Deforestation in Brazil

Ashley Wong Ping Rui, Calla Koh Wei Yu, Khalisah Binte Shari, Ong Jing Jun, Wayne Ng Jun Cheng

2025-10-14

```
library(tidyverse)
library(readxl)
library(stringr)
library(lubridate)
library(scales)
library(ggthemes)
library(countrycode)
library(ggrepel)
library(tidytext)
library(gridExtra)
```

1. Introduction

With deforestation being responsible for about 10% of global warming, it is imperative to put an end to it to tackle the climate crisis. (WWF). With deforestation being such a pressing issue, we aim to answer this question: **What are the predominant trends in deforestation, including the countries and activities primarily responsible for it?**

Deforestation trends will be identified using 3 plots. First, a facet wrap diverging bar plot will show deforestation trends from countries with the highest net change in forest area over time. Following our analysis of the first plot, the second plot will use a stacked area plot to examine the country that has contributed the most to deforestation and how deforestation drivers have affected it over time. Finally, using a map and time series graph, we will compare global soybean production and usage, one of the main deforestation drivers, to plot 1's deforestation trends.

In our research, we have used data from Our World in Data. The following are the data sets that we have used:

1. The 'forest' data set is comprised of data from the years 1990, 2000, 2010, and 2015, encompassing various countries and their changes in forest area (measured in hectares) between each time period. We did not use this data set as a whole for plot 1; we only used the data for countries we were interested in.
2. The 'brazil_loss' data set is comprised of data from 2001 to 2013 about the different drivers of deforestation and the amount of forest area lost (measured in hectares) for the activity. We used this data set as a whole to look at how the activities contribute to deforestation in Plot 2.
3. The 'soybean_use' data set is comprised of data from 1961 to 2013 about soybean production and use in each country. We added up the allocation of soybean products in this data set to find the total soybean production for each year and country. We did not use this data set as a whole for plot 3; we extracted data associated with the time frame we were interested in.

2. Data Cleaning and Visualization

```r
forest <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/20
brazil_loss <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/da
soybean_use <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/da
```

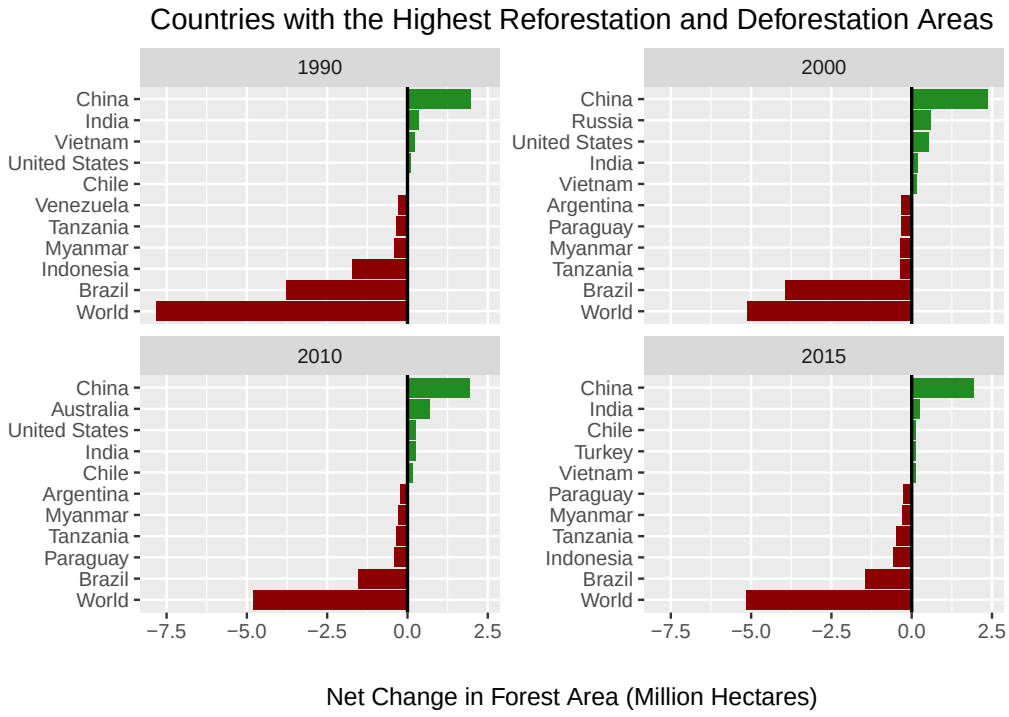
2.1 Plot 1

First, we are interested in identifying countries with the most significant net change in forest area over the years. This means that we will filter for the data of the top 5 countries for reforestation and deforestation, as well as the world's deforestation data, for each year that is available in the data set (1990, 2000, 2010, 2015). This lets us zoom in to countries where such phenomena are most pronounced and see the overall deforestation trend.

