World Defortestation: A visual analysis

Shah Ali Gardezi

1/14/2022

Introduction

The aim of this project is to visualize the worldwide trends of deforestation and report facts for the which will serve as a useful information for global authorities to take corrective measures before it is too late. The data used for this analysis is acquired from **TidyTuesday**. For the analysis I have used skills learned in the course *Data Visualization in R* and used themes like plotly and gganimate along with data.table for my analysis.

Data

The dataset consists of 4 different tables. Two of these provide factual information about the different aspect of forests such as *forest area* and *net forest conversion* for each country and continents. The other two data table provide information about the *causes* of deforestation and how much does each factor contribute to the overall subject. Interestingly, one of these datasets is specific to Brazil and reports how much forest has been cleared for different purposes in the country.

For this analysis, we will be analyzing; 1) How the deforestation trends change in each continent over the 30 year period? 2) Which are the top 10 countries with positive and negative *Net Forest Conversion* for a given year? 3) How the *Net Forest Conversion* has evolved in Asia and Europe geographically? 4) What is statistical trend in major countries of Europe, where deforestation is a practice to produce raw material (Soybean) for processed foods? 5) What is the trend of deforestation in Brazil by different category?

Set Up and Data Loading

```
## Clear environment
rm(list = ls())

## Loading Library
pacman::p_load(tidyverse,readr, data.table, kableExtra, leaflet, ggpubr, gganimate, magick,g

forest <- fread('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2
forest_area <- fread('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/d
brazil_loss <- fread('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/d
soybean_use <- fread('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/d
```

Explainatory Data Analysis (EDA)

We performed EDA to observe each dataset to check for the type of variables, if there were any missing values in our variables of interest, and see their distribution. The code below was used. We observe that the variables we were interested did not contain any missing values. With regards to the distribution we observed that there was skewness in the data, but since we are not interested in prediction, we do not need to take log of the skewed values.

```
summary <- datasummary( year + net_forest_conversion ~ Mean + Min + Max + N , data = forest,
summary %>% kableExtra::kable_styling(latex_options = "HOLD_position", position = "center")
```

Forest Summary Statistics

	Mean	Min	Max	N
year	2004.33	1990	2015	475
net_forest_conversion	−76931.03	−7818000.00	2360980.00	475

```
summary1 <- datasummary( year + forest_area ~ Mean + Min + Max + N , data = forest_area, tit
summary1 %>% kableExtra::kable_styling(latex_options = "HOLD_position", position = "center")
```

Forest Area Summary Statistics

	Mean	Min	Max	N
year	2005.22	1990	2020	7846
forest_area	1.81	0.00	100.00	7846

```
summary2 <- datasummary(entity + year + commercial_crops + selective_logging + pasture + fir
                        Mean + Min + Max + N , data = brazil_loss, title = "Brazil Loss St
summary2 %>% kableExtra::kable_styling(latex_options = "HOLD_position", position = "center")
```

Brazil Loss Statistics

	Mean	Min	Max	N
Brazil			−Inf	13
year	2007.00	2001	2013	13
commercial_crops	234846.15	52000	747000	13
selective_logging	104846.15	44000	166000	13
pasture	1561769.23	546000	2761000	13
fire	157692.31	26000	537000	13
small_scale_clearing	305769.23	232000	415000	13

```
summary3 <- datasummary( year + human_food + animal_feed + processed ~
                        Mean + Min + Max + N , data = soybean_use, title = "Soybean Use St
summary3 %>% kableExtra::kable_styling(latex_options = "HOLD_position", position = "center")
```

Soybean Use Statistics

	Mean	Min	Max	N
year	1987.74	1961	2013	9897
human_food	164194.28	0	10649000	9682
animal_feed	216220.00	0	17478000	4359
processed	3458245.96	0	227311000	6253

