Agriculture in India

By : Nityam Pareek
Roll Number : 210150021
Github : https://github.com/NityamPareek/Indian-Agriculture-Analysis

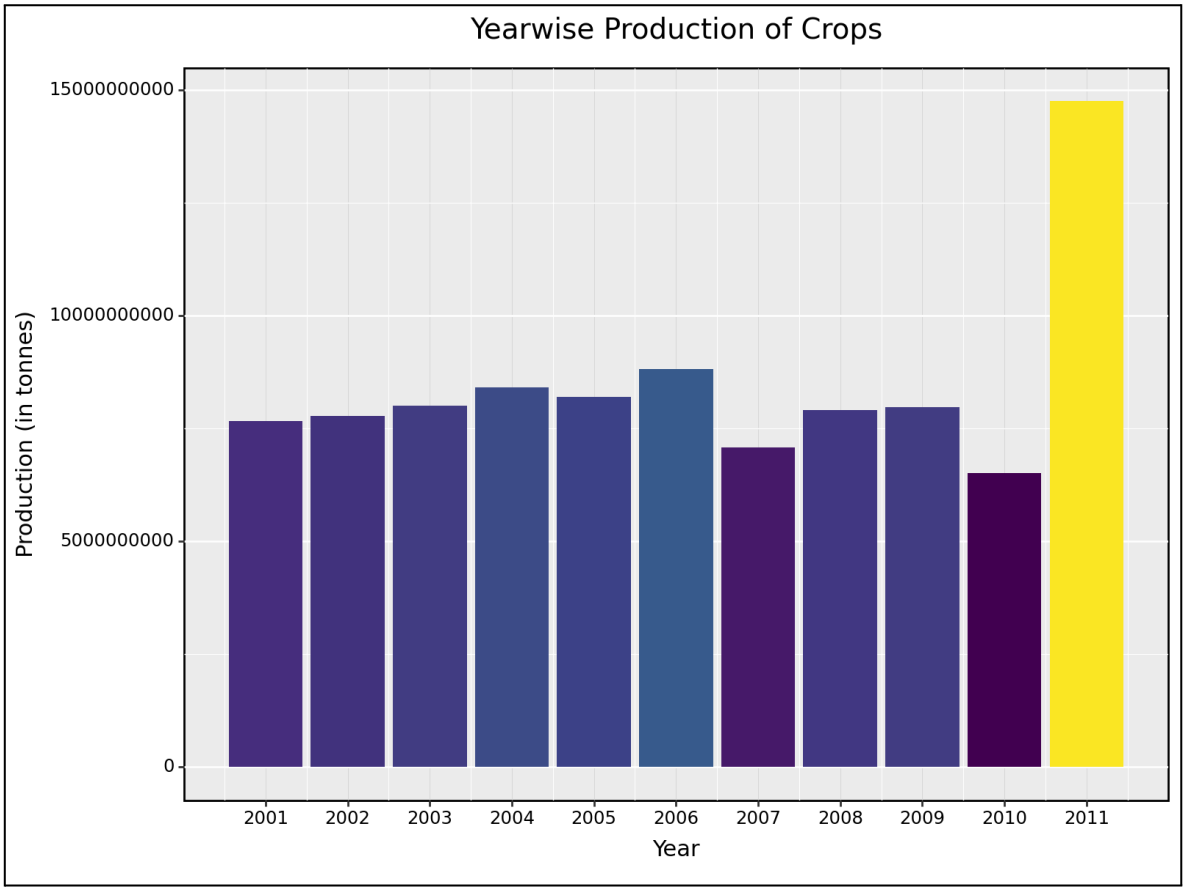
Introduction

Agriculture is the backbone of the Indian economy, providing livelihood to about 58% of the population and contributing about 17-18% to the country's GDP. The country has a diverse range of agro-climatic conditions, which allows for the cultivation of a wide variety of crops. The following report is an attempt at understanding the state of Indian agriculture and looking at it from multiple viewpoints using publicly available data.

Note : Due to inconsistencies across various publicly available datasets, our analysis will be based on the data available between 2001-2011.

