

Crop Production in India

Aman, MDS202305

2023-09-30

1 Introduction

India is a global agricultural powerhouse. It is the world's largest producer of milk, pulses, and spices, as well as having the largest area under wheat, rice, and cotton cultivation. India also ranks as the second-largest producer of rice, wheat, cotton, sugarcane, fruits, vegetables, and tea. The country cultivates approximately 195 million hectares of land, with roughly 63 percent being rainfed (approximately 125 million hectares) and 37 percent being irrigated (around 70 million hectares). Additionally, forests cover about 65 million hectares of India's land.

In the available dataset, we have access to a substantial amount of information on crop production in India, spanning multiple years (1997-2015). This data is also distributed across the districts of various states and union territories in India. In this data analytics project, we conduct exploratory Data Analysis on crop production in India using R.

2 Data Description

The '*crop production file*' contains data related to crop production in India. It includes information on various crops, their production quantities, the season of production, and the land area covered. This data is likely sourced from agricultural records or government reports.

2.1 Dataset Variables:

Below are the fields present in our crop production dataset:

- *State_Name* : Contains the name of the Indian state where the crop was produced (*look at notes 1 and 2 in 2.3*).
- *District_Name* : This section contains the names of districts within the state.
- *Crop_Year* : It includes the year of crop production, spanning over many years, allowing us to analyze trends and changes in crop production over time.
- *Season* : The season in which the crop was planted (e.g., Kharif, Rabi).
- *Crop* : The name of the crop.
- *Area* : The area of land used for crop cultivation.
- *Production* : For each crop type and year, the dataset has data on the quantity of that crop produced. The measurement unit is not mentioned but often these quantities are measured in units such as metric tons.

2.2 Data Sample:

Below is a sample of how the dataset in the CSV file looks like:

State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254	2000
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2	1
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102	321

2.3 Important Note

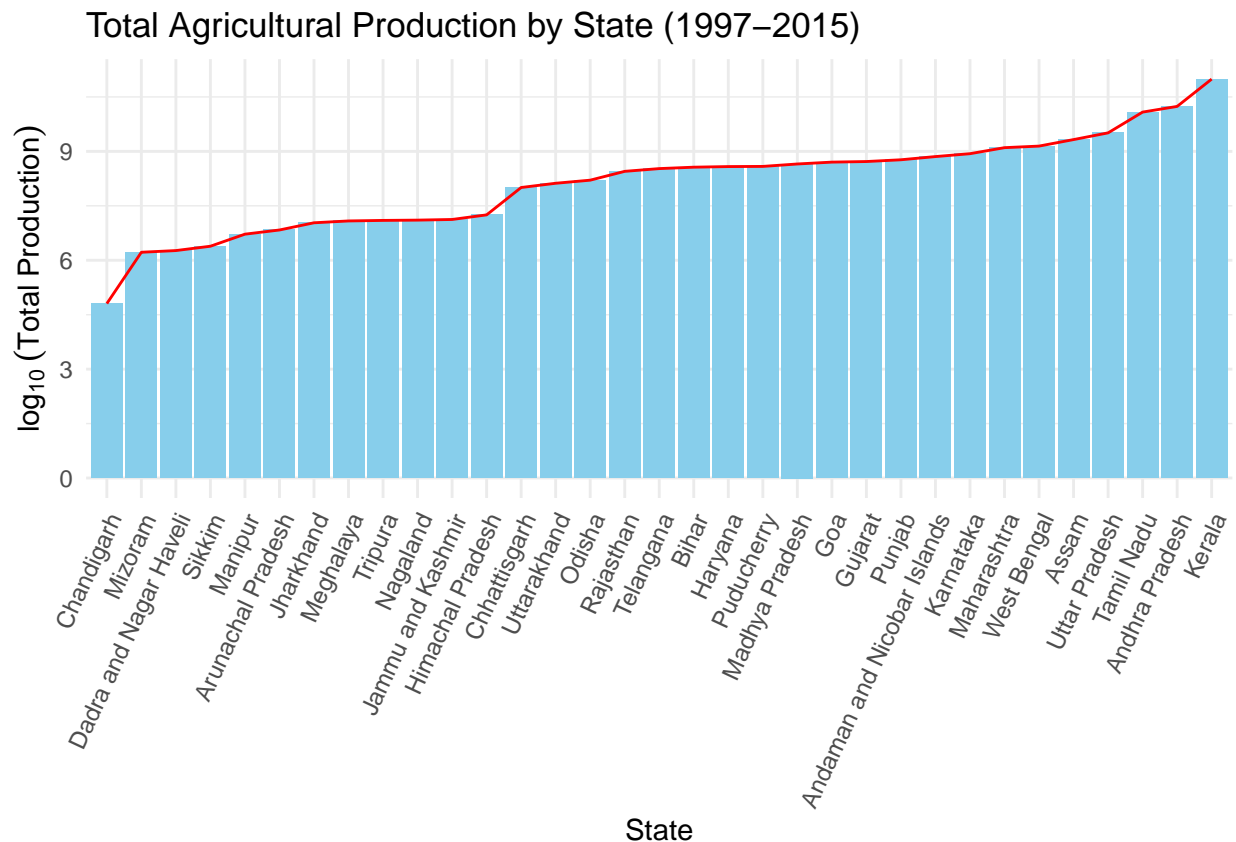
There are few things to note in the available data:

1. Ladakh was established as a union territory of India on 31 October 2019, therefore it is included as a district of Jammu and Kashmir in our data.
2. The Union territories Daman & Diu (Daman), Lakshadweep (Karavatti) and The Government of NCT of Delhi (Delhi) are not included in our data (1997-2015).
3. After data analysis, it was found that the given dataset has 3730 empty entries (1.5% of total sample size) in the production column.
4. There are 124 unique crops in the dataset.
5. Since the units of Area and Production is not specified, we will assume them to be in Hectare and tons respectively.
6. In the EDA part of the data analysis, many plots use a logarithmic scale (base 10) for the Production column. This is done to make the plots comparable and easier to read; otherwise, the plots wouldn't effectively represent the presence of other values and would make drawing conclusions more challenging. Additionally, using \log_{10} as a transformation is beneficial because it is an increasing function, which reduces the graph size while making comparisons more accessible

3 Exploratory Data Analysis

3.1 Total Crop Production by State

The barplot below visualizes the total agricultural production across various states in India. Each bar represents a different state, and the height of the bar corresponds to the \log_{10} of total agricultural production (refer to note 6 in 2.3) in that state. The states are sorted in descending order based on their total production.



As observed from the above plot :

- The top crop-producing state is *Kerala*.
- The lowest crop-producing state is *Chandigarh*. This can be attributed to the fact that Chandigarh has a very small land area for cultivation compared to other states and also the total land area under cultivation in Chandigarh has significantly decreased over the years (average cultivation area over 1997-2015 is 961.6923 Hectares).

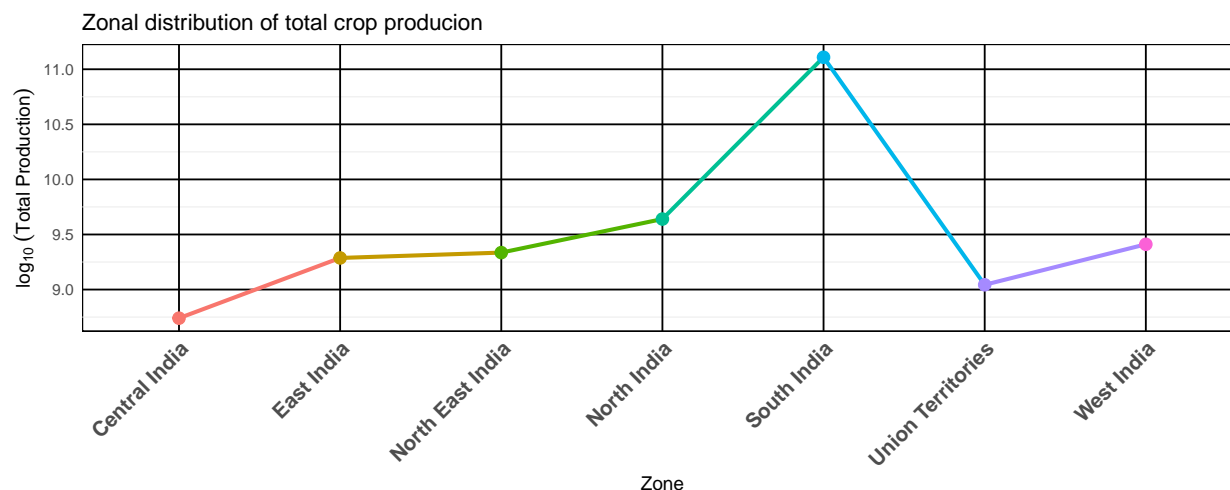
3.2 Zonal Distribution of crops:

To look at the zonal distribution of crop production, firstly the State_Name attribute was divided into the following seven zones.

