Analysis of agricultural production and land use in the world

Visualization Project

Rohan Karthikeyan

03 December 2022

Abstract

In this report, we look at agricultural production and land use in various regions and countries around the world. An in-depth analysis of a few selected regions is provided with additional commentary on correctly performing inter-country comparisons of agricultural output.

1 Introduction

As the world increasingly talks about climate change, the way we eat and produce our food is worth looking into. These are the days where we are witnessing the proliferation of public discourse around the distance food travels from farm to table, the rise of online calculators to determine carbon footprints of different foods, and mounting calls for the need to shift toward plant-based diets. Leveraging and redirecting consumer habits may prove crucial in combating climate change.

Agriculture is triply implicated in global climate change. While agricultural activities directly contribute to climate change, agricultural production is also strongly impacted by climate change. But even as agriculture contributes to, and is impacted by, climate change, it is on (and in) land and land use that one may find myriad possibilities for mitigating greenhouse gas emissions.

Now, the amount of land in use for agricultural production reached about 1.3 billion hectares in 1960, and has since expanded to nearer 1.6 billion hectares. Most analysts have suggested that the limit to arable land available is about 2 billion hectares, or a further 25% increase in land availability.

Of course, placing all of the globally available arable land in service to agricultural production would not prove desirable. There are many other land uses that are required to sustain the planet and speak against further expansion of the total acreage devoted to agriculture.

Biodiversity preservation is dependent upon land use allocations, as habitat conversion is one of the primary drivers of species loss. Also, carbon sequestration depends upon the retention of existing forests, and the expansion of these. In the coming century, with the onset of multiple global problems such as climate change in addition to food security, meeting food requirements from reduced land allocations proves crucial.

In Section 2, we will analyze how various regions and individual countries have used their land for agriculture. In Section 3, we will look at agricultural production levels and its association (if any) with land use patterns, once again on a regional and country level.

2 Agricultural land use over the long-run

Before further analysis, let's define 'agricultural area':

> Agricultural area is the sum of arable land, permanent crops, permanent meadows and pastures.

And the FAO definition for arable land is:

> Land under temporary agricultural crops (multiple-cropped areas are counted only once), temporary meadows for mowing or pasture, land under market and kitchen gardens and land temporarily fallow (less than five years).

Armed with these definitions, let's dig deeper into the data. We are going to analyze data in the time-period 1961 to 2007 provided by FAOSTAT, a freely accessible data platform.

```
land_data = read.csv("Databases/Land_data_regions.csv")
forest_data = read.csv("Databases/Forest_data_regions.csv")
```

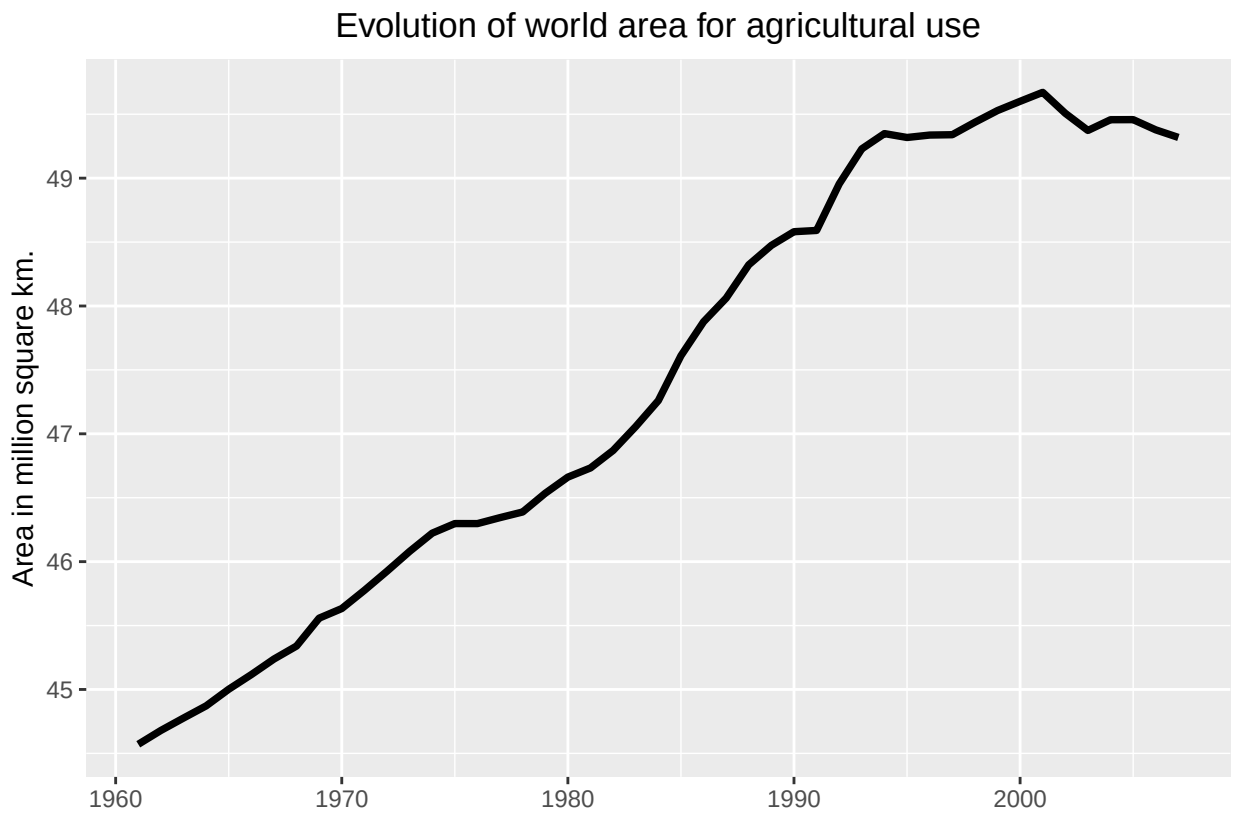
We will start by analyzing the data for the entire world:

```
# Take the agricultural component
agri_records = land_data[land_data$Category=='agricultural_area',c(1,2,3)]

world = agri_records[agri_records$Region == 'World',]
# Convert hectares into million square km.
world$Value = world$Value/1e8
world$Year = as.integer(world$Year)
head(world)
```

```
##      Region Year    Value
## 1521  World 2007 49.31862
## 1522  World 2006 49.37784
## 1523  World 2005 49.45770
## 1524  World 2004 49.45699
## 1525  World 2003 49.37312
## 1526  World 2002 49.50709
```

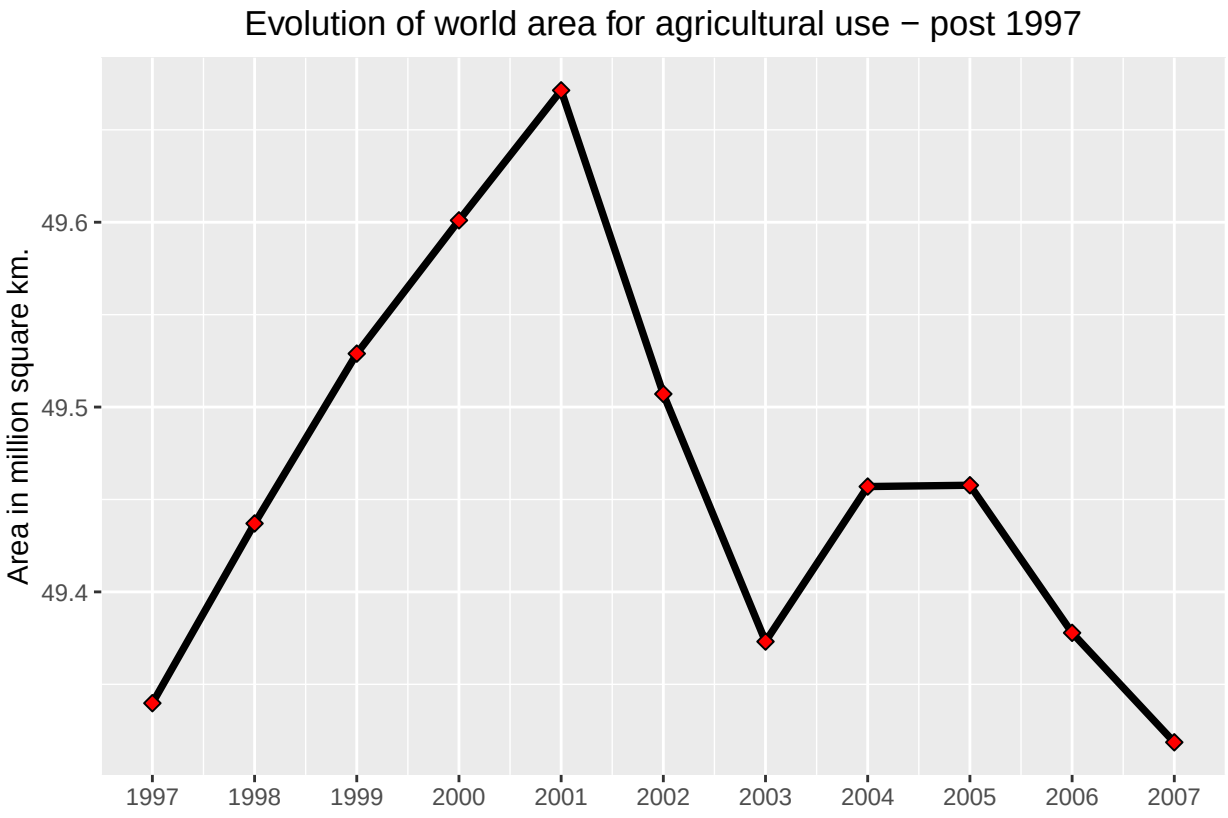
Now, let us plot a graph of the above data:

```
ggplot(world, aes(x = Year, y = Value)) +
  geom_line(size = 1.3) +
  ggtitle("Evolution of world area for agricultural use") +
  labs(x='', y='Area in million square km.')+
  theme(plot.title = element_text(hjust=0.5))
```



