Deforestation

DSA2101 Project 96671

2024-07-05

Contents

Introduction

The deforestation data set provides vital information about deforestation trends around the world, as well as a data set specifically about Brazil. This data is from *Our World in Data*, comprising of 5 CSV files. Breaking down these CSV files:

1. `forest.csv` shows the change in forest area every 5 years.
2. `forest_area.csv` shows the change in forest area as a percentage of the global forest area.
3. `brazil_loss.csv` shows the loss of Brazil's forests due to various reasons such as pasture and commercial use.
4. `soybean_use.csv` shows the use and production of soybeans by country and year (annually).
5. `vegetable_oil.csv` shows the trend of vegetable oil production by year and country.

The question that our group is examining is **Who is the main contributor of deforestation over time, and why?**. In our project, we are using 4 data sets mentioned above, namely `forest.csv`, `brazil_loss.csv`, `soybean_use.csv` and `vegetable_oil.csv`.

Our group aims to identify the country with the most extensive forest destruction and the highest deforestation rates across the years. Additionally, we intend to discover the underlying causes behind the high deforestation rates in that country.

Our group is interested in this question as we recognise the important role that forests play in our ecosystem. They act as the lungs of the earth, absorbing carbon dioxide and reducing heat trapped in the atmosphere *("Why Deforestation Matters—and What We Can Do to Stop It")*, thus lowering global temperature. They also provide habitats for indigenous people and a diverse range of wildlife species *(WWF, n.d.)*. Without them, air temperatures increase, contributing to global warming, and wildlife and indigenous people lose

their homes. Analysing the main contributors and causes of deforestation allows stakeholders and relevant agencies to intervene and take necessary actions.

To address this question, we will start off by using the `forest.csv` data set with the "net_forest_conversion" variable, indicating whether forest area has increased or decreased in five-year intervals. Using this data, we will generate a world map to illustrate the extent of forest destruction across the world, thus identifying the country with the highest total forest area destroyed till 2010. The use of `brazil_loss.csv`, `soybean_use.csv` and `vegetable_oil.csv` data set will be explained in the later part of the report.

Data cleaning and visualisation

Importing libraries

```
library(readr)
#install.packages("tidytuesdayR")
library(tidyverse)
#install.packages("maps")
library(maps)
library(scales)
library(ggthemes)
# install.packages("gridExtra")
library(gridExtra)
```

Importing datasets

```
forest <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/202
forest_area <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/dat
brazil_loss <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/dat
soybean_use <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/dat
vegetable_oil <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/c
```

First visualisation

Data cleaning for `forest.csv`

```
forest_chng = forest %>%
  filter(year >= 2000, year <= 2010, entity != "World", code != "NA") %>%
  group_by(entity) %>%
  summarise(net_forest_conversion = sum(net_forest_conversion)) %>%
  ungroup() %>%
  filter(net_forest_conversion <0) %>%
  slice_min(net_forest_conversion, n = 5) %>%
  mutate(net_forest_conversion = str_sub(net_forest_conversion,2)) %>%
  rename(forest_loss = net_forest_conversion) %>%
  mutate(forest_loss = as.numeric(forest_loss))

forest_chng
```

```
## # A tibble: 5 x 2
##   entity    forest_loss
##   <chr>           <dbl>
## 1 Brazil        5489970
## 2 Paraguay       756390
```

```
## 3 Tanzania       744000
## 4 Myanmar        632410
## 5 Argentina      539800
```

Based on the {forest_chng} data frame, Brazil has experienced the greatest forest loss between 2000 and 2010, with a loss of 725% greater than that of Paraguay, which came in second.