The data was presented using a facet wrap diverging bar plot. Each plot shows the net change in forest area in million hectares by country for each year. A facet wrap diverging bar plot allows for a clear and concise comparison of deforestation and reforestation trends and how they evolve over the years. It shows which countries are the biggest deforestation contributors quickly. To aid visualisation, we coloured the bars that show positive net change (reforestation) and negative net change (deforestation) green and red respectively.

```r
#we process the data to find the highest rates deforestation (forest_min) and highest rate of reforesta

forest_max <- forest %>% #using forest data set
  filter(entity != "World") %>% #filtering out the World entity
  group_by(year) %>% #group by year
  slice_max(net_forest_conversion, n = 5) %>% #slicing top 5 countries with highest positive net forest
  ungroup() #ungroup data

world_deforestation <- forest %>% #using forest data set
  filter(entity == "World") #filtering for the World entity

forest_min <- forest %>% #using forest data set
  filter(entity != "World") %>% #filtering out the World entity
  group_by(year) %>% #group by year
  slice_min(net_forest_conversion, n = 5) %>% #slicing top 5 countries with highest negative net forest
  ungroup() #ungroup data

forest_all <- rbind(forest_max, forest_min, world_deforestation) %>% #combining forest_max and forest_m
  mutate(fill_color = ifelse(net_forest_conversion > 0, "positive", "negative")) %>% #mutate a new col
  mutate(entity = reorder_within(entity, net_forest_conversion ,year)) #reordered entity within each ye
```

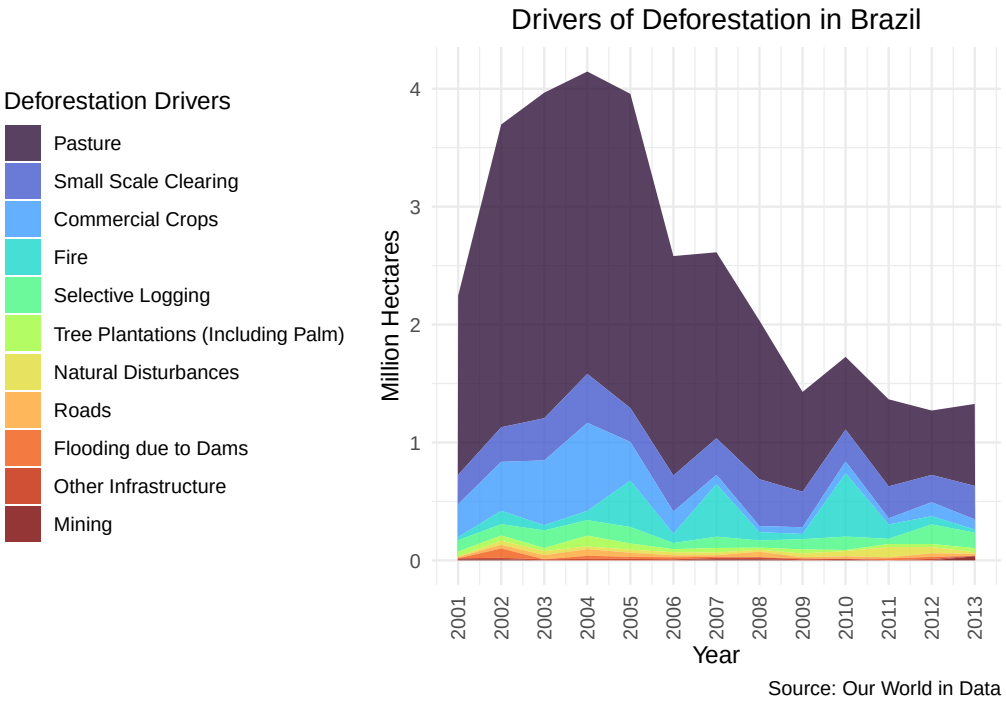
```r
ggplot(forest_all,aes(net_forest_conversion/1e6, entity,
             fill = fill_color)) + #creating a ggplot layer
  geom_col() + #creating a bar plot. x = net forest conversion in millions. y = entity. fill is based o
  scale_fill_manual(values = c("positive" = "forestgreen", "negative" = "darkred")) + #changing fill co
  facet_wrap(~ year, scales = "free_y") + #creates a bar plot faceted by year
  scale_y_reordered() + #trims off the "___year" for the y axis values
  geom_vline(xintercept = 0, color = "black", linetype = "solid", alpha = 1, lwd = 0.7) + #added a vert
  theme(legend.position = "none") + #remove legend box
  labs(x = "Net Change in Forest Area (Million Hectares)",
       y = "") + #add axis labels
  theme(axis.title.x = element_text(margin = margin(t = 20)), #adjusting position of axis and plot titl
        plot.title = element_text(hjust = 0.5)) +
  ggtitle("Countries with the Highest Reforestation and Deforestation Areas") #add plot title
```



