

Visualization Project on Agricultural Crop Production in India

Name: Meghna Mondal Roll No. : MDS202123

29/11/2021

Introduction:

India's agriculture is composed of many crops, with the foremost food staples being rice and wheat. Indian farmers also grow pulses, potatoes, sugarcane, oilseeds, and such non-food items as cotton, tea, coffee, rubber, and jute. Despite the overwhelming size of the agricultural sector, however, yields per hectare of crops in India are generally low compared to international standards. So, here we wish to analyse more about the crop cultivation, cost invested in cultivation in different parts of our country. Now we consider 2 datasets from the Kaggle dataset on Agricultural Crop Production in India (source: <https://www.kaggle.com/srinivas1/agriculture-crops-production-in-india/version/1?select=datafile+%282%29.csv>) namely datafile (1).csv and datafile (2).csv containing different variables described below.

```
Variable_name = c("Crop", "State", "Cost.of.Cultivation....Hectare..A2.FL",  
"Cost.of.Cultivation....Hectare..C2", "Cost.of.Production....Quintal..C2",  
"Yield..Quintal..Hectare.")  
Variable_type = c("Nominal", "Nominal", "Continuous", "Continuous",  
"Continuous", "Continuous")  
Variable_content = c("Different types of crops", "States where the crop is  
cultivated", "Expected cost of cultivation of the crop per Hectare", "Cost of  
cultivation of the crop per Hectare", "Cost of Production per Quintal", "Yield  
of crop in Quintal/Hectare")  
Table_1 = data.frame( Variable_name, Variable_type, Variable_content)  
knitr::kable(Table_1, "pipe", col.names = c("Variable Name", "Variable  
Type", "Variable Content"), align = c("l", "c", "c"))
```

Variable Name	Variable Type	Variable Content
Crop	Nominal	Different types of crops
State	Nominal	States where the crop is cultivated
Cost.of.Cultivation....Hectare..A2.FL	Continuous	Expected cost of cultivation of the crop per Hectare
Cost.of.Cultivation....Hectare..C2	Continuous	Cost of cultivation of the crop per Hectare
Cost.of.Production....Quintal..C2	Continuous	Cost of Production per Quintal
Yield..Quintal..Hectare.	Continuous	Yield of crop in Quintal/Hectare
Variable_Name = c("Crop", "Production.2006.07", "Production.2007.08", "Production.2008.09", "Production.2009.10", "Production.2010.11", "Area.2006.07", "Area.2007.08", "Area.2008.09", "Area.2009.10",		

```

"Area.2010.11", "Yield.2006.07", "Yield.2007.08", "Yield.2008.09",
"Yield.2009.10", "Yield.2010.11")
Variable_Type = c("Nominal", "Continuous", "Continuous", "Continuous",
"Continuous", "Continuous", "Continuous", "Continuous", "Continuous",
"Continuous", "Continuous", "Continuous", "Continuous", "Continuous",
"Continuous", "Continuous")
Variable_Content = c("Different types of crops cultivated", "Production of a
particular crop in Year 2006-07", "Production of a particular crop in Year
2007-08", "Production of a particular crop in Year 2008-09", "Production of a
particular crop in Year 2009-10", "Production of a particular crop in Year
2010-11"
, "Area of cultivation of particular crop in Year
2006-07", "Area of cultivation of particular crop in Year 2007-08", "Area of
cultivation of particular crop in Year 2008-09", "Area of cultivation of
particular crop in Year 2009-10", "Area of cultivation of particular crop in
Year 2010-11"
, "Yield of a particular crop in Year 2006-07",
"Yield of a particular crop in Year 2007-08", "Yield of a particular crop in
Year 2008-09", "Yield of a particular crop in Year 2009-10", "Yield of a
particular crop in Year 2010-11")

Table_2 = data.frame( Variable_Name, Variable_Type, Variable_Content)

knitr::kable(Table_2,"pipe",col.names = c("Variable Name","Variable
Type","Variable Content"), align = c("l","c","c"))