Towards the right-hand corner, we can notice that a visible lull in the amount of area used for agriculture. Let's look at that part specifically:

```
ggplot(world[world$Year >= 1997,],
       aes(x = Year, y = Value)) +
  geom_line(size = 1.3) + geom_point(fill='red', shape=23, size=2) +
  ggtitle("Evolution of world area for agricultural use - post 1997") +
  labs(x='', y='Area in million square km.') +
  theme(plot.title = element_text(hjust=0.5)) +
  scale_x_continuous(breaks=c(1997:2007))
```



Interestingly, we observe a dip in the land area starting from 2001. Would there be any disparity between the land used for agriculture in developed and developing regions? Let us look at data starting from 1997 for both types of regions.

2.1 Regional analysis

2.1.1 Developed regions

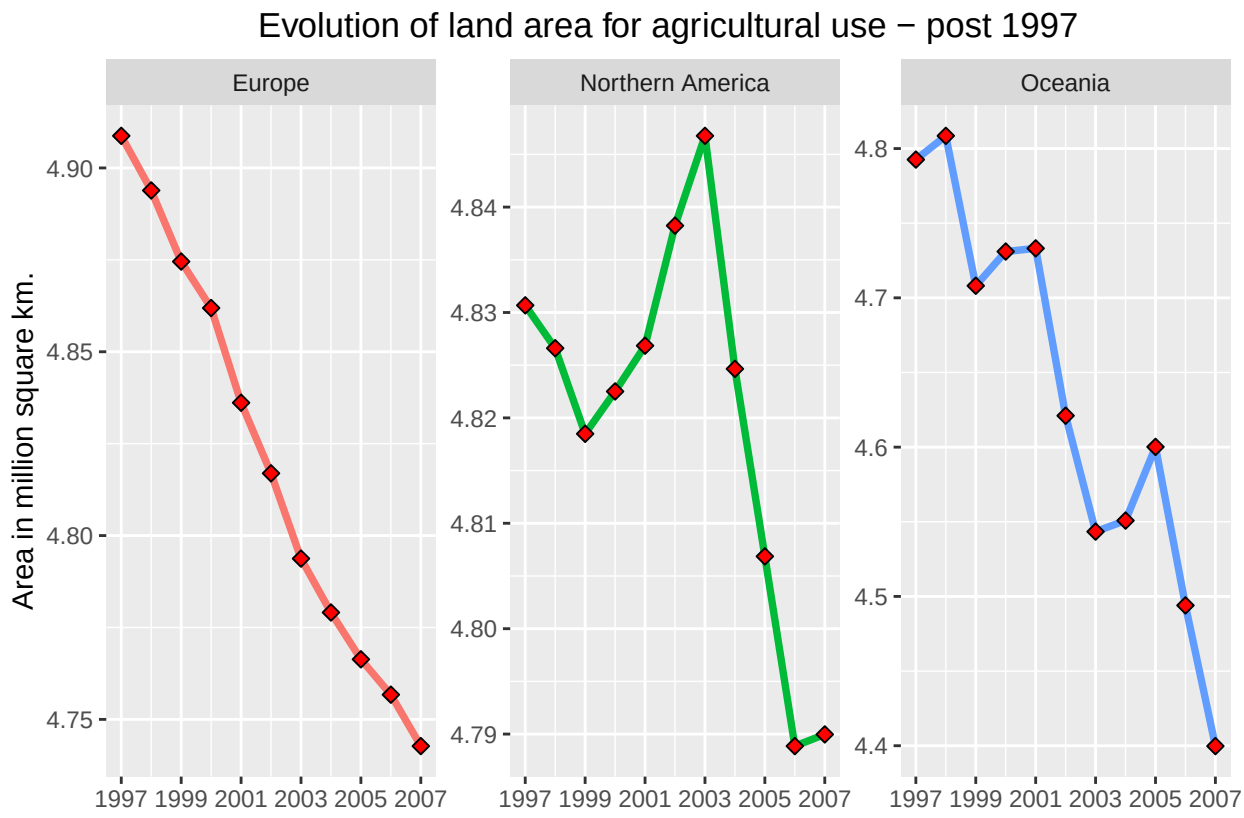
While the UN Statistics Division holds that there is no established convention for the designation of "developed" and "developing" countries or areas, the United Nations Conference on Trade and Development considers that this categorization can continue to be applied:

> The developed economies broadly comprise Northern America and Europe, Israel, Japan and the Republic of Korea, as well as Australia and New Zealand.

Let's analyze the data specifically looking at Northern America, Europe and Oceania.

```
developed = agri_records[agri_records$Region %in% c(
  'Northern America', 'Oceania', 'Europe'),]
developed$Value = developed$Value/1e8
```

```
our_plot = ggplot(developed[developed$Year >= 1997,],
                  aes(x = Year, y = Value)) +
  geom_line(aes(color = Region), size = 1.3) +
  facet_wrap(vars(Region), scales="free_y")

our_plot = our_plot + geom_point(fill='red', shape=23, size=2) +
  ggtitle("Evolution of land area for agricultural use - post 1997") +
  labs(x='', y='Area in million square km.') +
  theme(plot.title = element_text(hjust=0.5), legend.position = 'none') +
  scale_x_continuous(breaks=c(1997, 1999, 2001, 2003, 2005, 2007))
our_plot
```



