DS4A / Empowerment: Project Description

Team 39: Manas Ramesh, Zander Sargeant, Elliott Yoon, Colin Senat, Jotham Kriakos

---

We will explore the topic of deforestation, its industrial and environmental impacts, and why those can have disastrous negative effects. More specifically, we wish to draw connections between deforestation in a global scene and the economic and environmental context that surrounds it throughout time.

At the most basic level, forests are imperative for the health of our planet. Densely populated rainforests provide necessary habitats and food for millions of animals to survive. Trees convert our excess amounts of carbon dioxide into the oxygen necessary for living beings to survive and help regulate greenhouse gas levels. Moreover, deforestation contributes to the addition of greenhouse gases to the atmosphere and can increase the amount of soil erosion. We want to identify a correlation between deforestation and quantifiable benchmarks of societal well-being to make the argument that deforestation should be proscribed. Examples of such benchmarks include economic loss, health issues, and community displacement.

Introduction

Business Problem

Deforestation is a widespread practice often used as a means to an end in industrial, residential, corporate, and other settings. In particular, the deforestation of the world's largest rainforest - the Amazon rainforest in Brazil - is increasing each year. Will deforestation negatively impact the economy of Brazil over the next few decades?

Business Impact

With forests being razed every day in Brazil, economists say that the exponential decay of the GDP in Brazil will rise because of the natural resources, jobs, and other factors that deforestation provides. It provides the unemployed new jobs such as farming, forestry jobs and lumberjacking. The amount of natural resources such as palm oil, logs, and paper can bolster the dying economy

of Brazil. But this bolster in GDP will suddenly drop, potentially putting Brazil through a horrible national depression that they are not prepared for. Our purpose in predicting the loss in GDP and the economy will show how deforestation seems perfectly fine right now, but will be devastating in the future. This will not only show how deforestation negatively affects Brazil's economy, but for countries all around the world.

Data

- GDP dataset(s): Consist of hundreds of different measures of economic stability in Brazil by year from 1960s to 2020. This is extremely useful because it gives so many different ways to measure how the Brazilian economy is doing at a given time and can be easily compared to other datasets.
- Deforestation dataset: Measures deforestation (by percent change) per country per year. We can take the data from this that is relevant to Brazil for analysis, possibly in conjunction with the gdp datasets.

  From the dataset source: Deforestation area (km²) by year and state, from 2004 to 2019. The data are public and were extracted from INPE website on December 16th 2019. It was already aggregated, so, no data process was made. Program: PRODES (Programa de Monitoramento da Floresta Amazônica Brasileira por Satélite, or Brazilian Amazon Rainforest Monitoring Program by Satellite). Methodology: maps primary forest loss using satellite imagery, with 20 to 30 meters of spatial resolution and 16-day revisit rate, in a combination that seeks to minimize the problem of cloud cover and ensure interoperability criteria.

Methods

- Visualizations(Scatterplots, heatmaps, line plots,etc.)
- Telling a story, using quantitative measures (visualizations, statistical models, etc) to supplement and strengthen arguments.

Milestones

1. Cleaning the data and isolating data for Brazil per year such as GDP, rate of deforestation and measures that can help determine the association between deforestation and the economy
2. Experimenting with the data and looking for potential trends and patterns.
3. Focusing on a specific trend or a select few trends organizing them into presentable graphs or data and further researching and analyzing them.

4. Finalize conclusions with well thought out and written explanations that describe the data as and publish the analysis.

Concerns:

- Some concerns that arose in the planning and research for the project is how we would be able to determine the association between deforestation and Brazil's economy as well as attempting to predict the negative long term effects of deforestation.
- Additionally, finding enough accurate and relevant data is proving to be a challenge. Hopefully, the search for viable information will have been facilitated now that we have narrowed down the topic of exploration.

