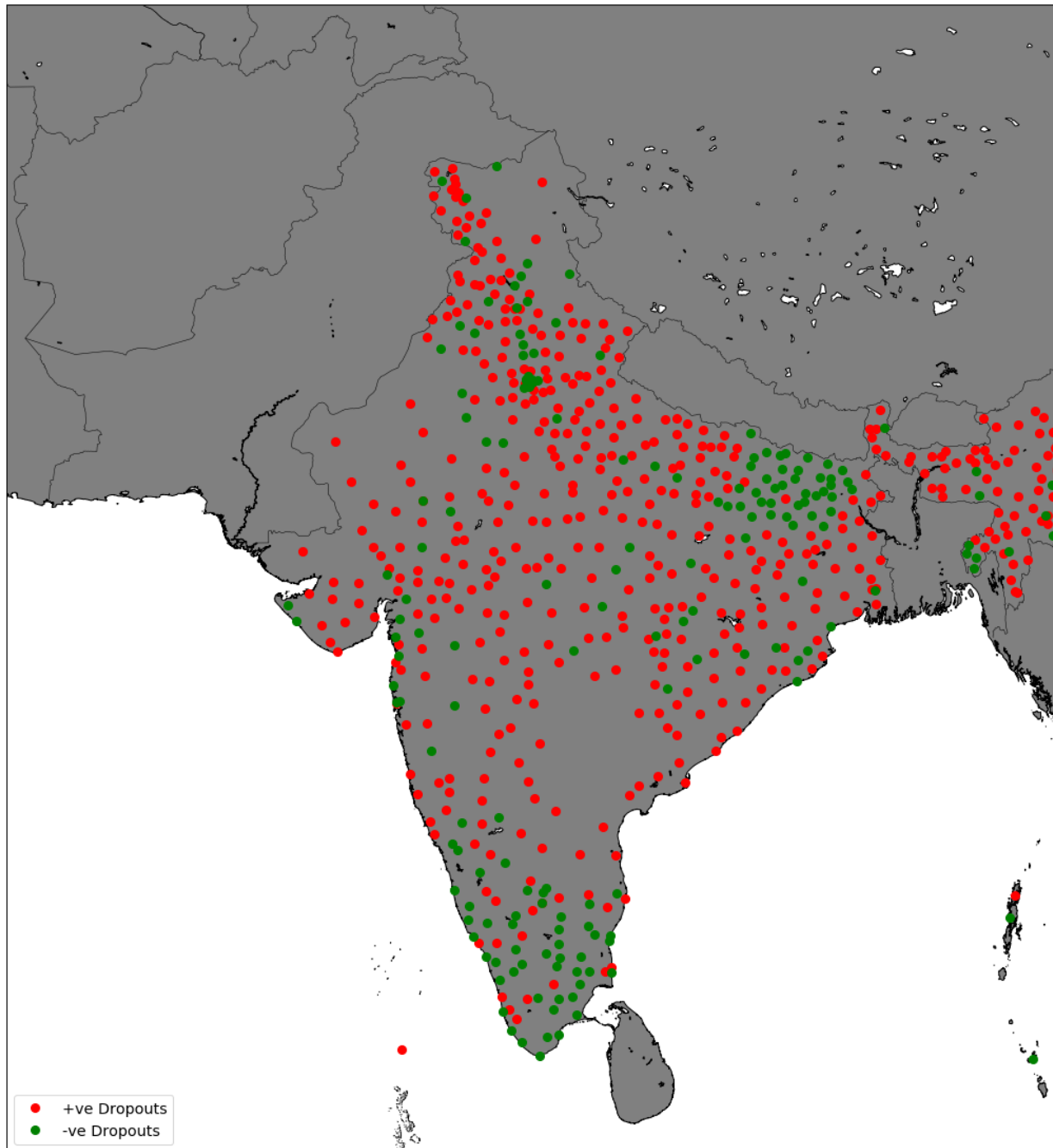
$$\text{Dropoutrate} = (\text{year14stdxpop} - \text{year15stdx} + 1\text{pop}) / \text{year14stdxpop}$$

where x

is the grade student enters. The dropout for each grade was calculated and summed up. The classification of dropouts from no dropouts/increased enrollment to +ve dropouts were obtained.

