

# A Visualization Tool for Asthma Prevalence

Carter Schmidt, Dr. Ebelechukwu Nwafor

## Abstract

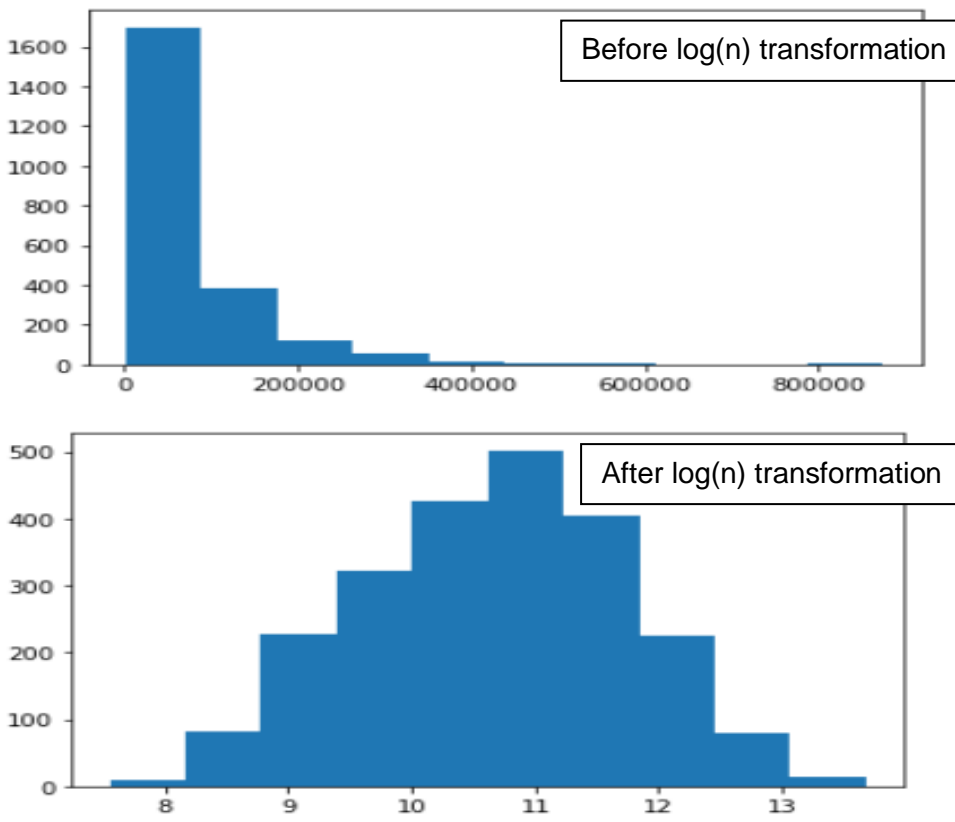
In an era of unprecedented progress in various areas such as technology and medicine, it can be difficult to understand the direction we are headed in as a society. Due to the high value of understanding our future, the collection and use of big data has been rapidly increasing. Despite increased medical innovation, one area that has historically remained an issue is the prevalence of asthma. Hundreds of thousands of people are diagnosed with asthma every year, and many are susceptible to severe asthma attacks. Therefore, it is important that we are able to visualize the future of asthma in our country in order to gain information about the growing prevalence of this debilitating condition.

Thankfully, with machine learning techniques, visualization software, and abundant data on the United States population, we can get a better idea of what the future will look like in many ways. In this project, we created a data visualization application to represent machine learning predictions on the future of asthma prevalence up to 2024. To make these predictions, a polynomial linear-regression model was trained over tens of thousands of data records, resulting in a high degree of accuracy. In order to visualize these predictions and sort based on features such as year, state, and age, we created responsive charts with the JavaScript library “D3”.

## Motivation

- To create an interactive visualization tool to understand trends in asthma through the near future.