Datasets

We focused on two main datasets, Brazil Economic dataset; which contained nearly 1,500 different economic measurements from the years 1960 to 2020 and Brazil Deforestation Dataset which showed the amount of deforestation in 9 different Brazilian states in Km from 2004-2019

Brazil Economic Dataset:

The original Brazil Economic Dataset we found contained nearly 1,500 different economic measures from the years 1960-2020. However, a large number of these metrics were the same, just measured in different ways (ex. GDP in constant US$ vs GDP(current LCU), furthermore a large number of these metrics did not make sense to use in comparison to deforestation. In the end we narrowed down the dataset to 35 different economic metrics that we believed were useful in helping to find a connection to deforestation.

```
Year                                              int64
Food_Exports(%_of_exports)                      float64
GDP                                             float64
GDP_per_capita                                  float64
Forest_Area(sq_km)                              float64
Good_Exports                                    float64
Net_Trade_Goods_&_Services                      float64
Imports_Goods_&_Services                        float64
Forest_area_%_of_land                           float64
International tourism, number of arrivals       float64
Unemployment_%_of_labor_force                   float64
Male_unemployment_with_advanced_edu             float64
GNI                                             float64
Gross_capital_formation                         float64
Gross_domestic_savings                          float64
Arable_land_(hectares)                          float64
%_male_unemployment                             float64
Gross_domestic_income                           float64
net_income                                      float64
GDP(current_LCU)                                float64
exports                                         float64
Food_imports                                    float64
Tourism                                         float64
Imports_of_goods_&_services(%_GDP)              float64
Total_Debt_Service                              float64
%_Agricultural_land                             float64
Inflation(annual%)                              float64
Trade_%_GDP                                     float64
Annual%_GDP_per_capita_growth                   float64
Gross_fixed_capital_formation(annual_%_growth)  float64
Exports_of_goods_&_services(%GDP)               float64
S&P_Global_Equity_Indices(annual_%_change)      float64
Trade_in_services(%_GDP)                        float64
Broad_Money                                     float64
Annual%_GDP_pre_capita_growth                   float64
total_deforestation                             float64
dtype: object
```

Brazil Deforestation Dataset:

The Brazil Deforestation Dataset has information from nine different Brazillian States; Acre(AC), Amazonas (AM), Amapa (AP), Maranhao (MA), Mato Grosso (MT), Para (PA), Rondonia (RO), Roraima (RR), Tocantins (TO). The dataset measures the amount of annual deforestation in km from each of these states from the year 2004-2016. The dataset also includes the total deforestation from all of the states per year.

| Year	AC	AM	AP	MA	MT	PA	RO	RR	TO	Total Deforestation
2004	728	1232	46	755	11814	8870	3858	311	158	27772
2005	592	775	33	922	7145	5899	3244	133	271	19014
2006	398	788	30	674	4333	5659	2049	231	124	14286
2007	184	610	39	631	2678	5526	1611	309	63	11651
2008	254	604	100	1271	3258	5607	1136	574	107	12911

Exploratory Data Analysis:

## Line plot of all Brazilian states deforestation per year
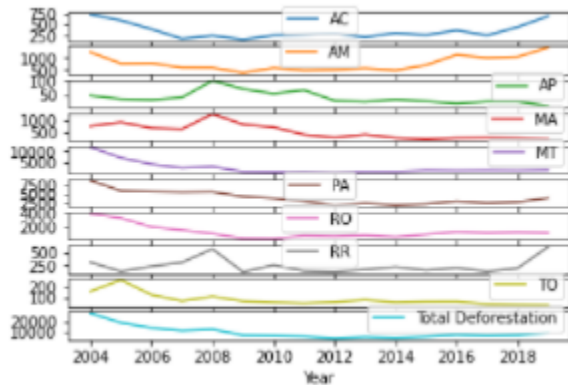
At first glance there is not a lot of information that we get from this picture. Since all of the plots used different scales on the y-axis it is difficult to analyze these graphs or begin to make observations.