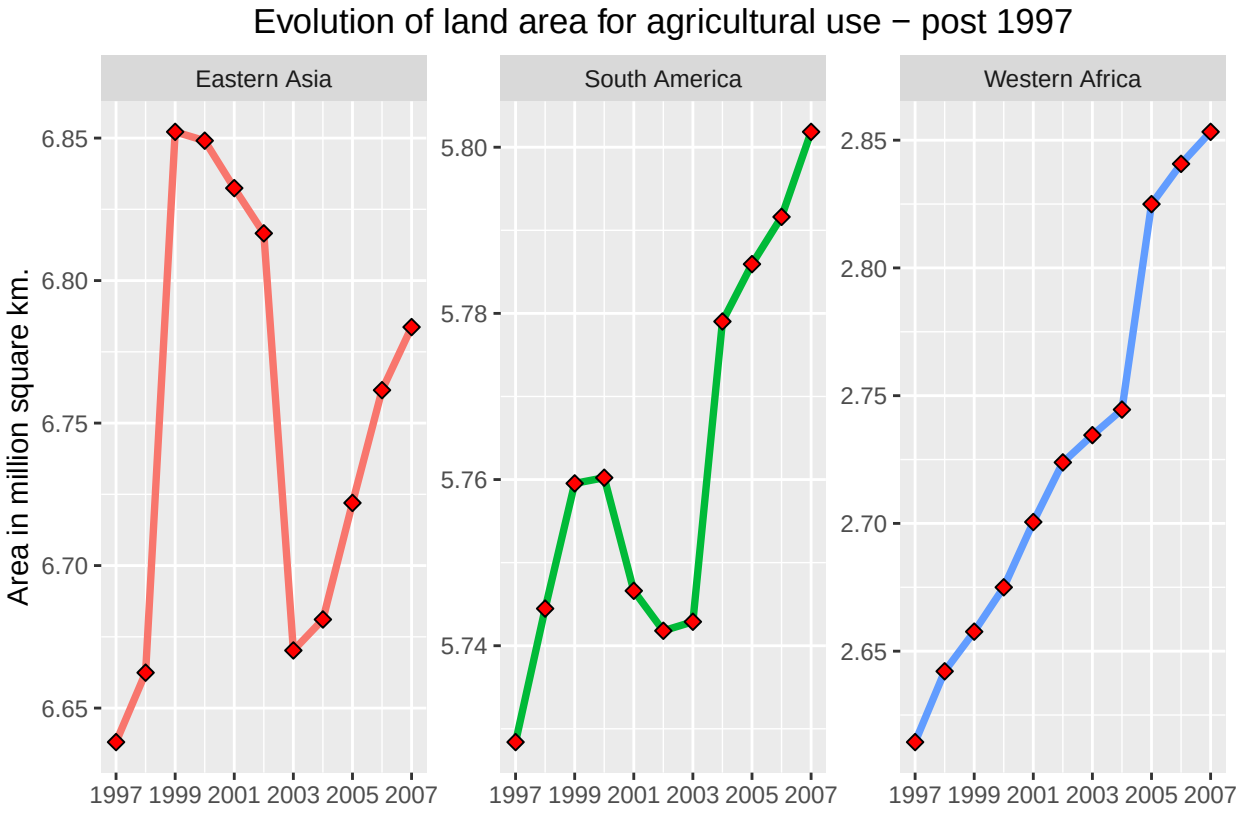
In developed regions, the rate of growth of agricultural land use has turned negative.

2.1.2 Developing regions

As part of developing regions, let's look at Africa and Latin America specifically:

```
developing = agri_records[agri_records$Region %in% c(
  'Eastern Asia', 'South America', 'Western Africa'),]
developing$Value = developing$Value/1e8
```

```
our_plot = ggplot(developing[developing$Year >= 1997,],
                  aes(x = Year, y = Value)) +
  geom_line(aes(color = Region), size = 1.3) +
  facet_wrap(vars(Region), scales="free_y")

our_plot = our_plot + geom_point(fill='red', shape=23, size=2) +
  ggtitle("Evolution of land area for agricultural use - post 1997") +
  labs(x='', y='Area in million square km.') +
  theme(plot.title = element_text(hjust=0.5), legend.position = 'none') +
  scale_x_continuous(breaks=c(1997, 1999, 2001, 2003, 2005, 2007))
our_plot
```



In line with population changes, the primary locations for further land use allocations are situated in developing regions, with sub-Saharan Africa being the main source of new agricultural lands. This imbalance between developed and developing regions is an important reason to look deeper for the source of future food security problems.

Now, with the total land area in the world just under 150 million square km., it's remarkable that we use roughly half of global habitable land for agriculture. But just how much of total land area is utilized for agriculture across the world?

2.1.3 Share of land area used for agriculture

To answer the above question, let's find the share of land used for agriculture as a fraction of the total land area.

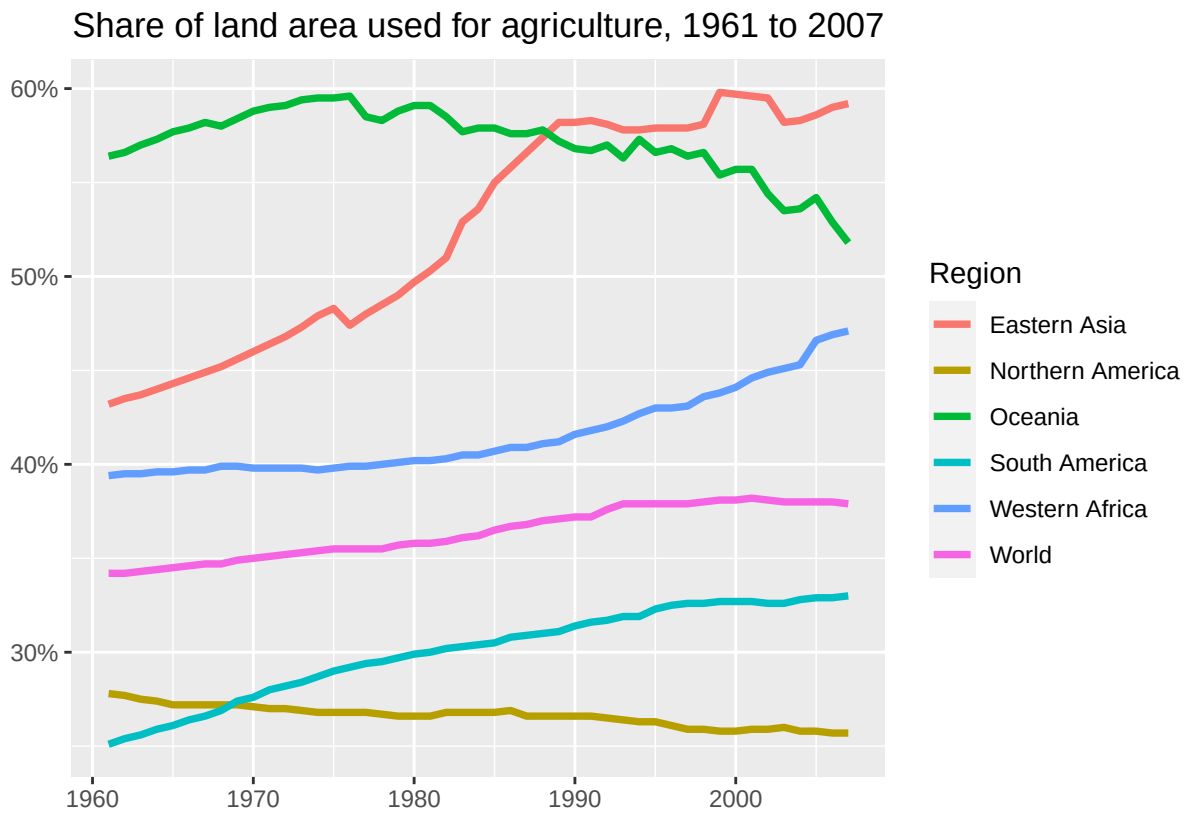
```
agri_df = land_data[land_data$Category=='agricultural_area',c(1,2,3)]
total_df = land_data[land_data$Category=='land_area',c(1,2,3)]
total_df = total_df[total_df$Region %in% agri_df$Region, ]
ratio_df = cbind(agri_df[-3],round(agri_df[3]/total_df[3],3))
```

Let's consider a selected list of regions, encompassing both developed and developing regions.

```
regions = c('Western Africa', 'South America', 'World',
            'Eastern Asia', 'Northern America', 'Oceania')
```

Let us now filter the data and plot it:

```
small_ratio = ratio_df[ratio_df$Region %in% regions,]
ggplot(small_ratio, aes(x = Year, y = Value)) +
  geom_line(aes(color = Region), size = 1.3) +
  scale_y_continuous(labels = scales::percent_format(accuracy = 1)) +
  labs(title = 'Share of land area used for agriculture, 1961 to 2007',
       x='', y='') +
  theme(plot.title = element_text(hjust=0.5))
```



There is large variability in the land area allocated to agriculture across regions. The share of land used for agriculture has been slowly increasing across the developing regions over the past few decades. However, land use for agriculture across North America and Oceania has been declining.