Zones	States Included
North India	Jammu and Kashmir, Punjab, Himachal Pradesh, Haryana, Uttarakhand, Uttar Pradesh, Chandigarh
East India	Bihar, Odisha, Jharkhand, West Bengal
South India	Andhra Pradesh, Karnataka, Kerala, Tamil Nadu, Telangana
West India	Rajasthan, Gujarat, Goa, Maharashtra
Central India	Madhya Pradesh, Chhattisgarh
North East India	Assam, Sikkim, Nagaland, Meghalaya, Manipur, Mizoram, Tripura, Arunachal Pradesh
Union Territories	Andaman and Nicobar Islands, Dadra and Nagar Haveli, Puducherry

Table 1: Dividing Indian States into Zones.

Below is a graphical representation that illustrates the agricultural output in various geographical regions or zones within the country. It provides valuable insights into the distribution of crop production across above mentioned zones.

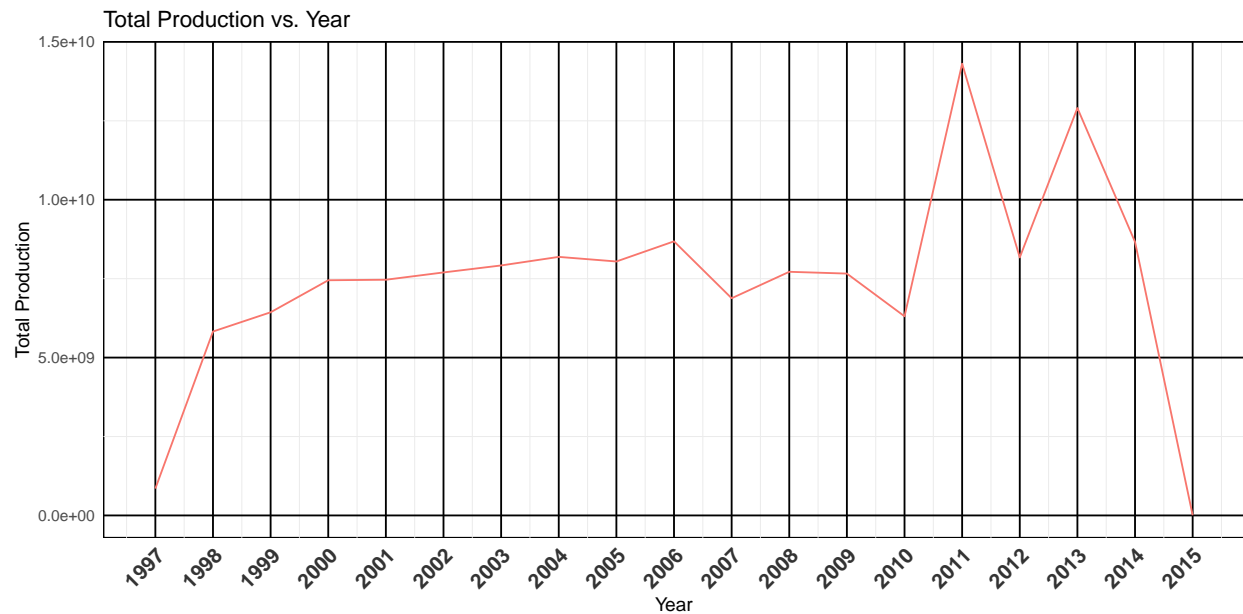


Observations :

- In the South Indian zone, the total crop production remains nearly a hundred times higher than that of any other zone. The tropical climate in South India is conducive to the cultivation of a wide variety of crops.
- Next, observe that Central India has the lowest total production, which can be attributed to the fact that this zone has the fewest number of states.

3.3 Plotting total production over different years

The given dataset contains crop data spread over 19 years (1997-2015). Below is a line plot showing total crop production over different years.