Line plot of all Brazilian states deforestation per year
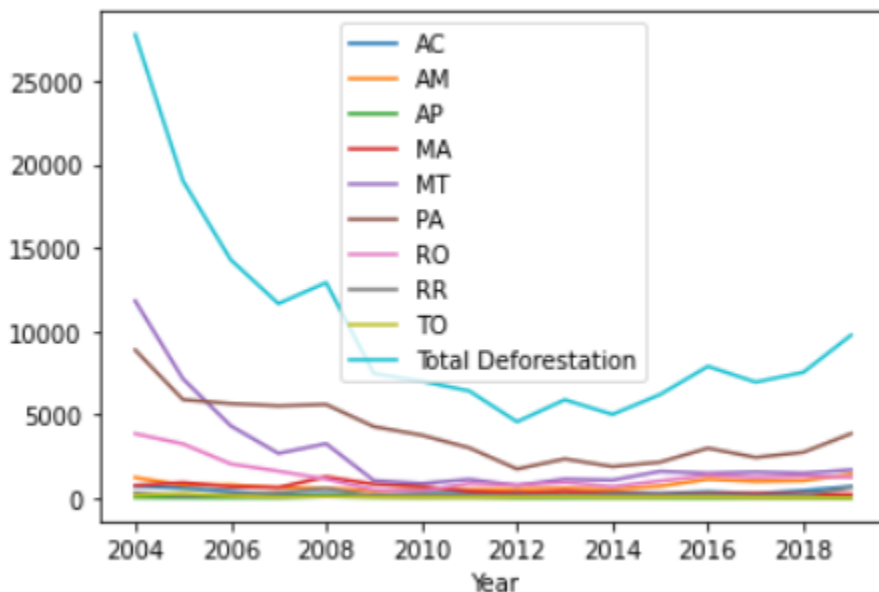
After combining all of the graphs into one larger graph we have a slightly clearer understanding of the data and the change of deforestation per year. Since all of the deforestation in the states seemed to follow a similar trend our group decided to primarily focus on total deforestation since we knew the aggregated data was not an instance of Simpson's paradox where the broken down data told a different story.

## Line plot of Total deforestation per year
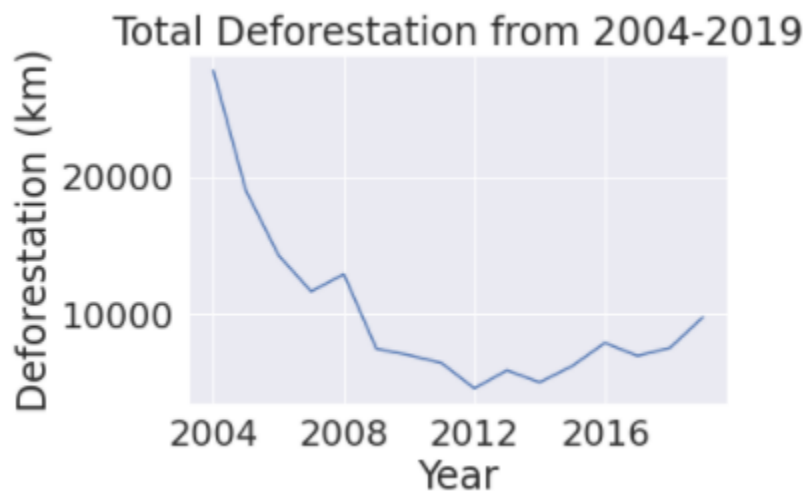
Looking at just the total deforestation we can begin to see that the amount of deforestation that was occurring during the early 2000's was much more than in recent years. We can see that there have been successful efforts by the government to reduce the amount of deforestation that occurs. However, there has been a slight uptick in the amount of deforestation within the last couple years,