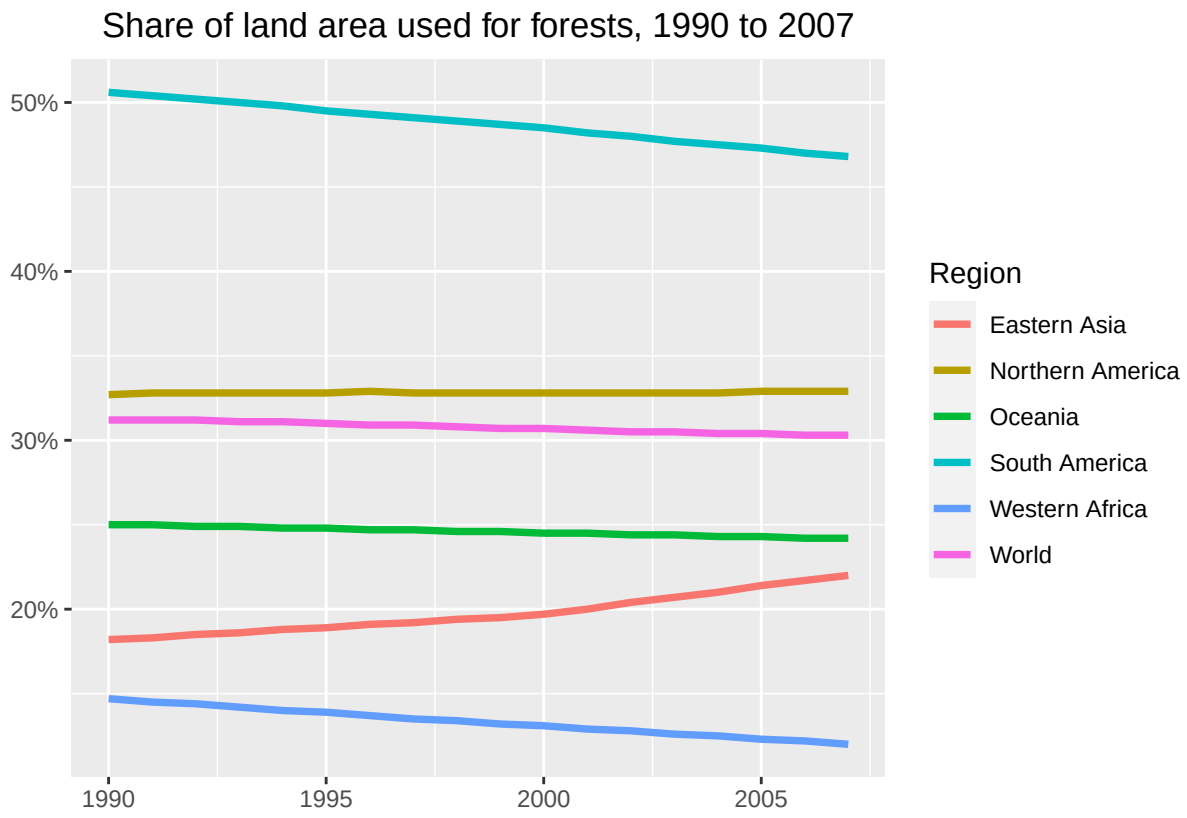
It's important to note that this metric includes both land used for arable (cropland) production and pasture land for livestock grazing; this means that agriculture can consume a large share of land area, even in arid and semi-arid regions where extensive arable farming is not possible.

2.1.4 Share of forest area

Though forests deserve a big tome of analysis, it's worthwhile to look at how forest area has fared in the aforementioned world regions.

```
forest_df = forest_data[forest_data$Category=='forest_area',c(1,2,3)]
ftotal_df = forest_data[forest_data$Category=='land_area',c(1,2,3)]
ftotal_df = ftotal_df[ftotal_df$Year %in% forest_df$Year, ]
fratio_df = cbind(forest_df[-3],round(forest_df[3]/ftotal_df[3],3))
```

```
fsmall_ratio = fratio_df[fratio_df$Region %in% regions,]

ggplot(fsmall_ratio, aes(x = Year, y = Value)) +
  geom_line(aes(color = Region), size = 1.3) +
  scale_y_continuous(labels = scales::percent_format(accuracy = 1)) +
  labs(title = 'Share of land area used for forests, 1990 to 2007',
       x='', y='') +
  theme(plot.title = element_text(hjust=0.5))
```



This plot throws up some differences as compared to our previous analysis. There is a noticeable decline in South America, reminding us of the deforestation in the Amazon rain forest. So is the situation in Western Africa. But, the Eastern Asian region shows a considerable rise in forested area. Coming to the developed regions, the forested area in North America is almost constant but that of Oceania shows a slight decline.

2.2 Country-wise analysis

A regional analysis has shown us the disparity between the share of land area for both agricultural purposes and forests. It's now time to see how specific nations have fared.

```
land_data = read.csv("Databases/Land_data_countries.csv")
```

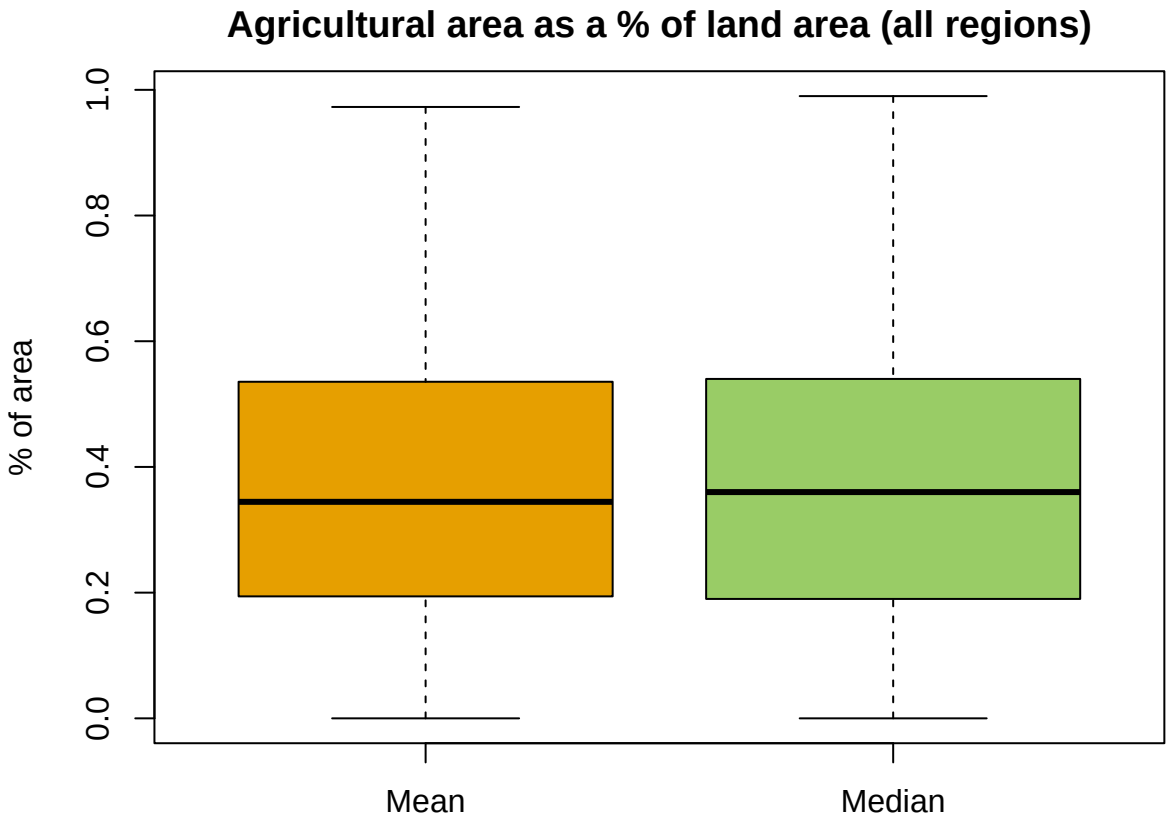
```
# Get agri. area as a fraction of land area
agri_df = land_data[land_data$Category=='agricultural_area',c(1,2,3)]
total_df = land_data[land_data$Category=='land_area',c(1,2,3)]
total_df = total_df[total_df$Region %in% agri_df$Region, ]
ratio_df = cbind(agri_df[-3],round(agri_df[3]/total_df[3],2))
```

Now, as a first analysis, we're going to get the 7 nations that have devoted most of their land area for agricultural purposes. And, as we mentioned before, this metric includes both cropland and pasture land. For our purposes, let's convert our data frame into a wide format.

```
ratio_df2 = pivot_wider(ratio_df, id_cols = Region,
                        names_from = Year,
                        values_from = Value)
ratio_df2 = na.omit(ratio_df2)
```