Plotting first visualisation

```r
# Retrieve the top 5 entities with highest net forest loss from 2000 to 2010
top5 = forest_chng %>%
  pull(entity)

# Inserting world data
world = map_data("world")

# Data cleaning for world
world = world %>%
  select(-subregion)
forest_chng_world = left_join(forest_chng, world, by = c("entity" = "region")) %>%
  na.omit()

# get mean latitude and longitude for top 5 entities
top5_lat_long <- forest_chng_world[,c("entity", "long","lat")]

top5_lat_long <- top5_lat_long %>%
  group_by(entity)%>%
  summarize(avg_long = mean(long, na.rm = TRUE),
            avg_lat = mean(lat, na.rm = TRUE))

# Visualisation of world map with deforestation
ggplot(data = world, aes(x = long, y = lat, group = group)) +
  geom_polygon(fill = "lightgrey", color = "white") +
  geom_polygon(data = forest_chng_world, aes(fill = forest_loss)) +
  geom_text(data = top5_lat_long, color="red", size = 3, aes(x = avg_long, y = avg_lat, label = entity,
  theme_minimal() +
  theme(legend.position = "bottom", axis.ticks = element_blank(), axis.text = element_blank()) +
  labs(x = "", y = "", fill = "Forest loss in hectares") +
  ggtitle("Map Plot of Forest Loss from 2000 to 2010") +
    scale_fill_continuous(labels = scales::comma_format(scale = 1e-6, suffix = "M"),
                      low = "lightblue", high = "darkblue")  + # change the legend title for better r
  guides(fill = guide_colorbar(reverse=FALSE)) # flip the legend around to order from lowest net forest
```
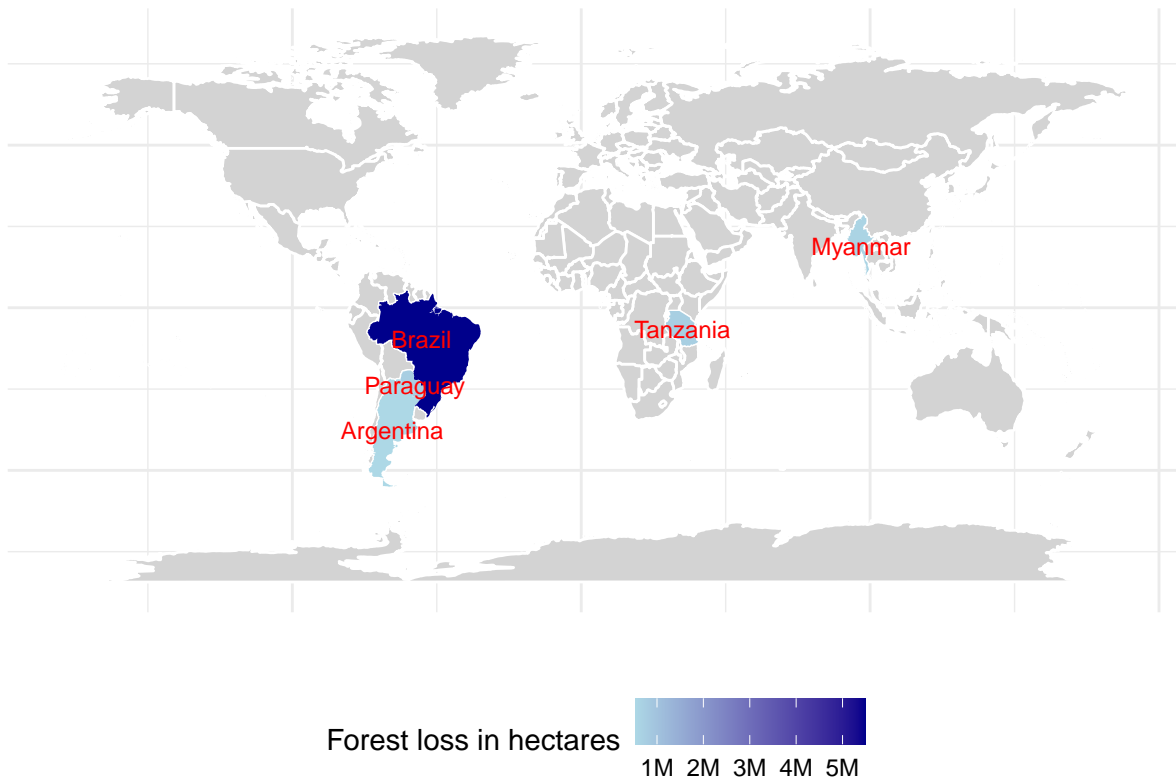
Map Plot of Forest Loss from 2000 to 2010

Forest loss in hectares
1M 2M 3M 4M 5M

The world map plot illustrates the forest loss from 2000 to 2010, using a gradient where lighter shades of blue indicate less forest loss and darker shades represent greater loss. Brazil appears to have the darkest shades of blue, indicating the highest forest loss in hectares compared to other countries.