Line plot of Brazil GDP per year

Now taking a look at the GDP of Brazil we can see almost the exact opposite of deforestation. Brazil's annual GDP started off very low in the early 2000's and gradually increased over the last two decades to nearly double what it was. However there was a slight dip in GDP in 2016.Line plot of annual GDP per capita growth
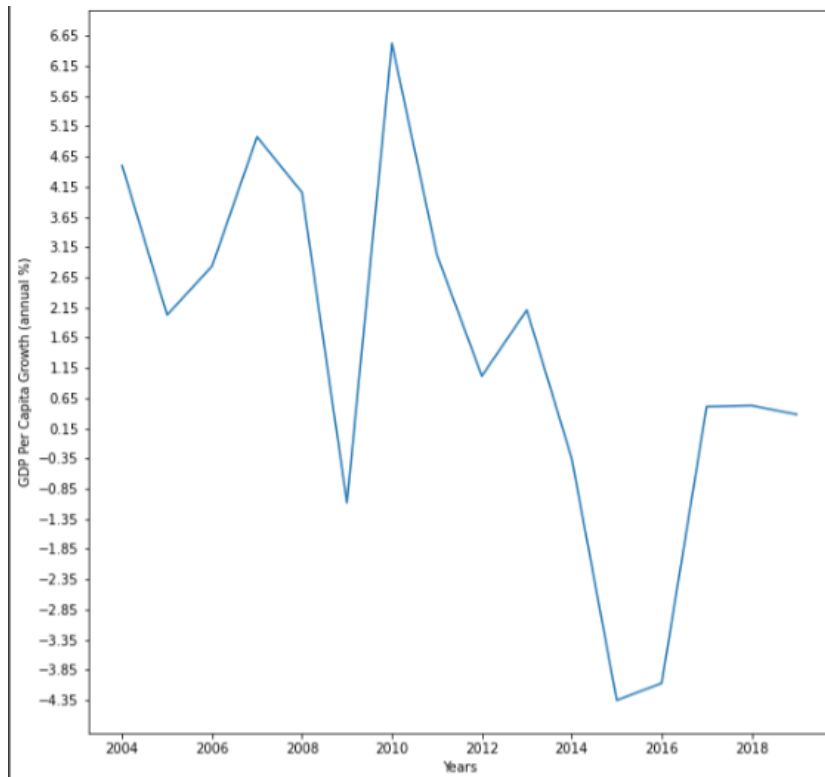
Looking at the annual gdp per capita growth we can see after 2010, the gdp per capita growth decreased significantly in a short amount of time and in 2016 there was an increase again. But the overall gdp per capita growth compared to where it started in 2004 dropped, which could be a cause of many different things, but it is good to note that when looking at the deforestation graph, when deforestation went down over the last decade, the gdp per capita growth went down. But at the end of the deforestation by 2018, deforestation numbers were going up, which is the same with this graph where per capita growth numbers went up.

T-test of GDP vs Deforestation

	T	dof	tail	p-val	CI95%	cohen-d	BF10	power
T-test	-37.300641	30	two-sided	1.072656e-26	[-2284196579671.14, -2047053400286.99]	13.187768	1.11e+23	1.0

The results of this t-test were extremely promising. We had hypothesized that there was a negative correlation between GDP and Deforestation based on our exploratory data analysis. This t-test helped prove that there is a statistically significant relationship between deforestation and GDP and from there we felt more confident in our project.

T-test of Total Deforestation vs Goods exports

	T	dof	tail	p-val	CI95%	cohen-d	BF10	power	name
T-test	-16.100276	30	two-sided	2.632705e-16	[-217427988044.2, -168477071820.55]	5.692307	1.137e+13	1.0	Goods exports (BoP, current US$)

This test helped prove that there were also connections between deforestation and other economic metrics. Based on this we can see that that the amount of goods that were exported were likely affected by the amount of deforestation that occurred,

T-test of Total Deforestation vs Imports of Goods and Services

	T	dof	tail	p-val	CI95%	cohen-d	BF10	power	name
T-test	-11.382106	30	two-sided	2.066189e-12	[-260108622909.42, -180967127949.83]	4.024182	2.243e+09	1.0	Imports of goods and services (BoP, current US$)

Similar to the t-test above we can see a connection between deforestation and the number of goods and services imports.

T-test of Total Deforestation vs GDP per Capita growth

	T	dof	tail	p-val	CI95%	cohen-d	BF10	power	name
T-test	6.556778	30	two-sided	2.974570e-07	[6898.69, 13140.35]	2.318171	3.589e+04	1.0	GDP per capita growth (annual %)

This t-test was also very useful because it showed that there was a statistically significant association and relationship with the amount of deforestation and the GDP growth per capita.

All of these t-tests were extremely useful in our exploration phase, they helped show that there was a statistically significant connection between the amount of deforestation and economic measures in Brazil. Going forward however we wanted to try a few more t-tests and then look for a correlation between the two variables since we know now that their relationship is statistically significant.

Statistical Data Analysis:

Heat Map of all the variables vs each other
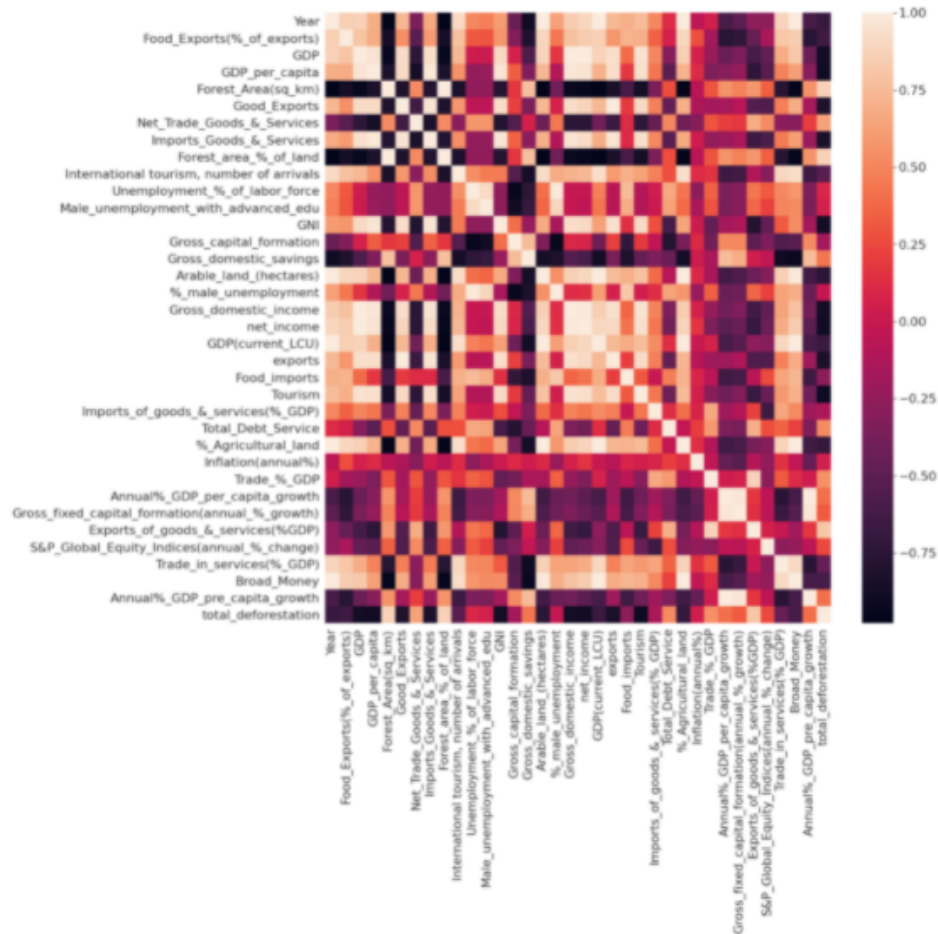
Heat Map of all statistically significant variables vs each other
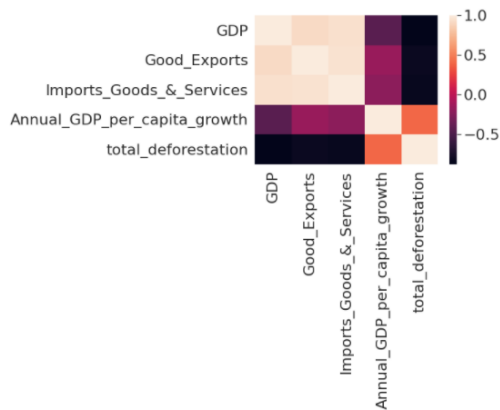
After conducting all the t-test and seeing that our variables weren't connected based on coincidence we felt confident in being able to make correlations between variables and knowing our results were accurate. While at first glance the heat map is a lot to digest the best course of action would be to focus on the very last line of correlations which are correlations of all of the variables vs total deforestation. Contrary to what we hoped, GDP per capita growth had a positive correlation with deforestation so as one went up so did the other meaning deforestation helped GDP growth per capita. However, the other variables that we looked at such as annual GDP, goods exports and imports of goods and services all had almost a perfectly negative correlation with total deforestation. Which is exactly what we were looking for, the negative correlation meant as one variable went up the other went down and vice versa. So as total Deforestation went down GDP, goods exports and goods and services imports went up. Based on this information we can begin to say with less deforestation the economy will likely be better and grow.
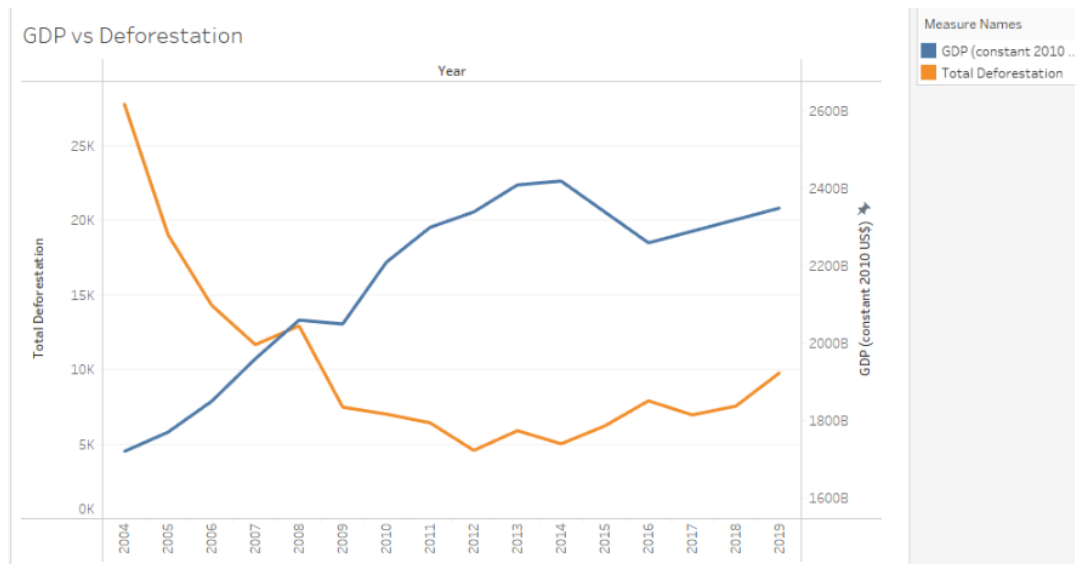
The chart above has GDP and total deforestation plotted on the same graph. Here we can clearly see the correlation between GDP and total deforestation.