We're interested in seeing whether there has been any drastic change in the share of land area in the countries. For this, let's find the mean and the median of the land use share and plot it:

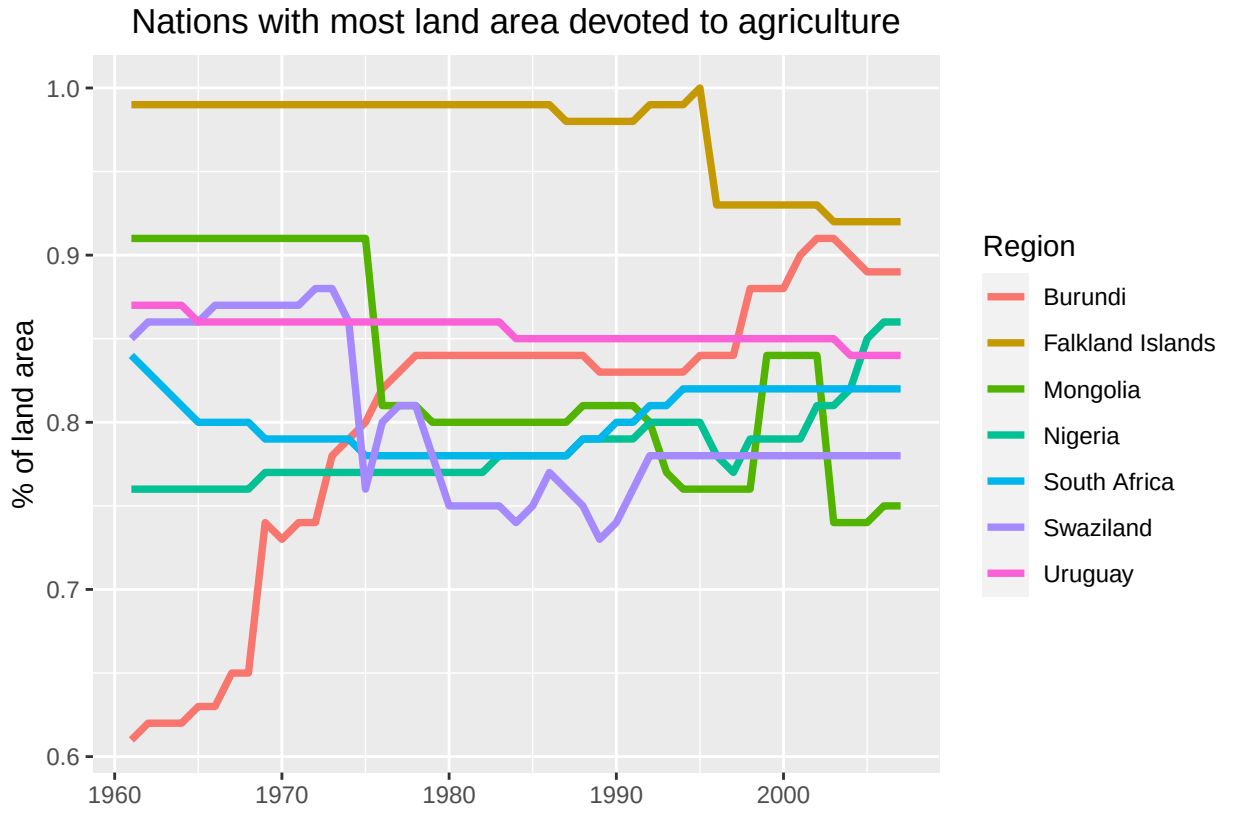
```
ratio_df2$Mean = rowMeans(ratio_df2[,-1], na.rm=TRUE)
ratio_df2$Median = apply(ratio_df2[,-1], 1, median)

par(mar=c(2.5,4.1,2.5,2.1))
boxplot(ratio_df2[, c("Mean", "Median")],ylab='% of area',
        main='Agricultural area as a % of land area (all regions)',
        col=c('#E69F00', '#99CC66'))
```



This plot shows us that we might not need to worry about using the mean statistic for further analysis because of the close proximity between the mean and the median. This possibly bolsters the statement that there hasn't been any drastic change within the countries.

```
ratio_df2 = ratio_df2[order(-ratio_df2$Mean), ]
agri_top7 = ratio_df2[1:7,]
```

```
long_df = pivot_longer(agri_top7[,c(-49, -50)], -1,
                       names_to="Year",
                       values_to="Ratio")
long_df$Year = as.integer(long_df$Year)

ggplot(long_df, aes(x = Year, y = Ratio)) +
  geom_line(aes(color = Region), size = 1.3) +
  ggtitle("Nations with most land area devoted to agriculture") +
  labs(x='', y='% of land area')+
  theme(plot.title = element_text(hjust=0.5))
```

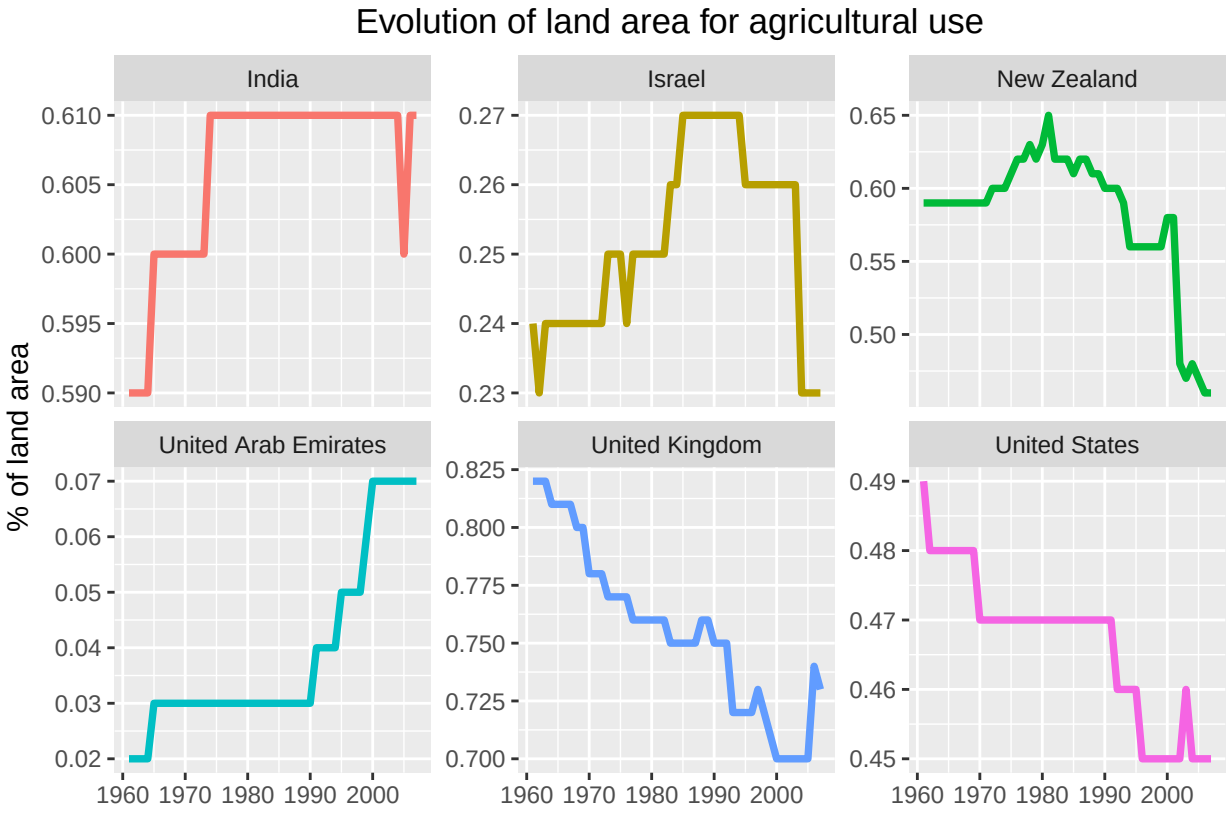


It's interesting to note that the maximum allocation is close to 80 percent in quite a few nations including Uruguay, South Africa, and Burundi.