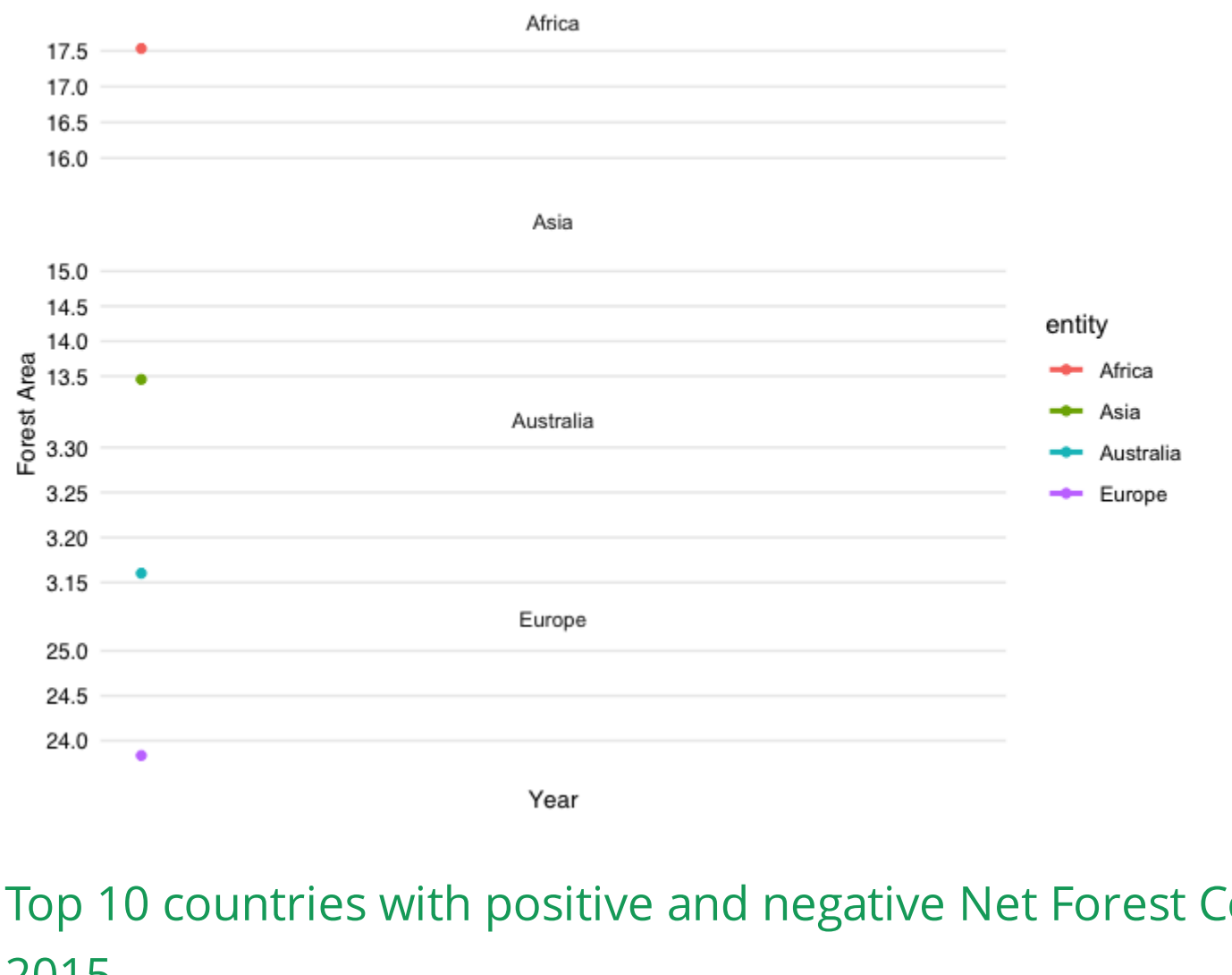
Creating customized theme

```
theme_atraf <- function(){
  font <- "sans"   #assign font family up front

  theme_minimal() %+replace%

    theme(

      #grid elements
      panel.grid.minor = element_blank(),
      panel.background = element_blank(),

      #text elements
      plot.title = element_text(
        family = font,
        size = 10,
        hjust = 0.5,
        vjust = 2,
        color = "Black"),

      plot.subtitle = element_text(
        family = font,
        size = 10,
        hjust = 0.5,
        color = "#2ca25f"),

      plot.caption = element_text(
        family = font,
        size = 6,
        hjust = 1),

      axis.title = element_text(
        family = font,
        size = 10),

      axis.text = element_text(
        family = font,
        size = 9),

      axis.text.x = element_text(
        margin=margin(5, b = 10))
    )
}
```

Desforesation trend in continents over 30 year period

This visualization is the trend of percentage of Forest Area of each continent, depicting how the area changes over the 30 year period. This visualization helps in comparing the continental deforestation side by side. We see that in Asia and Europe, the percentage of forest area increases gradually from 1990 to 2020, whereas in Africa the area is on a linear decline which is quite alarming. However the in Australia we see that that Forest Area has been somewhat constant, but after 2010, there has been an increase in forest cover owing to afforestation campaigns done there.