Bubble maps

We plotted two bubble maps that represent the amount of deforestation of a given Brazilian state per year. The size of the bubbles represent the amount of deforestation for the given state; the location of the bubbles are plotted based on the latitude and longitude of the center of each Brazilian state and not precisely where the deforestation is taking place.

There are a few reasons we plotted two bubble maps instead of just one. Firstly, the plot was two convoluted with 16 different rings per state and it was extremely difficult to comprehend, let alone make meaningful conclusions, from the singular aggregate bubble map. So, with a need to split the plot into smaller time series pieces, some research led to this article on Brazilian anti-deforestation policies (https://www.sciencedirect.com/science/article/pii/S2530064418301263), in which it is said that the greatest decline in deforestation took place from 2004-2012). Some reasons for this decline

include legislations prohibiting timber commercialization from deforested areas, the exchange rate between USD and Brazilian Real falling and reducing the value of exporting soy and beef (commodities whose supply would increase with increase in deforestation). Thus, with a splitting year of 2012, we plotted two bubble maps, one spanning the years 2004-2012 and the other 2013-2019.

It can be readily seen that there was a massive decrease in deforestation from 2004-2012 and a marginally smaller amount and decline of deforestation in the years after. To interpret the plots, refer to the legend included with each plot that matches years with ring colors.

```
# color key
dict(zip([x for x in range(2004, 2020)], colors))

{2004: 'red',
 2005: 'blue',
 2006: 'green',
 2007: 'purple',
 2008: 'orange',
 2009: 'darkred',
 2010: 'lightred',
 2011: 'beige',
 2012: 'darkblue',
 2013: 'darkgreen',
 2014: 'cadetblue',
 2015: 'darkpurple',
 2016: 'white',
 2017: 'pink',
 2018: 'lightblue',
 2019: 'lightgreen'}
```

