Data Visualization Project: To study the production of major crops in various states of India in the last 50 years(from 1966 to 2017) and visualize some pattern between them.

Akshat Jain,Roll no. 204161003

4/13/2021

Motivation:

Agriculture is the most important sector of Indian Economy. Indian agriculture sector accounts for 18 per cent of India's gross domestic product (GDP) and provides employment to 50% of the countries workforce.Despite this,it is considered highly inefficient, wasteful, and incapable of solving the hunger and malnutrition problems. The use of modern technology such as automated machines, crop sensors, etc. can significantly improve the agriculture sector.

Dataset Source: http://data.icrisat.org/dld/src/crops.html ICRISAT-District Level Data.csv Sector: Agriculture Sub-sector: Crops production Target states/UTs: 20 major states #Objective 1. states which contribute to maximum crop production in quantity for year 2017 2. states which have maximum agricultural land for producing crops 3. comparing the production of crops from 1966 to 2017 4. comparing the area under production from 1966 to 2017 5. relation between quantity of production and area of production 6. crops which are produced in maximum quantity 7. crops which are grown in maximum area and analysing their trend,discussing on ways to improve the agriculture sector

```
library(dplyr)
library(sf)
library(rgdal)
library(units)
library(ggplot2)
library(gganimate)
library(zoo)
library(tidyverse)
library(dplyr)
library(av)
library(reshape2)
library(scales)
library(plotly)
library(treemapify)
data <- read.csv("ICRISAT-District Level Data.csv", header = TRUE)
data[data < 0] <- NA
data<-na.locf(na.locf(data), fromLast = TRUE)
data$total_production <- rowSums(data[,c(7,10,13,16,19,22,25,28,31,34,37,40,43,46,49,52,55,58,61,64,67,
```

1

1. states which contribute to maximum crop production in quantity for the year 2017

Plot used: Bar plot, TreeMap, Pie chart, Proportional Symbol map Considering total crop production quantity: 1. Bar plot

```
newdata <- data[ which(data$Year=='2017'),]
ggplot(data = newdata, aes(x = reorder(State.Name,-newdata$total_production),y= newdata$total_production
  geom_bar(aes(fill= State.Name),stat = "identity",size = 2) +
  ggtitle("Comparision of total crop production between
                Indian states for the year 2017")+
   labs(y=" Total production in 1000 metric tons", x = "States Name")+
  theme_bw() +
  theme(legend.position = "none")+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```
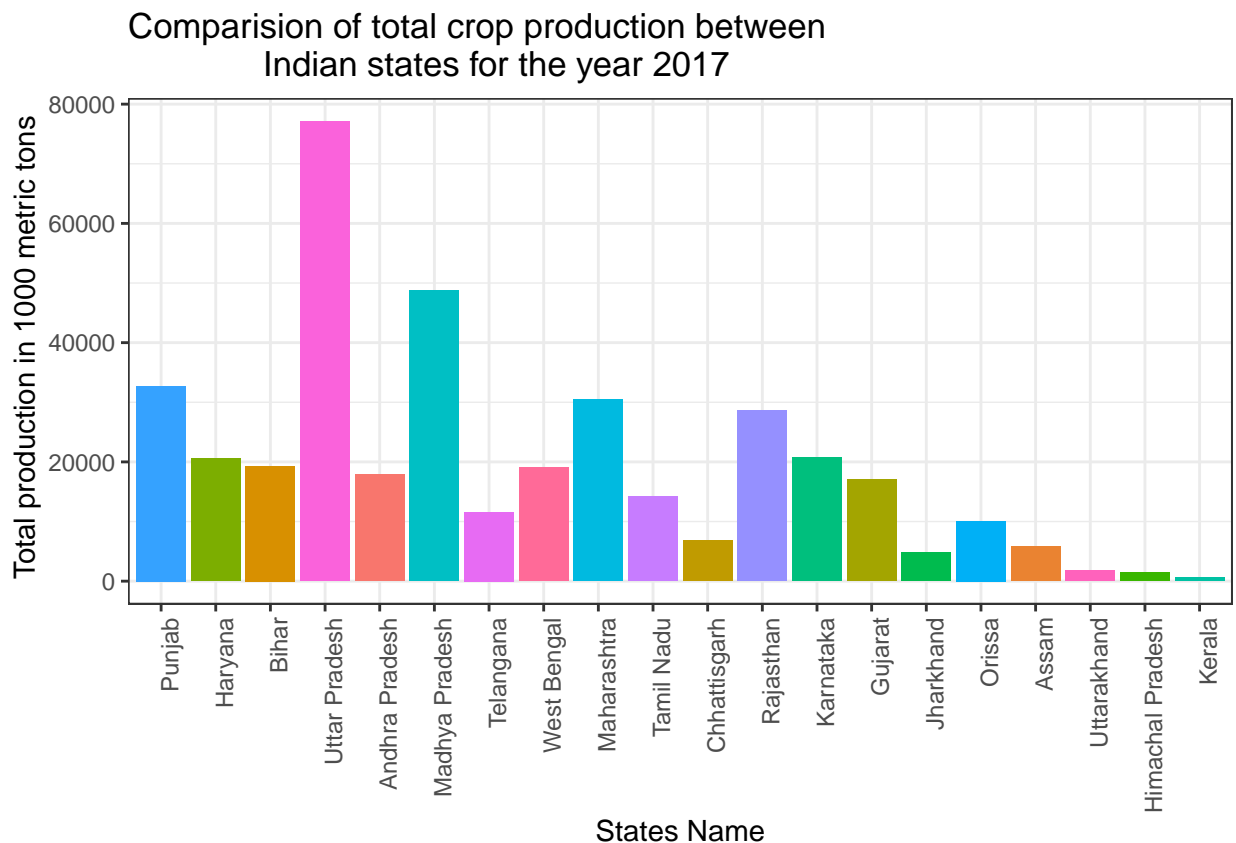

2. TreeMap

```
data$total_area <- rowSums(data[,c(6,9,12,15,18,21,24,27,30,33,36,39,42,45,48,51,54,57,60,63,66,69,72,75

newdata <- data[ which(data$Year=='2017'),]

dat <-newdata %>%
  group_by(State.Name) %>%
    summarise(across(c(total_production, total_area), sum) )
```