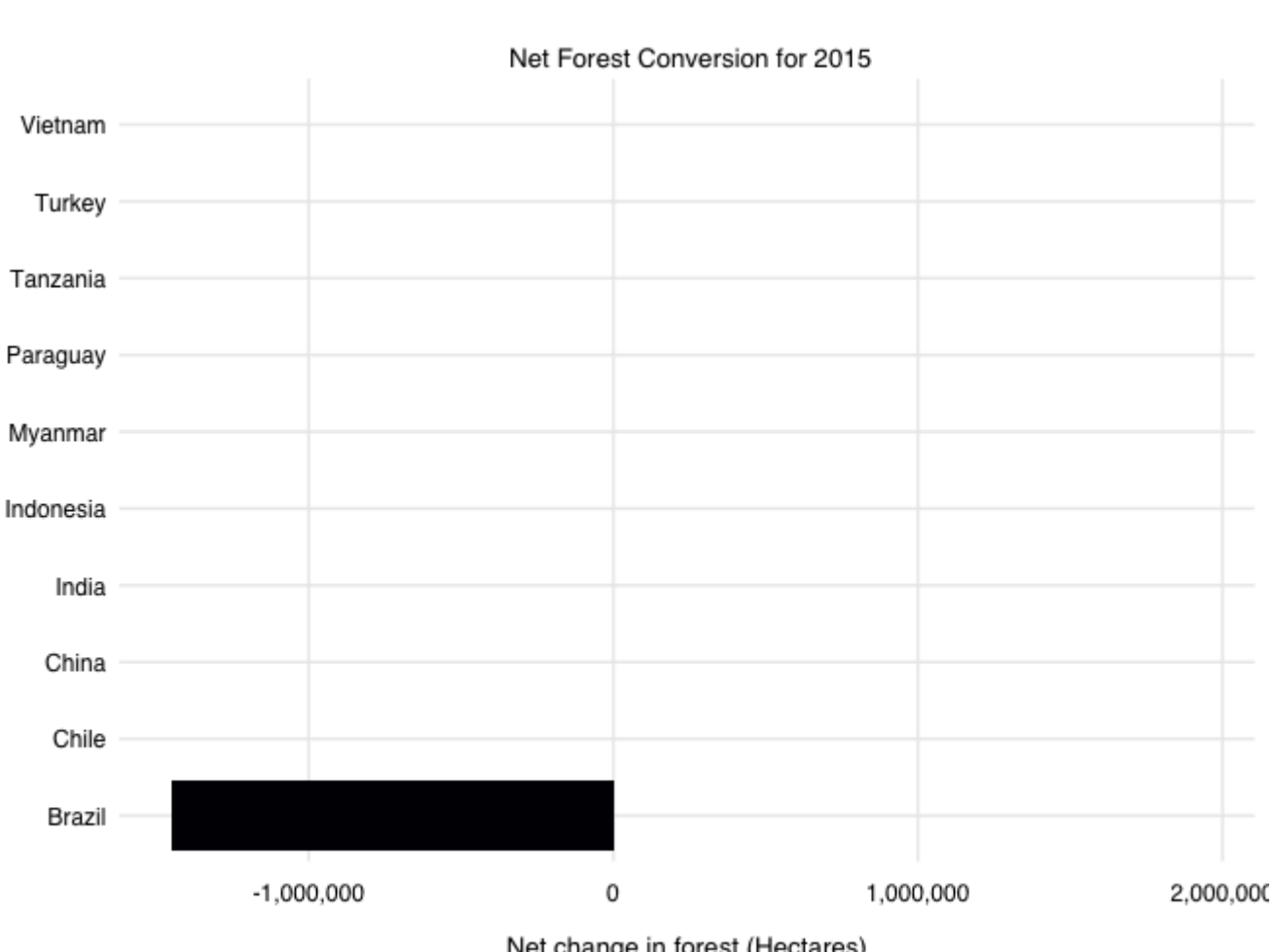
```
a <- forest_area[entity== "Africa"|entity=="Asia"|entity=="Australia"|entity=="Europe"|entit

viz <- ggplot(a) +
  aes(x = year, y = forest_area, colour = entity) +
  geom_line(size = 1.05) +
  scale_fill_viridis_d(option = "inferno", direction = 1) +
  labs(
    x = "Year",
    y = "Forest Area",
    title = "Changes in Forest Year over 30 Years"
  ) +
  theme_atraf()+facet_wrap(vars(entity), scales = "free",ncol=1)+
  geom_point()+
  scale_x_continuous(breaks = 0:200)

viz + transition_reveal(year)
```



Top 10 countries with positive and negative Net Forest Conversion in 2015

This visualization is the pyramid bar graph. It shows a total of 10 countries for 2015, 5 countries with highest positive net forest conversion (depicting afforestation) and 5 countries with highest negative net forest conversion (depicting deforestation).

