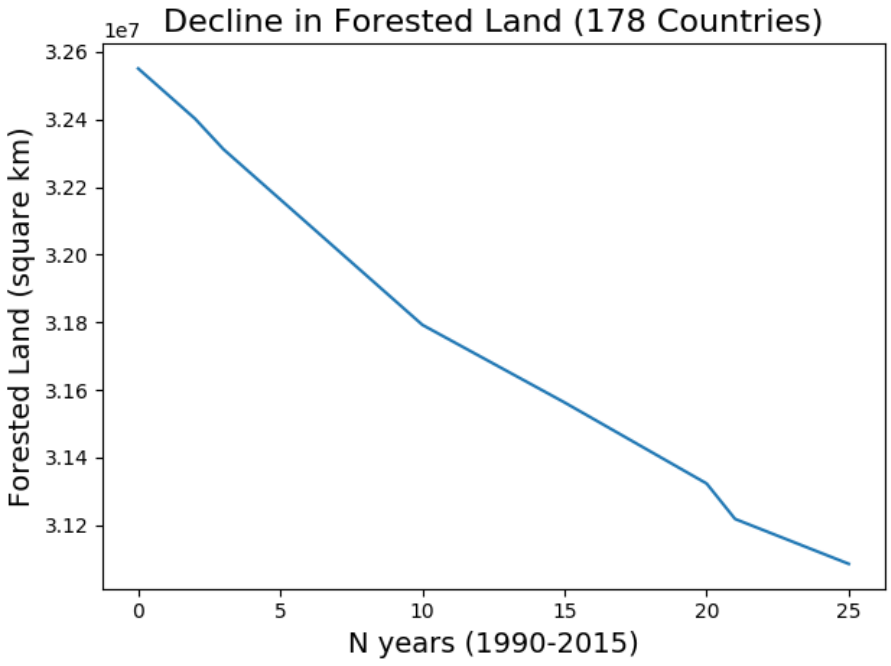
Investigation of Factors Leading to Deforestation
Data Incubator Fellowship Project Proposal
Jeremy R Jacobsen, July 2018

Due to human energy consumption, atmospheric carbon dioxide concentration recently surpassed 400 ppm. The trajectory of atmospheric CO2 has been positive since the industrial revolution and 400 ppm amounts to over a 30% increase in atmospheric CO2 since the 1950s. Although mitigation efforts have been devoted primarily to reducing carbon and methane emissions, there is another important factor that has arguably not garnered sufficient attention. In addition to the release of carbon dioxide, humans are also busy decimating carbon sinks that could otherwise be actively sequestering atmospheric carbon; namely trees.

It's not uncommon to hear alarming statistics about rain forest destruction in Brazil, but deforestation is in fact a global phenomenon. The question I'll attempt to answer is: What are the most important factors impacting global deforestation? The motivation for working on the project is to determine if there are factors which could be influenced by grass roots movements, governments, or companies to reduce the rate of deforestation. As Americans, we've been told that we can be effective by avoiding red meat and products which contain palm oil but I'd like to quantify how much of an impact this can actually have, and determine whether we can affect other factors that may have an equal or greater impact. I'd really like to know what a person like myself could do about it without traveling thousands of miles and dedicating my life's work to a singular cause.

The World Bank provides indicators on Agriculture & Rural Development which can be accessed here (https://data.worldbank.org/topic/agriculture-and-rural-development). In addition to historic data on deforestation they also offer co-temporal historic data on potentially related factors such as (land area, poverty, and agricultural development). Before analyzing anything else, let's take a look at global deforestation data.