```
ggplot(dat, aes(area = total_area, fill = total_production, label = State.Name)) +
geom_treemap() +
geom_treemap_text(fontface = "italic", colour = "black", place = "centre", grow = TRUE) +
  scale_fill_gradient(low="white", high="darkgreen")
```
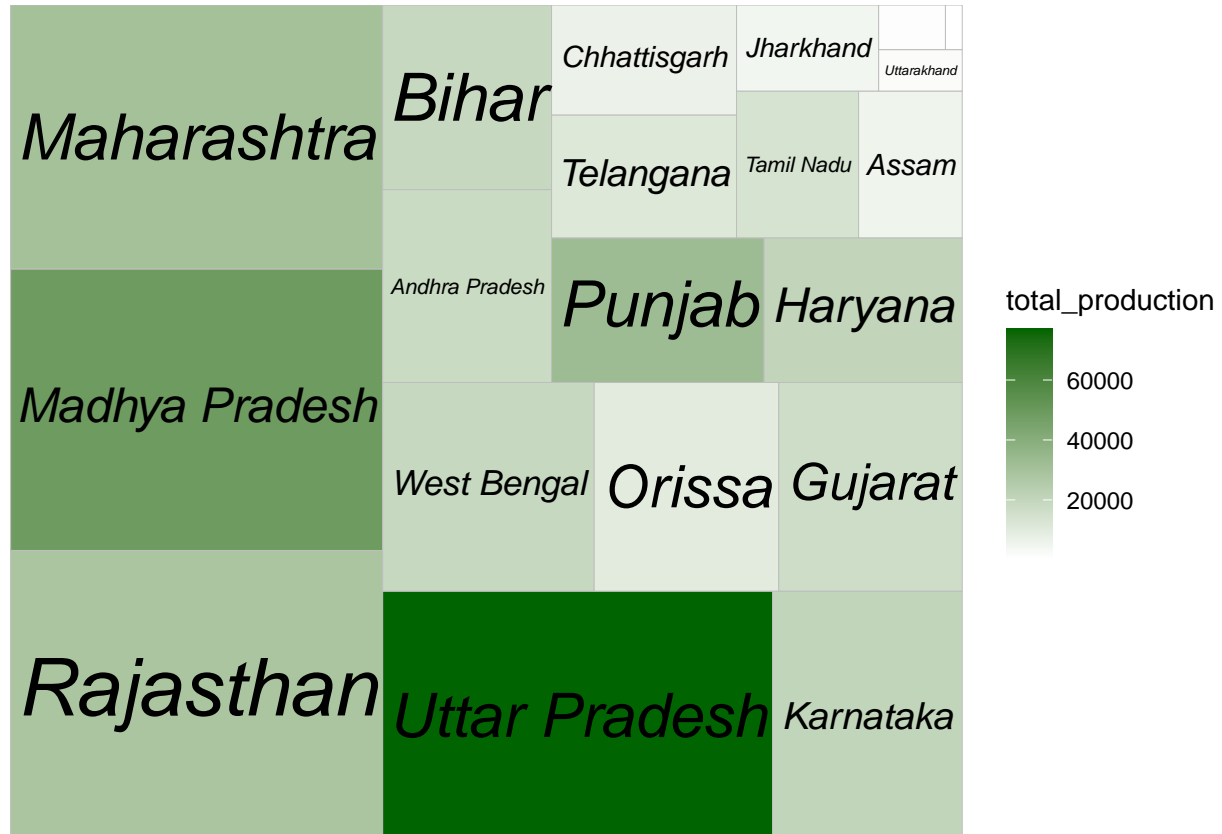

3. Pie Chart

```
newdata <- data[ which(data$Year=='2017'),]
fig <- plot_ly(newdata, labels = ~State.Name,
               values = newdata$total_production, type = 'pie',
               textposition = 'inside',
               textinfo = 'label+percent',
               insidetextfont = list(color = '#FFFFFF'),
               marker = list(line = list(color = 'black', width = 1)),
               showlegend = FALSE)


fig %>%
  layout(title = 'Percentage share of state in total Production of crop for the year 2017')
```

entage share of state in total Production of crop for the year 2

Madhya Pradesh 12.5% **Uttar Pradesh 19.8%** **Punjab 8.41%** **Maharashtra 7.83%** **Rajasthan 7.33%** **Karnataka 5.33%** **Haryana 5.29%** **Bihar 4.95%** **West Bengal 4.9%** **Andhra Pradesh 4.6%** **Gujarat 4.37%** **Tamil Nadu 3.62%** **Telangana 2.96%** **Orissa 2.6%** **Chhattisgarh 1.76%** **Assam 1.49%**

Observations: Uttar Pradesh contributed the maximum crop production for the year 2017 followed by Madhya Pradesh and Punjab. 4. Proportional symbol map:

```
library(rgdal)
library(sp)
shp = readOGR(dsn = ".",layer = "IND_adm0")
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\Hcl\Desktop\2nd sem\Lab CS595 Data Visualisation Lab- Anand Sir-3\project", layer:
## with 1 features
## It has 70 fields
## Integer64 fields read as strings:  ID_0 OBJECTID_1
```

```
plot(shp)
```

```r
states <- fortify(shp)
```

I tried merging the shape file with csv data file, but could not do it because of having only 20 states in the csv data file.

Considering production of rice for the year 2017: 1. Bar Plot:

```r
ggplot(data = newdata, aes(x = reorder(State.Name,-RICE.PRODUCTION..1000.tons.),y = RICE.PRODUCTION..10
  geom_bar(stat = "identity",size = 2) +
  ggtitle("Comparision of total rice production between
              Indian states for the year 2017")+
  labs(y=" Total production in 1000 metric tons", x = "States Name") +
  theme_bw() +
  theme(legend.position = "none")+