```

Variable Name	Variable Type	Variable Content
Crop	Nominal	Different types of crops cultivated
Production.2006.07	Continuous	Production of a particular crop in Year 2006-07
Production.2007.08	Continuous	Production of a particular crop in Year 2007-08
Production.2008.09	Continuous	Production of a particular crop in Year 2008-09
Production.2009.10	Continuous	Production of a particular crop in Year 2009-10
Production.2010.11	Continuous	Production of a particular crop in Year 2010-11
Area.2006.07	Continuous	Area of cultivation of particular crop in Year 2006-07
Area.2007.08	Continuous	Area of cultivation of particular crop in Year 2007-08
Area.2008.09	Continuous	Area of cultivation of particular crop in Year 2008-09
Area.2009.10	Continuous	Area of cultivation of particular crop in Year 2009-10
Area.2010.11	Continuous	Area of cultivation of particular crop in Year 2010-11
Yield.2006.07	Continuous	Yield of a particular crop in Year 2006-07
Yield.2007.08	Continuous	Yield of a particular crop in Year 2007-08
Yield.2008.09	Continuous	Yield of a particular crop in Year 2008-09
Yield.2009.10	Continuous	Yield of a particular crop in Year 2009-10

Yield.2010.11

Continuous

Yield of a particular crop in Year 2010-11

In datafile(1) we have production cost, cultivation cost and yield of some specified crops in particular states. We have a series of data for a nominal attribute. So, we will use Bar Diagrams for analyzing those Cross-Sectional Data.

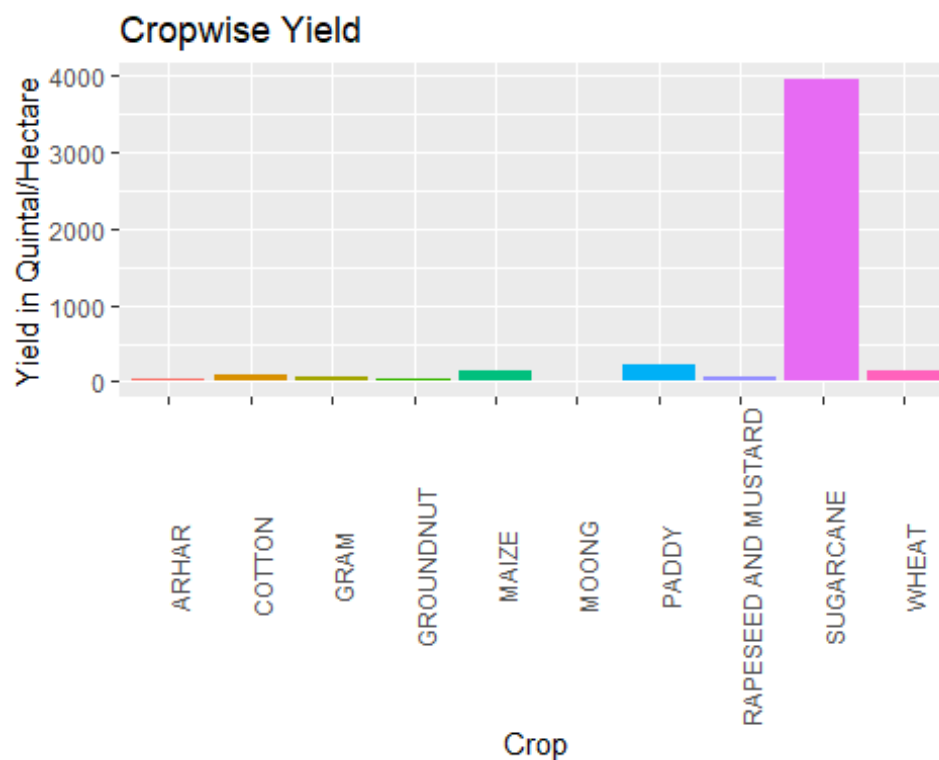
In datafile(2) we amount of production, area of cultivation and yield of some specified crops for five years. We have a series of data for a nominal attribute for varying time. So, we will use Bar Diagrams for analyzing those Panel Data.