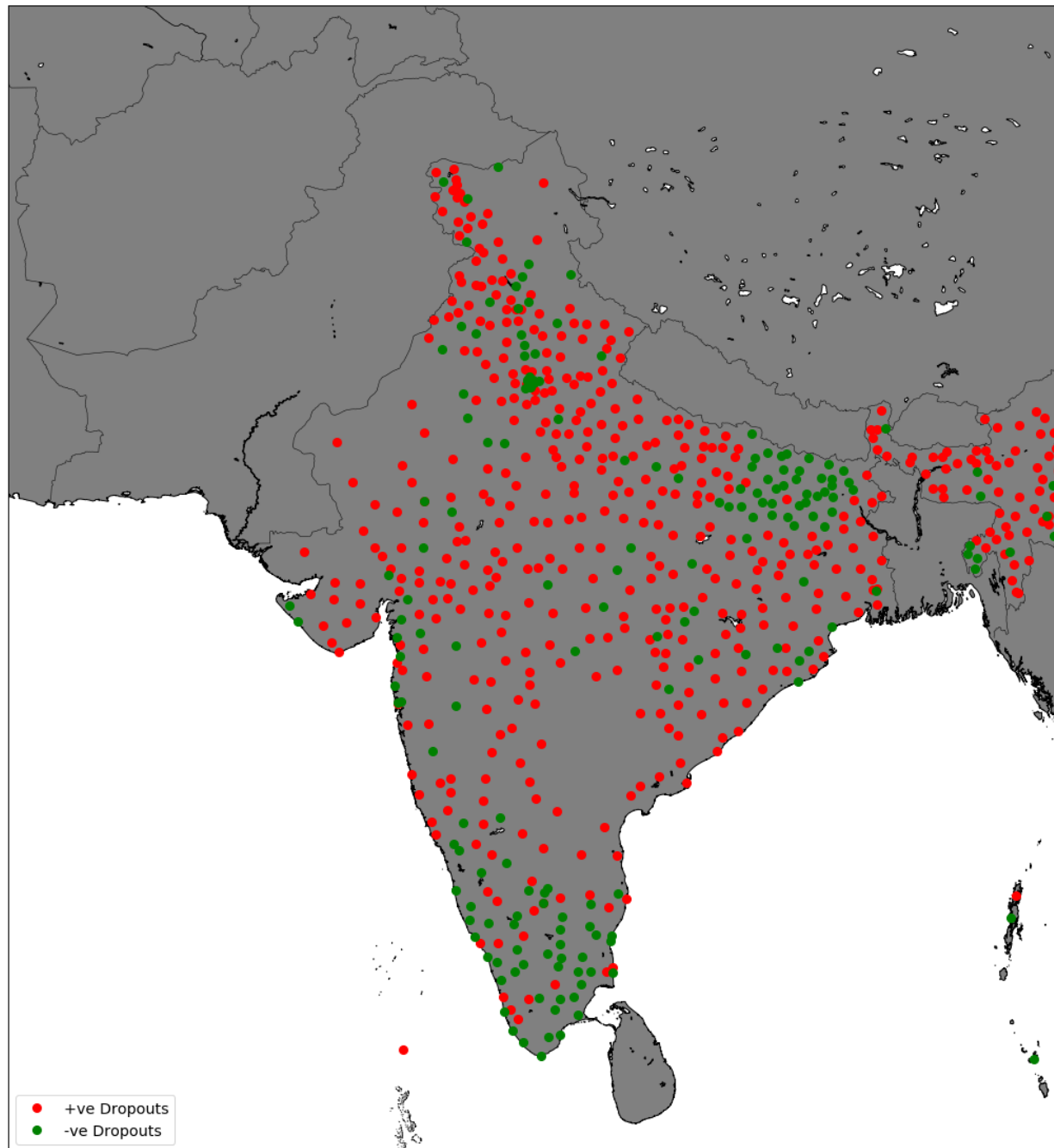
The obtained Dropout rate were marked in the form of markers in the Indian plot below.