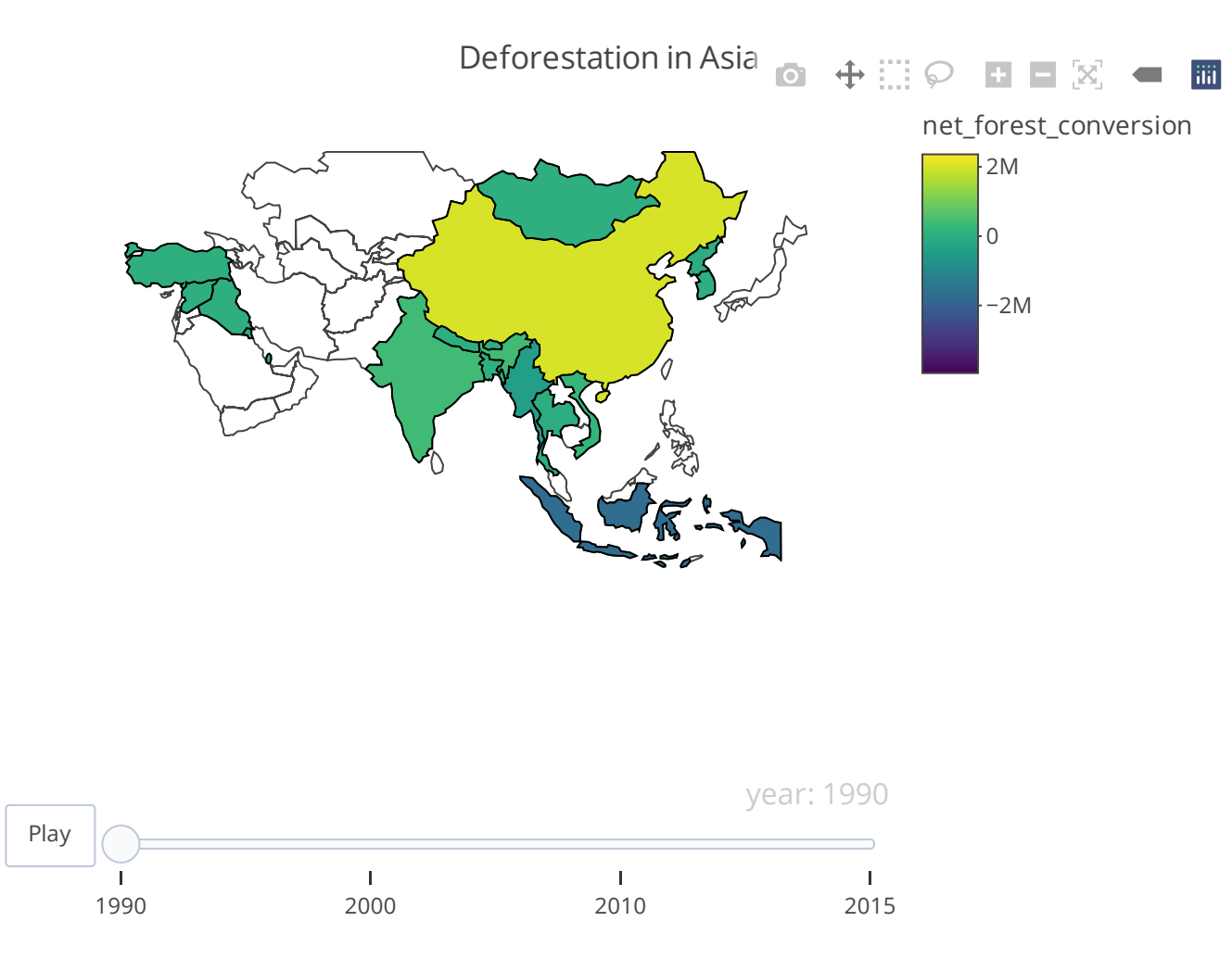
```
b <- forest[year==2015]
b <- b[order(-rank(net_forest_conversion))]


b1 <- b[1:5]
b2 <- b[116:120]
b3 <- rbindlist(list(b1, b2), fill=T)


viz1 <- ggplot(b3,aes(net_forest_conversion, entity, fill= net_forest_conversion > 0 ))+
  geom_col() +  scale_fill_viridis_d(option = "inferno", direction = 1)+
  scale_x_continuous(labels = scales::comma)+
  labs(x="Net change in forest (Hectares)", y= "")+
  ggtitle(" Net Forest Conversion for 2015")+
  theme_atraf()+
  theme(legend.position = "none")


viz1+ transition_states(entity) +shadow_mark(alpha=0.8)+
  ease_aes("linear")
```



