Final Project Report

Brazil Deforestation

01.19.2022
DATA 2010

Brazil Deforestation

Introduction

Deforestation has been an ongoing problem for the past few decades. While climate change, an increase in global warming, and an increase in greenhouse gas emissions are major factors, humans have had just as big of an impact on deforestation around the world. Some "easy" fixes to help reduce deforestation would be to plant a tree, use less paper, recycle, and buy sustainable wood products. However, trees aren't only used for paper and wood products - their oils and seeds are used for human or animal consumption. Our goal for this data analysis was to dig down and see which consumable products have had the biggest impact on deforestation.

In our initial exploration of the data, we noticed that Brazil had the highest rate of deforestation, and China had the highest rate of afforestation. Brazil, but in particular the Amazon Rainforest, saw the highest level of deforestation during the 2020-21 period, where it lost 13,235 square kilometers. The Amazon Rainforest is the home to around three million species of plants and animals, and one million indigenous peoples. It's also a vital carbon store that slows the pace of global warming. Beef, soy, palm oil, and wood production are the top four drivers of the majority of the recorded habitat loss in the Amazon. This was evident in our exploration as soy and palm oil had the highest production in Brazil. At the COP26 climate summit, Brazil promised to end and reserve deforestation by 2030. Starting in 2021 until 2025, China will plant 36,000 square kilometers of new forest each year in an effort to combat climate change and to protect natural habitats. Tree planting is a major part of their plan to bring their carbon emissions down to net zero by 2060.

Our main focus of our analysis became Brazil. We wanted to see how the rate of deforestation has increased over the years, what the main causes are, and how the causes are related to each other.

Data

This dataset comes from Hannah Ritchie and Max Roser published in Our World in Data. The data is split into 5 separate datasets:

- `forest.csv` is the change in forest area in 1990, 2000, 2010, and 2015 for 130 countries.
- `forest_area.csv` is the change in forest area every year from 1990 to 2020 for 224 countries.
- `brazil_loss.csv` is the loss of Brazilian forest every year from 2001 to 2013 by 11 different causes.
- `soybean_use.csv` is soybean production every year from 1961 to 2013 by different uses (*human food*, *animal feed* and *processed* (into vegetable oil, biofuel, and processed animal feed)) in 167 countries.

- `vegetable_oil.csv` is vegetable oil production every year from 1961 to 2014 by 13 different crops (Coconut (copra), Cottonseed, Groundnut, Linseed, Maize, Olive (virgin), Palm, Palm kernel, Rapeseed, Safflower, Sesame, Soybean, Sunflower) in 192 countries.

Methods

Data Preprocessing The first step in analyzing data is to clean it up and make it easier to work with. Due to the nature of the data, we first filtered out any regions (such as *North America*, or *World*) and only considered countries. We also changed the shape of some of the datasets by pivoting wider or longer, and removing or renaming columns and variables. Next, we created sub-datasets that only included data from Brazil, since we wanted to analyze that data specifically. Lastly, we combined the Brazil-focused datasets so that all of the data was together.

Summary Statistics For a general overview of the data, and to see possible trends and averages, we calculated the mean and median of several features. To see how spread out our data points were, we used five-number summaries/quantiles, which finds the minimum, maximum, range and quartiles (boundaries for the lowest, middle, and upper quarters of data). To display our summary statistics, we used tables, boxplots, and scatterplots.

Rankings To see how different countries matched up against each other, we arranged them according to a certain variable (ex. soybean production) in ascending or descending order. In some cases, we considered both extremes; those ranked at the top and those that ranked at the bottom. Then, we could see which countries are thriving, and which may be struggling.

Modeling In order to look at the relationships between variables in particular datasets, such as forest area and the drivers of deforestation, we used several techniques. We measured the correlation of different features between the datasets, which measures the association/relation between two variables. We also ran several models to try to either predict a variable and to analyze the structures and trends of the data. We mainly used linear regression models to try and predict forest area from various covariates.

Results

Countries with the Largest Global Forest Area

Figure 1 looks at the top five countries with the highest global percentage of forest area. If the coloured line for each country is above its black reference line, it means that the amount of forest area has increased since 1992. If the coloured line is below, it means that the forest area has decreased since 1992.