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.5

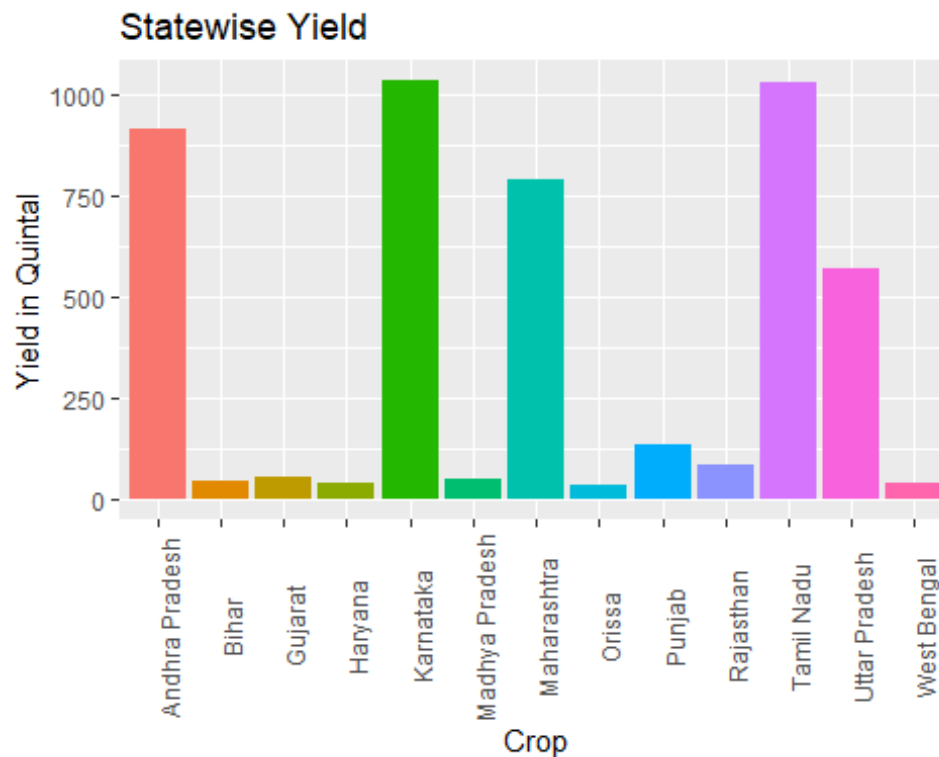
setwd("C:/Users/MONDAL/Desktop/CMI/VISU")
crop_prod<-read.csv("datafile (2).csv")
crop_cost<-read.csv("datafile (1).csv")