*NOTE: The map used below is showing error/irregularities in the Jammu Kashmir region. I apologize for using this map but I had to use it as it was the only map that was available for this 'basemap' library. I request Indian government to provide a proper map for these developers so this error can be rectified.*



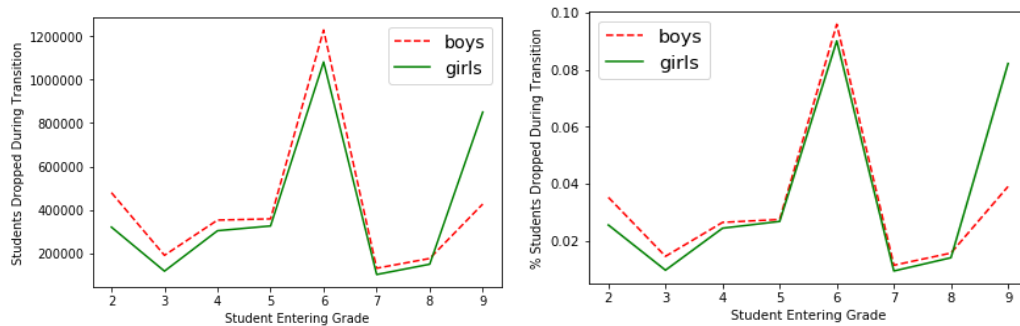
The above map shows the overall dropout rate of girl students. The **green** region suggests that the drop has not occurred or even the enrollment has increased. Whereas the **red** region suggests

the districts where the students has dropped out. The similar phenomena can be seen in the plot below which is for the boys student.



**Analysis and Initial Intuition:** Most of the states are covered with red spots which is bad. The situation isn't all bad as even for a single dropout, the red marker is placed. This is within a normal possibility that happens due to the migrant workers. The dropout situation is improved in the southern states that are Kerla and Tamilnadu. The literacy rate of those states is 93.91% and 80.3% which is higher than most other states and that could be the reason behind a good

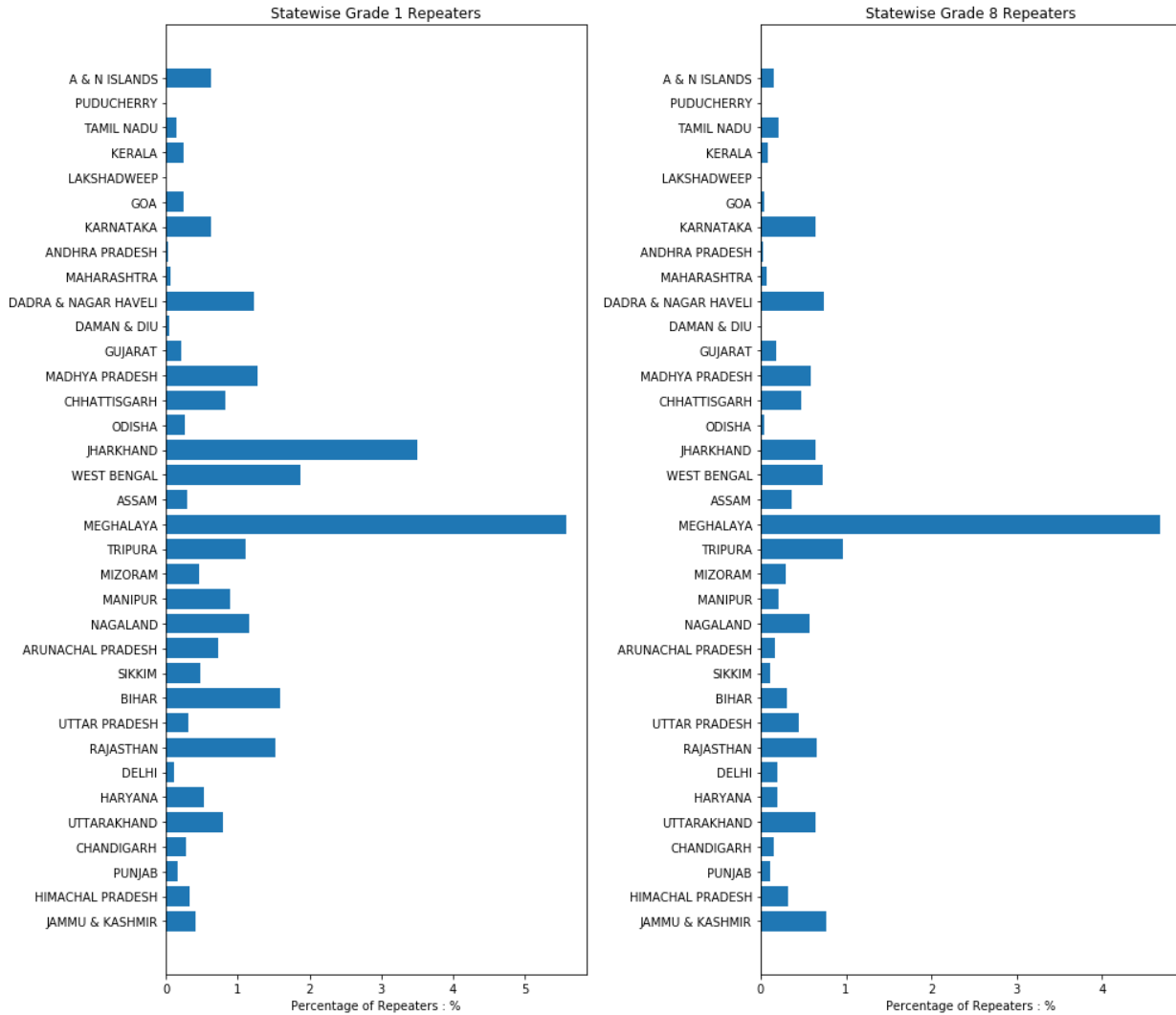
performance. Similar tendency is observed in Delhi with literacy rate of 86.34. An Anomalous behavior is observed in the state of Bihar which is not following this trend with the literacy rate of 63.82%. The reason behind that is unknown and it can be analysed later in this scope. For each passing grade, lakhs of boys and girls drop out from school. The plots below denotes this trend. It is scary to see that the trend is consistent. Also, when moving from 8th to 9th grade, a lot of girls quit schools. The number is almost twice than the drop out rate of boys.