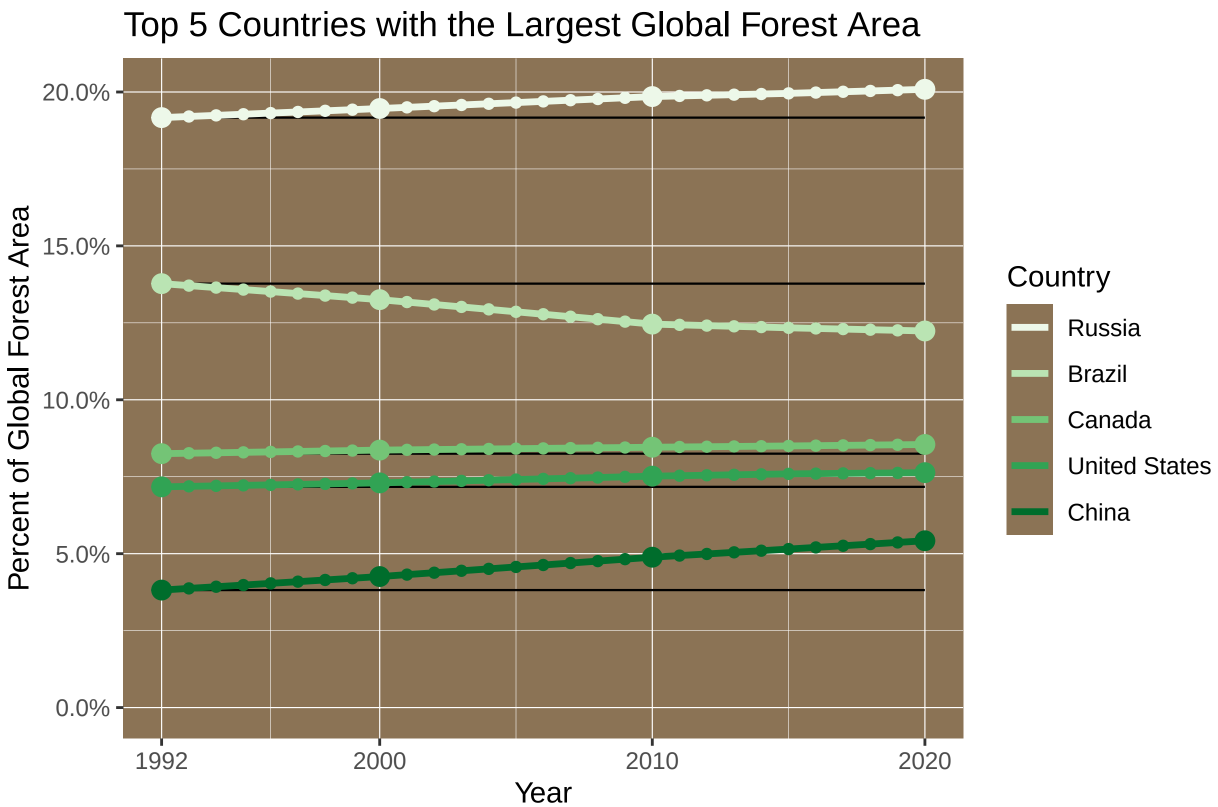


Figure 1

Of the top five countries, Brazil is the only one that has decreased since 1992. We can see that China has had the biggest positive increase in percentage of global forest area out of the five countries, followed by Russia, United States, and then Canada.

One important (and possibly misleading) aspect of this, is that this is the percent of global forest area. Although Russia, Canada, United States, and China are gaining global forest area, that does not mean that they are all growing forest space faster than they are losing it. In Figure 2, we explore this further.

Net Change in Forest Area

Figure 2 looks at the top ten countries that had the most extreme net change in forest area in 1990, 2000, 2010, and 2015.