National Trends

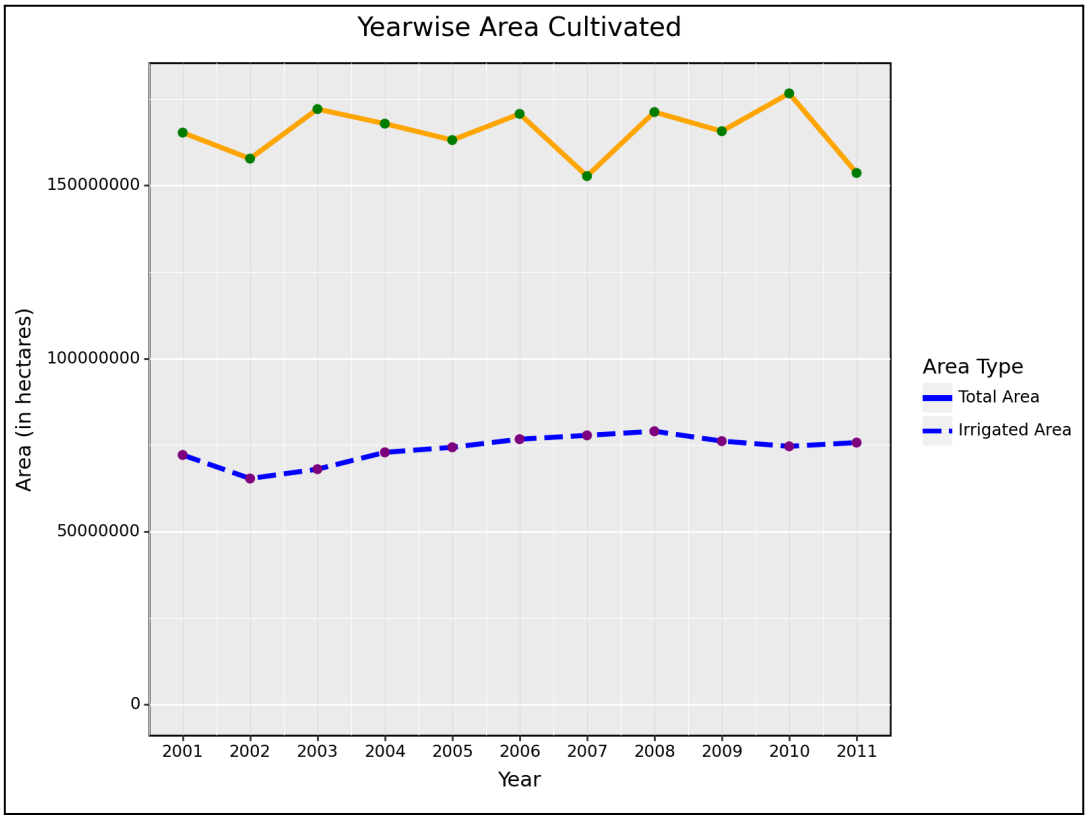
Let's start by seeing the total agricultural production by India over time:



The data shows that while production had a decreasing trend from 2006, it picked up its pace from the year 2011.

A bar graph has been used here to easily interpret the magnitude. The colour of the bars also indicates the magnitude.

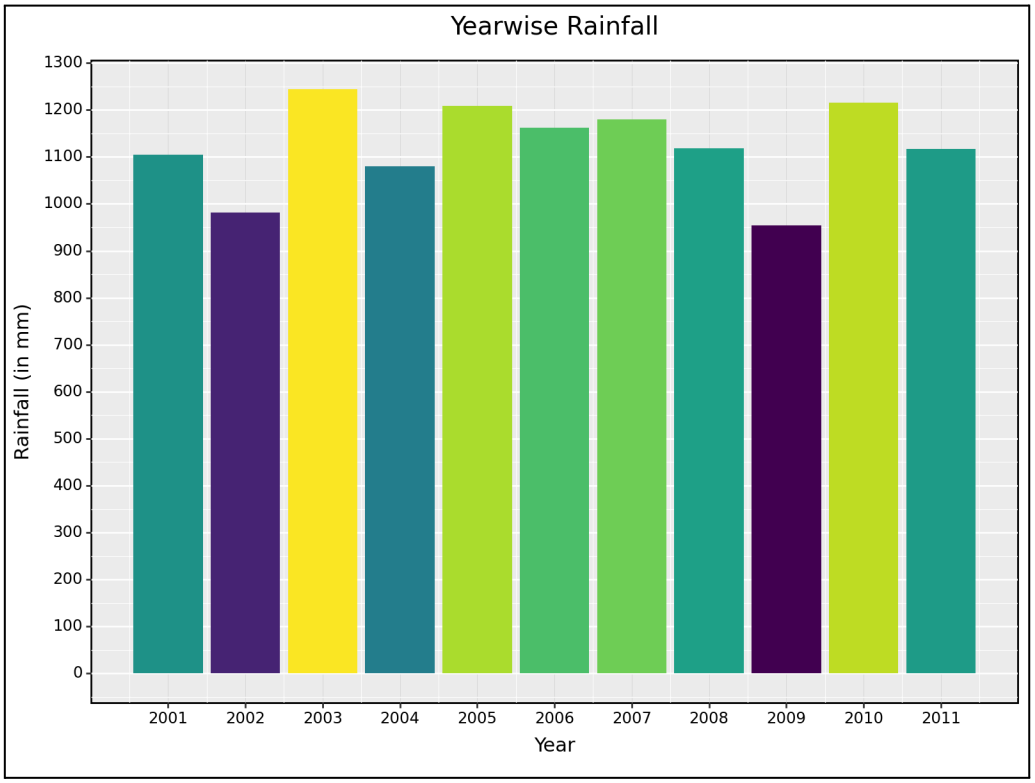
We can also see the yearwise area cultivated:



The area under cultivation each year shows a steadily increasing trend. The area under cultivation dips in 2007 and 2009, which is also consistent with a dip in production.