  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Comparision of total rice production between Indian states for the year 2017
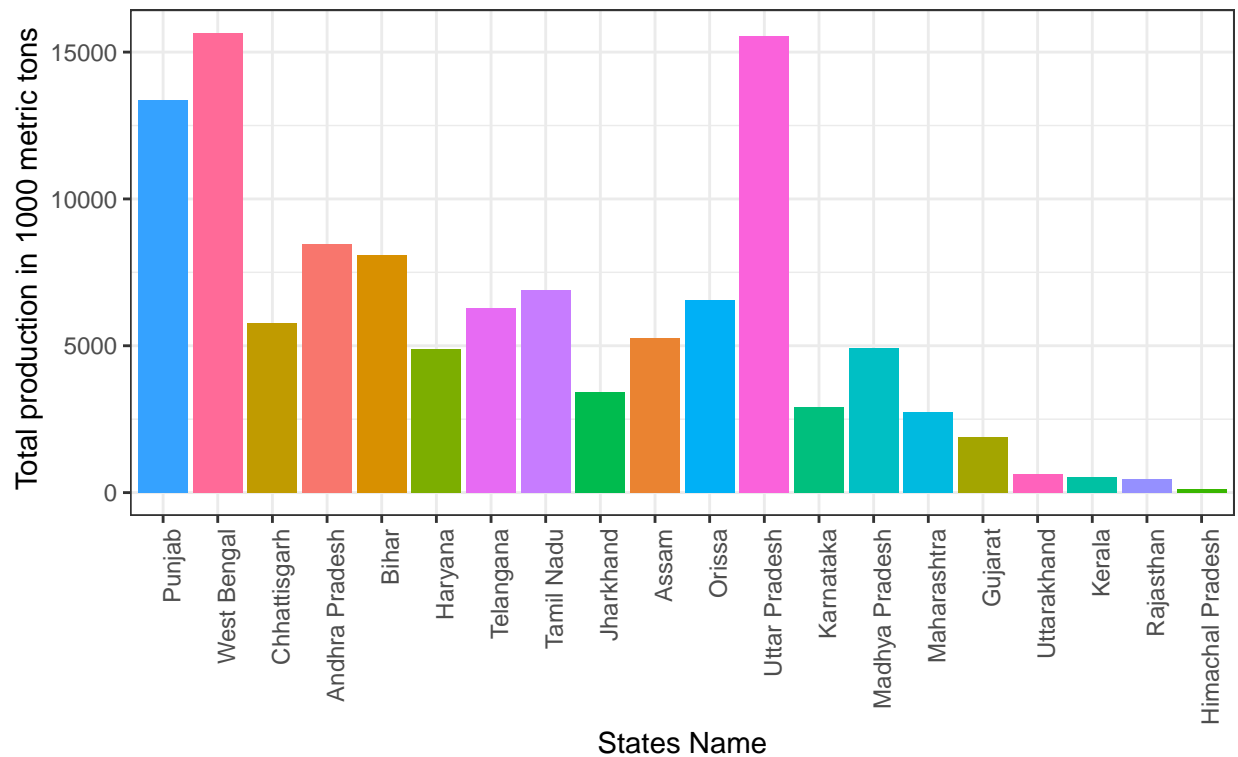
2. Pie Chart

```r
newdata <- data[ which(data$Year=='2017'),]
fig <- plot_ly(newdata, labels = ~State.Name,
               values = newdata$RICE.PRODUCTION..1000.tons, type = 'pie',
               textposition = 'inside',
               textinfo = 'label+percent',
               insidetextfont = list(color = '#FFFFFF'),
               marker = list(line = list(color = 'black', width = 1)),
               showlegend = FALSE)


fig %>%
  layout(title = 'Percentage share of state in Production of rice
                           for the year 2017')
```

Percentage share of state in Production of rice for the year 2017
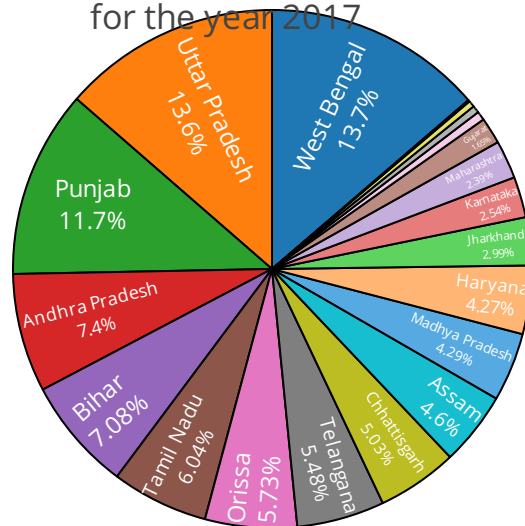
Observation: West Bengal produced maximum rice quantity followed by Uttar Pradesh and Punjab

Pie chart is more convenient for comparing the quantity of crop production among various states