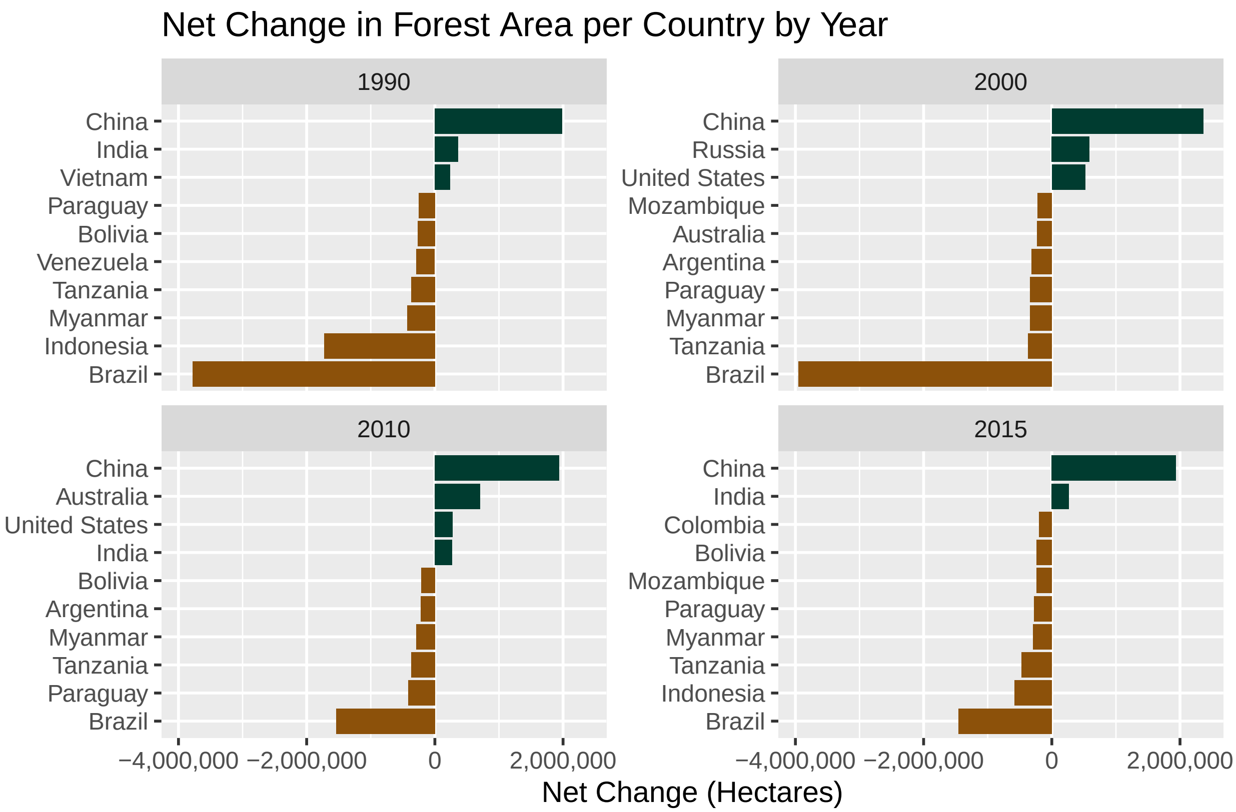


Figure 2

China is always at the top, having the most positive net change, or afforestation, in all four years. On the other hand, Brazil is always at the bottom, having the most negative net change, or deforestation, in all four years.

Continuing with the discussion from Figure 1, while Canada has had an increase in forest area, in all 4 years measured, we have had a negative change, i.e. losing more than we're able to restore. There was no data provided from the United States in 2015, but they did have a positive change in 2010, just not an extreme amount. There was no data provided for Russia in 1990, but in 2010 they had a negative net change, and in 2015 they did have a positive net change, but not at the level of the top ten most extreme changes.

Vegetable Oil

The following table illustrates the median, mean, and the country with the most production along with its corresponding production value for each of the thirteen vegetable oil crop types.

Table 1: Vegetable Oil Production from 1961 to 2014 (In Tonnes)

crop_oil	Median	Mean	Top Producer	Total Production
Palm	22,000	395,282.77	Malaysia	368,205,292
Rapeseed	16,350	195,424.19	China	116,824,150
Soybean	10,000	247,185.59	United States	315,507,763
Sunflower	7,000	96,622.31	Argentina	53,415,039
Maize	5,500	30,490.08	United States	35,909,560
Olive, virgin	4,825	72,388.84	Spain	37,511,931
Palm kernel	3,087	41,046.73	Malaysia	44,256,068
Groundnut	3,055	46,110.52	India	79,278,000
Cottonseed	2,502	36,167.86	China	42,750,820
Sesame	2,448	13,996.63	India	8,126,700
Coconut (copra)	1,536	32,193.76	Philippines	57,229,438
Linseed	1,260	12,285.71	United States	6,281,764
Safflower	1,176	13,031.88	India	2,876,900

The median and the mean are quite different from each other for each crop type, thus, there is a presence of extreme outliers in the production values across all types. Note that China is the top producer of Rapeseed, and Brazil doesn't make the list.

Figures 3 and 4 look at vegetable oil production in Brazil. We looked at the summary statistics/quantiles with and without Soybean, we noticed it was an outlier. In Figure 3, it is difficult to see the production of the other crops. By removing Soybean in Figure 4, it is easier to see which crops were largely produced after Soybean.