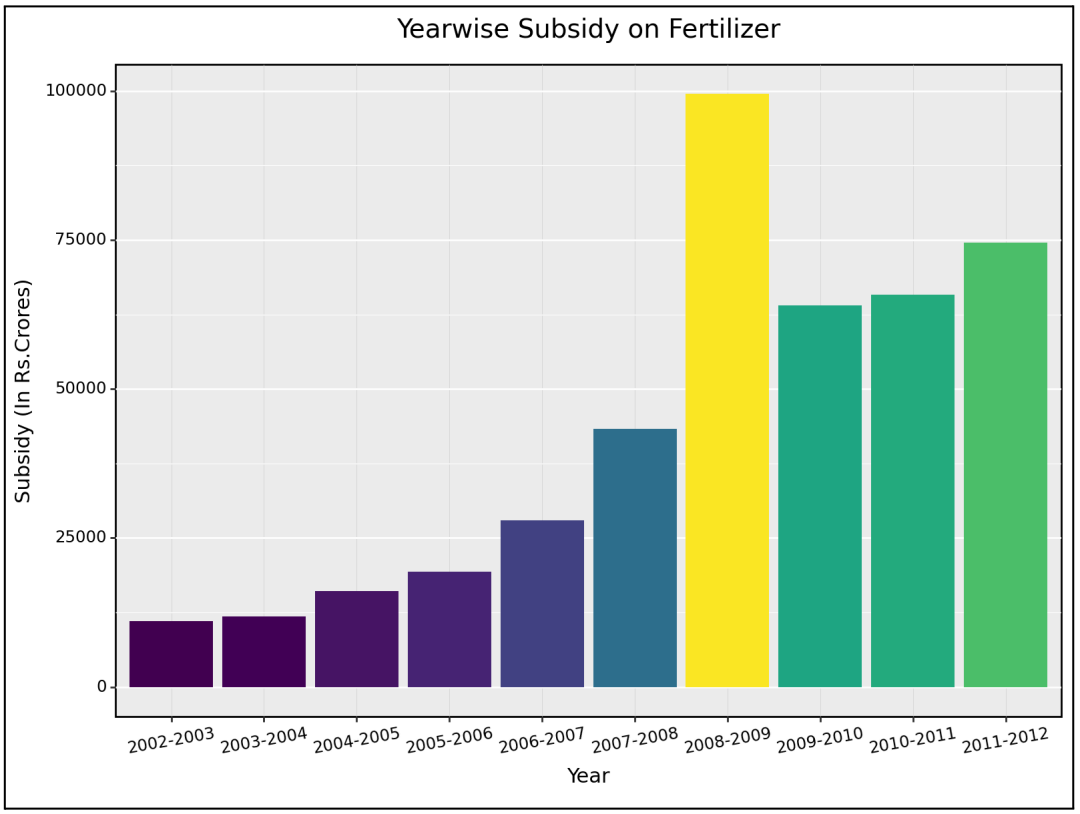
The use of a line plot enables us to compare 2 values in the same plot, and interpret them easily.

Another useful metric would be the rainfall across the years:



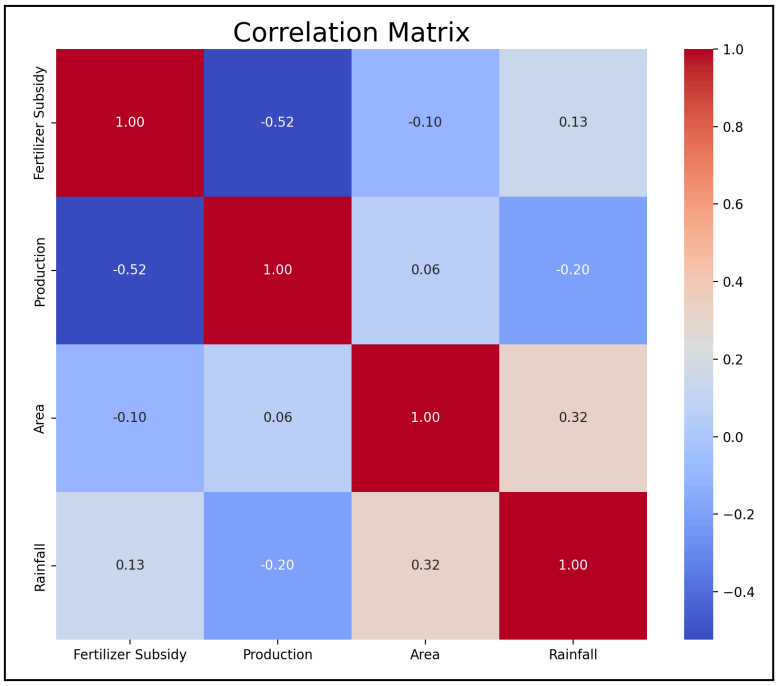
Since rainfall was less in the years 2002 and 2009, the production was also affected. However, we also see a drop in production in the years 2007 and 2010 despite there being adequate overall rainfall. This could indicate the influence of other factors, or an imbalance in the distribution of rainfall which causes variation in production.

Another interesting angle is looking at the subsidies on fertilizers:



The data indicates that due to a fall in production post 2006, the government had to significantly increase the subsidy on fertilizers in an attempt to boost production. However, the production values post subsidy, do not show promising results except for in the year 2011.

Finally, we can plot a heatmap of the correlations between variables:



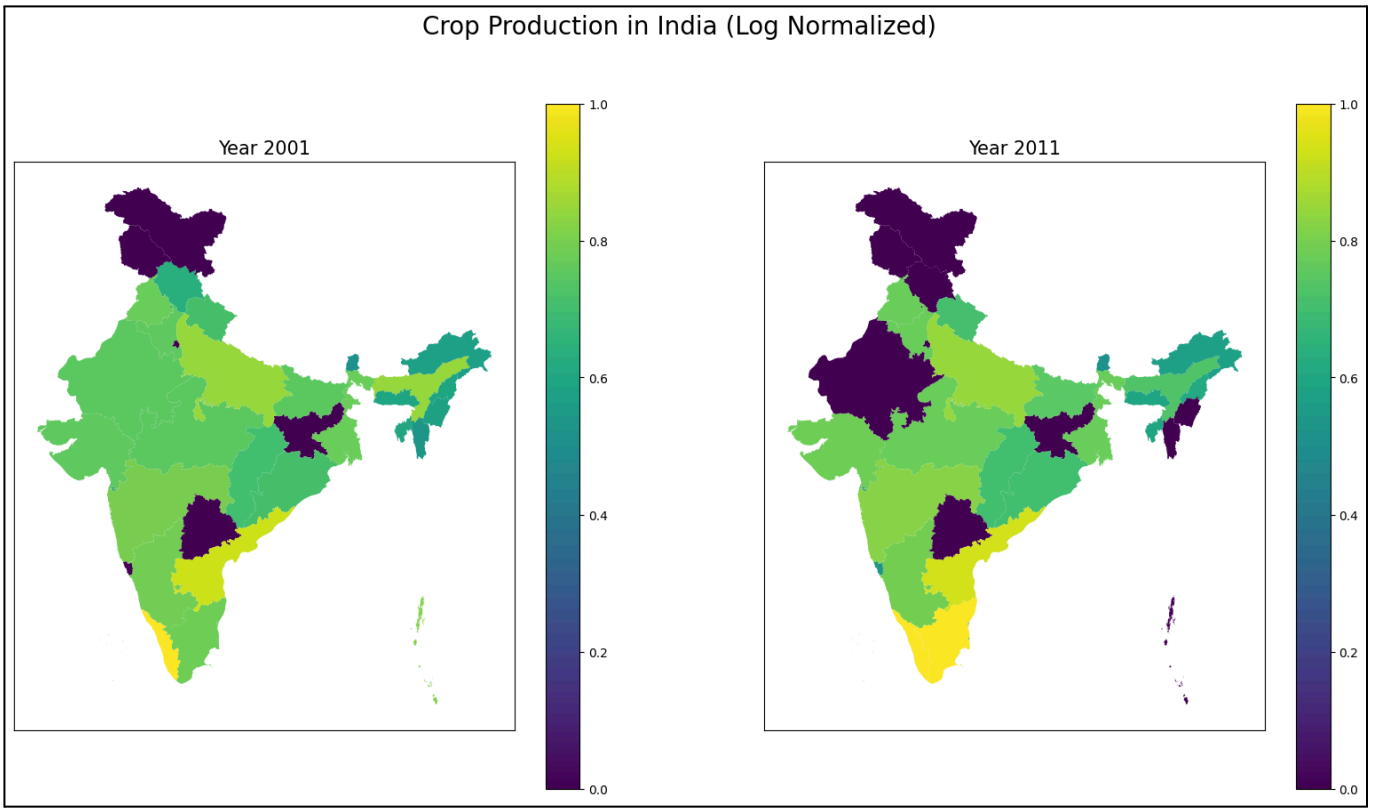
The heatmap shows that Production and Fertilizer Subsidy are negatively correlated, which could be explained by the phenomenon of subsidy being increased with falling production. We also see a positive correlation between Area and Rainfall, signalling an increase in rainfall could lead to an increase in area being cultivated.

An interesting data point to consider is production in the year 2011. The production has risen significantly even though cultivated/irrigated areas and rainfall have not. This could be due to better genetically modified seeds leading to higher yield, or promoting crops which give larger weight in smaller areas.

These national statistics have answered our questions about India's production over time. We also looked at rainfall, area under cultivation and irrigation, and subsidies on fertilizers to try to understand the trend in production. Now, let us zoom in and shift our attention to zonal trends in agriculture.

Zonal Trends

We move forward with our analysis by seeing the distribution of production across states in India. Log normalisation is done to enhance visibility of the plot:

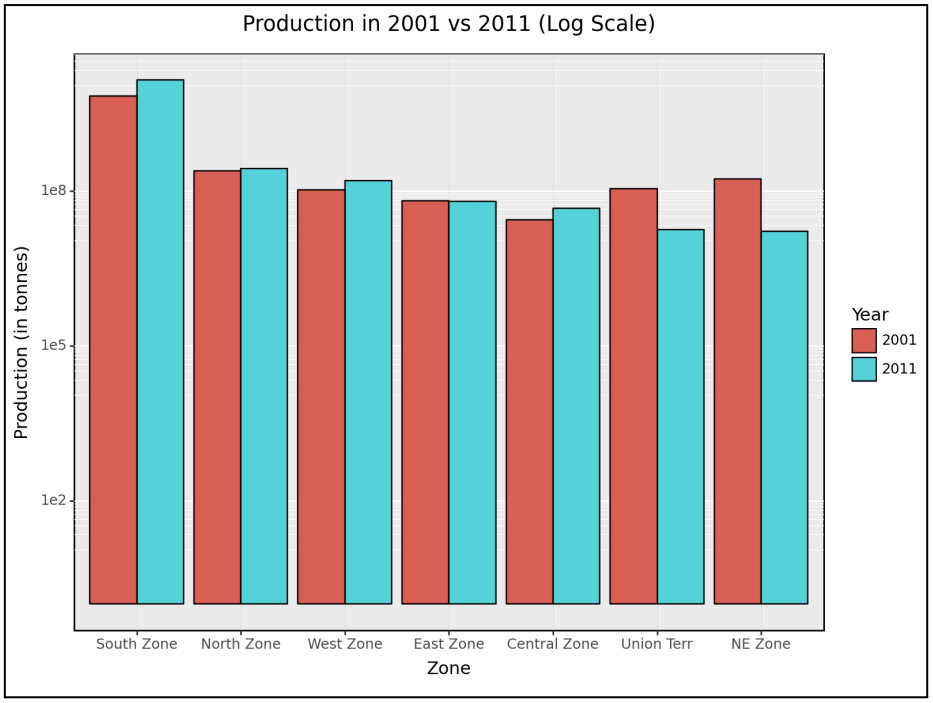


For this, we use choropleth map because it enables us to gain a comprehensive understanding of production across states at a particular point in time.