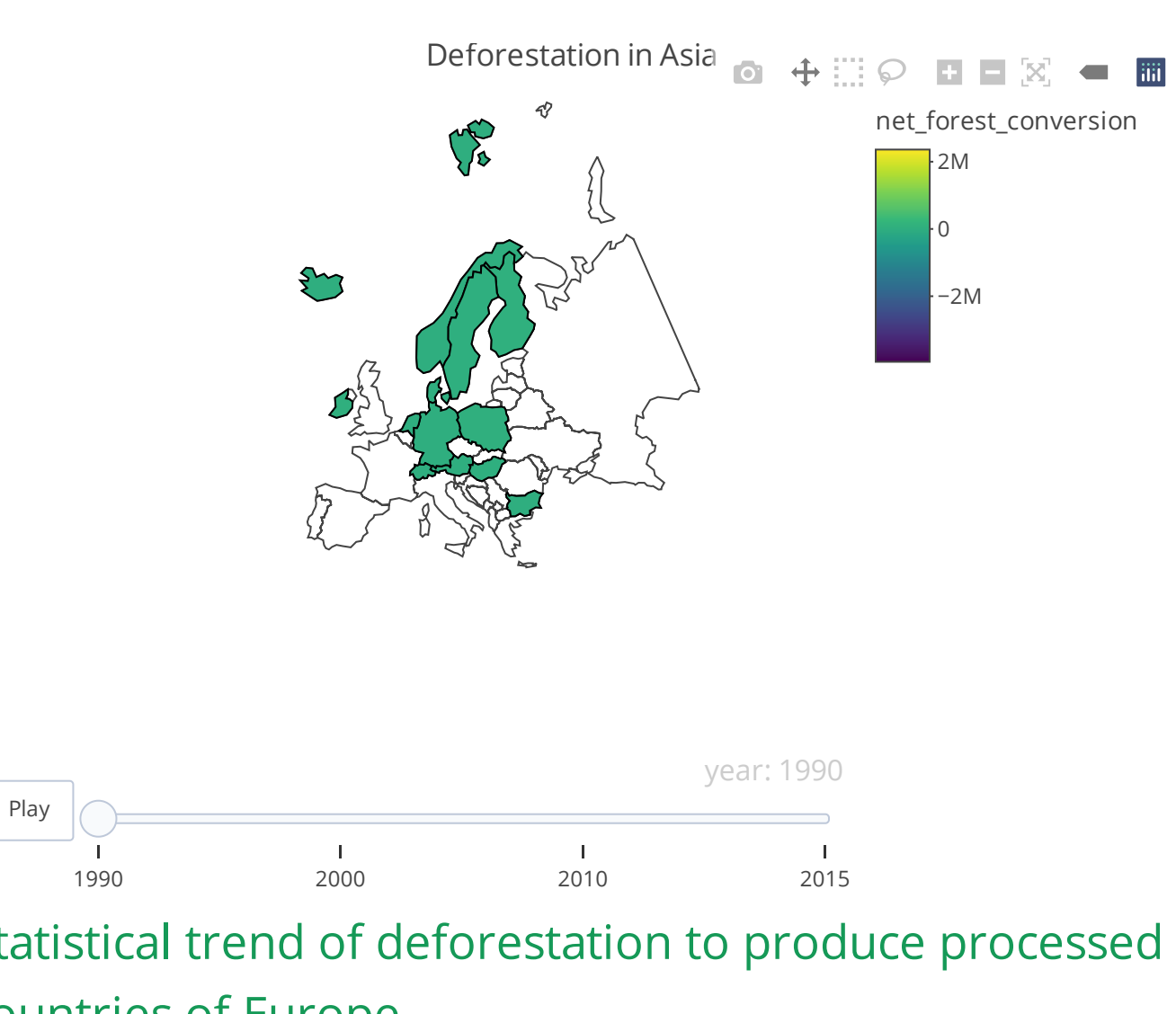
Forest evolution in Asia and Europe

The map below helps us to visualize the Net Forest Conversions in Asia and Europe. This geoprahical map produced using tools like *plotly*. It is an interactive map tools such as lasso-select help us to pinpoint the net forest conversion in the region of choice by hovering over the cursor. The data is in 10 year interval from 1990 to 2010 and then a 5 year interval from 2010 to 2015

```
## Asia ##
mapdata <- forest[code != ""]

mapdata <- mapdata[str_length(code)==3]
mapdata$hover <- paste0(mapdata$entity, "\n", mapdata$net_forest_conversion)

map2<- plot_geo(mapdata,
                locationnode= 'world',
                frame=~year) %>%  add_trace(locations= ~ code ,
                                            z= ~ net_forest_conversion,
                                            zmax=max(mapdata$net_forest_conversion),
                                            zmin=min(mapdata$net_forest_conversion),