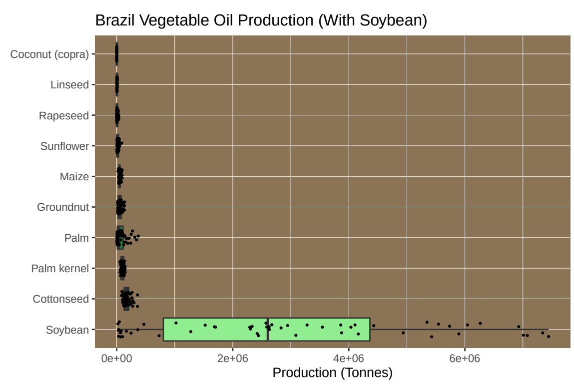


Figure 3

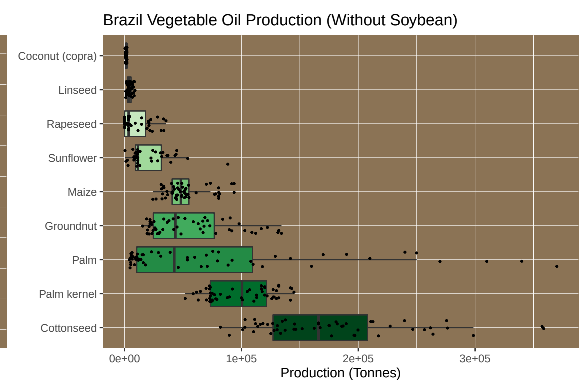


Figure 4

We note that Soybean has the highest production overall, followed by Cottonseed. Palm has a bigger spread, but Palm Kernel has a higher mean. We also note that Soybean is the outlier.

Soybean Production

Soybean production has been broken down into three categories: human food, animal feed, and processed. In our data, processed soybeans are used for vegetable oils, biofuel, and processed animal feed. Figure 5 looks at soybean production around the world by type from 1961 to 2013.