Deforestation per state from 2004-2012
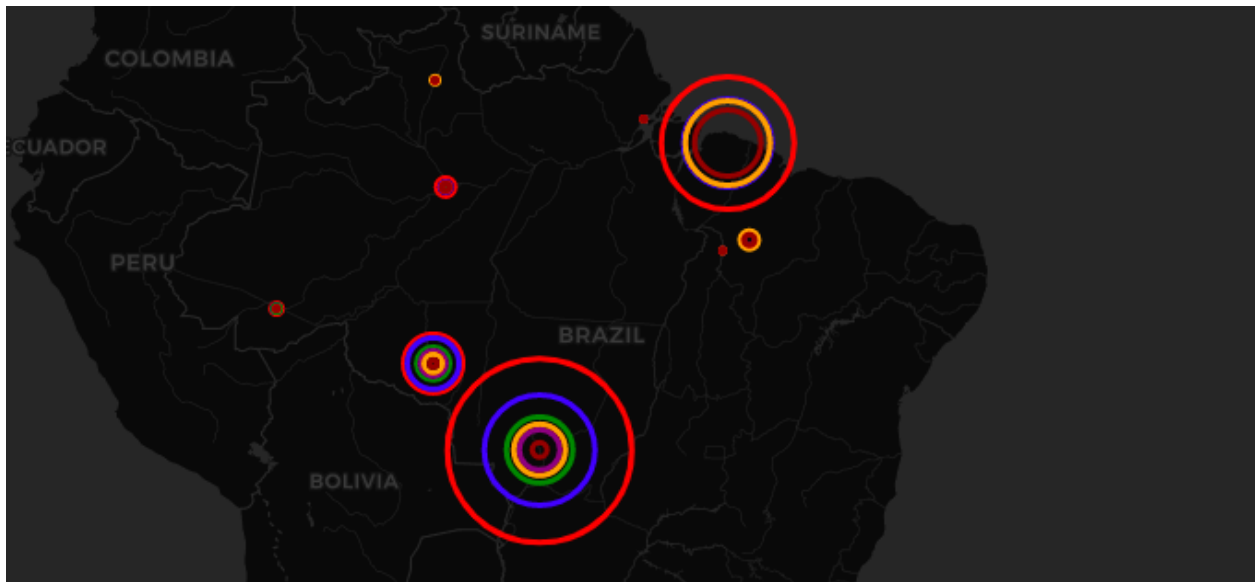
As you can see, the red rings (2004) were by far the largest indicating the greatest amount of deforestation. From there on, the amount of deforestation declines rapidly each year with each bubble getting smaller and smaller as the years progress.

```
# color key
dict(zip([x for x in range(2004, 2020)], colors))

{2004: 'red',
 2005: 'blue',
 2006: 'green',
 2007: 'purple',
 2008: 'orange',
 2009: 'darkred',
 2010: 'lightred',
 2011: 'beige',
 2012: 'darkblue',
 2013: 'darkgreen',
 2014: 'cadetblue',
 2015: 'darkpurple',
 2016: 'white',
 2017: 'pink',
 2018: 'lightblue',
 2019: 'lightgreen'}
```

Bubble map of Deforestation per State from 2012-2019

As you can see, the rings are much smaller and closer together, indicating both smaller amounts of deforestation and decline in deforestation. This is to be expected, since the aforementioned article explains that by far the largest amounts of deforestation took place before the year 2012 and this plot visualizes deforestation after 2012, further showing the magnitude (or lack thereof) of deforestation compared to that of the years before 2012.

Description of Application:

Conclusions:

- What we've found is that total deforestation and GDP almost had a negative correlation (when GDP went up, total deforestation went down)
- Some of our t-testing wasn't fully correct because of our sample size and the variables we were testing. However, most of the t-test that were correct, proved that there was 99% correlation between GDP and other economic factors with deforestation
-

- We were able to prove a correlation between the two and the heat map thoroughly explains that when there is an increase in deforestation, the economy actually drops which proves that deforestation would actually be a burden to Brazil in the future, especially with the amazon rainforests.

- Initially, we were trying to compare South American GDP and total deforestation rates. We realized that there was too much data to work with, so we focused our research on one country, being Brazil. We chose Brazil because there was a lot of news pertaining to deforestation and the Amazon Rainforest there.

- We learned how to use python through the cases to compare our datasets with different EDAs. We experimented with several representations to figure out how to best present our data. We also learned that datasets are empirical and are the hardest part of the

whole project. We needed more information because we don't have a sufficient sample size. Finally we learned how to analyze our t-tests and make sure they are fully sufficient to use

What we wanted to do better in this project:

- We would have wanted a much larger dataset from the get-go, to have more concrete evidence as well as to avoid potentially misleading results due to aggregated data.

How we can further this in the future (also what we would do if we had more time):

- We would have wanted to try some predictive modelling and regression/ r-square testing.
- We would also try to look into other factors that influence GDP and/or deforestation and get a better understanding of how specifically deforestation affected GDP.

How we can inspire people:

- We are young students (high school and college) who figured out that deforestation affects the economy in order to somewhat prove that deforestation should not happen. We hope this inspires not only people of our age, but people older that they can do what we did as well and educate people on why deforestation must not happen