2.2 Plot 2

Since Brazil has been the highest contributor to deforestation, we dive deeper into Brazil's deforestation activities to find out which activities are the main drivers. We have chosen a stacked area plot that shows us the area of Brazilian forest loss in hectares across the years 2001 to 2013. The x-axis represents the years, and the y-axis represents the total area of Brazilian forest loss in hectares. To aid visualisation, we have coloured the areas for the different driving factors in different colours, and included a legend that depicts their corresponding colours. We have also arranged the different driving factors according to size to further aid visualisation. This plot would enable us to easily identify the main driver of deforestation in Brazil by simply identifying the activity with the largest area.

```
options(scipen = 999) #removes scientific notations
plot_2 = brazil_loss %>% #using brazil loss data set
  pivot_longer(commercial_crops:small_scale_clearing, names_to = "type", values_to = "value") %>% #comb
  mutate(type = reorder(type, desc(value))) %>% #reorder type based on the descending order of value.
  ggplot() + #create ggplot layer
  geom_area(aes(x = year, y = value/1e6 ,fill = type) ,alpha = 0.8, position = "stack") + #adding a lay
  scale_x_continuous(breaks = seq(2001,2013, by = 1)) + #show all years on the x axis ticks.
  theme_minimal() + #theme
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1), #Rotate year label on the x axis
        legend.position = "left", #Shift legend position to the left
        plot.title = element_text(hjust = 0.5)) +   #Adjust position of plot title.
  labs(title = "Drivers of Deforestation in Brazil", x = "Year", y = "Million Hectares",
       caption = "Source: Our World in Data") + #add axis labels
  scale_fill_viridis_d(name = "Deforestation Drivers", option = "H", direction = 1,
                       labels = c("commercial_crops"="Commercial Crops", "fire"="Fire", "flooding_due_t

plot_2
```