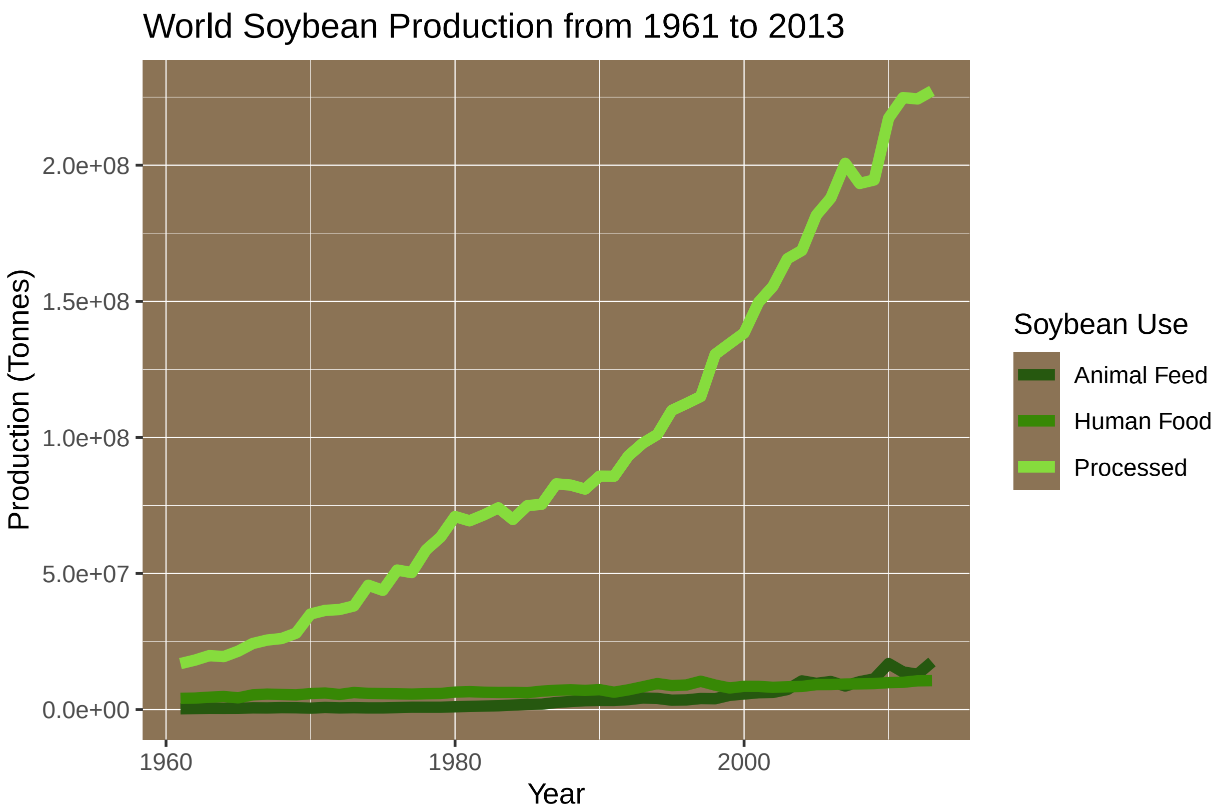


Figure 5

We can see that soybean production for processed goods constantly increased over time. The use for human food was next highest, but then dipped down below animal feed around 2005.

Vegetable Oils and Soybean Production

Figure 6 shows soybean production (by type) and vegetable production (as a whole) in Brazil from 1961 to 2013. Figure 7 shows the change in forest area in Brazil from 1990 to 2020.

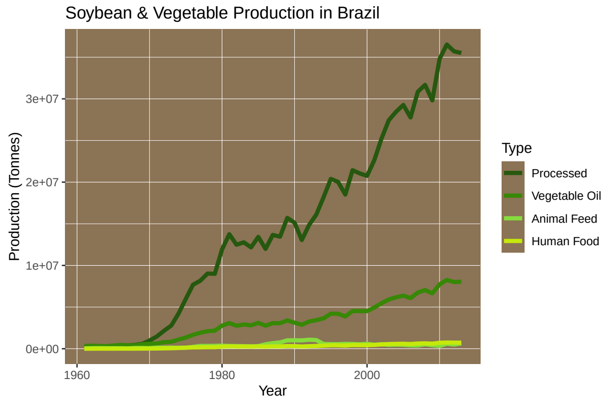


Figure 6

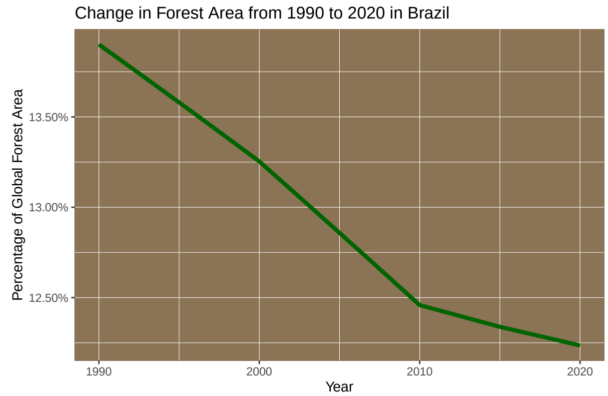


Figure 7

Similar to Figure 5, soybean production for processed goods increased constantly over time. Vegetable oils were produced second highest, also increasing constantly. Soybeans produced for human food and animal feed are near the bottom, staying at about the same level as each other. Figure 7 shows a constant decrease in forest area over time. Thus, we can say that soybean and vegetable oil production have a negative relationship with forest area, that is, as soybean and vegetable production increased, forest area decreased. This is logical as soybeans have been proved to be a major factor in deforestation.

Drivers of Deforestation

Figures 8 and 9 show the area of Brazilian forest lost (in Hectares) by cause from 2001 to 2013. When we looked at the summary statistics with and without Pasture and Commercial Crops, we noticed they were outliers. In Figure 8, it is difficult to see the area of forest lost by the other causes. By removing Pasture and Commercial Crops in Figure 9, it is easier to see which causes destroyed the most Brazilian forest.