Approximate percentage dropout from 1st to 9th grade: (a) Boys 26.5604267387 (b) Girls: 26.5604267387

This is a bad thing as every year about 2.61 crore students enroll in 1st grade of the schools. and only about 74% of them are going to reach 9th grade is a scary situation. The number is over 70 lakh. When that much amount of population can not finish schooling, the nation is really wasting a lot of potential right there. If this dropout rate is reduced, it will help a lot towards building a better nation.

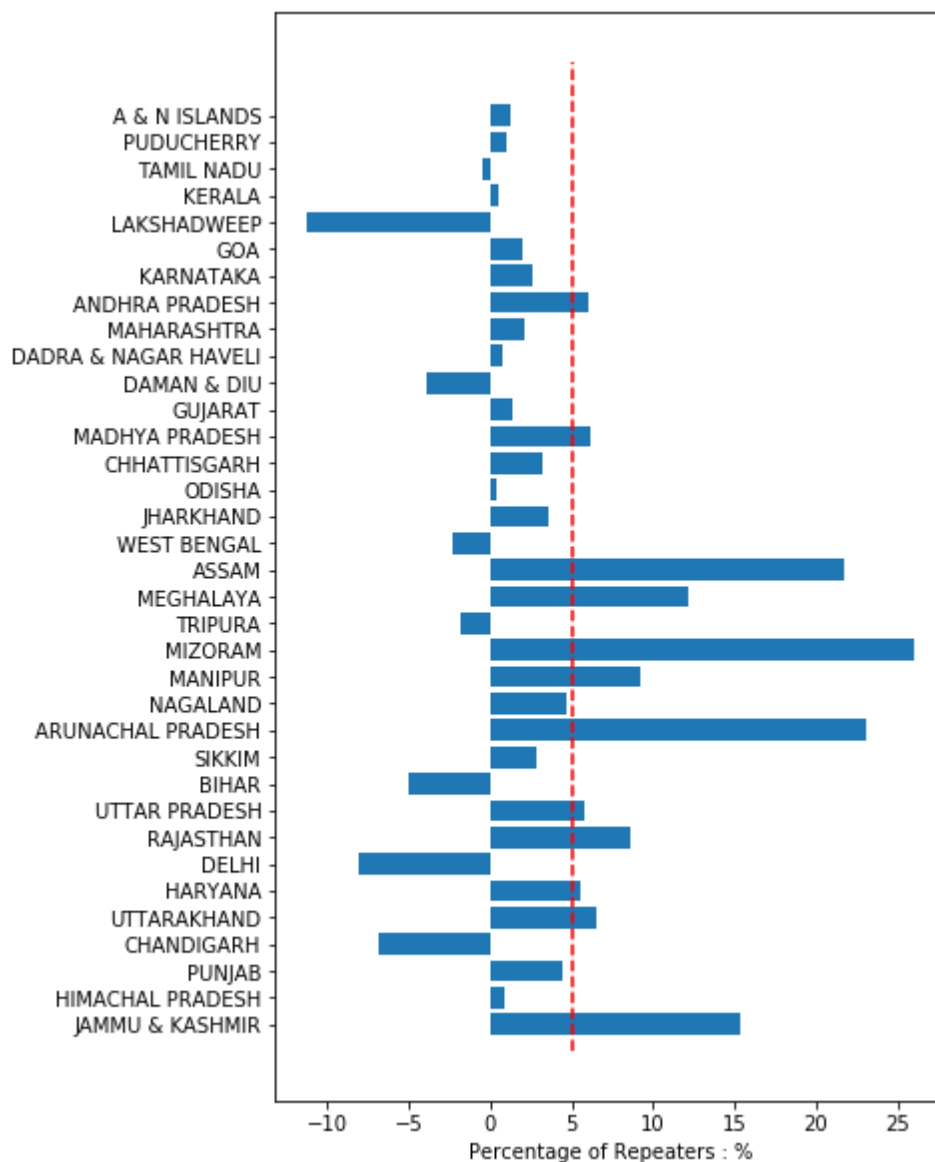
In the next few code snippets, we will analyze the dropouts and repeaters. Then we will find the factors that can help reduce the dropout or repeaters. Repeaters are still less harmful as they do not quit school.