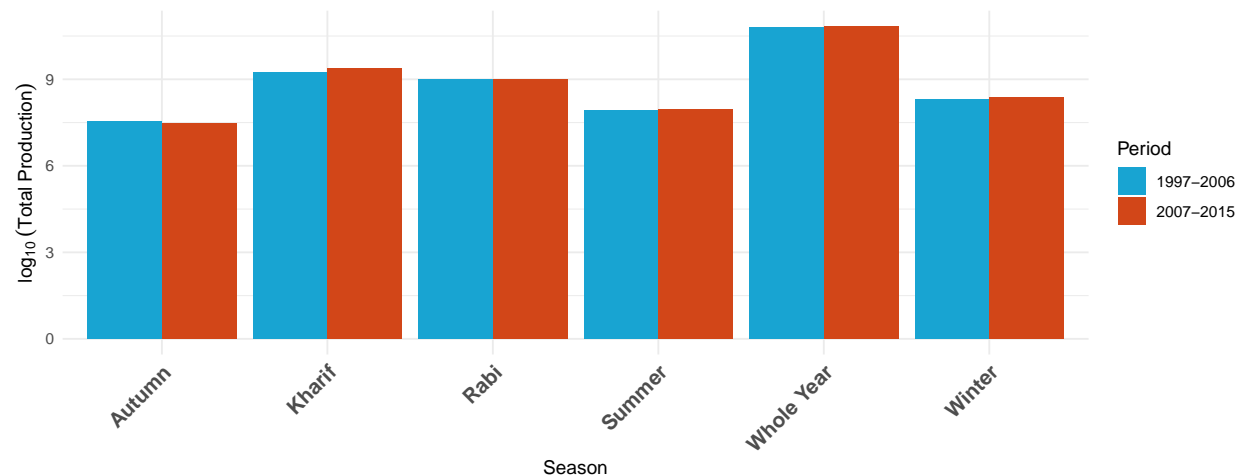


Following are the observations from the above plot :

- There is a significant increase in total crop production just before 2000, but there aren't any big increases or decreases until 2010.
- The years between 2010-14 show great fluctuations, with 2011 showing the highest production over all years.

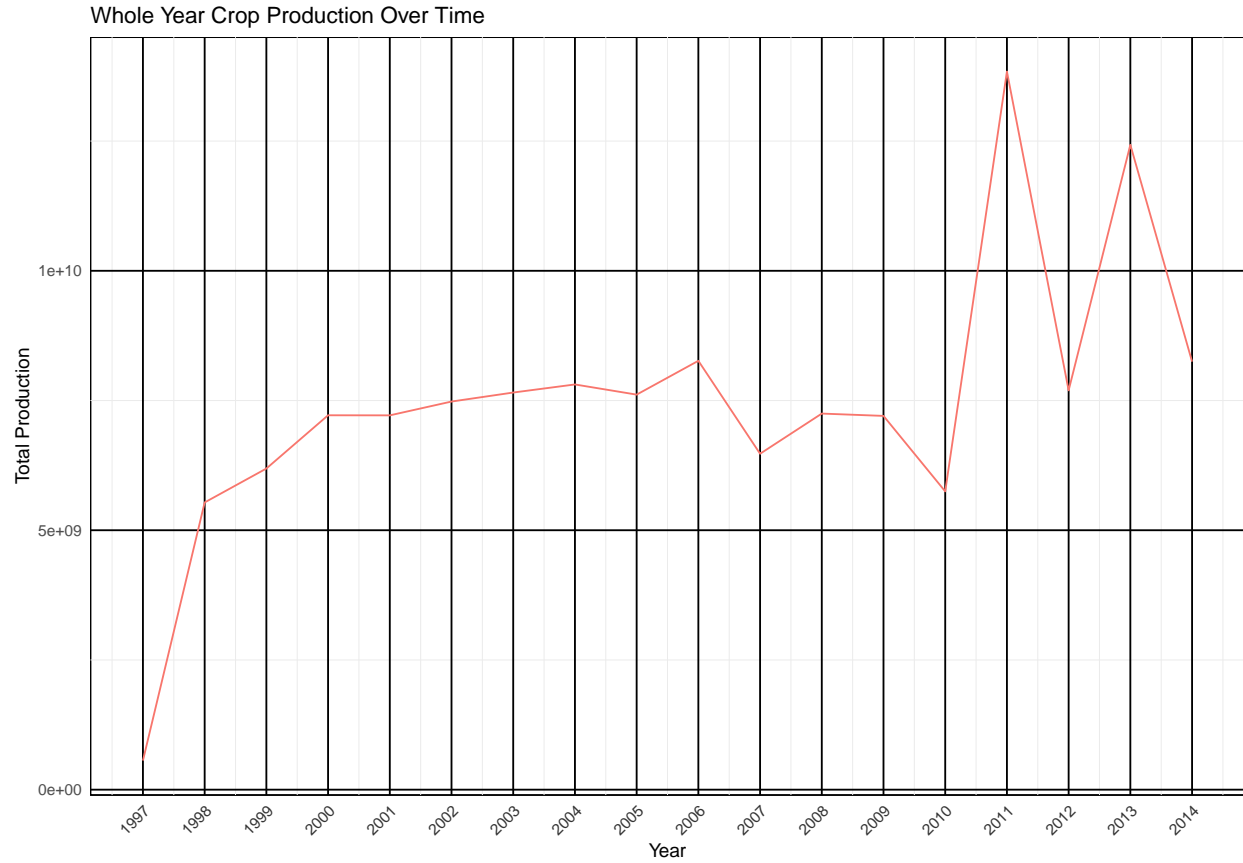
3.4 Total Production by Season over years (1997-2006 vs. 2007-2015)

In the dataset, there are six different seasons, namely Kharif, Annual, Autumn, Rabi, Summer, and Winter crops. Below is a double barplot that visualizes the total agricultural production across these different seasons, comparing two distinct time periods: 1997-2006 (represented by blue bars) and 2007-2015 (represented by red bars). Each season is displayed on the x-axis, while the y-axis represents the \log_{10} of the total production quantity (refer to note 6 in 2.3).