This map plot clearly shows that Brazil is by far the largest contributor to deforestation with around 5 million hectares of forest loss in 10 years while the runner up comes in around 1 million hectares.

A map plot was chosen as it allows us to immediately spot and identify any obvious outliers. Since our first visualisation aims to only spot outliers, we reserve the information gathering abilities of other plots for identifying causes and trends when drilling down into Brazil.

Second visualisation

Having established Brazil as the primary contributor to deforestation, our focus now shifts to uncovering the underlying causes. For this purpose, we will utilise the `brazil_loss.csv` data set, comprising 11 distinct causes potentially contributing to deforestation in Brazil.

Data cleaning for `brazil_loss.csv`

```
# Data cleaning for brazil_loss dataset
brazil_long = brazil_loss %>%
  select(-code, -entity)
brazil_long = pivot_longer(brazil_long, cols = -year, names_to = "Loss_type", values_to = "Losses")

# Renaming variables
brazil_long <- brazil_long %>%
  mutate(Causes = recode(Loss_type, "pasture" = "Pasture",
                         "small_scale_clearing" = "Clearing Land",
                         "selective_logging" = "Selective Logging",
```

```
                            "commercial_crops" = "Commercial Crops",
                            "tree_plantations_including_palm" = "Tree Plantations",
                            "mining" = "Mining",
                            "fire" = "Fire",
                            "roads" = "Roads",
                            "natural_disturbances" = "Natural Disturbances",
                            "other_infrastructure" = "Other Infrastructure",
                            "flooding_due_to_dams" = "Flooding")) %>%
  select(-Loss_type)

# Data cleaning to filter causes of Brazil by percentage from 2001-2010
Causes_in_Brazil = brazil_long %>%
  filter(year <= 2010) %>%
  group_by(year) %>%
  mutate(pct = Losses/sum(Losses)*100,
         pct = round(pct, 2)) %>%
  mutate(Causes = reorder(Causes, pct))
```

Plotting second visualisation

```
ggplot(Causes_in_Brazil, aes(x = pct, y = Causes)) +
  geom_col(aes(fill = Causes),
           position = "dodge") + #it will produce side by side bars
  facet_wrap(~ year) + #produce different graphs for years
  labs(title = "Percentage of Causes of Deforestation in Brazil from 2001 to 2010",
       x = "Percentage", y = "Causes", fill = "") +
  scale_color_discrete() +
  xlim(0,75)
```

Percentage of Causes of Deforestation in Brazil from 2001 to 2
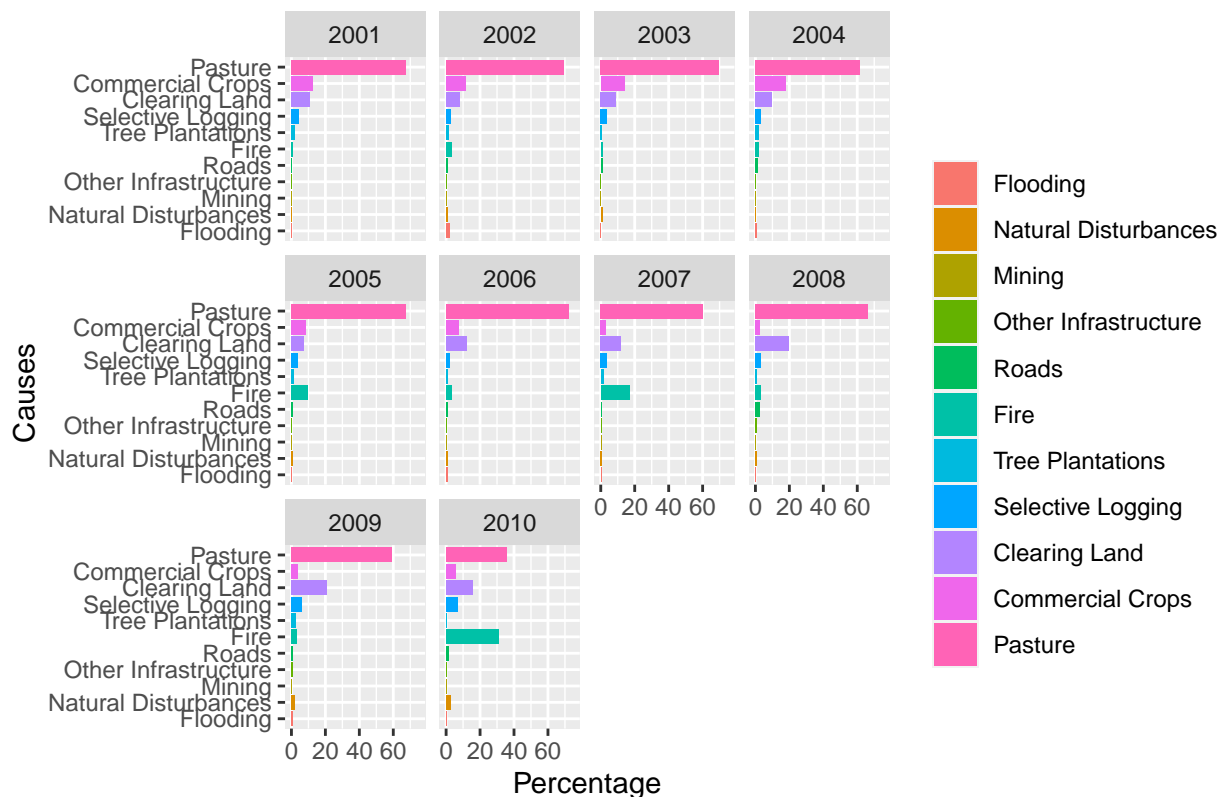
The bar plot reveals that Pasture is consistently the primary cause of deforestation from 2001 to 2010. It also shows that in 2010 there was a sharp decrease in percentage of deforestation attributed to pasture and commercial crops while the percentage of deforestation attributed by fire significant increased, particularly from 2009 to 2010.