                                            color= ~net_forest_conversion,
                                            text = ~hover,
                                            hoverinfo = 'text') %>%
  layout(geo=list(scope="asia"), title="Deforestation in Asia")
map2
```



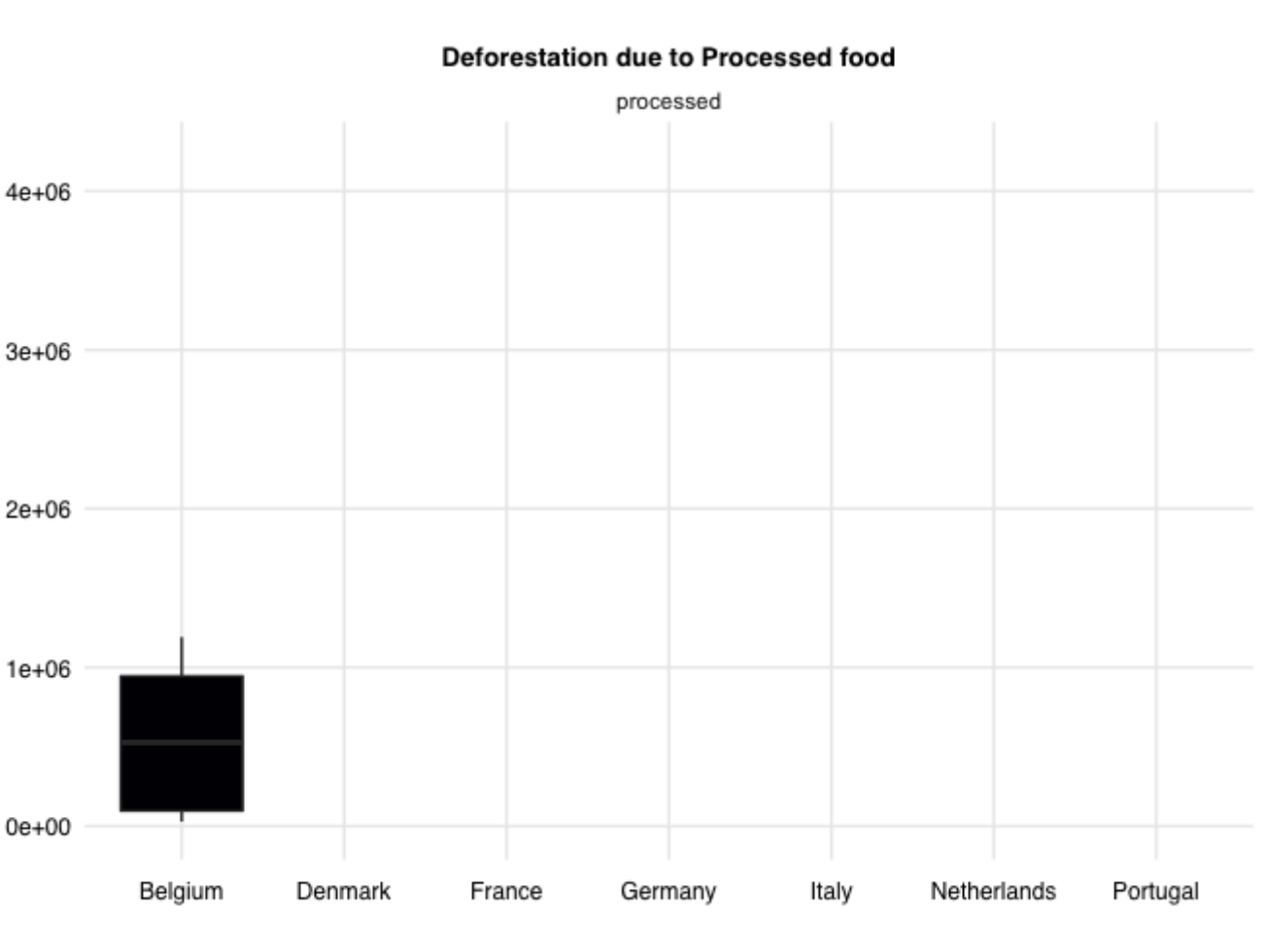

```
## Europe ##

map3<- plot_geo(mapdata,
                locationnode= 'world',
                frame=~year) %>%  add_trace(locations= ~ code ,
                                            z= ~ net_forest_conversion,
                                            zmax=max(mapdata$net_forest_conversion),
                                            zmin=min(mapdata$net_forest_conversion),
                                            color= ~net_forest_conversion,
                                            text = ~hover,
                                            hoverinfo = 'text') %>%
  layout(geo=list(scope="europe"), title="Deforestation in Asia")
map3
```