Observations :

- The graph provides a clear comparison of how agricultural production has evolved over time. By observing the height of the bars, viewers can discern trends and variations in production for each season between the two time periods.
- If we look closely at each bar in above plot it is clear that crop production in most crop season has increased (except Autumn season) with Kharif season showing the highest change .
- It is noticed that crops of Whole Year season type give the highest production.
- The season giving the second highest total production is Kharif, which clearly points to the dependence on monsoon rains. Therefore, to look at trends of crop production in Season = “Whole Year” below is a line plot.



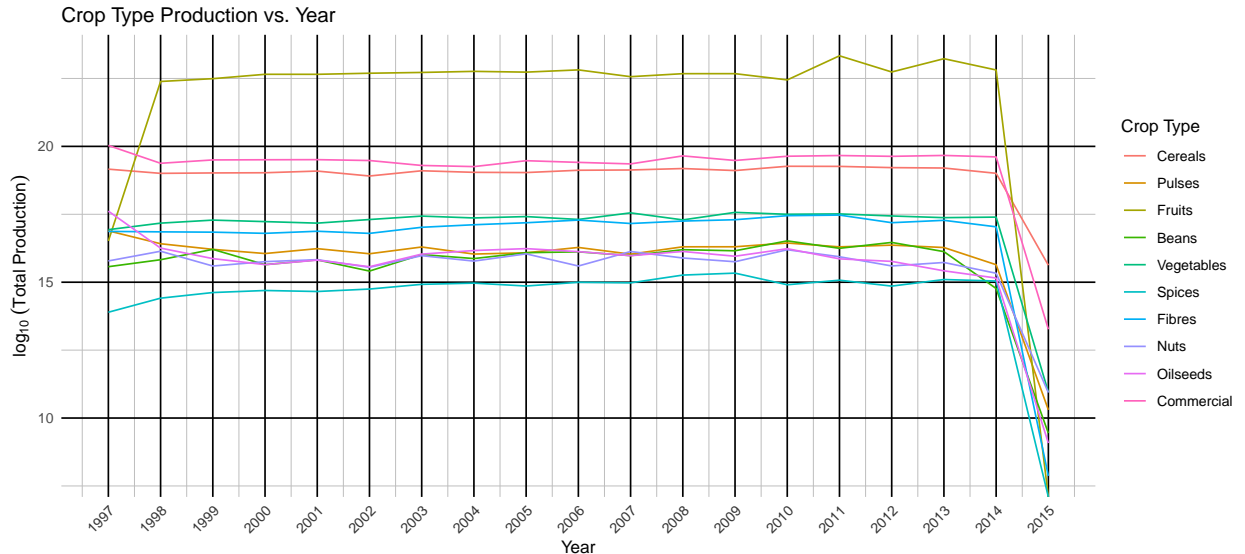
Observe that this plot is quite similar to *the plot in 3.3*. This is majorly because the total production value for “Whole Year” season type crops is more than 10^{11} and other season have total production values less than or near 10^9 . Therefore, the Whole Year Season crops dominate the total production value.

3.5 Plotting production of different type of crops over different years

Crop Types	Crops Included
Cereals	Rice, Maize, Wheat, Barley, Varagu, Other Cereals and Millets, Ragi, Small millets, Bajra, Jowar, Paddy, Total foodgrain, Jobster
Pulses	Moong, Urad, Arhar/Tur, Peas and beans, Masoor, Other Kharif pulses, other misc. pulses, Ricebean (nagadal), Rajmash Kholar, Lentil, Samai, Blackgram, Korra, Cowpea(Lobia), Other Rabi pulses, Other Kharif pulses, Peas and beans (Pulses), Pulses total, Gram
Fruits	Peach, Apple, Litchi, Pear, Plums, Ber, Sapota, Lemon, Pomegranate, Other Citrus Fruit, Water Melon, Jack Fruit, Grapes, Pineapple, Orange, Pome Fruit, Citrus Fruit, Other Fresh Fruits, Mango, Papaya, Coconut, Banana
Beans	Bean, Lab-Lab, Moth, Guar seed, Soyabean, Horse-gram
Vegetables	Turnip, Peas, Beet Root, Carrot, Yam, Ribbed Guard, Ash Gourd, Pumpkin, Radish, Snak Guard, Bottle Gourd, Bitter Gourd, Cucumber, Drum Stick, Cauliflower, Beans and Mutter(Vegetable), Cabbage, Bhindi, Tomato, Brinjal, Khesari, Sweet potato, Potato, Onion, Tapioca, Colocasia
Spices	Perilla, Ginger, Cardamom, Black pepper, Dry ginger, Garlic, Coriander, Turmeric, Dry chilies, Cond-spcs other
Fibres	Other fibers, Kapas, Jute & mesta, Jute, Mesta, Cotton(lint), Sannhamp
Nuts	Arecanut (Processed), Areca nut (Raw), Cashewnut Processed, Cashewnut Raw, Cashewnut, Arecanut, Groundnut
Oilseeds	Other oilseeds, Safflower, Niger seed, Castor seed, Linseed, Sunflower, Rapeseed and Mustard, Sesamum, Oilseeds total
Commercial	Tobacco, Coffee, Tea, Sugarcane, Rubber