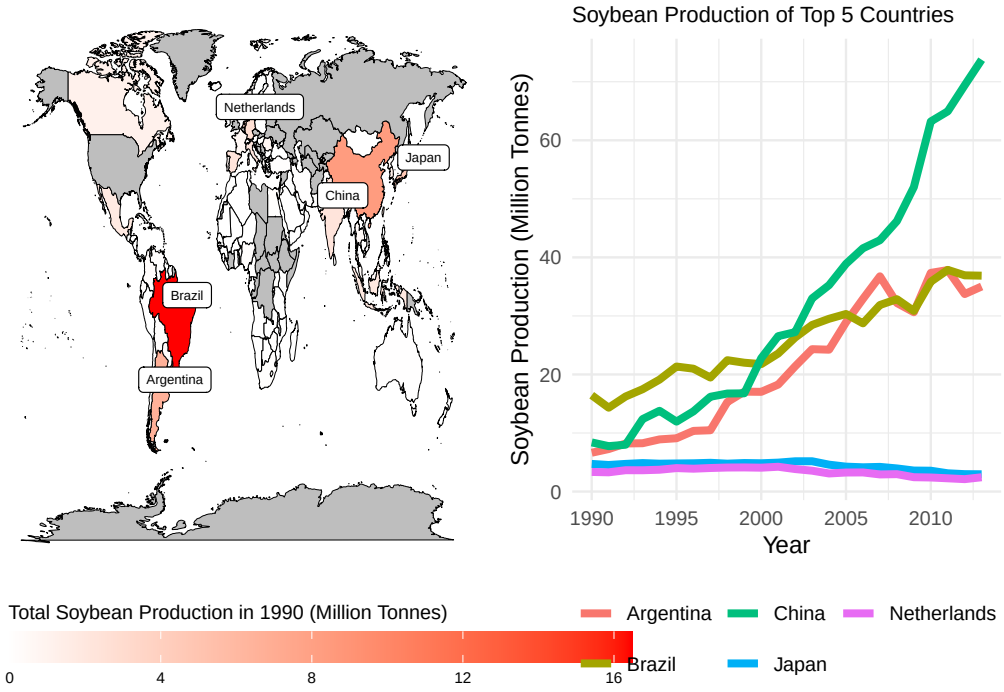


2.3 Plot 3

Soybean is one of the main deforestation drivers (Our World in Data). Therefore, we wish to see if the countries that produce and use more soybean indeed correspond to the countries with high deforestation rates in plot 1. To obtain the data we want for our plot, we added up the allocation of soybean products in the 'soybean_use' data set to find the total soybean production for each year and country. In the time series plot, we kept the data between 1990 to the latest year of data available 2013 to match the years for plot 1 for ease of comparison. Additionally, we added a world map that captures a point of soybean production in the different countries for the starting year 1990. The map is coloured by intensity, with darker regions representing higher soybean production. We also labelled the top 5 soybean producers. This makes the map helpful in visualising where the top soybean producers are in the world at a glance.

```
world <- map_data("world") # loading World df from countrycode library, containing coordinates and coun

countries_vector <- unique(world$region) # get the name of different countries as a vector that will be

soybean_used_filtered <- soybean_use %>%  # summing soybean_use for respective year and countries from
  filter(year >= 1990, year <= 2013, entity %in% countries_vector ) %>% # filter out non-country data a
  group_by(entity, year) %>% # group the data by country and year
  mutate(total_use = sum(human_food, animal_feed, processed, na.rm = TRUE)) # sum up the human_food, an

soybean_used_1990 <- soybean_use %>% # data frame for 1990
  filter(year == 1990) %>% # filter data for year 1990
  select(c(entity, human_food:processed)) %>% #select specific cols for the data frame.
  filter(entity %in% countries_vector) %>% # filter out continents, only include countries in the data
  group_by(entity) %>% # group by the countries
  mutate(total_use_1990 = sum(human_food, animal_feed, processed, na.rm = TRUE)) # sum the human_food,

soybean_used_plot <- soybean_used_filtered %>%
  left_join(soybean_used_1990, by = ("entity")) %>% # left join by country name
  ungroup() %>%
  rename(Countries = entity) %>%  #  changed the name of entity col to Countries
```

```r
  slice_max(total_use_1990, n = 120) # we find the top 5 countries in soybean use from 1990, for each o

#time series plot
plot3 = ggplot(data = soybean_used_plot, aes(x=year, y=total_use/1000000)) + #line graph for the soybea
  geom_line(aes(color=Countries), lwd = 1.7) + # line plot layer, edited the line width
  labs(title = "Soybean Production of Top 5 Countries", x = "Year", y = "Soybean Production (Million To
  theme_minimal() + # theme
  theme(plot.title = element_text(size = 10), legend.position = "bottom") + # legend position
  guides(color = guide_legend(nrow = 2)) #edit arrangement of legend labels

#label for chloropeth map
plotlab <-soybean_used_1990 %>% # generate coordinates for choropleth map for the top 5 countries in so
  ungroup() %>% # ungroup
  slice_max(total_use_1990, n=5) %>% # find the top 5 countries for production in 1990
  left_join(world, by = c(entity = "region")) %>% # left join for coordinates for the respective countr
  group_by(entity) %>%
  summarise(long=mean(long),lat=mean(lat),group=mean(group),total_use_1990=mean(total_use_1990)) # coor

# choropleth map
world_plot <- world %>%
  left_join(soybean_used_1990, by = c(region = "entity")) %>% # left join with world df
  ggplot(aes(x = long, y = lat, group = group, fill = total_use_1990/1e6)) + #plotting the world map, f
  geom_polygon(color = "black", size = .05) +
  scale_fill_gradient2(low = "white", high = "red", na.value = "grey",
                       labels = comma, guide = guide_colorbar(title.position = "top")) +
  theme_map() +
  labs(fill = "Total Soybean Production in 1990 (Million Tonnes)") + # labels for legend
  theme(legend.position = "bottom",  # Set legend position at the bottom of the plot
        legend.direction = "horizontal", # Make the legends horizontal
        legend.key.width = unit(2, "cm"), # Set legend key width
        legend.key.height = unit(0.5, "cm"), # Set legend key height
        legend.box.margin = margin(t = 0, r = 0, b = 0, l = 0, unit = "cm")) + # Box margins for legend
  geom_label_repel(data = plotlab, aes(label = entity), # add labels in the plot
                   box.padding = 0.1, point.padding = 0.2, segment.size = 0.1, size = 2,fill="white")

grid.arrange(world_plot,plot3, nrow = 1) # arranging the time series graph and chloropeth
```



