Predicting Factors of Teen Vape Use

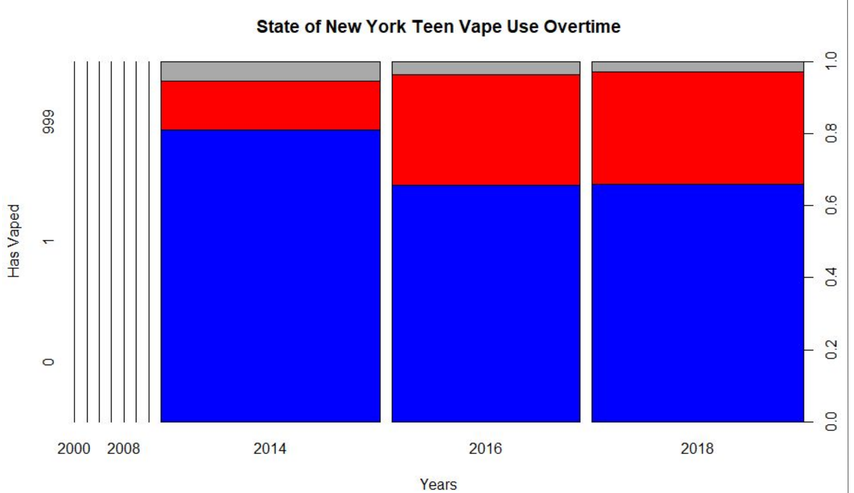
Marisa Papagelis, Chandler Pettigrew, Sophie Rosas-Smith, Viki Zygouras
Wellesley College

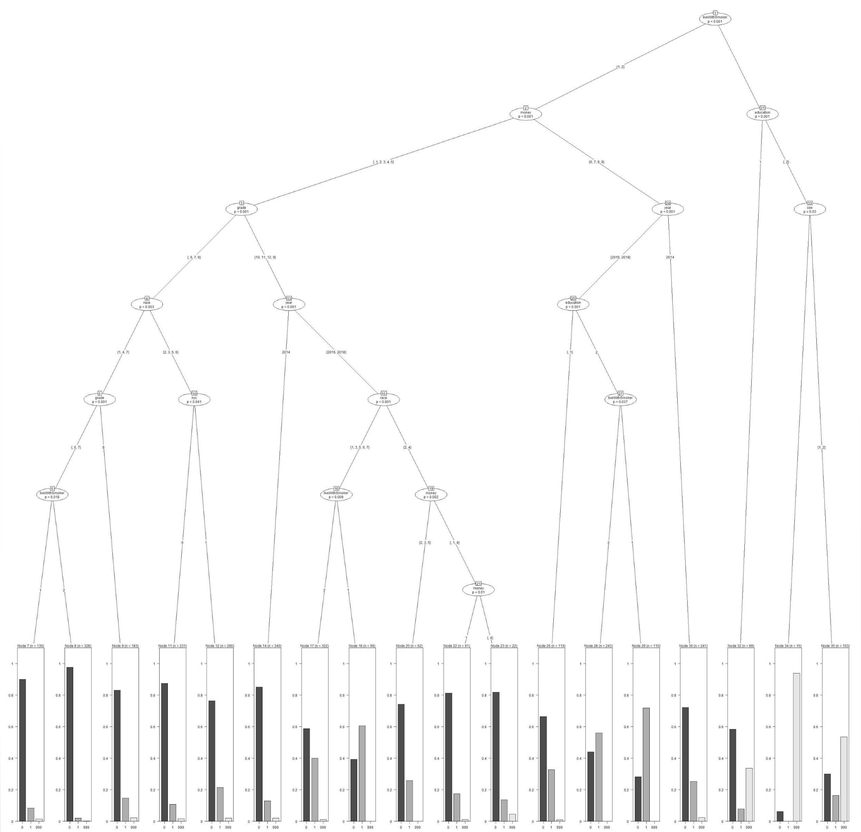
Motivation

- Vape use and related illnesses among teenagers in the United States have been rising
- It is a matter of public health to investigate and understand the factors that drive teen vape abuse

Data Cleaning Methods

- Used student responses from the New York Youth Tobacco Survey
- Chose eight categorical variables hypothesized to contribute to vape use
- Cleaned and dealt with missing values
- Limited the dataset to include survey information from 2014 - 2018



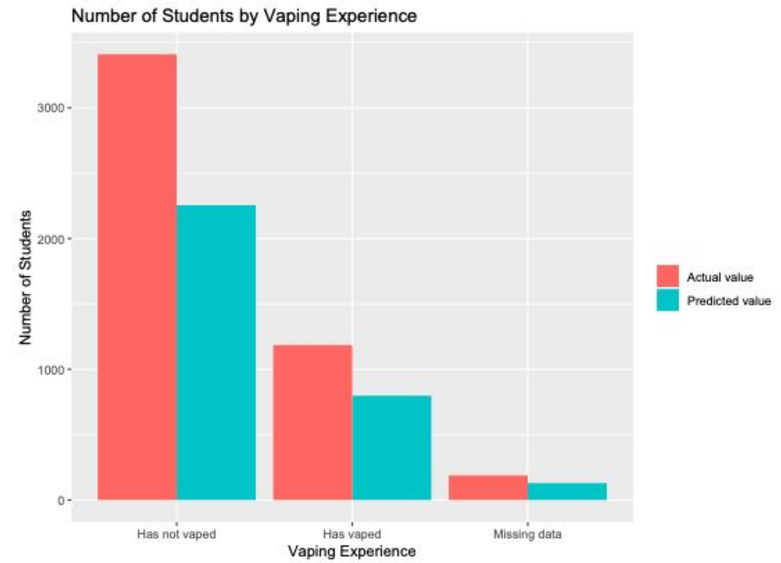


Model

- Converted the eight variables to factors
- Trained and tested a binary classification tree with randomly sampled groups of data
- Found that the most significant factor for predicting teen vape use was **living with a smoker** (p-value<=0.001), among other influential factors

Validation

- Model Prediction Success Rate: ~74%



$$Accuracy = \frac{Number\ of\ correctly\ predicted\ values}{Total\ number\ of\ predicted\ values}$$

Conclusion

- Most influential factors:
  1. Living with a smoker
  2. Attending High School vs. Middle School
  3. Student Income Per Week
- Results are not generalizable to the larger United States
- Future Work: adding more factors to the model, collecting new data