Table 2: Crops divided by their type.

Using this division of crop types, below is a multiple-line plot showing the \log_{10} of total crop type production (see 6 in 2.3) for different crop types over 1997-2015 dataset.

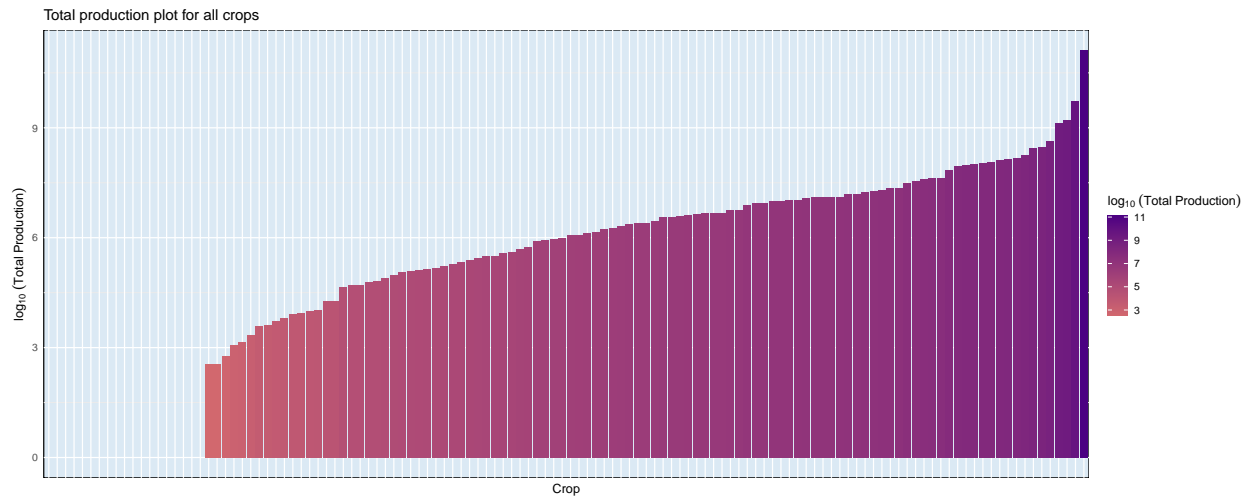


Observations:

- In the given dataset, there were no production values for any kind of fruits and spices in 2015. So, the total production for these type is shown as zero in 2015.
- For most part of 1997-2015 fruits were the most produced crop and spices remained the least produced crop type.

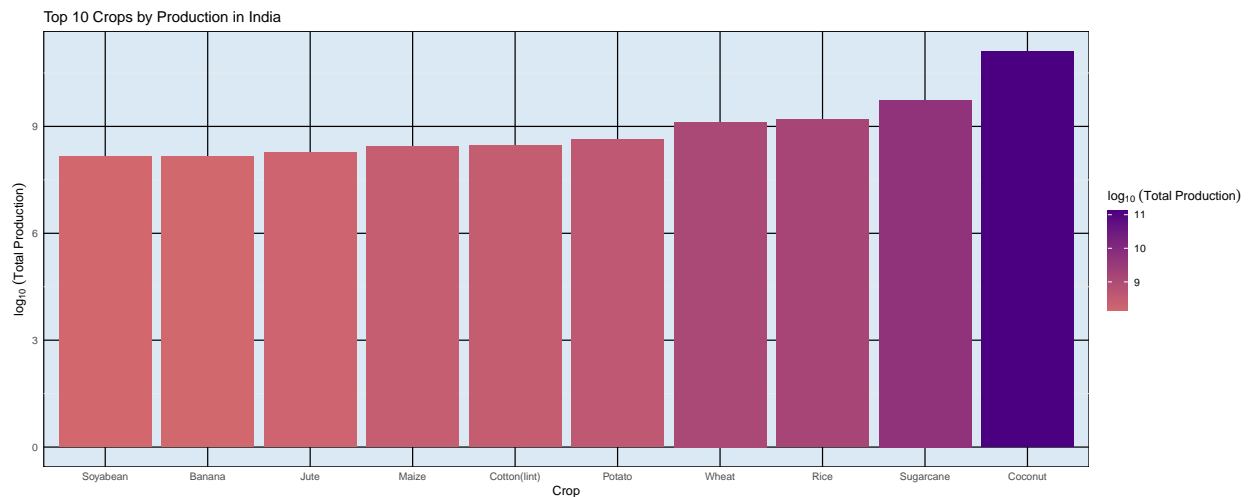
3.6 Crops produced:

As pointed in 4 of 2.3, there total 124 different crops present in the given dataset. Below is a plot consisting of total production of all crops.



In the above scaled gradient plot:

- There is a big difference in total production of different crops with the highest being produced almost 10^8 times the lowest produced crop.
- Since, there are total 124 crops, therefore the crop ticks on x-axis were removed. Rather we have plot for top 10 crops produced.

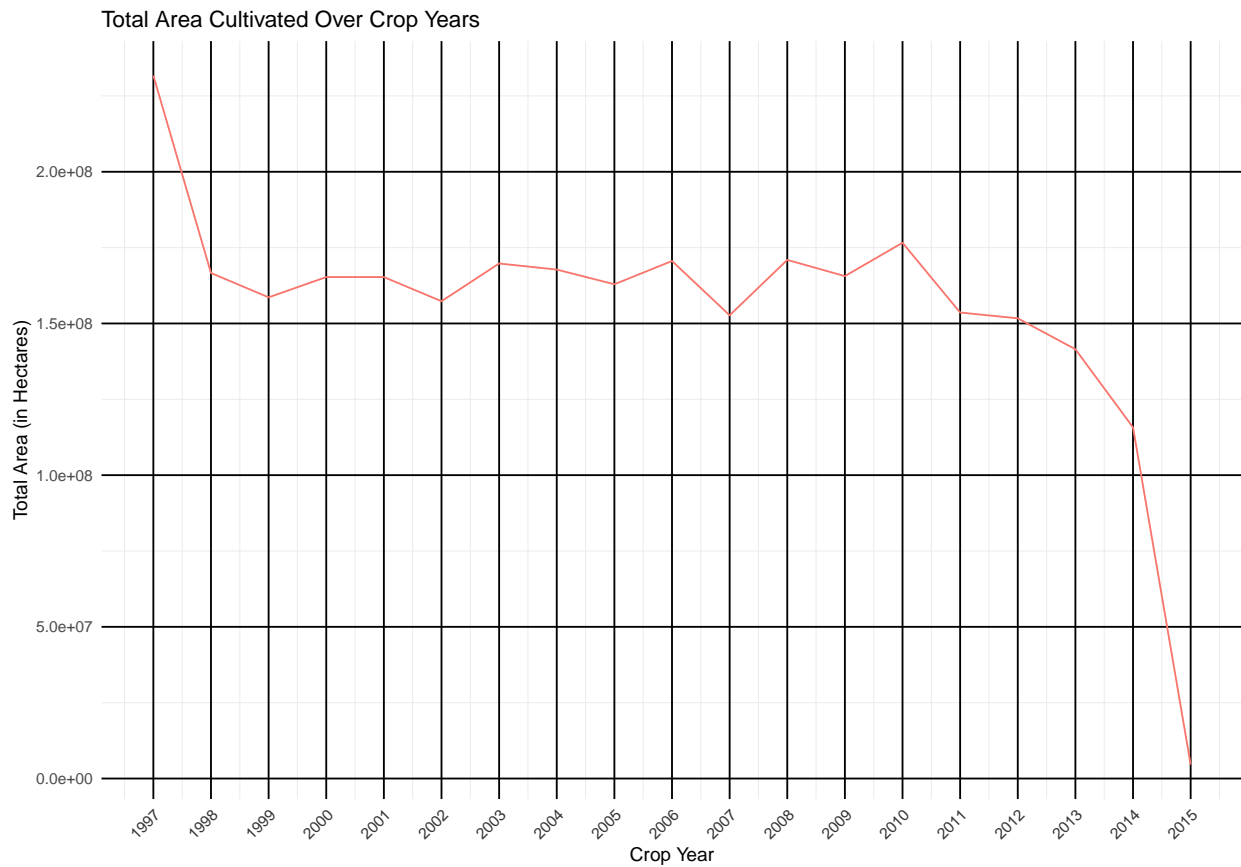


Observations:

- We see that the top 3 produced crop in India for 1997-2015 is are Coconut, Sugarcane and Rice.
- Using the crop type table in 3.5 and this plot we see that:
The most produced fruit in India from 1997 to 2015 was coconut.
The most produced commercial crop in India from 1997 to 2015 was Sugarcane.
The most produced cereal in India from 1997 to 2015 was Rice.
The most produced vegetable in India from 1997 to 2015 was Potato.
The most produced fibres crop in India from 1997 to 2015 was Cotton(lint).
The most produced bean crop in India from 1997 to 2015 was Soyabean.