Statistical trend of deforestation to produce processed food in major countries of Europe

The box-plot shows top 7 countries in Europe, where deforestation is done to make the land available for the production of raw materials such as Soybean for processed foods. Germany and Netherlands were the most effected.

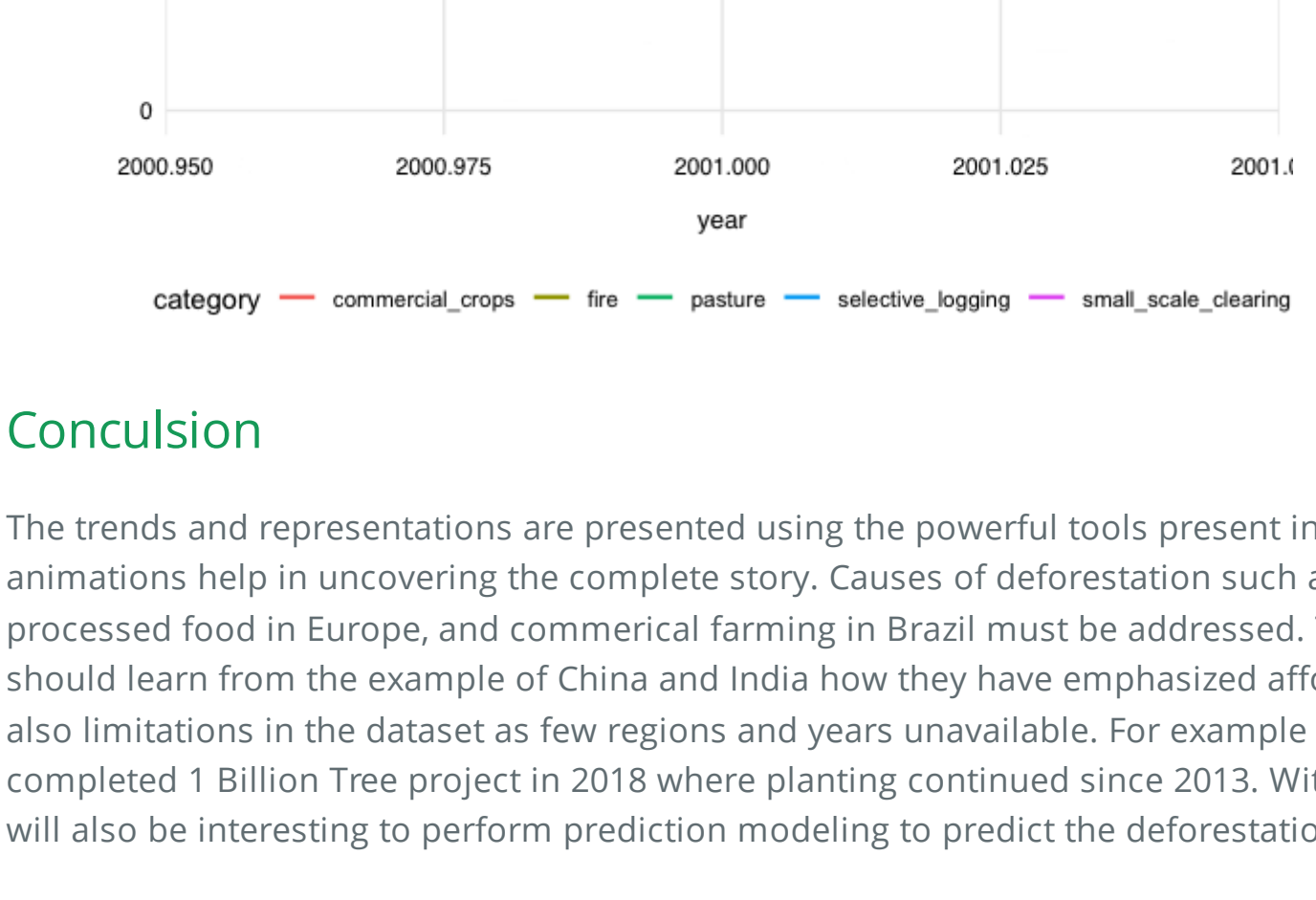
```
s <- soybean_use[code== "AUT"|code=="BEL"|code=="BGR"|code=="HRV"|code=="CYP"|
              code== "CZE"|code=="DNK"|code=="EST"|code=="FIN"|code=="FRA"|
              code== "DEU"|code=="GRC"|code=="HUN"|code=="ITA"|code=="LVA"|
              code== "LTU"|code=="MLT"|code=="NLD"|code=="POL"|code=="PRT",]
bc <- gather(s,"category","value",4:6)

gandu <- bc %>%
  filter(entity %in% c("Belgium", "Denmark", "France", "Germany", "Italy", "Netherlands", "P
  )) %>%
  filter(!(code %in% "EST")) %>%
  filter((category %in% "processed")) %>%
  ggplot() +
  aes(x = entity, y = value, fill = entity) +
  geom_boxplot(shape = "circle") +
  scale_fill_viridis_d(option = "inferno", direction = 1)+
  theme_atraf() +
  theme(plot.title = element_text(face = "bold",
                                hjust = 0.5))+
  facet_wrap(vars(category), scales = "free")+
  theme(legend.position = "none")+
  ggtitle(" Deforestation due to Processed food ")+
  labs(x="",y="Deforestation in Hectares")

gandu+transition_states(entity, wrap =FALSE)+ shadow_mark(alpha=0.5)+
  enter_grow()+
  exit_fade()+
  ease_aes("back-out")
```



What is the trend of deforestation in Brazil by different category?