Considering production of wheat for the year 2017: Bar Plot:

```
ggplot(data = newdata, aes(x = reorder(State.Name,-WHEAT.PRODUCTION..1000.tons.),y = WHEAT.PRODUCTION..1
  geom_bar(stat = "identity",size = 2) +
  ggtitle("Comparision of total wheat production between
            Indian states for the year 2017")+
  labs(y=" Total production in 1000 metric tons", x = "States Name") +
  theme_bw() +
  theme(legend.position = "none")+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Comparision of total wheat production between Indian states for the year 2017
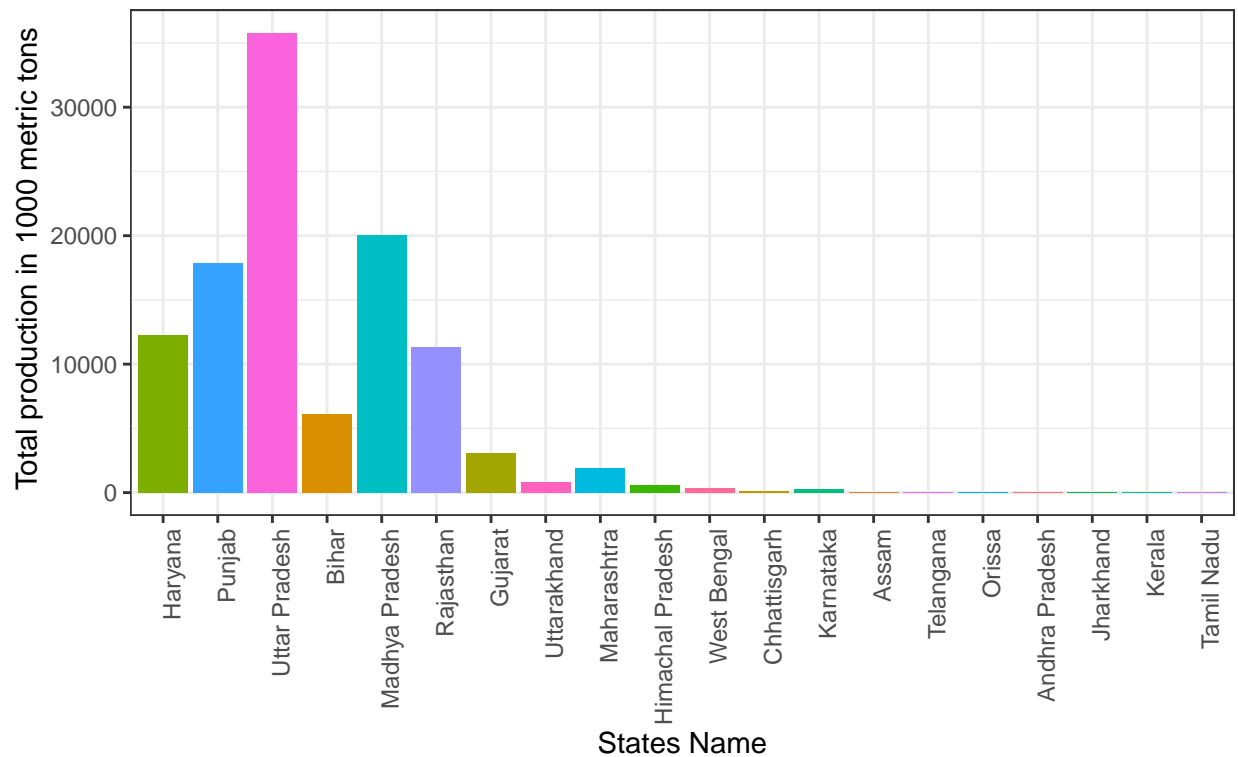
Pie Chart

```
newdata <- data[ which(data$Year=='2017'),]
fig <- plot_ly(newdata, labels = ~State.Name,
               values = newdata$WHEAT.PRODUCTION..1000.tons., type = 'pie',
               textposition = 'inside',
               textinfo = 'label+percent',
               insidetextfont = list(color = '#FFFFFF'),
               marker = list(line = list(color = 'black', width = 1)),
               showlegend = FALSE)


fig %>%
  layout(title = 'Percentage share of states in wheat Production
         of crop for the year 2017')
```

Percentage share of states in wheat Production of crop for the year 2017
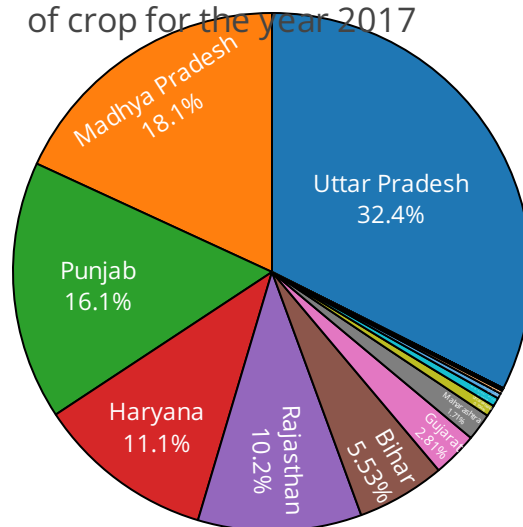
Observation: Uttar Pradesh produced the highest wheat quantity for the year 2017, contributing nearly 1/3rd of the total wheat production. It is followed by Madhya Pradesh and Punjab.

2.

states which have maximum agricultural land for producing crops

Bar plot:

```
newdata <- data[ which(data$Year=='2017'),]
ggplot(data = newdata, aes(x = reorder(State.Name,-newdata$total_area),y= newdata$total_area))+
  geom_bar(aes(fill= State.Name),stat = "identity",size = 2) +
  ggtitle("Comparision of total area under crop production between
          Indian states for the year 2017")+
  labs(y=" Total area in  Hectares", x = "States Name")+
  theme_bw() +
  theme(legend.position = "none")+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Comparision of total area under crop production between
Indian states for the year 2017
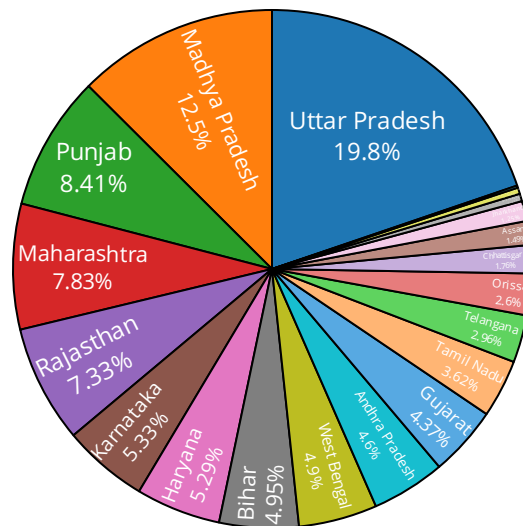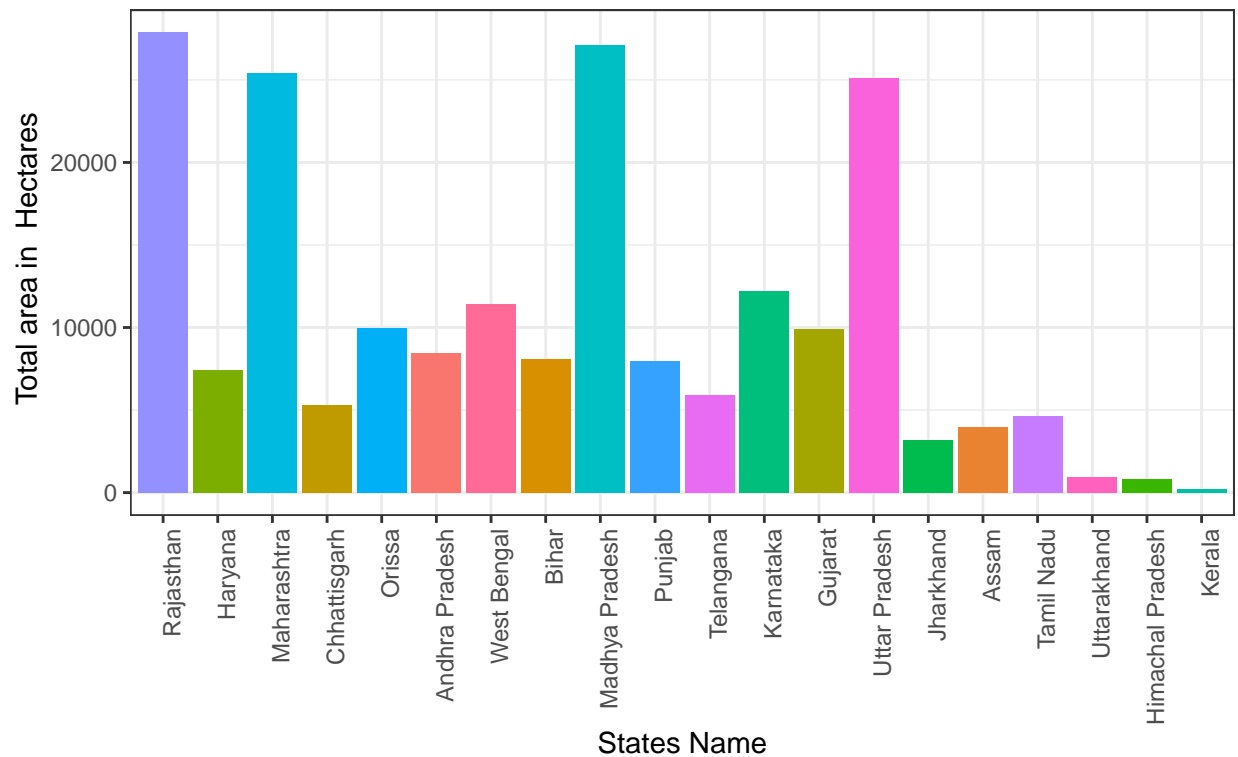

Pie chart

```
newdata <- data[ which(data$Year=='2017'),]
fig <- plot_ly(newdata, labels = ~State.Name,
               values = newdata$total_area, type = 'pie',
               textposition = 'inside',
               textinfo = 'label+percent',
               insidetextfont = list(color = '#FFFFFF'),
               marker = list(line = list(color = 'black', width = 1)),
               showlegend = FALSE)


fig %>%
  layout(title = '% share of state in total area under Production of crop for the year 2017')
```

re of state in total area under Production of crop for the year
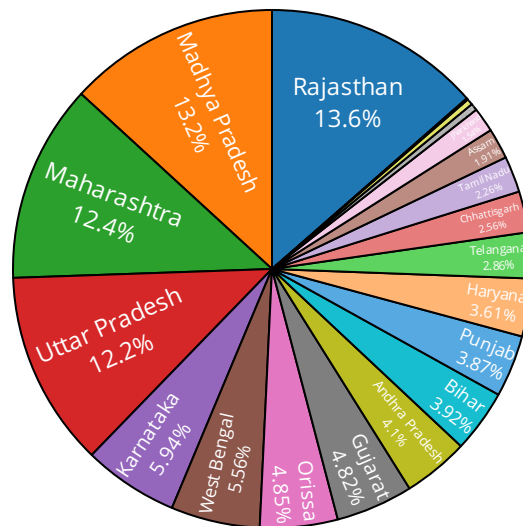
Observation: Rajasthan has maximum area under production of crops followed by Madhya Pradesh and Maharashtra #3. Comparing the total crop production of crops yearwise from 1966 to 2017 Bar Plot:

```
data$total_production <- rowSums(data[,c(7,10,13,16,19,22,25,28,31,34,37,40,43,46,49,52,55,58,61,64,67,

ggplot(data = data, aes(x = Year, y = total_production, fill = Year))+
  geom_bar(stat = "identity",position="stack") +
  scale_y_continuous(name="Total production of crops in tons", labels = comma)+
  ggtitle("Comparing the total crop production of crops yearwise
          from 1966 to 2017")+
  theme_bw()
```

Comparing the total crop production of crops yearwise from 1966 to 2017
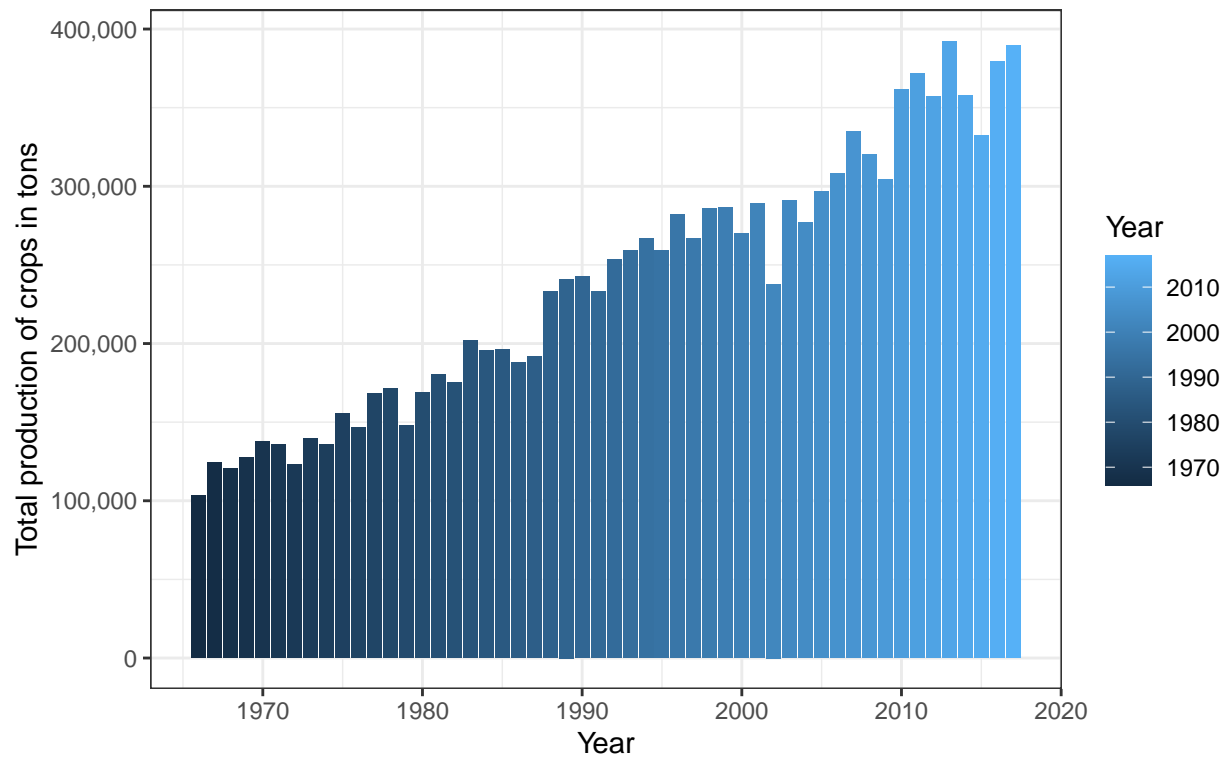
Observation: Total crop production has increased almost 4 times from 1966 to 2017.