4. Discussion

Plot 1 shows us that Brazil is the country most responsible for deforestation. Brazil has the longest bar in the negative direction throughout 1990 to 2015. Myanmar and Tanzania have also consistently been one of the greatest contributors, indicated by their appearance as top countries contributing to deforestation throughout the given time period. Deforestation appears to be decreasing over the years in the top countries. Despite this, while overall deforestation has fallen since 1990, it has remained roughly stagnant from 2000 to 2015. Hence, it is likely that besides the top countries, more countries are contributing significantly to deforestation as well. We can also see that despite the net deforestation globally, there are countries that have been working on reforestation instead of contributing towards deforestation. In contrast to Brazil, China has the longest bar pointing in the positive direction throughout the years, illustrating how China's reforestation rates have remained consistently high through the years. India has also been one of the top contributors to reforestation in recent years, appearing throughout the graphs of 2000-2015.

Plot 2 shows us that out of all the activities in Brazil, pastures have been the greatest contributor to deforestation. It is also significantly larger than all the other drivers of deforestation. This is reflected by the fact that the area in purple representing pasture is the largest across the years from 2001 to 2013. This suggests that deforestation rates for pastures have remained consistently high. We can also see that deforestation for pastures in Brazil peaked in 2004 and has had a generally decreasing trend since. Notably, there are large falls in hectares of deforestation from 2004-2005, and 2007-2009, due to the introduction of deforestation policies in those periods (Time).

Plot 3's map reveals that Brazil, Argentina, China, Japan and the Netherlands are the top soybean producers in the world. Having the darkest area, Brazil topped soybean production in 1990. Plot 3's time series plot further reveals that the top 3 producers remain the same throughout the years. Brazil's soybean production increased from 1990 to 2013, remaining in the top 3 countries for soybean production. Correspondingly, Brazil was the country with the highest deforestation rate during those years. Similarly, Argentina has also appeared in both Plots 1 and 3; it is amongst the top soybean producers and deforestation contributors. Hence, the trend of high soybean production generally corresponds with high deforestation rates. However, there seems to be a discrepancy for China. China's soybean production increased significantly from 1990 to 2013, and rose from second to first in soybean production. Yet, China had the world's highest reforestation rate during those years. This discrepancy can be explained from the fact that China has been importing raw soybeans for soybean production from places like Brazil, where raw soybeans are actually grown (OEC

World). Thus, China does not have to clear land to make way for soybean plantations. Hence, we maintain that the overarching trend remains valid. Additionally, we note that although soybean production in top countries has increased, deforestation rates have fallen. This trend seems to contradict the point that soybean is one of the main deforestation drivers. This discrepancy can be explained by efficiency improvements that require less land use per unit of production (The Conversation), and the presence of other factors that drive deforestation more (for instance, pasture for Brazil as seen from Plot 2). Hence, this trend does not show the full picture, and we maintain that soybean is still one of the main drivers of deforestation.

5. References

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

https://www.wwf.org.uk/learn/effects-of/deforestation#:\~:text=Trees%20absorb%20and%20store%20carbon,we%20don't%20

https://ourworldindata.org/soy

https://oec.world/en/profile/hs/soybeans?yearSelector1=2010

https://time.com/amazon-rainforest-disappearing/

https://theconversation.com/strict-amazon-protections-made-brazilian-farmers-more-productive-new-research-shows-105789

6. About

This project was developed for DSA2101 Essential Data Analytics Tools: Data Visualization at the National University of Singapore (NUS). The team is comprised of the following members:

Name	Year	Major	Responsibilities
Ashley Wong Ping Rui	2	Data Science and Economics	Code
Calla Koh Wei Yu	2	Economics	Write-up
Khalisah Binte Shari	2	Data Science and Economics	Write-up
Ong Jing Jun	2	Data Science and Analytics	Code
Wayne Ng Jun Cheng	3	Data Science and Analytics	Code