A bar plot is used to compare the percentages of deforestation for various causes, with each bar representing the percentage attributed to the specific cause. By observing which bar is the longest with different colours, we can better differentiate them and easily discern the primary cause of deforestation in Brazil is pasture. Adding `facet_wrap` layer with allows us to analyse and compare the causes of deforestation across different years to identify trends over time.

Third visualisation

In addition to the 11 causes documented in the `brazil_loss.csv` data set, our investigation will extend to explore additional potential contributors to Brazil's high deforestation rate. As of 2022, Brazil ranked 25th globally in terms of total exports *(The Observatory of Economic Complexity, 2022)*. Considering this, we hypothesise that the elevated deforestation rate in Brazil may be partially due to the cultivation of soybeans and production of vegetable oil. Both activities necessitate significant land use, which may consequently lead to heightened deforestation.

To validate this assumption, we will now delve into the `vegetable_oil.csv` and `soybean_use.csv` data sets.

Data cleaning for `vegetable_oil.csv`

```
vegetable_oil_clean = vegetable_oil %>% # Change all NA values to 0   .
  mutate(production = coalesce(production, 0))
# Pivot wider to combine production values of all categories into 1 .
vegetable_oil_wide = pivot_wider(vegetable_oil_clean, id_cols = "entity", names_from = "year", values_f
```

6

```r
# Revert back to long data
vegetable_oil_long <- pivot_longer(vegetable_oil_wide, cols = -entity, names_to = "year", values_to = "
  mutate(production = sapply(production, sum)) %>%
  mutate(year = as.double(year)) %>%
  rename(vege_oil_production_tonnes = production)
```