4.

Comparing the area under production of crops yearwise from 1966 to 2017

```
data$total_area <- rowSums(data[,c(6,9,12,15,18,21,24,27,30,33,36,39,42,45,48,51,54,57,60,63,66,69,72,75

ggplot(data = data, aes(x = Year, y = total_area, fill = Year))+
  geom_bar(stat = "identity",position="stack") +
  scale_y_continuous(name="Total area under production of crops in hectare", labels = comma)+
  ggtitle("Comparing the area under production of crops yearwise
          from 1966 to 2017")+
  theme_bw()
```

Comparing the area under production of crops yearwise
from 1966 to 2017
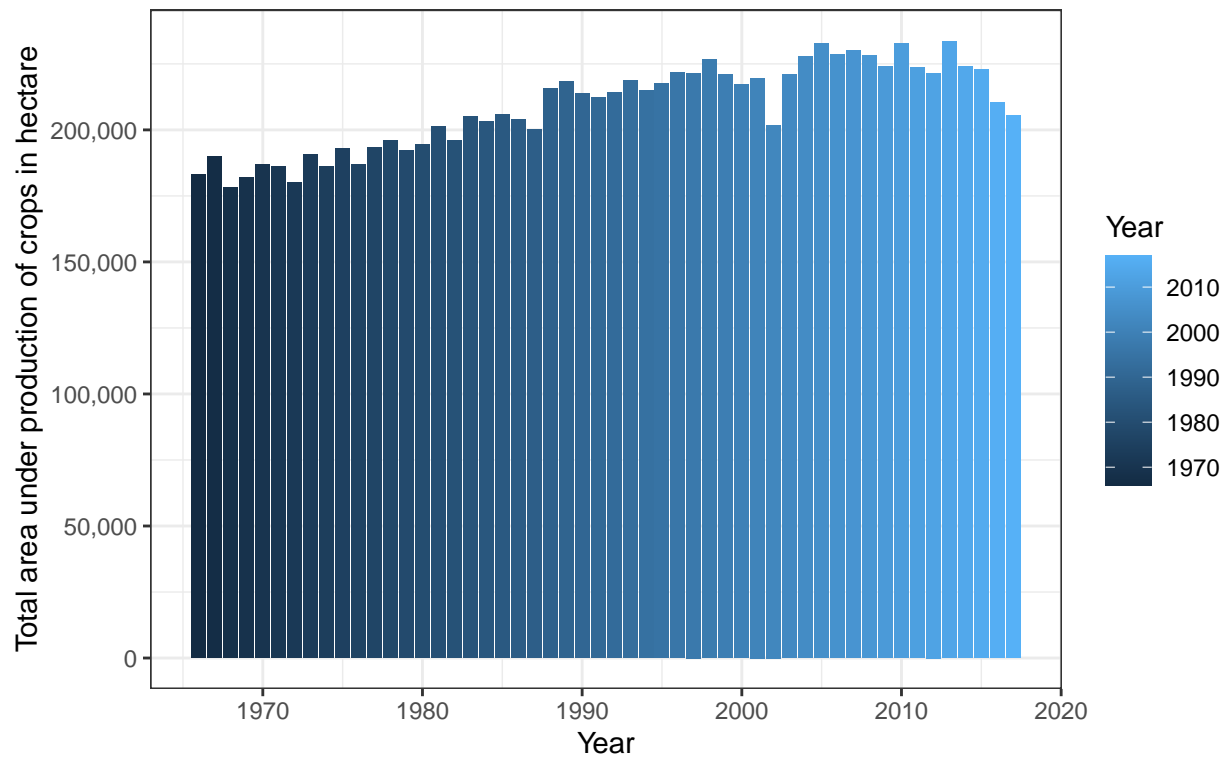
Observation: The area under production of crops increased gradually from 1966 to 2010 but has been declining rapidly since 2010. Urbanization and industrialization can be one of the reasons for its decline.