We can see that from 2001 - 2011, the production in southern states has grown more than other regions, while the production in north-eastern states has reduced.

Note: The states in purple did not release data in that particular year.

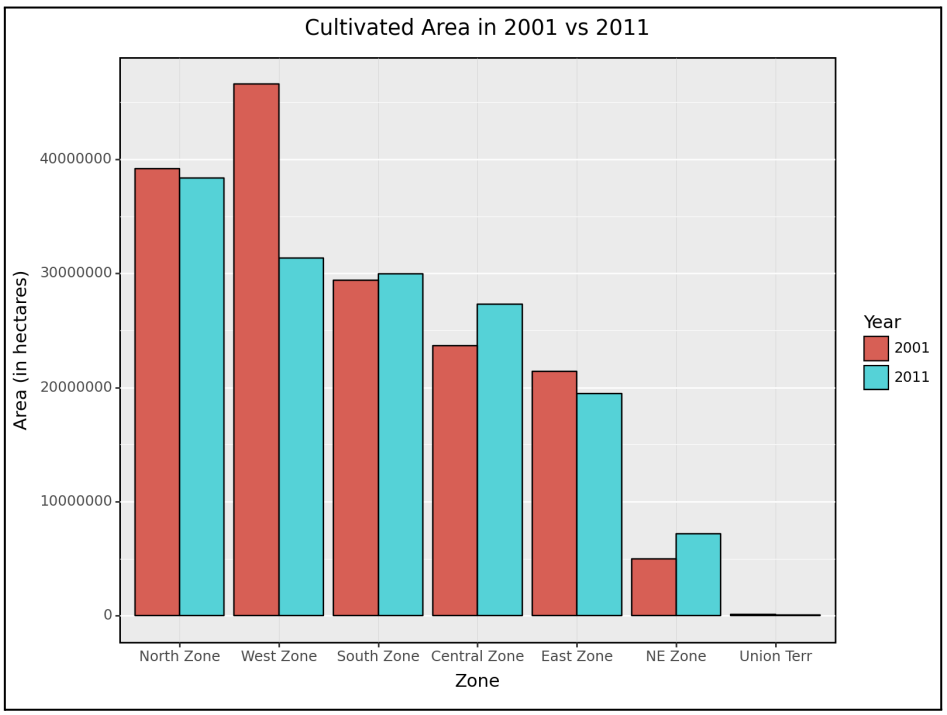
To make it easier to interpret our data, we group states into zones and compare the production in the year 2011, with the year 2001 using the log scale:



We observe that the south zone has always been a major contributor to India's agricultural output, having the highest contribution in both years and also growing the most. Production has also risen in the North, West, and Central zones and fallen in the East, North-east, and union territories. This also explains the contribution of southern states in the sudden rise in production in 2011, as observed earlier.

The use of a grouped bar graph allows for easy comparison and using the log scale enables us to better visualise relative values because skewness in the original data makes the other bars too small to draw meaningful comparison.

This comparison also begs the question about change in cultivated area across regions:

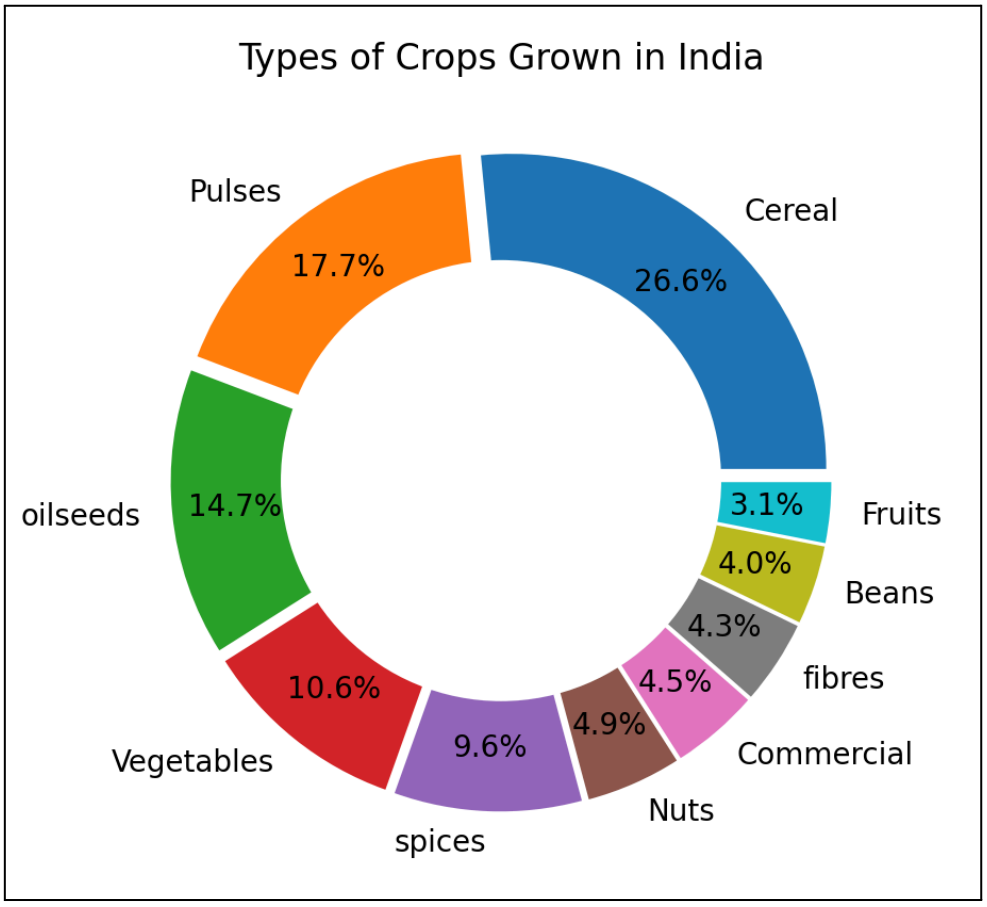


Even though cultivated area has not changed by much in any zone and shows a decreasing trend from 2001 to 2011, we do not see a dip in production. This could indicate better methods of agriculture leading to higher yields, and an increase in production of heavy crops like coconut, which could explain the increase in production in southern states.

Now instead of speculating, let us zoom in further and look at Indian agriculture from a microscopic view by looking at the distribution of crops, and see whether that explains the national and zonal trends we have seen so far.

Crop-Wise Trends

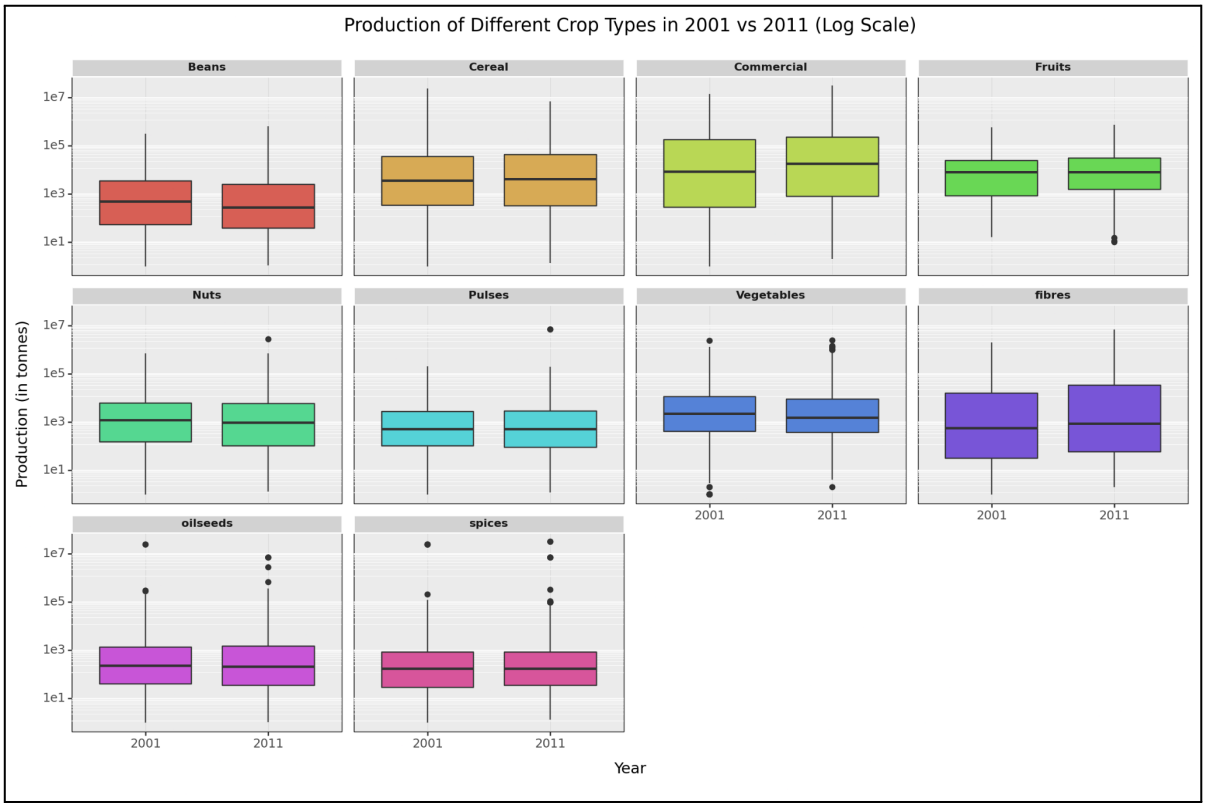
Let's begin this section by looking at the types of crops grown in India:



We observe that most commonly grown types of crops are cereals, pulses, and oilseeds.

We use a donut plot for the same to visualise part-to-whole relationships in a good way.