This visualization is time series analysis showing deforestation in Brazil due to several reasons. Pasture was the main reason but showed a decreasing trend over the period after 2005. The commercial crop and bush fires are the other main contributors to deforestation in Brazil.

```
list <- c("entity","year","commercial_crops","selective_logging","pasture","fire","small_sca
brazil <- brazil_loss[ ,colnames(brazil_loss) %in% list, with=FALSE]


brazil <- gather(brazil,"category","loss",3:7)


loss <- ggplot(brazil) +
  aes(x = year, y = loss, colour = category, group = category) +
  geom_line(size = 0.85) +
  scale_fill_viridis_d(option = "inferno", direction = 1) +
  scale_y_continuous(labels = scales::comma)+
  labs(
    y = "Loss of Lands in Hectare ",
    title = "Loss of land due reasons"
  ) +
  theme_atraf() +
  theme(
    legend.position = "bottom",
    plot.title = element_text(face = "bold",
                              hjust = 0.5)
  )
loss + transition_reveal(year)+
  view_follow(fixed_y = T)
```



Conculsion

The trends and representations are presented using the powerful tools present in the R studio. The animations help in uncovering the complete story. Causes of deforestation such as land clearing for processed food in Europe, and commerical farming in Brazil must be addressed. The world economies should learn from the example of China and India how they have emphasized afforestation. There were also limitations in the dataset as few regions and years unavailable. For example Pakistan recently completed 1 Billion Tree project in 2018 where planting continued since 2013. With addition datasets it will also be interesting to perform prediction modeling to predict the deforestation trends in the future.