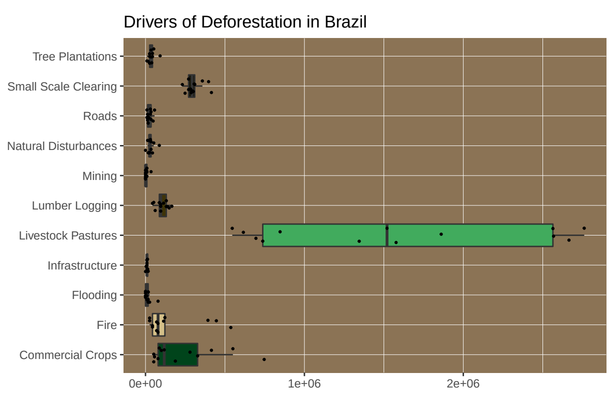


Figure 8

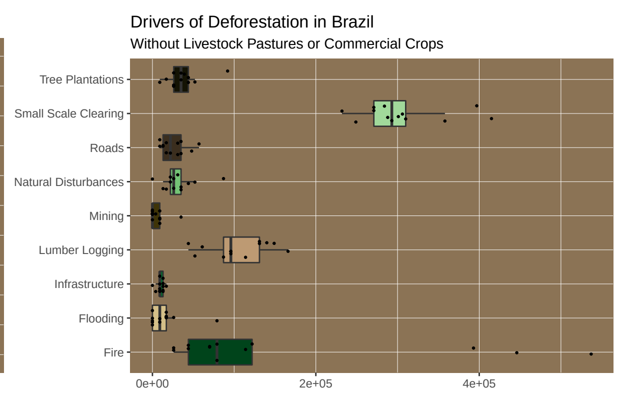


Figure 9

We note that after Pasture and Commercial Crops, Small Scale Clearing, Fires, and Lumber Logging are the next three causes of deforestation in Brazil.

Figure 10 takes Figures 8 and 9 and displays the main causes of deforestation in Brazil by year from 2001 to 2013.

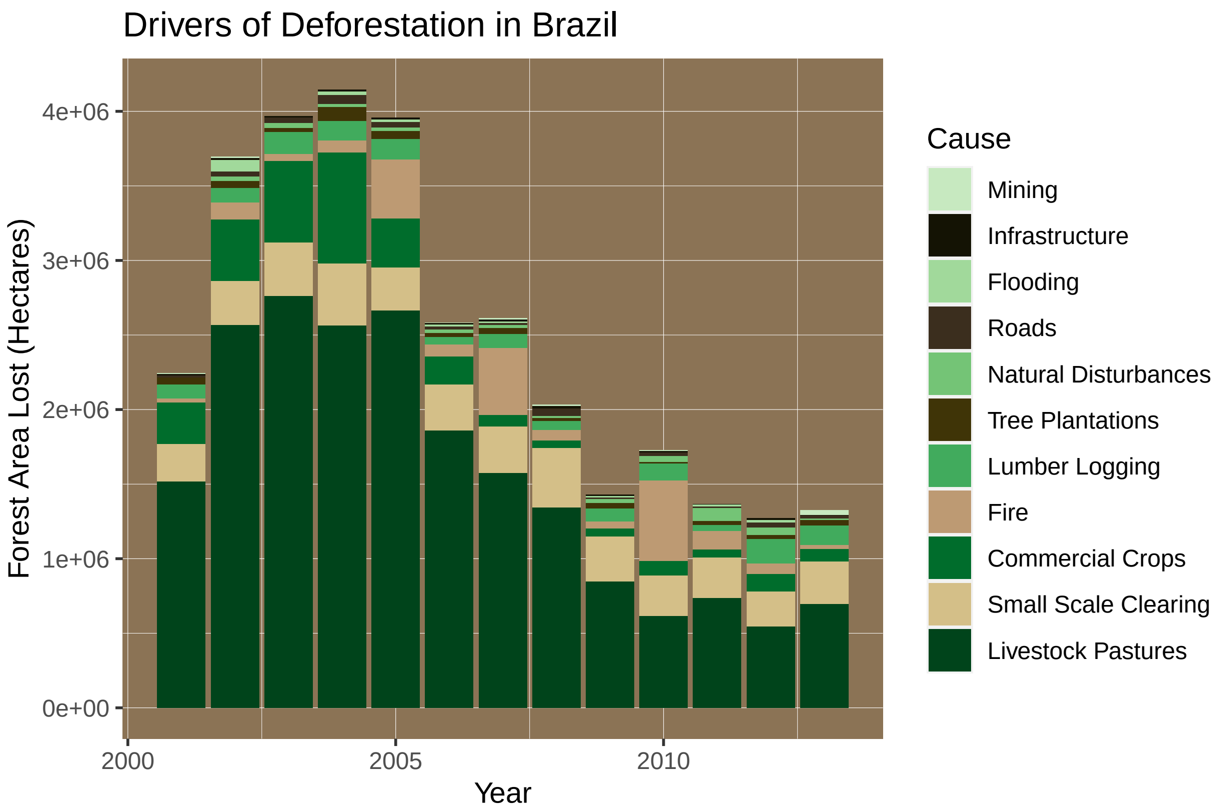


Figure 10

We can see that in each year, Pastures are the main cause of deforestation. The second cause is either Small Scale Clearing or by Fire, depending on the year. It is clear that Brazilian forest took the most damage in 2004, and took the least damage in 2012.