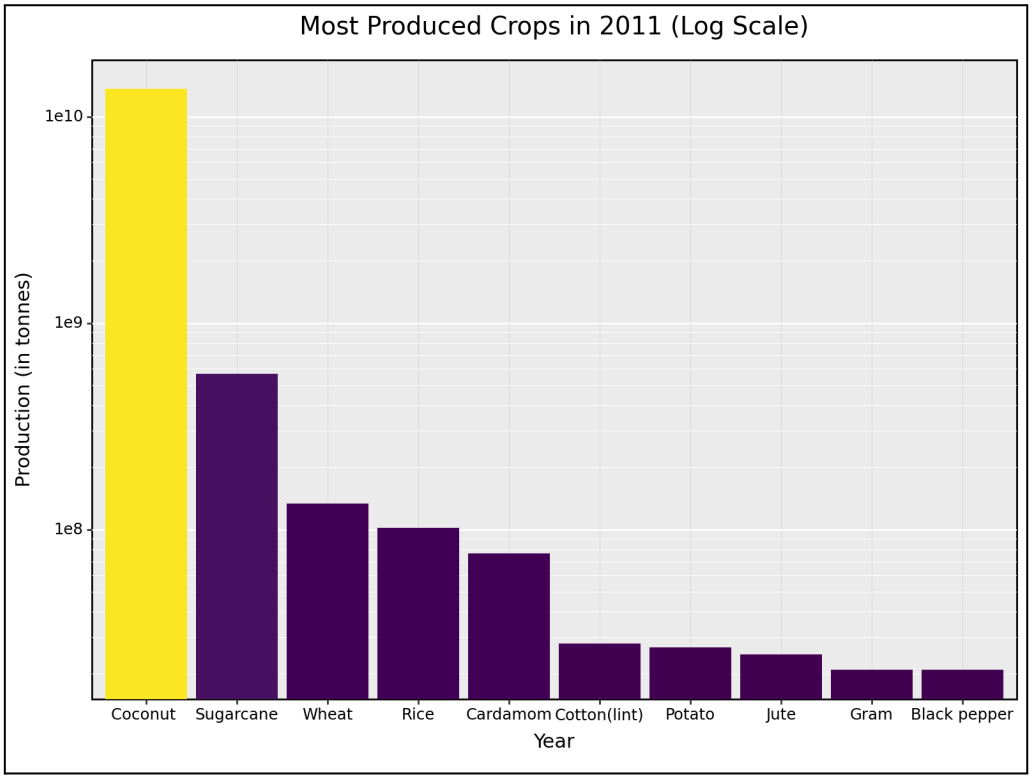
However, this does not reflect their contribution to production. To see that, we use the following:



While all other types of crops show growth in production, we see a dip in production of beans.

Using the log scale enables us to better visualise relative values and the statistical measures of the overall data because the outliers have a huge effect on data and cannot be ignored. The boxplot has fixed scales which greatly simplify observing the distribution and contribution of various crops.

Now, let us see the individual crops which contribute most to production in 2011:



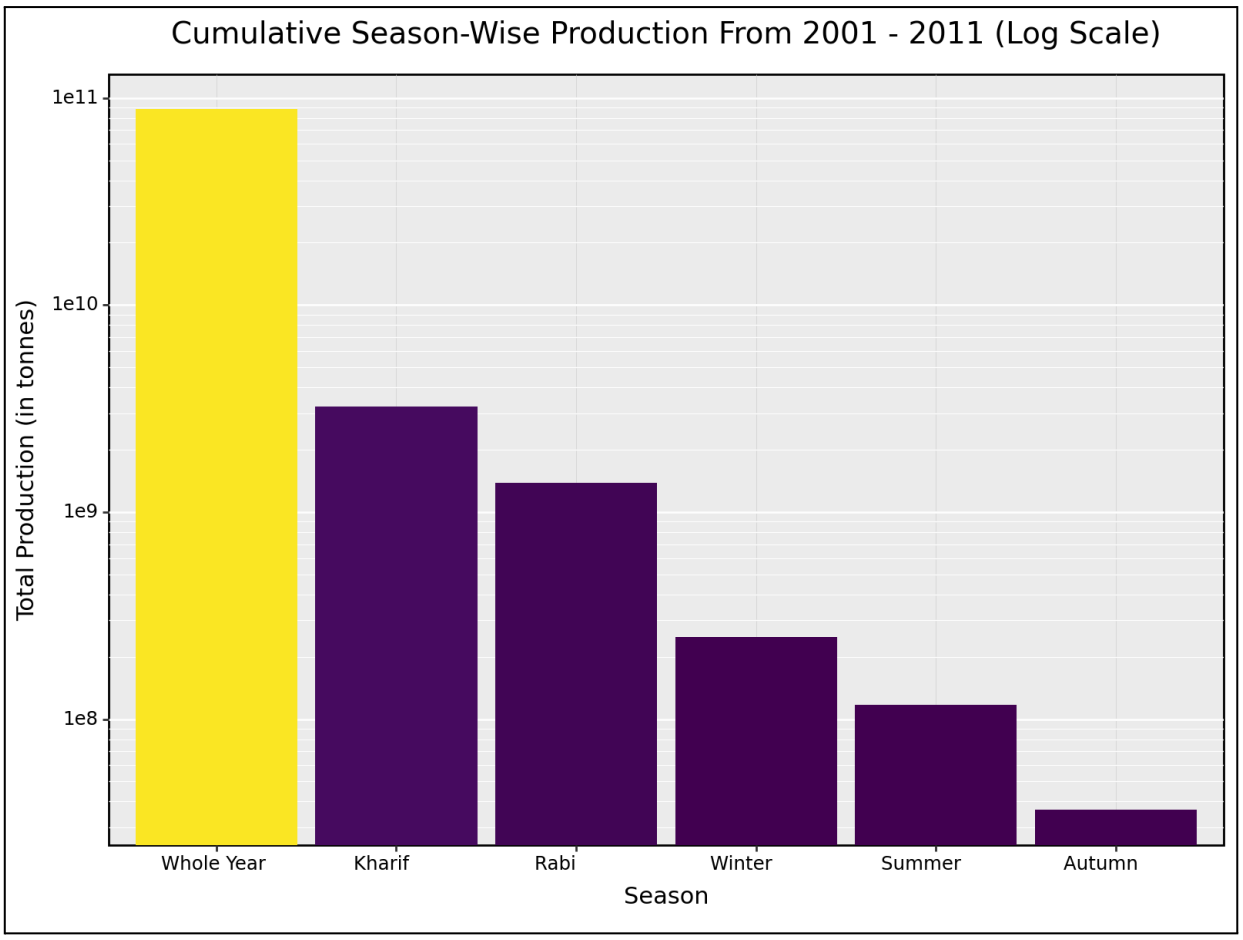
We can clearly see that coconut dominates other crops in terms of production. With coconut, cardamom, and sugarcane being among crops commonly grown in south India, this plot explains the contributions of the south zone in total production.