## Analysis of Repeaters

The plots above shows statewise representation of repeaters in that state. Most of the states have least amount of repeaters. The State Meghalaya has almost 5% of students as repeaters, whereas the states of Maharashtra, Andhra Pradesh have the least.

Now, one can not judge the performance of state from only this factor. The State of Maharashtra refrains from failing students till 8th grade so the lower number isn't really surprising. Also, higher number of repeaters can also be considered as a better curriculum.



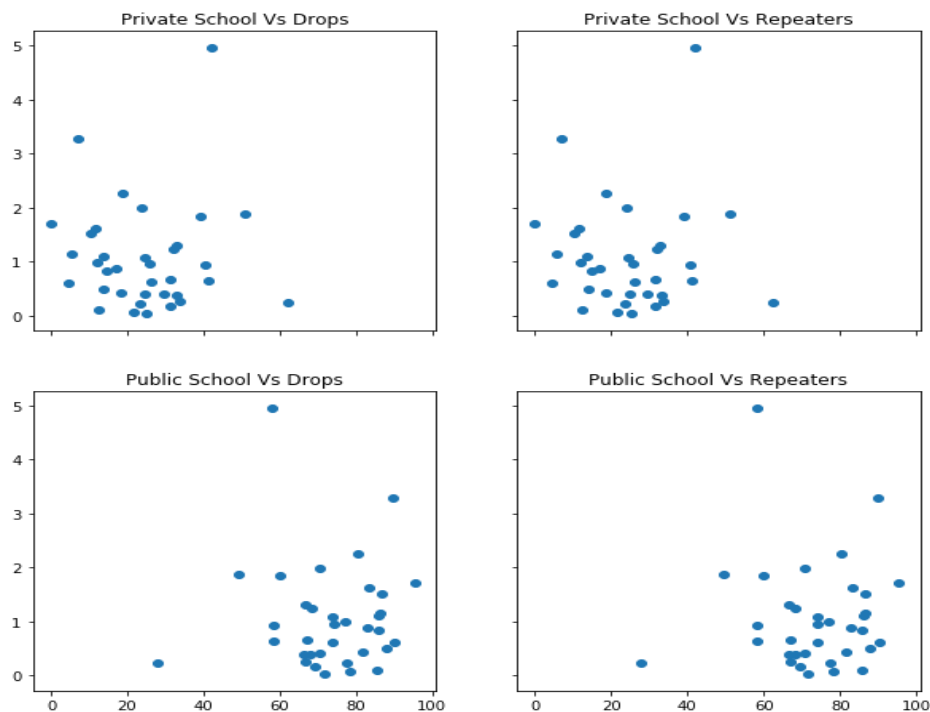
## Statewise Analysis of Dropouts

The above plot shows a statewise dropout rate. There are many states which shows no drop out or even a positive enrollment. Let us consider these states states with good educational support. 5% drop is the average drop and thus, anything above 5% is an anomaly. The states which shows above 5% drop needs some fix.