The following table illustrates the absolute change in forest area lost in Brazil.

Table 2: Absolute Difference of Forest Area Lost in Brazil from 2001 to 2013

Cause	Absolute Change (Hectares)
Livestock Pastures	-825,000
Commercial Crops	-193,000
Lumber Logging	35,000
Small Scale Clearing	35,000
Mining	26,000
Natural Disturbances	13,000
Tree Plantations	-9,000
Infrastructure	-5,000
Roads	4,000
Fire	0
Flooding	0

As it was clear in the previous figures, Pastures have destroyed the most Brazilian forest from 2001 to 2013. When considering the "positive" values, such as with mining, that does not mean that mining increased the forest area, but instead means that the amount of forest area lost due to mining was greater in 2001 than in 2013. It is interesting to note that Fires and Flooding due to Dams have an absolute change of zero.

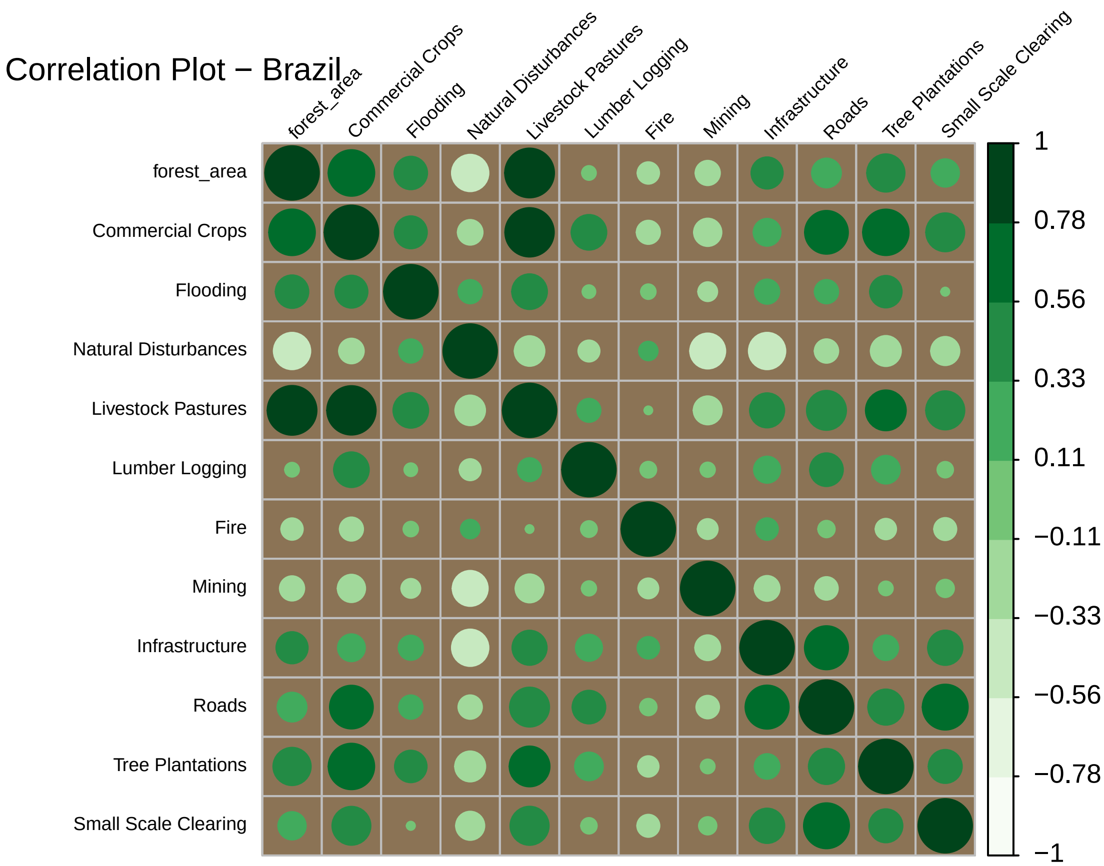
The following table looks specifically at the correlation between forest area and drivers of deforestation in Brazil.