5.

relation between quantity of production and area of production Plot: Scattered plot

```
newdata <- data[ which(data$Year=='2017'),]

dat <-newdata %>%
  group_by(State.Name) %>%
   summarise(across(c(total_production, total_area), sum) )

ggplot(dat, aes(y = total_production, x = total_area, color = State.Name)) + geom_point(size = 4) +
  labs(y = "Production in 1000 tons", x = "Total area in hectares",
   title = "Comparing production of crops with total area
       for various states for year 2017") +
  theme(plot.title = element_text(hjust = 0.5, size = 13))
```

Comparing production of crops with total area for various states for year 2017
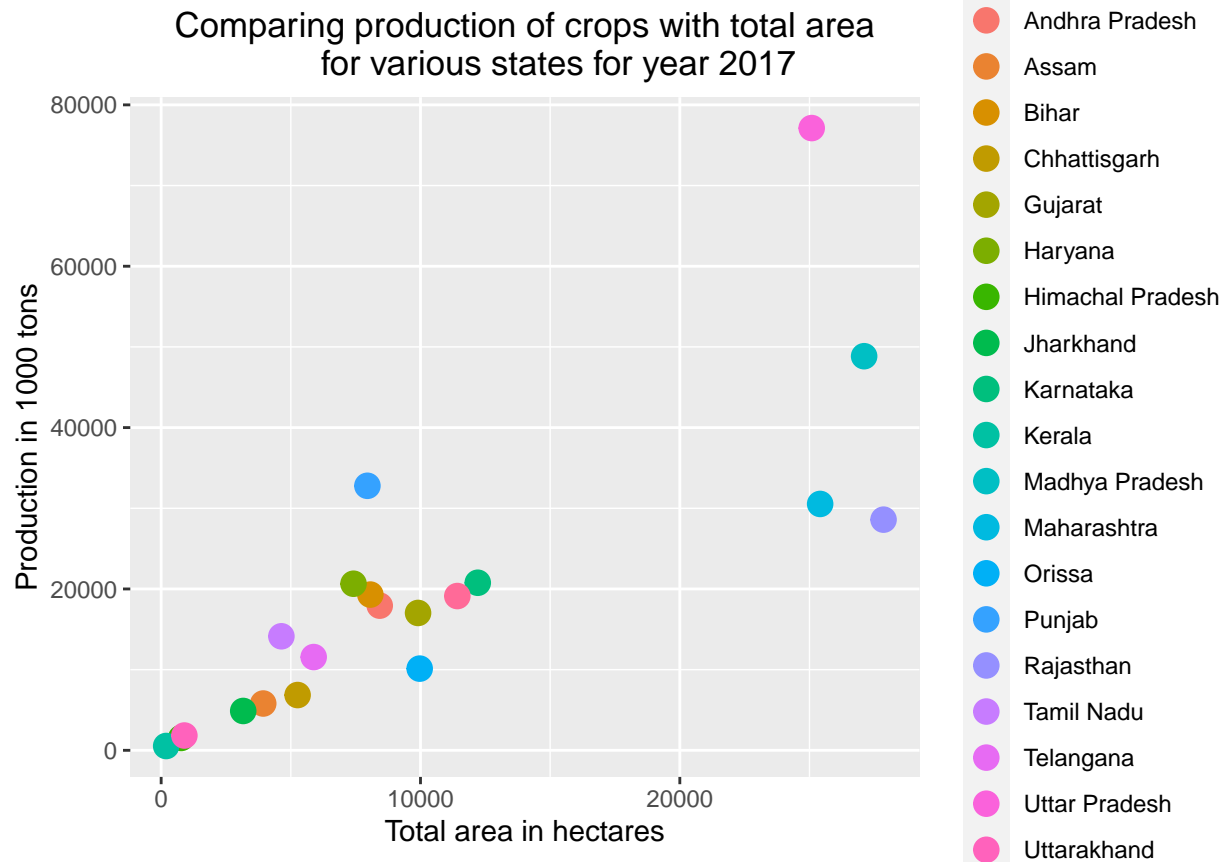
observation: Production of crops increases almost linearly with the increase in area.States such as Rajasthan, Odisha have very low yield output. Punjab has very high yield output because of use of modern techniques

6.

major crops produced and their trend Plot: Stacked bar plot

```
dfm <- melt(data[,c(2,7,10,40,67,70,73)],id.vars = 1)
ggplot(dfm,aes(x = Year,y = value)) +
    geom_bar(aes(fill = variable),stat = "identity",position = "stack")+

    scale_y_continuous(name="Production in Tons", labels = comma)+
  theme(legend.position="bottom")+
  scale_fill_discrete(name = "Crop", labels = c("Rice",      "Wheat","pulses","oilseeds", "Sugarcane","Co
  ggtitle("Comparing various crops production from 1966 to 2017")+
  theme(text = element_text(size = 25))
```

Comparing various crops production from 1966 to 2017
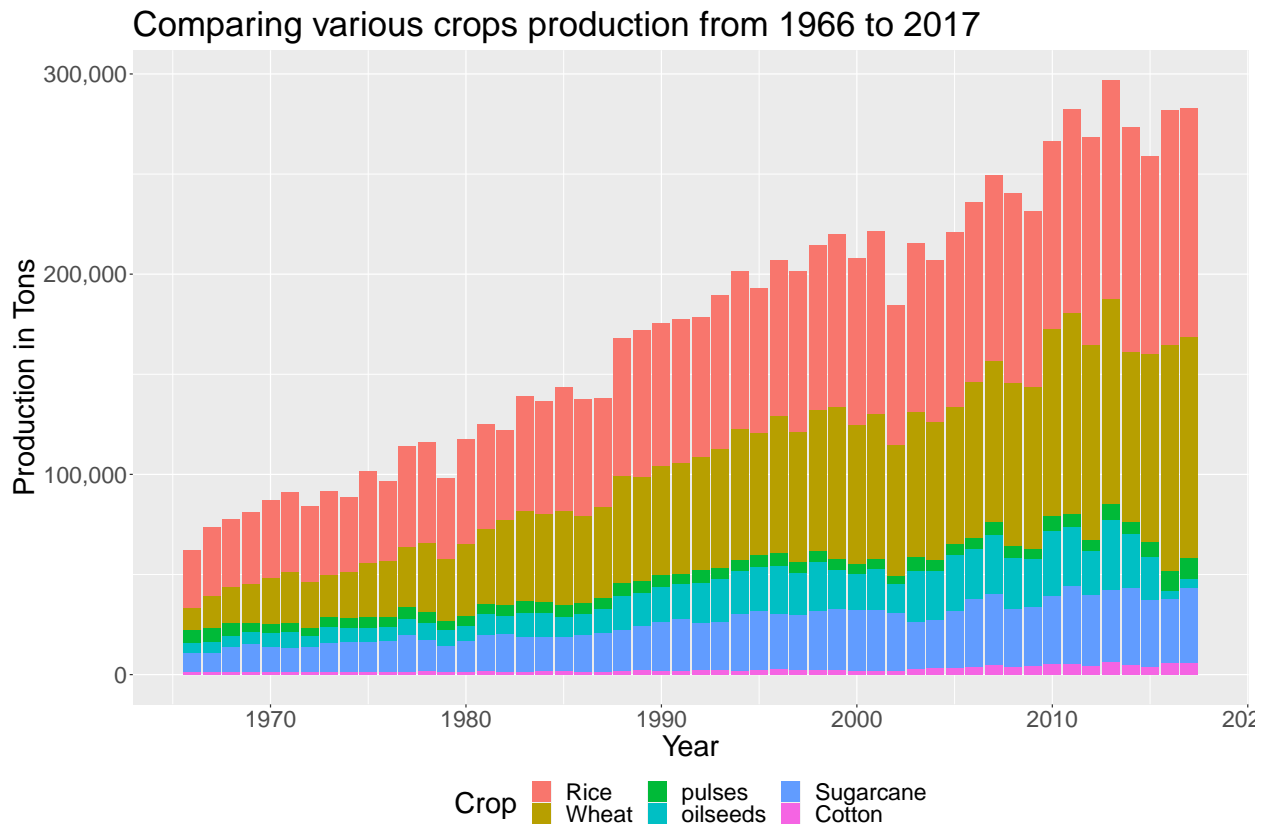
Observation: It is seen that there is continuous rise in the production of staple crop like wheat and rice. On the other hand, production of oil seeds and pulses has been declining gradually. Production of sugarcane and cotton has been constant over the years.

