

# Identifying State Literacy Rates

Using CART Models

Tristan Larose | Faculty Mentor: Dr. Chantal Larose

## Introduction

This project investigates factors that impact an area's education rate. This includes:

- Population
- Population Density
- Average Annual temperature in Fahrenheit
- Poverty Rate
- Crime rate

If we were to determine a solid relationship between an area's knowledge and other factors, we could engineer some of the factors to better suit education and/or see what other effects a rise in education could have in an area. For example, if there is a connection between temperature and education, adjusting the temperature in class could help improve it.

## Data Organization

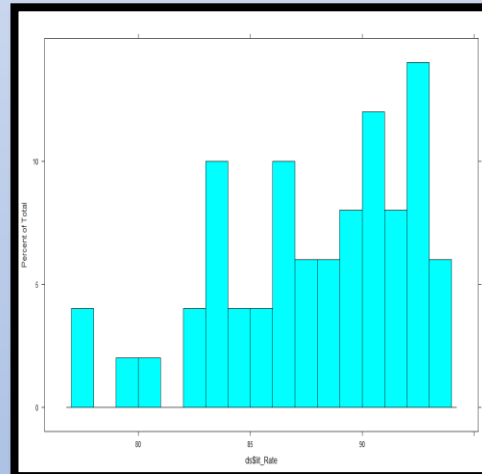
We created a categorical variable from a numeric one by binning the data. Based on the data distribution (shown, right), we made bins that are meant to label the literacy rates with measures of "low", "mid", or "high". The bins are defined as:

- $x \leq 85\%$  for "Low"
- for "Mid"
- $x > 90\%$  for "High".

We also had to calculate a baseline of performance for our model. In this case, the most accurate guess someone could make is "high" therefore the accuracy is:

$$\frac{\text{high}}{\text{total}} = \frac{20}{50} = 40\%$$

In order to be considered useful, the resulting model needs to have an accuracy higher than this.



## Conclusions

The overall accuracy for this model is a 26% improvement on the baseline model, a significant improvement over our baseline model.

Looking at the resulting model (shown, bottom left), we can see:

- Any state with an average temp of 59F or higher is about 78% likely to have a low literacy rate
- Of the states with temperatures below 59, any with populations higher than 2,700,000 are 53% likely to have a literacy rate and any with lower populations 75% are likely to have a high one

It's worth noting that these rules are not absolute. Alaska, for example is a state with a low literacy that breaks both rules.

It's also worth noting that only two variables were used in the final model. Recall that this model took multiple other variables into account which could have had an effect. This includes all of the variables mentioned in the introduction, but only temperature and total population were decent predictors.

Overall, this model is a useful method of determining a state's literacy rate via its population and annual temperature. Now there's the question "Could this be of use in today's education?"

## Statistics

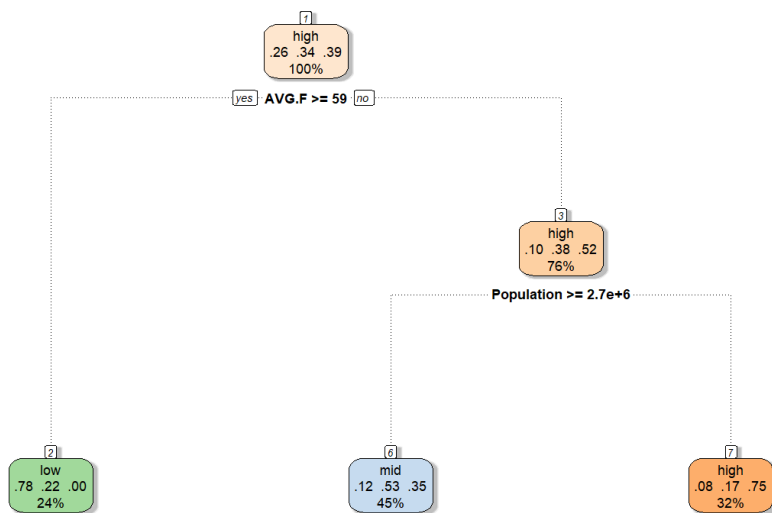
Overall Accuracy: 66%

Low: Accuracy = 77%  
Sensitivity = 70%  
Specificity = 93%  
Mid: Accuracy = 53%  
Sensitivity = 69%  
Specificity = 68%  
High: Accuracy = 75%  
Sensitivity = 60%  
Specificity = 87%

	pred.10		
	low	mid	high
low	7	2	1
mid	2	9	2
high	0	6	9

The specificity rates were usually the highest of the three statistics and the sensitivity was always the lowest statistic of the data. High specificity with low sensitivity means these models are biased toward false positives, this means our model is better at predicting what literacy rate a state isn't rather than the rate it is.

The main problem with this model is that it is geared toward labeling a state's literature rate "mid" when it's actually "low" or "high". This might mean that the "mid" section that was defined when binning the variable was too large and if there was more allocated room to the two extremes, the overall accuracy would increase. The overall accuracy for the extremes, however, was higher than the total accuracy. Given that extremes in literacy rates are slightly more important to track, this seems ok.



Rattle 2020-Dec-10 20:51:51 Family

## References

Population and literacy rate:

<https://nces.ed.gov/naal/estimates/StateEstimates.aspx>

Population Density:

[https://en.wikipedia.org/wiki/List\\_of\\_states\\_and\\_territories\\_of\\_the\\_United\\_States\\_by\\_population\\_density](https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States_by_population_density)

Poverty Rate:

[https://en.wikipedia.org/wiki/List\\_of\\_U.S.\\_states\\_and\\_territories\\_by\\_poverty\\_rate](https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_poverty_rate)

Temperature: <https://www.currentresults.com/Weather/US/average-annual-state-temperatures.php>