### The states are:

1.Jammu and Kashmir 2. Uttarakhand 3. Haryana 4. Rajasthan 5. Uttar Pradesh 6. Arunachal Pradesh 7. Manipur 8. Meghalaya 9. Assam 10. Chattisgarh 11. Andhra Pradesh

Now, except Tripura all the other states in North-East show a high dropout rate. It just tells us that the government has been ignorant about these states when considering the education. So More attention to these states is necessary.

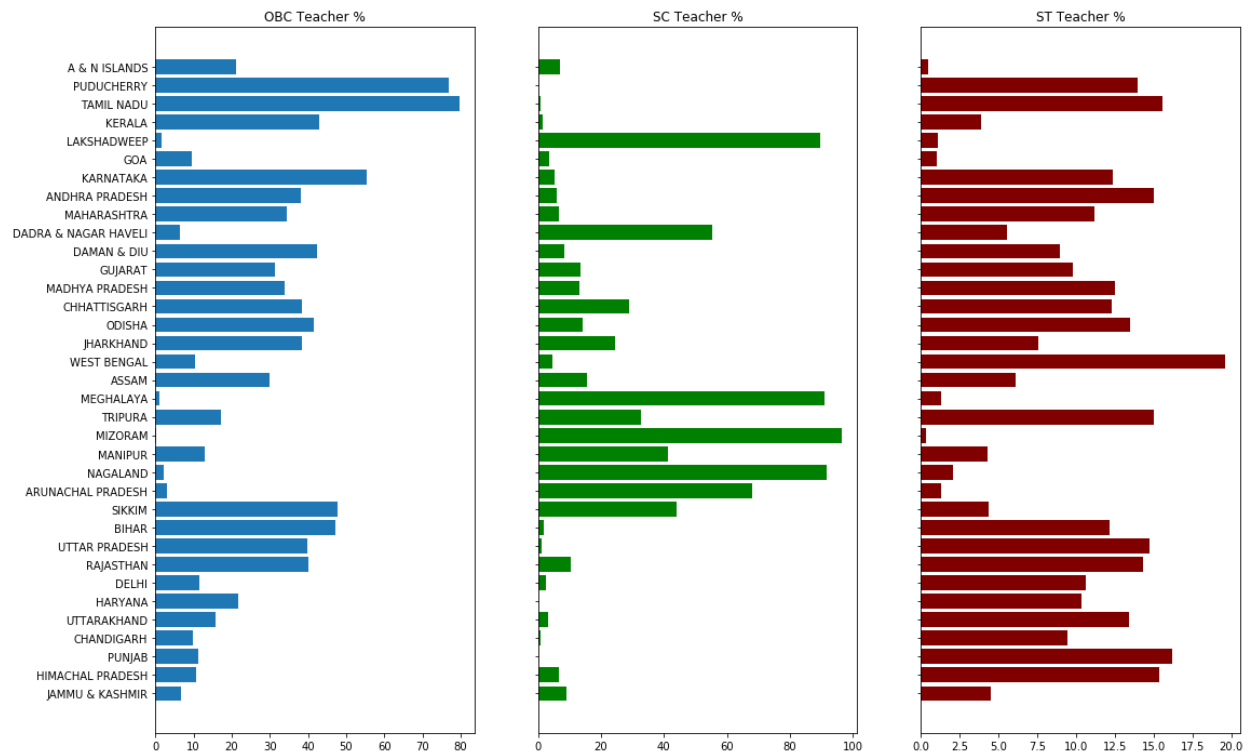


### Is there a correlation between some features.

The above plots show the features plotted against one another. There is some weak correlation, both positive and negative amongst these features but it isn't really clear. We will see more about it in the later sections.