Data merging for `vegetable_oil.csv` and `soybean_use.csv`

We merged the soybean use and vegetable oil data sets to investigate the underlying reasons for Brazil's
significant contribution to global deforestation.

```r
# merging the soybean_use dataset with vegetable_oil dataset
soy_vege = inner_join(soybean_use, vegetable_oil_long) %>%
  mutate(soy_production_tonnes = human_food + animal_feed + processed) %>%
  filter(code != "NA", entity != "World") %>%
  filter(year > 1999, year < 2011) %>%
  mutate(across(everything(), ~ifelse(is.na(.), 0, .)))

soy_vege_total_soy = soy_vege %>%
  group_by(entity) %>%
  summarise(vege_oil_production_tonnes = sum(vege_oil_production_tonnes),
            soy_production_tonnes = sum(soy_production_tonnes)) %>%
  ungroup() %>%
  arrange(desc(soy_production_tonnes)) %>%
  slice_max(soy_production_tonnes, n = 5) %>%
  pull(entity)

soy_vege_total_vege = soy_vege %>%
  group_by(entity) %>%
  summarise(vege_oil_production_tonnes = sum(vege_oil_production_tonnes),
            soy_production_tonnes = sum(soy_production_tonnes)) %>%
  ungroup() %>%
  arrange(desc(vege_oil_production_tonnes)) %>%
  slice_max(vege_oil_production_tonnes, n = 10) %>%
  pull(entity)

soy_vege_top_5_soy = soy_vege %>%
  filter(entity %in% soy_vege_total_soy)

soy_vege_top_5_vege = soy_vege %>%
  filter(entity %in% soy_vege_total_vege)

# summary of soy_vege_top_5
summary(soy_vege)
```

```
##     entity              code                year          human_food
##  Length:1765        Length:1765        Min.   :2000   Min.   :      0
##  Class :character   Class :character   1st Qu.:2002   1st Qu.:      0
##  Mode  :character   Mode  :character   Median :2005   Median :      0
##                                        Mean   :2005   Mean   :  55997
##                                        3rd Qu.:2008   3rd Qu.:   4000
##                                        Max.   :2010   Max.   :5378000
##   animal_feed        processed      vege_oil_production_tonnes
##  Min.   :      0   Min.   :      0   Min.   :      0
```

```
##  1st Qu.:       0   1st Qu.:         0   1st Qu.:     6417
##  Median :       0   Median :      2000   Median :    51330
##  Mean   :   58474   Mean   :   1107341   Mean   :   733732
##  3rd Qu.:    1000   3rd Qu.:    137000   3rd Qu.:   280383
##  Max.   : 9016000   Max.   :  49152000   Max.   : 25618981
##  soy_production_tonnes
##  Min.   :       0
##  1st Qu.:       0
##  Median :       0
##  Mean   : 1100760
##  3rd Qu.:    6000
##  Max.   :63214000
```

From the summary, we can observe that soy and vegetable oil productions are substantial components of the data. These figures underscore the importance of soy and vegetable oil production, suggesting their significant contribution to the agricultural landscape. We will now proceed to visualize the data to gain further insights into the trends and patterns of soybean and vegetable oil production over time.