Table 3: Correlation of Forest Area and Drivers of Deforestation in Brazil

Cause	Correlation
Livestock Pastures	0.8346338
Commercial Crops	0.7293284
Tree Plantations	0.4872314
Natural Disturbances	-0.4639226
Flooding	0.3714019
Infrastructure	0.3413813
Roads	0.2922923
Small Scale Clearing	0.2630907
Mining	-0.2067163
Fire	-0.1612540
Lumber Logging	0.0669774

The two strongest positive linear relationships with Forest Area is Pasture (with a value of 0.8346) and Commercial Crops (with a value of 0.7293). Thus, we can say as Forest Area increases, so does the loss of Brazilian forest due to Pastures or Commercial Crops, and vice-versa. This proves that our conclusions we were able to draw from the above boxplots were in fact, true.

Next we look at a correlation plot to see the relationships between the drivers as well.



Aside from the relationships with forest area, commercial crops and livestock pastures, roads and small scale clearing, commercial crops and tree plantations, and roads and infrastructure were all highly positively correlated. Forest area and natural disturbances, natural disturbances and mining, and natural disturbances and infrastructure were highly negatively correlated.

When looking at creating some models for our Brazil-focused data, it was difficult because we didn't have enough observations to separate them into test and train data. We tried several methods with little success. The following model uses Pasture and Commercial Crops to predict Forest Area, as they had a strong positive correlation (greater than 0.70) with Forest Area.

```
#
# Call:
# lm(formula = forest_area ~ Livestock_Pastures + Commercial_Crops,
#     data = data1)
#
# Coefficients:
#        (Intercept)  Livestock_Pastures    Commercial_Crops
#          1.232e+01           2.342e-07           1.672e-07
```

```
#                             2.5 %       97.5 %
```

```
# (Intercept)          1.208487e+01 1.255276e+01
# Livestock_Pastures  1.623303e-08 4.521622e-07
# Commercial_Crops   -6.733693e-07 1.007802e-06
```

The intercept is Forest Area, and 12.3188 is the predicted value of Forest Area (in a percentage of global forest area) if every other variable was zero. The value for pasture, 0.0000002342, is the predicted difference in Forest Area for each loss of Forest Area due to Pastures (with all other variables constant). The value for Commercial Crops, 0.0000001672, is the predicted difference in Forest Area for each loss of Forest Area due to Commercial Crops (with all other variables constant). The only confidence interval that contains zero is Commercial Crops, so we conclude that there is no significant evidence of a linear relationship between Forest Area and Commercial Crops.

Conclusion

Deforestation around the world is due to multiple causes. Specifically in Brazil, Pastures and Commercial Crops have caused the most damage. To put things in perspective, we saw were Brazil ranked among all other countries in terms of deforestation. We saw that of the top five countries with the largest global forest area, Brazil was the only one that decreased since 1992. Furthermore, Brazil had the most negative net change - or deforestation - in 1990, 2000, 2010, and 2015. To our surprise, Brazil was not a top producer of any of the thirteen kinds of vegetable oil.

Brazil's vegetable oil production was majorly on Soybean, with Cottonseed and Palm further behind. By breaking down soybean production, processed soybeans (used for vegetable oils, biofuel, and processed animal feed) were always produced the most. It was also evident that as soybean and vegetable oil production increased, the change in forest area decreased.

When diving into the overall causes of deforestation in Brazil, the use of Pastures was at the top, followed by Commercial Crops. The next causes that destroyed Brazilian forests were Small Scale Clearing, Fire, and Selective Logging - where their ranking differed based on the year. In our correlation plot, we saw that Forest Area has a negative relationship with Pastures, and Commercial Crops. Our model showed that there's no linear relationship between Forest Area and Commercial Crops.

In our initial research, we found that beef, soy, palm oil, and wood production are the top four drivers of recorded habitat loss in the Amazon Rainforest in Brazil. Our data analysis further proved that: pastures were the top cause of deforestation, soybean production constantly increased over the years, palm oil was the third highest produced vegetable oil, and selective logging was in the top five causes of deforestation.

This data analysis could be used to bring more attention to the factors that are causing the most deforestation in Brazil, and to then find alternatives or solutions to slow any further destruction of forests.

References

https://github.com/rfordatascience/tidytuesday/blob/master/data/2021/2021-04-06/readme.md

https://www.reuters.com/world/china/china-step-up-tree-planting-campaign-help-reach-net-zero-2021-08-20/

https://www.bbc.com/news/world-latin-america-59341770

https://www.worldwildlife.org/magazine/issues/summer-2018/articles/what-are-the-biggest-drivers-of-tropical-deforestation