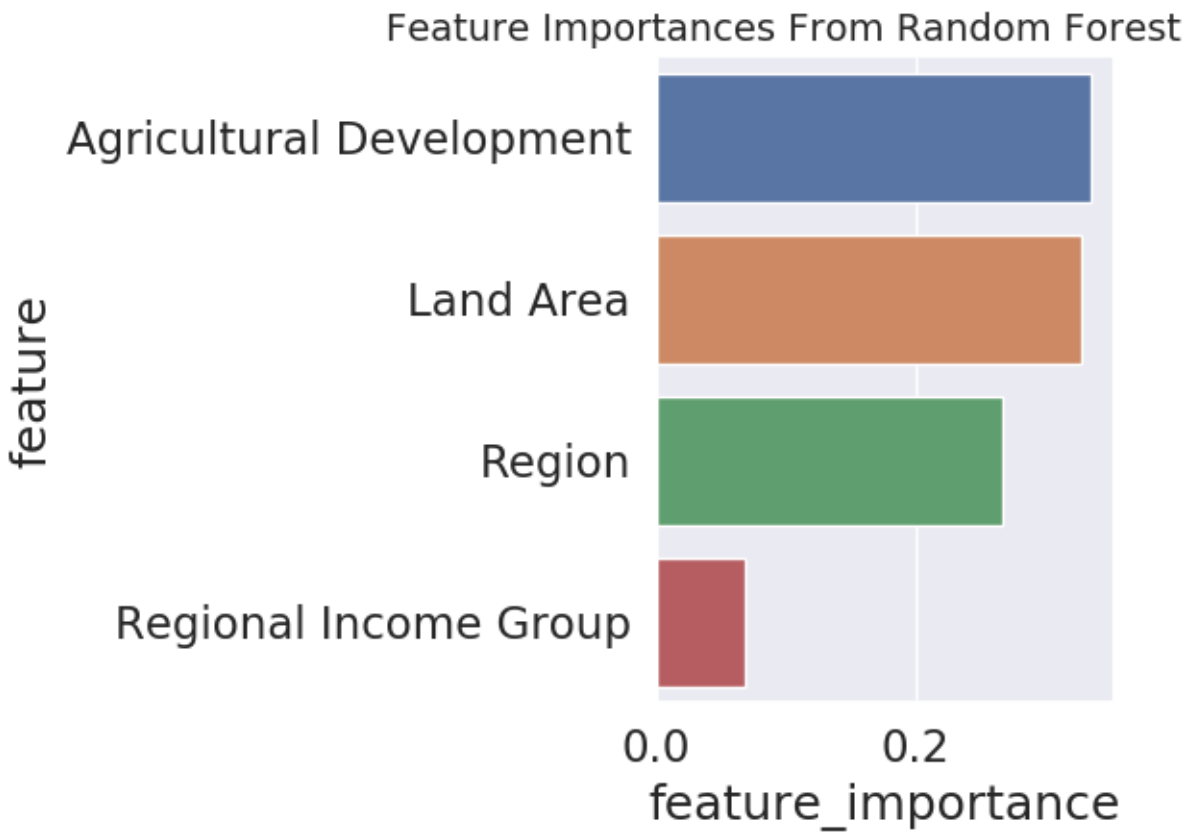


To put the data into perspective, the decline in forested land (~1.4 million square kilometers) is approximately twice the area of the state of Texas. That's pretty alarming for anyone who has driven across that state. It is important to note that countries with any missing data between 1990 and 2015 were excluded from the analysis, so the figure is almost certainly an underestimate. If we assume a standard tree is a few meters across this probably works out to hundreds of billions of trees and other carbon wicks. It's unlikely that everyone on earth can be convinced to plant several trees a year though that could offset the problem.

Now let's see if we can work out what factors are important for determining the rate of decline in forests. Since we're looking at deforestation it seems fitting to use a random forest model to investigate. Sklearn RandomForestRegressor class can return a feature importances metric so I'll use this to rank the most important factors. For time series variables (percent forestation, agriculture, and land area), I've simply calculated a slope for each country. The target variable in the model is the forestation slope (rate) calculated by

$$\frac{\%forestation(2015) - \%forestation(1990)}{25}$$

I've used numerical mapping to encode the income and region variables



It would seem that agricultural development is a primary factor in the rate of deforestation. This is already known to be the case. Land area is probably a shoo-in because countries with more land can more easily justify its exploitation without leading to perilous outcomes. It also seems fairly obvious that regional location would be a good predictor of deforestation rate since there aren't any forests in for example Qatar, but there are extensive forests in Brazil. There is some mystery surrounding this indicator though, since some countries with large forests such as Canada practice reforestation (note: Canada is still experiencing deforestation, albeit at a much slower rate than Brazil).

Identifying factors that the average American can impact in order to reduce deforestation is a difficult question to answer without having domain specific intuition. From crunching the data above I feel I'm now a bit closer to acquiring intuition for the problem. As a next step I would like to investigate the relationship between deforestation and region in greater detail. For that I will need to source additional data from each region and stack that into the training set.