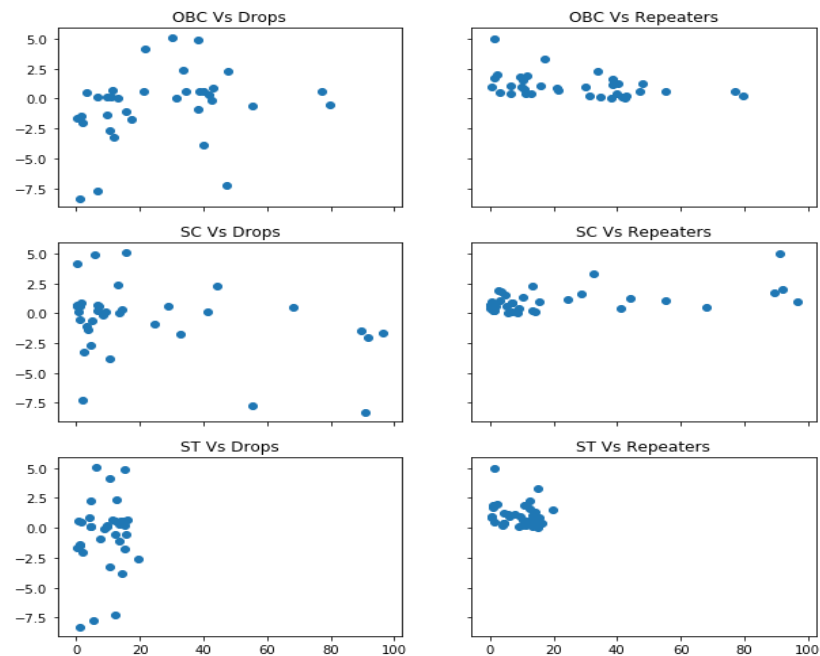
### Caste wise Distribution of teachers

The plots below show percentage of caste wise teachers populations in each state. The North Eastern states have most SC teachers. But other states have almost no SC teachers. SC population in India is 16.6% but the same amount in teachers is not there. The ST population though represents the national average. 41% OBC population is also not being represented in the teacher population. These factors may or may not represent the Dropout rates. We will see more later



## Correlation of Drops with Caste

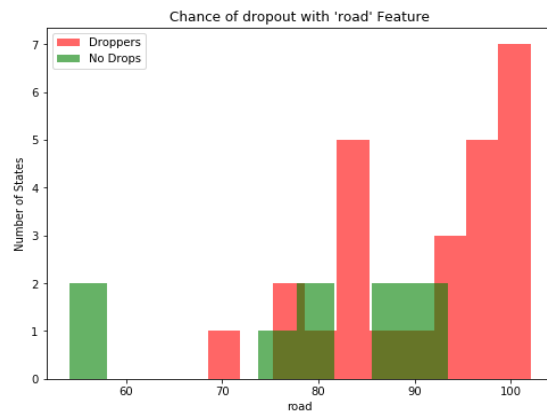
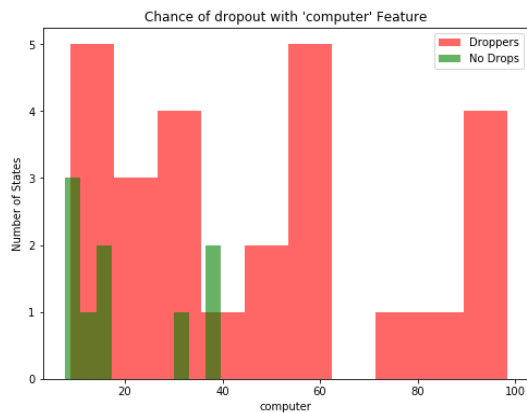
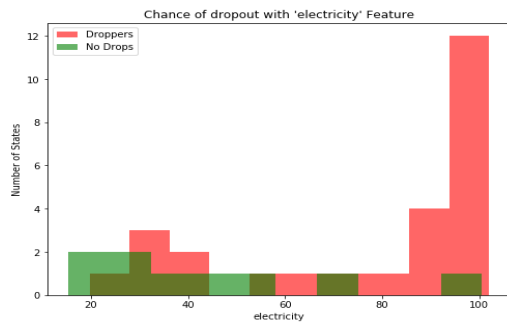
From the plots below it is seen that there is no correlation between Caste and the Dropouts and Repeaters. Thus, the Caste Representation can be improved without hindering the results.