In the last part of our analysis, we will explore the factors that affect the growth of these crops in order to understand the trends in their production.

Seasonal Trends

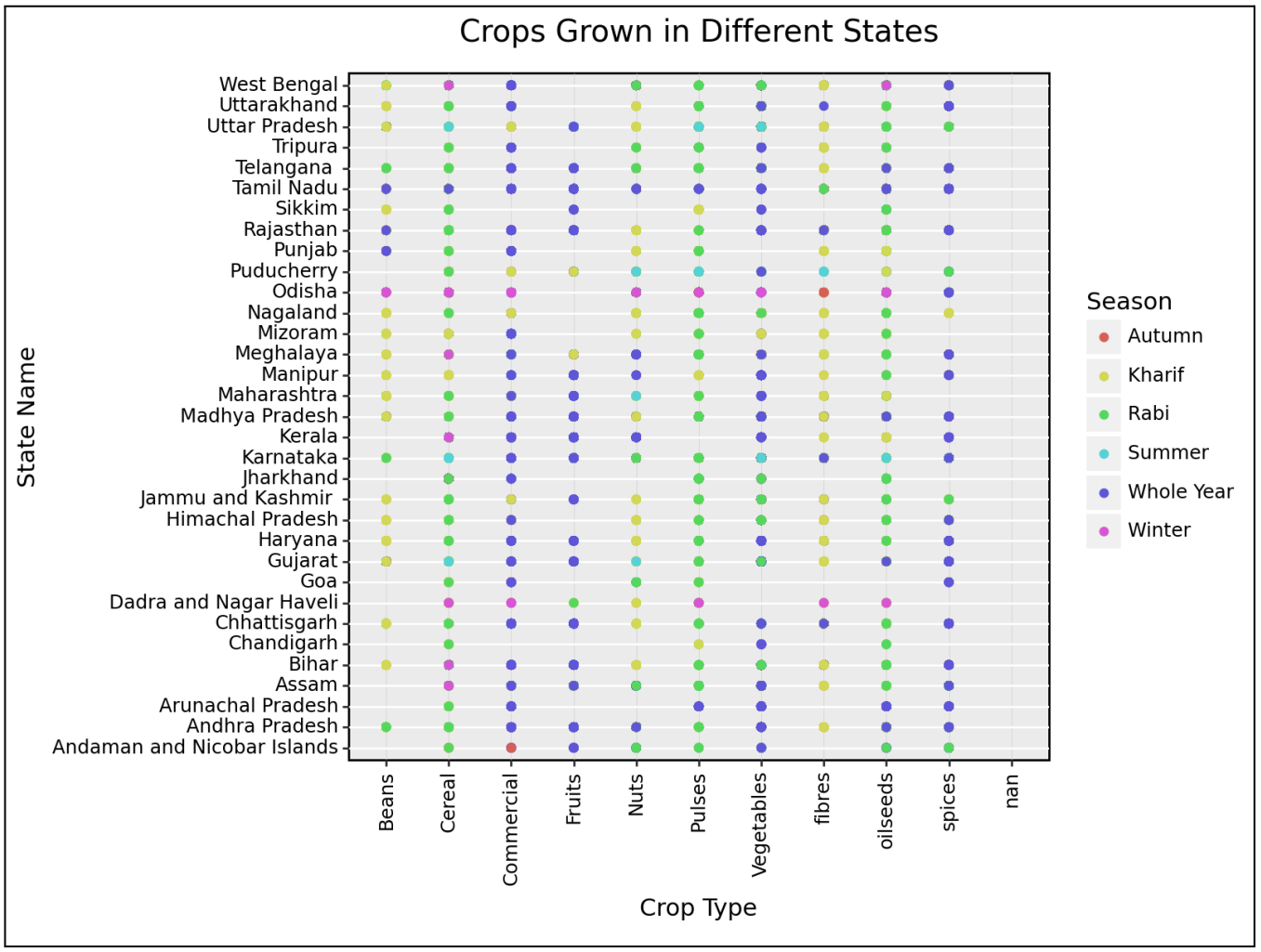
The crops in India can be divided into 6 classes based on the seasons in which they are grown, namely - whole year, kharif, rabi, winter, summer and autumn.

Let's start by seeing which seasons have given maximum production from 2001-2011. We once again use the log scale for better visualisation:



We see that most production is given by crops grown year round (as expected), followed by Kharif, Rabi, Winter, Summer, and lastly autumn.

Lastly, we now want to collectively see the influence of crops and their respective seasons across states to get an even better understanding of the disparity between production across states:



Top producing state Kerala shows an abundance of whole year seasonal crops. Top producing state Uttar Pradesh shows abundance of Kharif, Rabi and Summer crops.