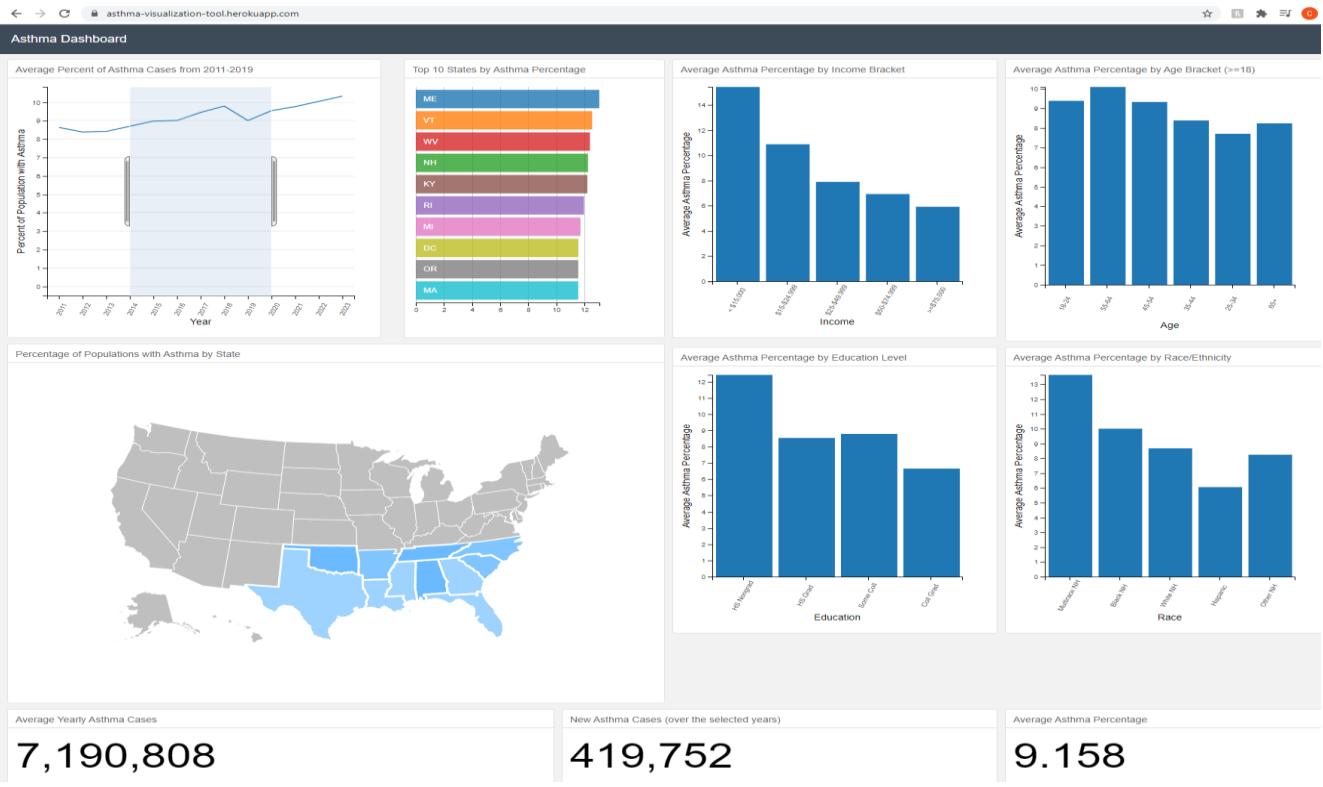
## Machine Learning Model



One-Hot Encoded Data														
Year	Percent_Cases	Num_Cases	AK	AL	AR	AZ	CA	CO	CT	DC	DE	FL	GA	HI
2011	16.1	76125.0	0	1	0	0	0	0	0	0	0	0	0	0
2011	9.5	63125.0	0	1	0	0	0	0	0	0	0	0	0	0
2011	5.0	39507.0	0	1	0	0	0	0	0	0	0	0	0	0
2011	5.9	24203.0	0	1	0	0	0	0	0	0	0	0	0	0
2011	5.5	33777.0	0	1	0	0	0	0	0	0	0	0	0	0
2011	14.1	5680.0	1	0	0	0	0	0	0	0	0	0	0	0
2011	9.2	5505.0	1	0	0	0	0	0	0	0	0	0	0	0
2011	7.6	7289.0	1	0	0	0	0	0	0	0	0	0	0	0
2011	5.9	4556.0	1	0	0	0	0	0	0	0	0	0	0	0
2011	7.5	12474.0	1	0	0	0	0	0	0	0	0	0	0	0
2011	14.4	79977.0	0	0	0	1	0	0	0	0	0	0	0	0
2011	11.4	97681.0	0	0	0	1	0	0	0	0	0	0	0	0
2011	8.7	96049.0	0	0	0	1	0	0	0	0	0	0	0	0
2011	7.6	45565.0	0	0	0	1	0	0	0	0	0	0	0	0
2011	7.0	70871.0	0	0	0	1	0	0	0	0	0	0	0	0
2011	16.1	61283.0	0	0	1	0	0	0	0	0	0	0	0	0
2011	9.4	40211.0	0	0	1	0	0	0	0	0	0	0	0	0

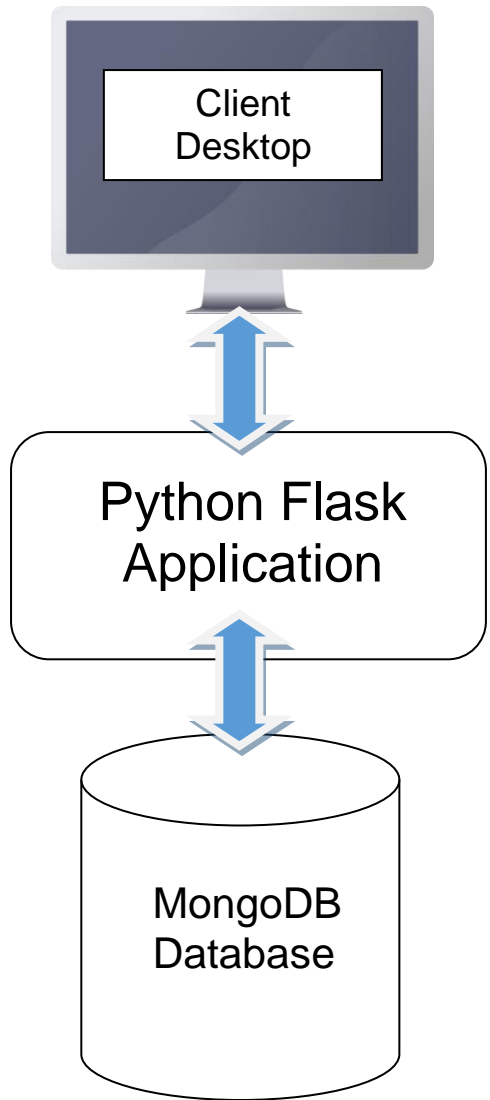
- One-Hot encoder is used to allow for a multiple linear regression on the feature set.
- Data is transformed to resemble a normal distribution by transforming the prediction values with log(n).
- Polynomial features are added to the data with the PolynomialFeatures method from sklearn.
- The Linear Regression model is fit on the modified data, then used to predict future asthma data.

## Implementation



- Dashboard contains the asthma cases and asthma percentages for the features: year, age, income, education, race, and state.
- Includes relevant calculations like “New Cases” based on the year selected, and the “Average Asthma Percentage” based on selected features.

## System Architecture



- The visualization tool is hosted on Heroku’s cloud service
- Flask application then gathers data from MongoDB, and allows the JavaScript code to access the data in JSON format
- Front-end created with HTML, CSS, and JavaScript (D3.js, dc.js, crossfilter)

## Conclusion

- Successfully developed the interactive visualization tool for asthma prevalence
- Predicted an overall increasing trend in asthma with regression models

## References

- Centers for Disease Control and Prevention. (2020, August 31). CDC - BRFSS. Centers for Disease Control and Prevention. <https://www.cdc.gov/brfss/index.html>.

Scan the QR code to access the visualization tool website!