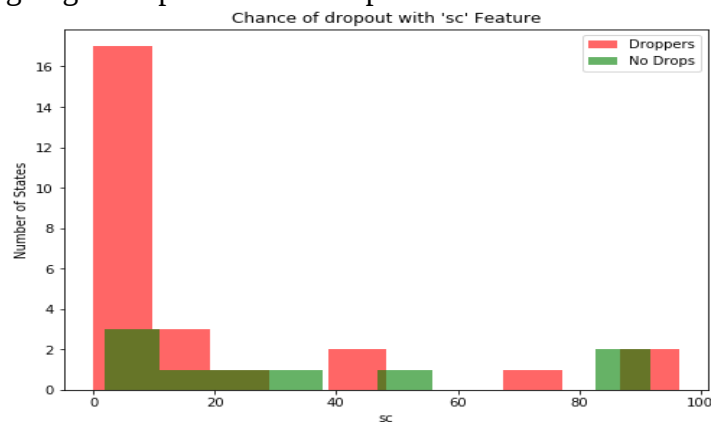
## Comparison of other factors

The plot below shows the highest drop outs have occurred even with 100% Electrified Schools though lower electrified states have lower dropouts.

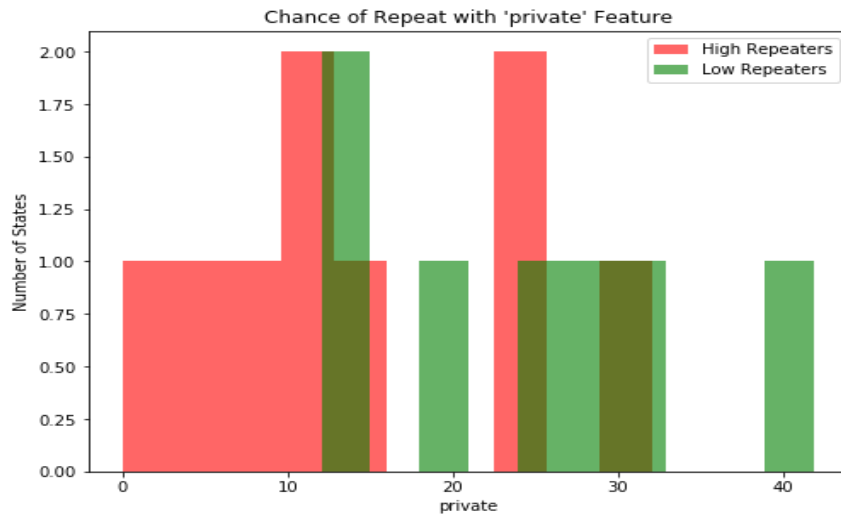


## Progressive Facilities do not help Dropout¶

The states which had low electrification, low computers, schools without roads still have lower drops than the schools which had all of these good. So funding more in these facilities is not going to help reduce the dropouts.



The plot above show the Schools with Lowest SC Teachers have highest amount of Dropout Rates.



Coming to the Private schools, the states which hav more private schools and also more SC teachers, the Repeater rate is low. The same trend is not observed with OBC's and ST's.

## Conclusions and Suggestions

The following things can be improved if schools has to be improved both socially and with quality.

- 5 Dropouts is an important issue. 30% Population is out of the schooling system and the important man power is being wasted.
- 6 The factors that government has been doing such as Mid-day meal, electrification are good things but are not enough to reduce the dropouts or repeaters. The other factors must be researched to reduce these problems.
- 7 The states which are not performing well are from North East and Kashmir. North East has been ignored by government most of the time. More attention has to be given to these states.
- 8 The caste wise distribution of Teachers does not represent the actual population representation in most cases and thus this balance has to be re-established. The Caste is not having any direct correlation with the performance and thus it is safe to use some work to balance out these things. On the other hand, the states with higher dropout has least number of SC teachers. Fixing that is a necessity.

*Note: The regression analysis was not completed due to lack of time. Though it is possible to find an equation to correlate these features with the dropout rate.*