7.

Comparing area under various crop production from 1966 to 2017

```
dfm <- melt(data[,c(2,6,9,39,66,69,73)],id.vars = 1)
ggplot(dfm,aes(x = Year,y = value)) +
    geom_bar(aes(fill = variable),stat = "identity",position = "stack")+

     scale_y_continuous(name="Area in Hectares", labels = comma)+
  theme(legend.position="bottom")+
  scale_fill_discrete(name = "Crop", labels = c("Rice",    "Wheat","pulses","oilseeds", "Sugarcane","C
  ggtitle("Comparing area under various crop production from 1966 to 2017")+
  theme(text = element_text(size = 25))
```

Comparing area under various crop production from 1966 to 2017
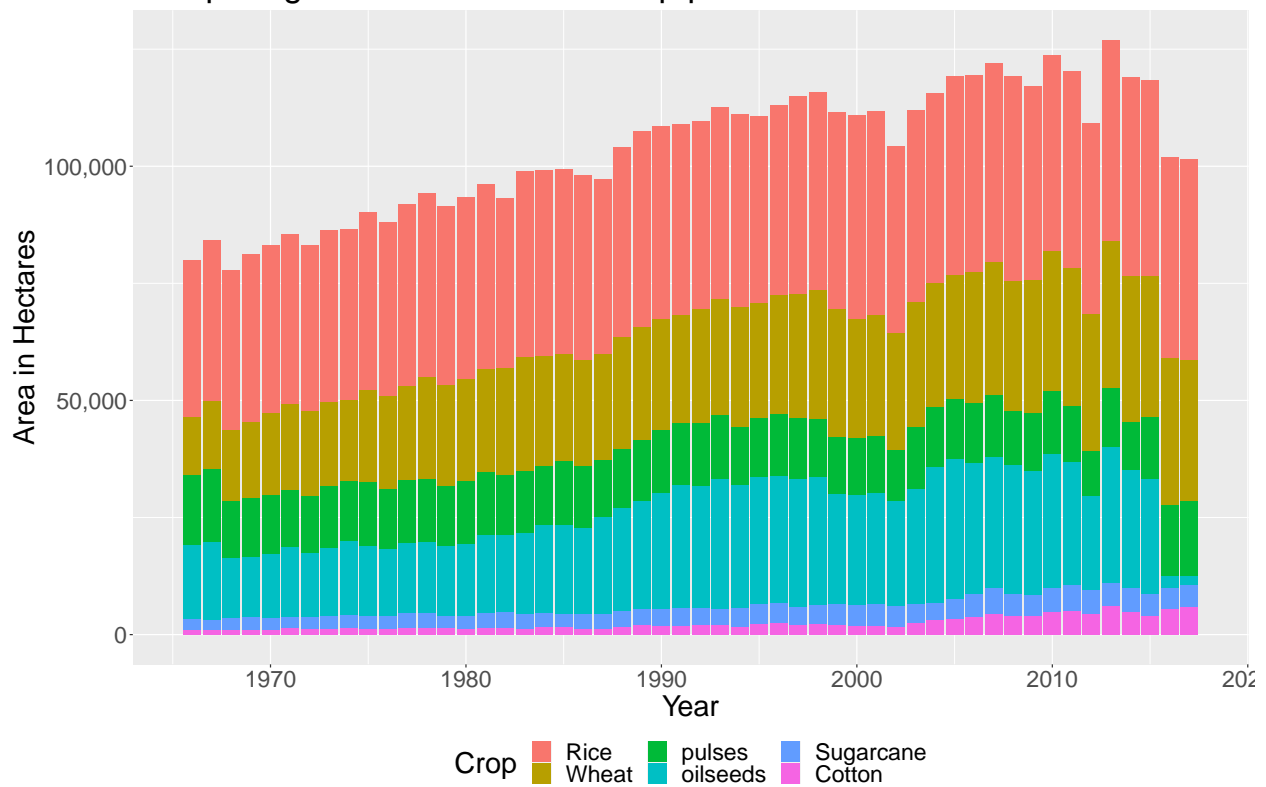
Observation: Area under wheat and rice has remain constant. There is sharp decline in the area under oilseeds in the last decade.

Conclusion: One pattern is seen among the farmers from the above visualization. Most of the farmers rely on traditional crops like wheat and rice. They tend to avoid growing oilseeds mainly because of its low productivity yield and high competition from the international market. India imports large quantity of edible oil such as soyabean oil and Palm oil from neighbouring countries. Government can provide subsidies to the farmer for growing such crops which will help curb imports on such items and also provide farmers new opportunities. This will be a win-win situation for both Indian Economy and the farmers.