Plotting third visualisation

```r
top5soyvege_combined = unique(rbind(soy_vege_top_5_soy, soy_vege_top_5_vege))

top5soyvege_combined_long = top5soyvege_combined %>%
  rename(`Vegetable Oil` = vege_oil_production_tonnes, `Soy Products` = soy_production_tonnes, Country =
  pivot_longer(cols = c(`Vegetable Oil`, `Soy Products`),  # Columns to pivot
            names_to = "production_type",  # Name of the new column for production type
            values_to = "production_amount") # Name of the new column for production amount


ggplot(top5soyvege_combined_long, aes(x = year, y = production_amount, color = Country)) +
  geom_line() + geom_point() + # it will produce side by side bars
  facet_wrap(~ production_type) + #produce different graphs for years
  labs(title = "Top producers of Soy products and Vegetable Oils from 2000 to 2010",
      x = "Year", y = "Commercial Crops (Tonnes)") +
  scale_color_discrete() +scale_x_continuous(breaks = seq(2000, 2010, by = 2)) +
  scale_y_continuous(labels=c('0M', '20M', '40M', '60M')) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Top producers of Soy products and Vegetable Oils from 2000 to 2010
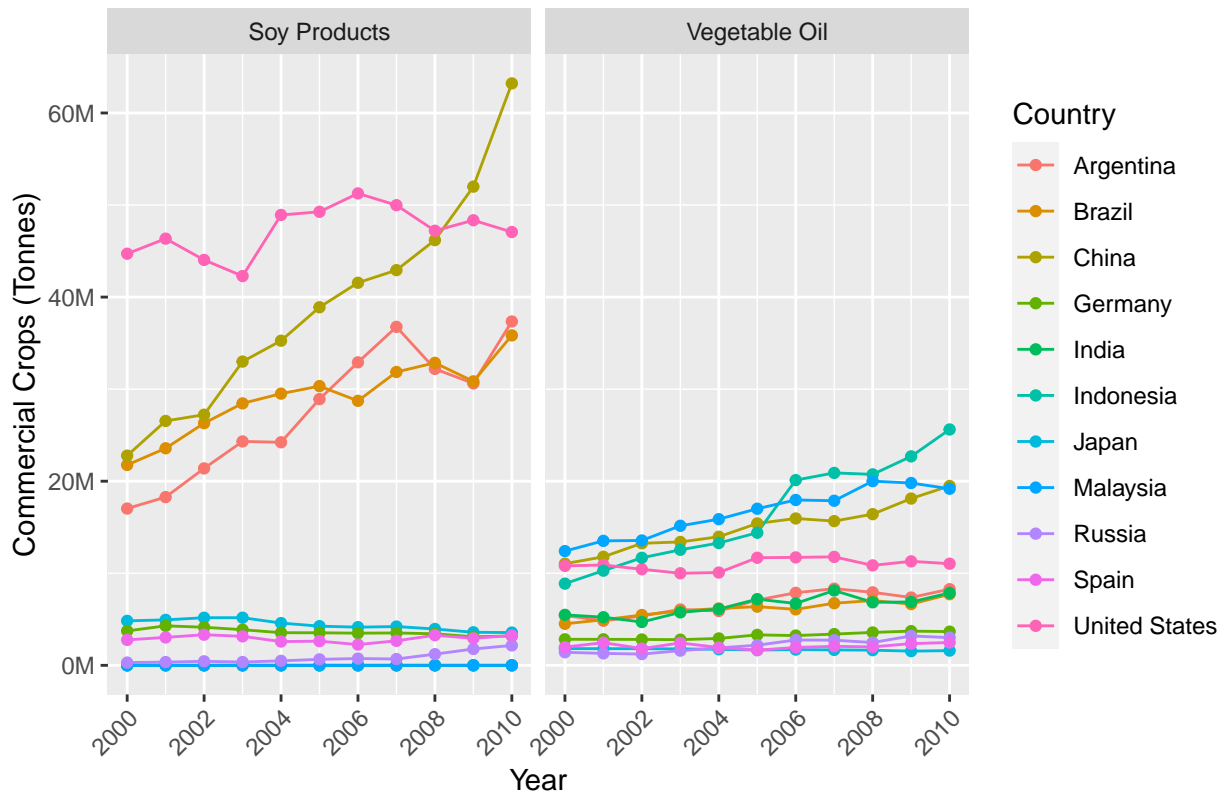
The line graphs illustrates a gradual increase in Brazil's soy and vegetable oil production from 2000 to 2010. Initially, Brazil held the third position globally in soy production in 2000. However, Argentina surpassed Brazil by a slight margin during the decade, causing Brazil to drop to the fourth position by 2010.

As for vegetable oil, Brazil consistently maintains its 7th position throughout the decade. This stable ranking indicates that Brazil consistently holds a significant position in the global vegetable oil market.

Moreover, the line graphs, as shown in Visualisation 3, aligns with the trends observed in Visualisation 2. Both soy and vegetable oil production in Brazil experienced a decline around 2005. This decline coincided with a decrease in deforestation rate for commercial crops, as seen in Visualisation 2.

In summary, the line graphs were chosen to facilitate easy comparison of soy and vegetable oil production for each country. The use of facet wrap also allows for separate observation of soy and vegetable production on distinct graphs. The use of colours helps differentiate between multiple lines when they overlap. Adding `geom_point` facilitates easy identification of individual data points for each country in each year. This makes it easier to identify trends across the entire time period, as well as within that duration.

Discussion

Commencing with the first visualisation (map plot), our aim was to identify the top 5 countries contributing to deforestation to address our main question. Brazil emerged as the main contributor of deforestation by a large margin, accounting for approximately 5 million hectares of deforestation. This prompts a deeper investigation into the factors driving deforestation within Brazil, to further answer our question as to why Brazil has the highest deforestation rate globally.

The second visualisation (bar plot) delved into the primary drivers of deforestation in Brazil annually, revealing pasture as a significant driver, consistently outweighing other contributors. Pasture is defined to be 'lands that are primarily used for the production of adapted, domesticated forage plants for livestock' *(US EPA, 2024)*. Given Brazil's status as the world's second largest producer and largest exporter of beef *(SEI, 2022)*, the

reliance of pasture lands for live stocks emerges as a key driver of deforestation in Brazil. Notably, a significant reduction in pasture-related deforestation was observed in 2010, conciding with a significant increase in fire incidents. Further investigation unveiled that 2010 experienced unusual environmental conditions, leading to a three-fold increase in fire incidents in 5 Brazilian states *(Earth Observatory, 2010)*. Up till here, we have largely answered part of our question as to why Brazil has such high deforestation rates.