Let's plot graphs for a handful of countries:

```
nations = c('India', 'Israel', 'United Arab Emirates',
            'New Zealand', 'United Kingdom', 'United States')
small_ratio = ratio_df[ratio_df$Region %in% nations,]

ggplot(small_ratio, aes(x = Year, y = Value)) +
  geom_line(aes(color = Region), size = 1.3) +
  facet_wrap(vars(Region), scales="free_y") +
```

```
  ggtitle("Evolution of land area for agricultural use") +
  labs(x='', y='% of land area') +
  theme(plot.title = element_text(hjust=0.5),
        legend.position = 'none')
```



2.3 Why a decrease in agricultural land use?

It's interesting to take a step back and ponder the reasons for the developed countries to start using lesser land for agriculture. One reason could be that those regions have started seeing a slowing down of population growth.

But Swanson[1] argues that the most important contribution to this trend has been the impact of R&D in the agricultural sector. Agricultural R&D, he argues, has been the primary contributor to the growth in yields on the production frontier over the past half-century. And, surprisingly, he asserts that land use has decreased in many developed countries while overall production has multiplied.

[1] *Endangered Earth.* Global Challenges, **6** (2019), Article 2

3 Analysis of agricultural production

3.1 On performing inter-country comparisons

Historically, inter-country economic aggregates have been undertaken by economists since the time of Adam Smith. Now, each country produces a range of agricultural commodities, valued at prices prevailing in each country and in their own national currency unit. Consequently, researchers have devised strategies to obtain meaningful comparisons. Some methods used previously include:

1. Converting national currency units to US Dollars;
2. Expressing output of different commodities into wheat units.

In this analysis, we'll employ the use of the FAO Gross **Production Index Number (PIN)**. The FAO indices of agricultural production show the relative level of the aggregate volume of agricultural production for each year in comparison with a base period. In our case, the base period is considered to be 1999-2001.

All the indices at the country, regional and world levels are calculated by the Laspeyres formula. Production quantities of each commodity are weighted by the 1999-2001 average international commodity prices and summed for each year. To obtain the index, the aggregate for a given year is divided by the average aggregate for the base period.

Why is this index useful? Although prices differ over time and across space, the important feature of commodity prices to measure output growth in agriculture is their value relative to each other, which tends to be fairly stable over time. Thus, the growth rate in agricultural output reported by an individual country (using annual domestic price data appropriately deflated) is generally close to the growth rate in the FAO output index for that country[2].

3.2 Regional analysis

Now, recall from our previous observations on land used for agricultural purposes, that a yet-to-be-verified hypothesis was made concerning agricultural production. It stated that:

> Developed regions have seen greater agricultural output while land used for agriculture has reduced.

So, let's start with the same regions we analyzed before:

```
pin_regions = read.csv("Databases/Gross_PIN_regions.csv")
```

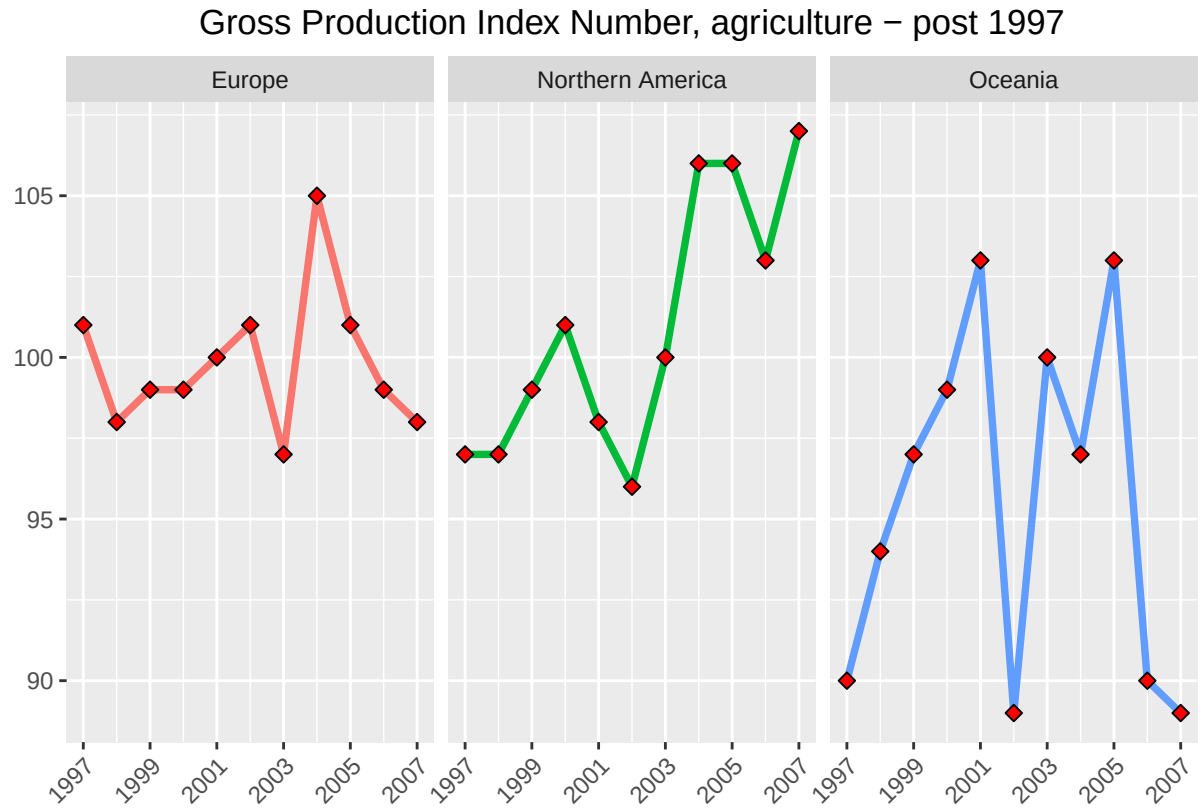
We will initially look at the production indices for agriculture. As per the FAO documentation[3]:

> The commodities covered in the computation of indices of agricultural production are all crops and livestock products originating in each country. Practically all products are covered, with the main exception of fodder crops.

[2]Fuglie, Keith. (2010). Total factor productivity in the global agricultural economy: Evidence from FAO data. The Shifting Patterns of Agricultural Production and Productivity Worldwide. pp. 63-95

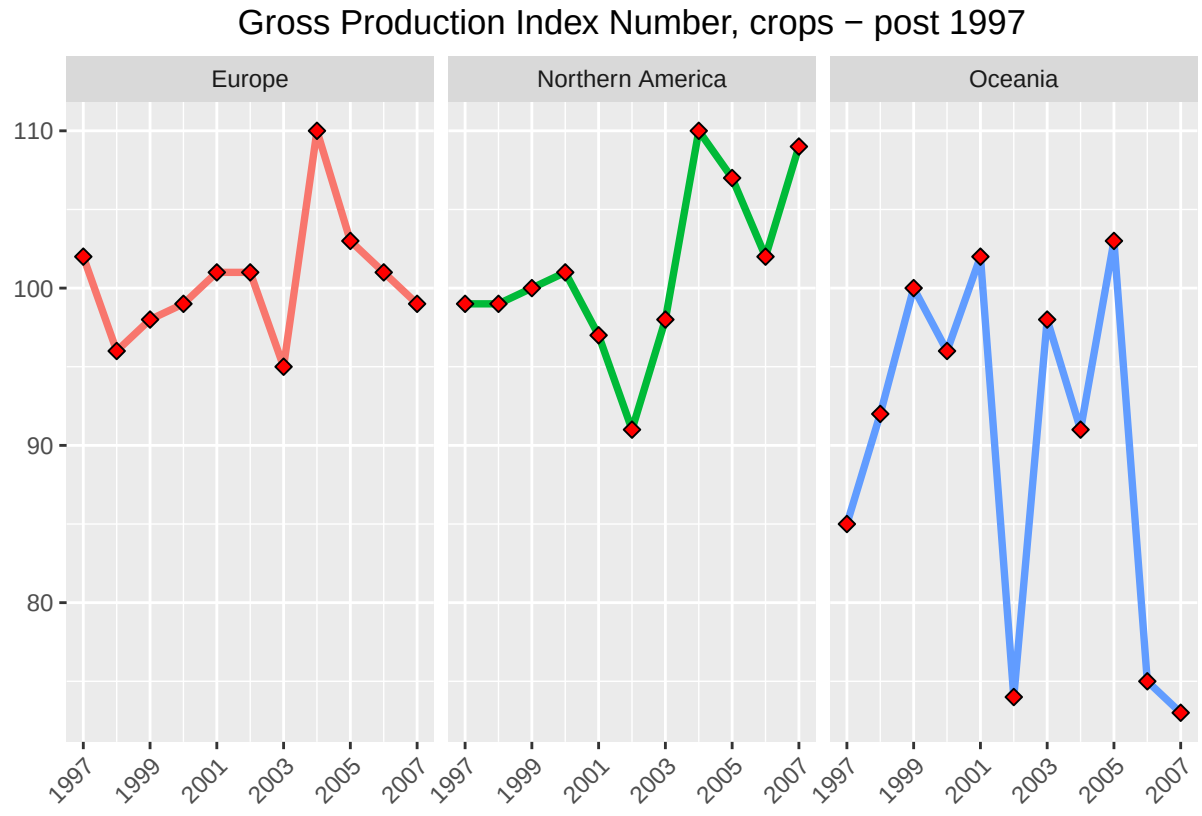
[3]https://fenixservices.fao.org/faostat/static/documents/QI/QI_e.pdf

```
agri_pin_regions = pin_regions[pin_regions$Category == 'agriculture',c(1,2,3)]
developed = agri_pin_regions[agri_pin_regions$Region %in% c(
  'Northern America', 'Oceania', 'Europe'),]

our_plot = ggplot(developed[developed$Year >= 1997,],
                  aes(x = Year, y = Value)) +
  geom_line(aes(color = Region), size = 1.3) +
  facet_wrap(vars(Region))

our_plot = our_plot + geom_point(fill='red', shape=23, size=2) +
  ggtitle("Gross Production Index Number, agriculture - post 1997") +
  labs(x='', y='') +
  theme(plot.title = element_text(hjust=0.5), legend.position = 'none') +
  scale_x_continuous(breaks=c(1997, 1999, 2001, 2003, 2005, 2007)) +
  theme(axis.text.x = element_text(angle = 45, hjust=1))
our_plot
```