3.7 Total land cultivated over the years

As mentioned in 5 of 2.3, assuming area to be in Hectare below is a plot of crop year vs total area of land cultivated.



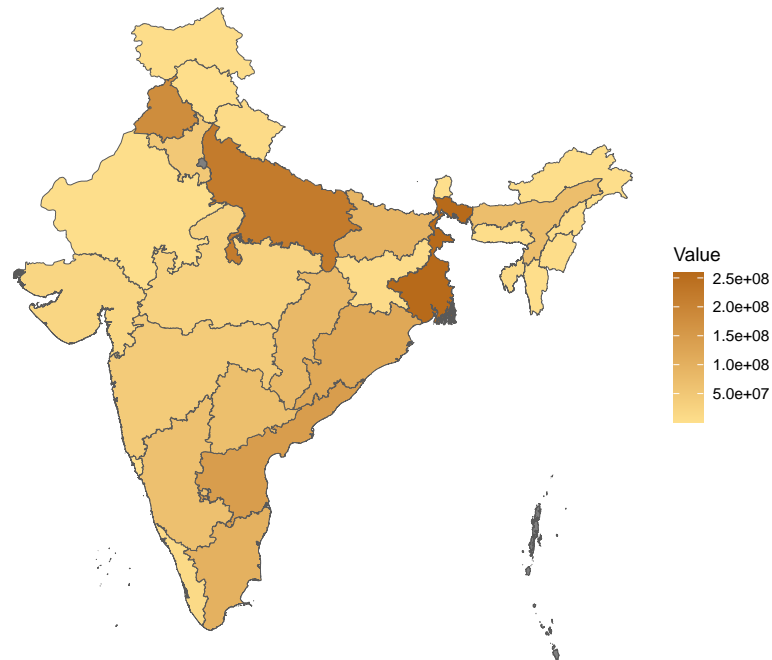
Observations :

- There is significant fall in total cultivated area from 1997 to 1998.
- We can see many small fluctuations from 1998-2010, with 2010 having the highest area cultivated in this duration.
- After 2010, there is continuous downfall in the total land area cultivated.

3.8 Plots related to rice production

India is the second highest rice producing country in the world. Using data on total rice crop production for each state (consisting all data from 1997-2015) a density map plot is made and later the image containing geological map of India is taken to make comparisons.

Density map plot for rice production in India



Observations from rice map plot:

- Rice is grown in varied soil but deep clayey and loamy soil provides the ideal conditions and this can be observed from the two plots.
- Rice is primarily grown in plain areas, with West Bengal being the top producer and Uttar Pradesh the second highest.
- Chandigarh produces the lowest amount of rice due to limited land available for cultivation.

4 Conclusion

In conclusion, our exploratory data analysis has provided valuable insights into the trends and patterns of crop production in India. We observed significant regional variations in production over the given years (1997-2015), with the South Indian states (Andhra Pradesh, Karnataka, Kerala, Tamil Nadu, Telangana) showing the highest production data. Seasonal analysis revealed a clear dependence on monsoon rains, with a peak production during the Kharif season, if the whole year type crops are not considered.

The total production plot over different years also shows the increase in total crop production compared to the early years which can be attributed to the fact that India is a developing nation with which the policies keep getting better for crop production.

Further investigation is needed to fully understand the causes of anomalies and outliers. Based on our findings, it's recommended to increase investment in modern irrigation infrastructure and climate-resilient farming practices, particularly in regions prone to drought.

Agriculture's share in India's economy has progressively declined to less than 15% due to the high growth rates of the industrial and services sectors, and this decline can be observed from our plot *in 3.7*. Understanding crop production trends in India is of paramount importance, given the country's heavy reliance on agriculture for livelihoods and the economy. Future research could delve deeper into the impact of climate change on crop yields and explore strategies for sustainable agriculture.

It's also important to note that many of the visualizations done in this project are valuable for policymakers, analysts, and stakeholders in the agriculture sector to understand regional disparities and trends in agricultural output across India's states. It can help inform decisions related to resource allocation, infrastructure development, and agricultural policies.

5 References

Link to the crop production CSV file : [*Crop Production file*](#)

Link to the India's map in rice production : [*Geological Map*](#)

Additional resources for data used : [*World Bank Data*](#)

Reference for future work : [*Wikipedia page for Green Revolution in India*](#)