Subsequently, our interest turned to investigating whether soybean cultivation and vegetable oil production could also contribute to deforestation in Brazil. The third visualisation (line plot) depicted the annual production of soy and vegetable oil by top producers from 2000 to 2010. Although Brazil is not the top producer, it remains as a significant player in both sectors. Research revealed that Brazil's lower ranking in soy and vegetable oil production is partially attributed to China's aim for self-sufficiency in food production and innovative practices in the USA *(South China Morning Post, 2023)*.

Moreover, insights derived from the third visualisation indicated a gradual increase in soy production since 2010, accompanied by a decline in the land used for pasture. This suggests that soy production is replacing deforested pasture lands, contributing to Brazil's agricultural landscape *(SEI, 2022)*.

As a country who is heavily reliant on agricultural exports, Brazil must enhance its agricultural output to fuel economic growth and remain competitive globally. However, achieving these objectives often involves clearing land for pasture and crop cultivation, leading to alarming deforestation rates *(Ferreira, Biazzin, & Hong, 2024)*. Thus, Brazil experience alarming rates of deforestation, perpetuating its position as the largest contributor to deforestation globally.

In conclusion, stringing the information gathered through the 3 visualisations, we have hence reached the answer for the question: Who is the main contributor of deforestation over time, and why? The answer is then as follows: **Brazil emerges as the largest contributor to deforestation, primarily driven by pasture-related activities fueled by cattle production, alongside increasing demands for soy and vegetable oil products.**

Reference

"Agricultural Pasture, Rangeland and Grazing | US EPA." (2024).

https://www.epa.gov/agriculture/agricultural-pasture-rangeland-and-grazing#:~:text=Pastures%20are% 20those%20lands%20that,pastures%2C%20and%20croplands%20producing%20forages.

"Brazil (BRA) Exports, Imports, and Trade Partners | the Observatory of Economic Complexity." (2022).

https://oec.world/en/profile/country/bra.

"China Food Security: Focus on Increased Domestic Soybean Production Amid Self-Sufficiency Drive | South China Morning Post." (2023).

https://www.scmp.com/economy/china-economy/article/3217744/china-food-security-focus-increased-domestic-soybean-production-amid-self-sufficiency-drive.

"Connecting Exports of Brazilian Soy to Deforestation | SEI." (2022).

https://www.sei.org/features/connecting-exports-of-brazilian-soy-to-deforestation/.

Ferreira, F. C. M., Biazzin, C., & Hong, P. C. (2024). Transition paths of Brazil from an agricultural economy to a regional powerhouse: A global supply chain perspective. Sustainability, 16(7), 2872. https://doi.org/10.3390/su16072872

"Fires in Brazil| Earth Observatory." (2010).

https://earthobservatory.nasa.gov/images/45349/fires-in-brazil.

"How United States Became World's Top Soybean Producer, What It Does with Them, and Why US-China Trade War Makes Crop's Future Uncertain | South China Morning Post." (2019).

https://www.scmp.com/lifestyle/food-drink/article/3042549/how-united-states-became-worlds-top-soybean-producer-what-it.

"Learn the Effects of Deforestation | WWF." (n.d).

https://www.wwf.org.uk/learn/effects-of/deforestation.

"Tidytuesday/Data/2021/2021-04-06/Readme.md at Master Rfordatascience/Tidytuesday."

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

"Trase: Brazilian Beef Exports and Deforestation | SEI." (2023).

https://www.sei.org/features/trase-brazil-beef-exports-deforestation/.

"Why Deforestation Matters—and What We Can Do to Stop It." (2022).

https://www.nationalgeographic.com/environment/article/deforestation.

About

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

Name	Year	Major	Responsibilities
Xiao Xuan	2	Data Science and Analytics	Visualisations
Nodoka	2	Data Science and Analytics	Formatting of Rmd and Discussion
Jia Yi	2	Data Science and Analytics	Visualisations, Discussion & Descriptions
Ryan	2	Data Science and Analytics	Cleaning and Visualisations
Min Yi	2	Psychology	Descriptions