This plot provides a mixed picture of the agricultural developments in these developed regions. While Northern America has surely seen a rise as compared to the 1999-2001 base region, the same cannot be convincingly said of Europe and Oceania. Could it be because of the fact that we are looking at both crops and livestock combined (recall the blockquote above)?

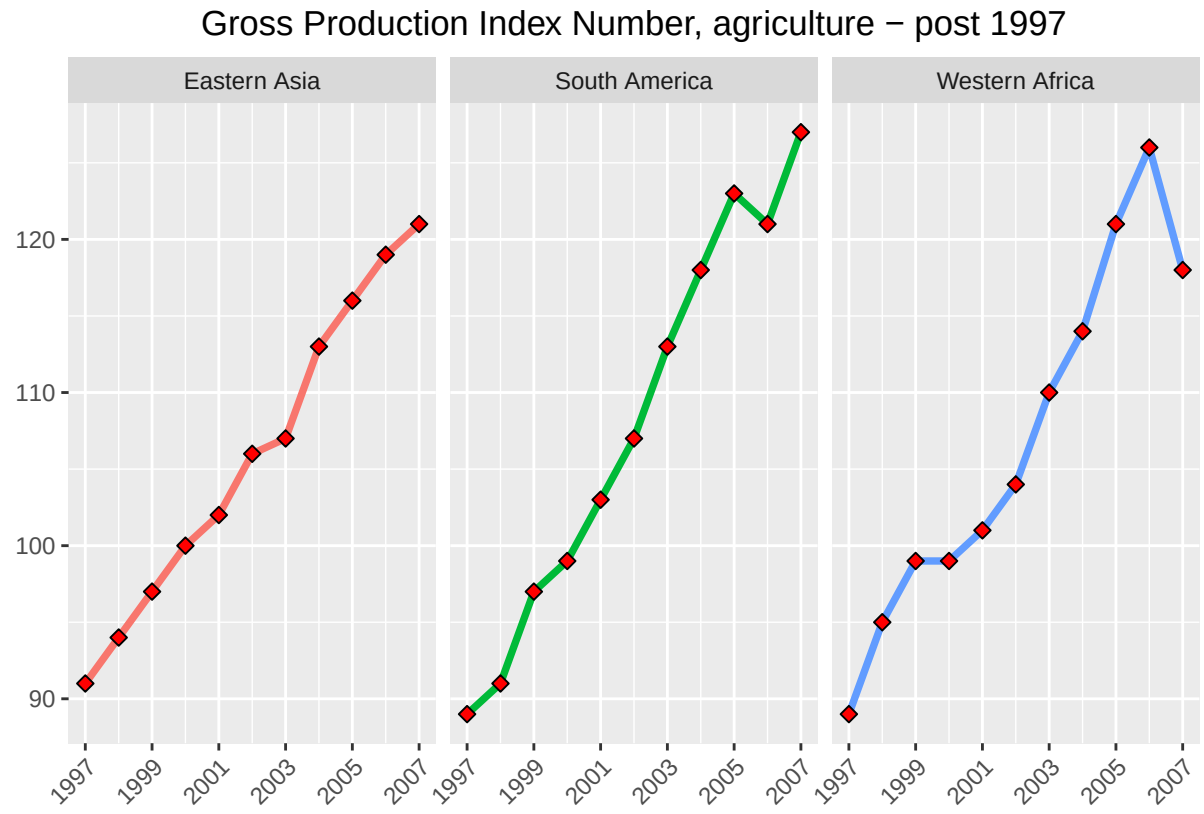
```
crop_pin_regions = pin_regions[pin_regions$Category == 'crops',c(1,2,3)]
c_developed = crop_pin_regions[crop_pin_regions$Region %in% c(
  'Northern America', 'Oceania', 'Europe'),]

ggplot(c_developed[c_developed$Year >= 1997,],
       aes(x = Year, y = Value)) +
  geom_line(aes(color = Region), size = 1.3) +
  facet_wrap(vars(Region)) +
  geom_point(fill='red', shape=23, size=2) +
  ggtitle("Gross Production Index Number, crops - post 1997") +
  labs(x='', y='') +
  theme(plot.title = element_text(hjust=0.5), legend.position = 'none') +
  scale_x_continuous(breaks=c(1997, 1999, 2001, 2003, 2005, 2007)) +
  theme(axis.text.x = element_text(angle = 45, hjust=1))
```



The above graph also doesn't convince us of the truth of our hypothesis. It seems that it would be worthwhile to look into individual nations; though the above analysis is casts doubt on Swanson's claim. But, we can provide a benefit of doubt in the sense that our data is limited to 2007; while Swanson has the advantage of being 12 years ahead, i.e., in 2019.

This consideration notwithstanding, let's check the scenario for developing regions:

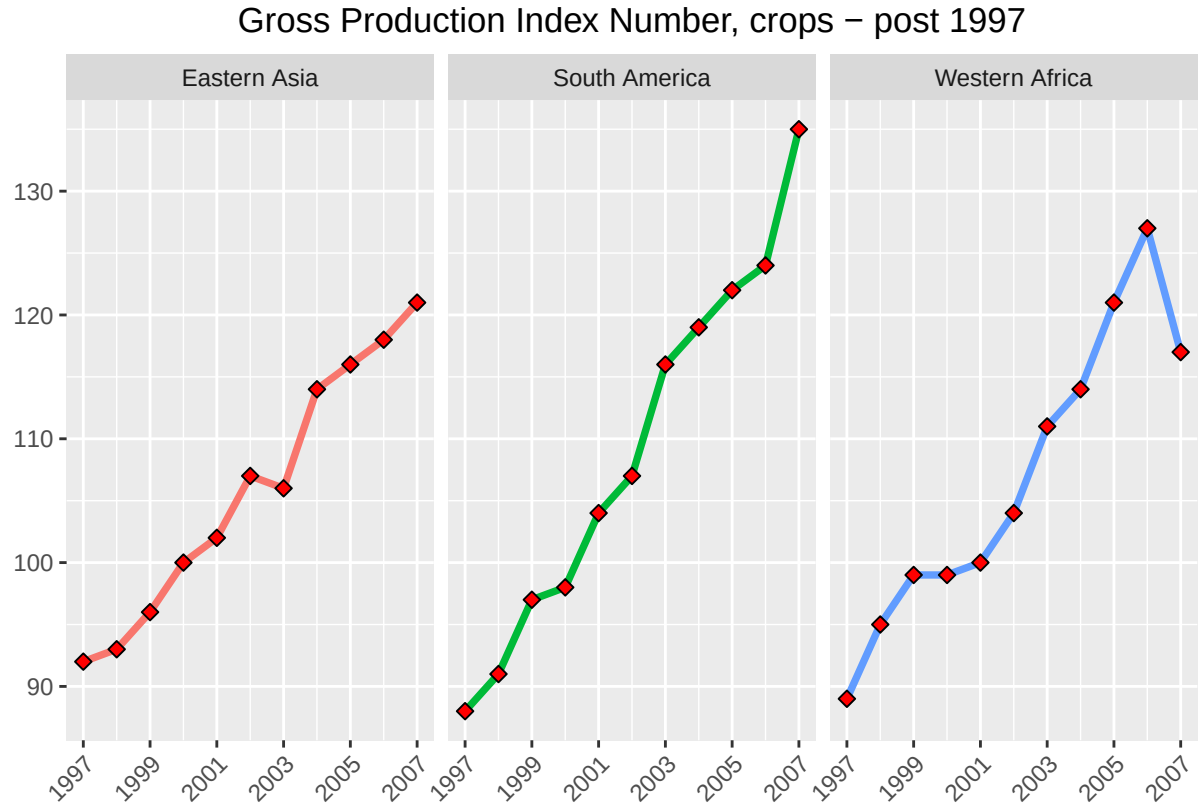
```
developing = agri_pin_regions[agri_pin_regions$Region %in% c(
  'Eastern Asia', 'South America', 'Western Africa'),]

our_plot = ggplot(developing[developing$Year >= 1997,],
                  aes(x = Year, y = Value)) +
  geom_line(aes(color = Region), size = 1.3) +
  facet_wrap(vars(Region))

our_plot = our_plot + geom_point(fill='red', shape=23, size=2) +
  ggtitle("Gross Production Index Number, agriculture - post 1997") +
  labs(x='', y='') +
  theme(plot.title = element_text(hjust=0.5), legend.position = 'none') +
  scale_x_continuous(breaks=c(1997, 1999, 2001, 2003, 2005, 2007)) +
  theme(axis.text.x = element_text(angle = 45, hjust=1))
our_plot
```



This shows us that more land being used for agriculture is translating to greater agricultural output. A similar trend is observed even if we plot the PIN for crops only:

```
c_developing = crop_pin_regions[crop_pin_regions$Region %in% c(
  'Eastern Asia', 'South America', 'Western Africa'),]

ggplot(c_developing[c_developing$Year >= 1997,],
```

```
                  aes(x = Year, y = Value)) +
  geom_line(aes(color = Region), size = 1.3) +
  facet_wrap(vars(Region)) +
  geom_point(fill='red', shape=23, size=2) +
  ggtitle("Gross Production Index Number, crops - post 1997") +
  labs(x='', y='') +
  theme(plot.title = element_text(hjust=0.5), legend.position = 'none') +
  scale_x_continuous(breaks=c(1997, 1999, 2001, 2003, 2005, 2007)) +
  theme(axis.text.x = element_text(angle = 45, hjust=1))
```



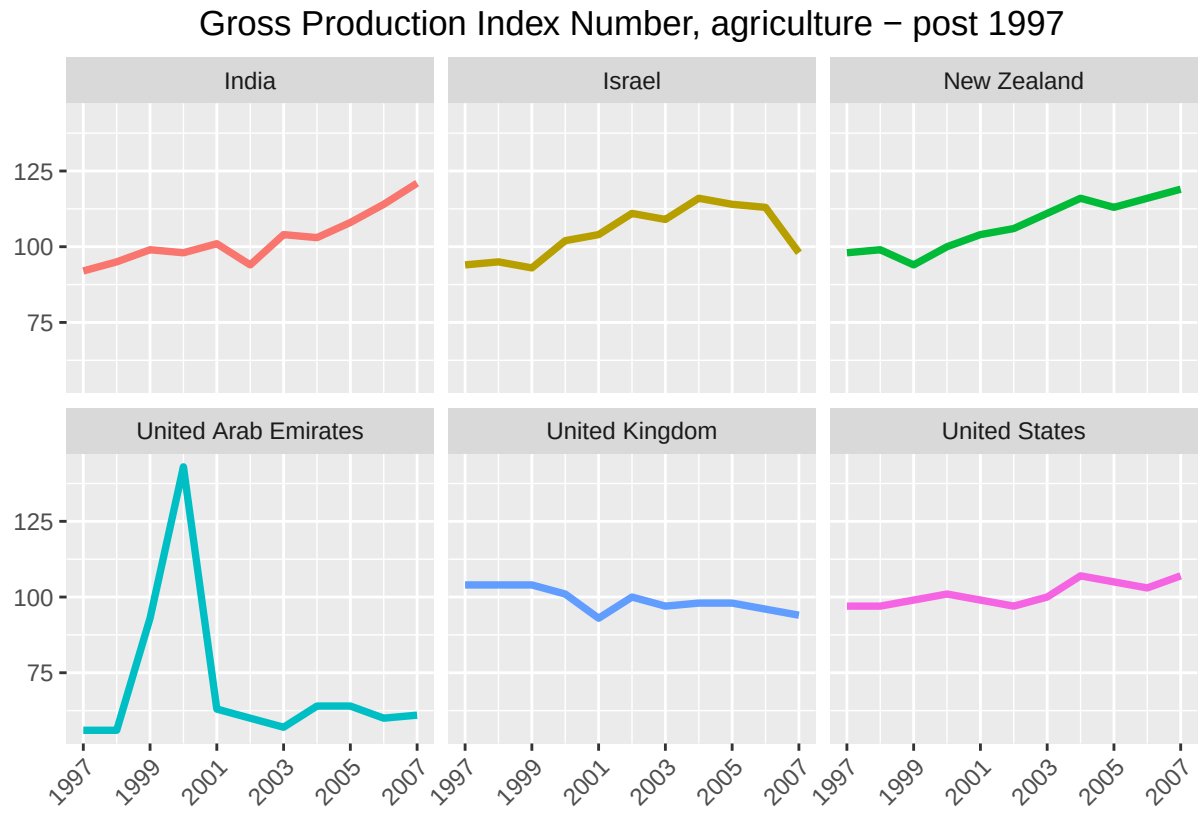
3.3 Country-wise analysis

Let's analyze the same set of countries we used previously for checking out how agricultural production has fared there:

```
pin_nations = read.csv("Databases/Gross_PIN_countries.csv")
```

```
agri_pin_nations = pin_nations[pin_nations$Category == 'agriculture',c(1,2,3)]

nations = c('India', 'Israel', 'United Arab Emirates',
            'New Zealand', 'United Kingdom', 'United States')
```

```
small_nations = agri_pin_nations[agri_pin_nations$Region %in% nations,]

ggplot(small_nations[small_nations$Year >= 1997,],
                  aes(x = Year, y = Value)) +
  geom_line(aes(color = Region), size = 1.3) +
  facet_wrap(vars(Region)) +
  ggtitle("Gross Production Index Number, agriculture - post 1997") +
  labs(x='', y='') +
  theme(plot.title = element_text(hjust=0.5), legend.position = 'none') +
  scale_x_continuous(breaks=c(1997, 1999, 2001, 2003, 2005, 2007)) +
  theme(axis.text.x = element_text(angle = 45, hjust=1))
```



Once again, we observe similar trends at the country level. India, a developing country, for instance, is reporting increased production compared to the 1999-2001 base value. The value for the United Arab Emirates in the year 2000 seems to be too high. And talking about the developed nations, the values for the US and New Zealand are rising, while those of Israel and the UK are falling.

4 Conclusion

In this study, we've aimed to present a picture of the food production and land use in different regions of the world. While we've been successful in explicitly bringing out a divide in the developed

and developing regions on both fronts, the root cause behind this seems to be shrouded under a cloud.

Nevertheless, while it's good news that some nations are able to produce more even when using fewer land area, it is advisable to replicate the techniques that have enabled such growth in the developing nations with a pinch of salt. The main reason is because each nation has been endowed with different climates, regional geographies and the like. However, it's encouraged for scientists to pursue R&D in the agricultural sector to both protect the land and ensure sufficient production.

5 Additional resources

One can access the FAO Analysis app using this link: https://fao-app.shinyapps.io/fao-app/.

Additionally, a YouTube video describing the app has been made here: https://youtu.be/aOW-1WXpFqY.