ggplot(crop_cost,aes(y=Yield..Quintal..Hectare.,x=Crop ,fill=Crop
))+geom_bar(stat="identity")+ylab("Yield in
Quintal/Hectare")+xlab("Crop")+theme(axis.text.x=element_text(angle=90),legende.position="none")+ggtitle("Cropwise Yield")
```

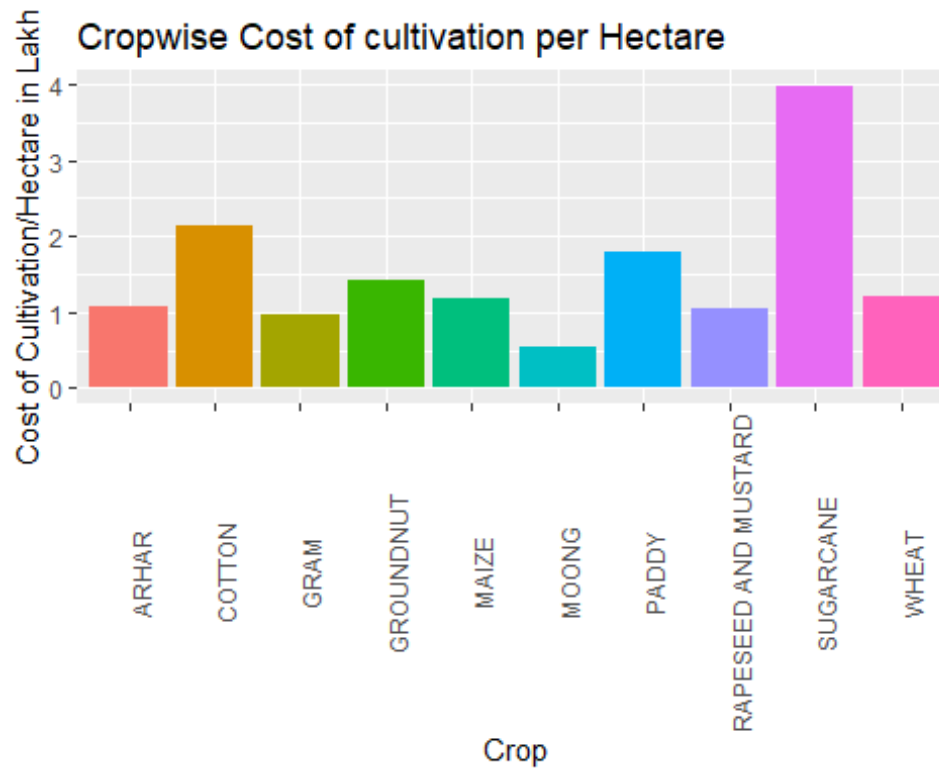


```
ggplot(crop_cost,aes(y=Yield..Quintal..Hectare.,x=State))+geom_bar(stat="identity",aes(fill=State))+ylab("Yield in
```

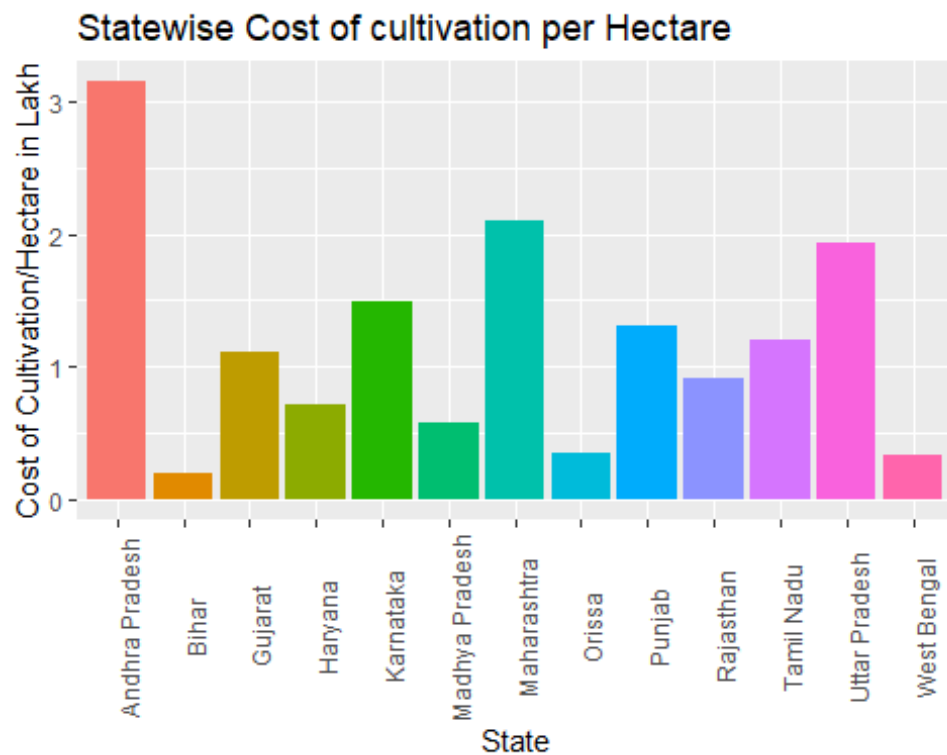
```
Quintal")+xlab("Crop")+theme(axis.text.x=element_text(angle=90),legend.position="none")+ggtitle("Statewise Yield")
```



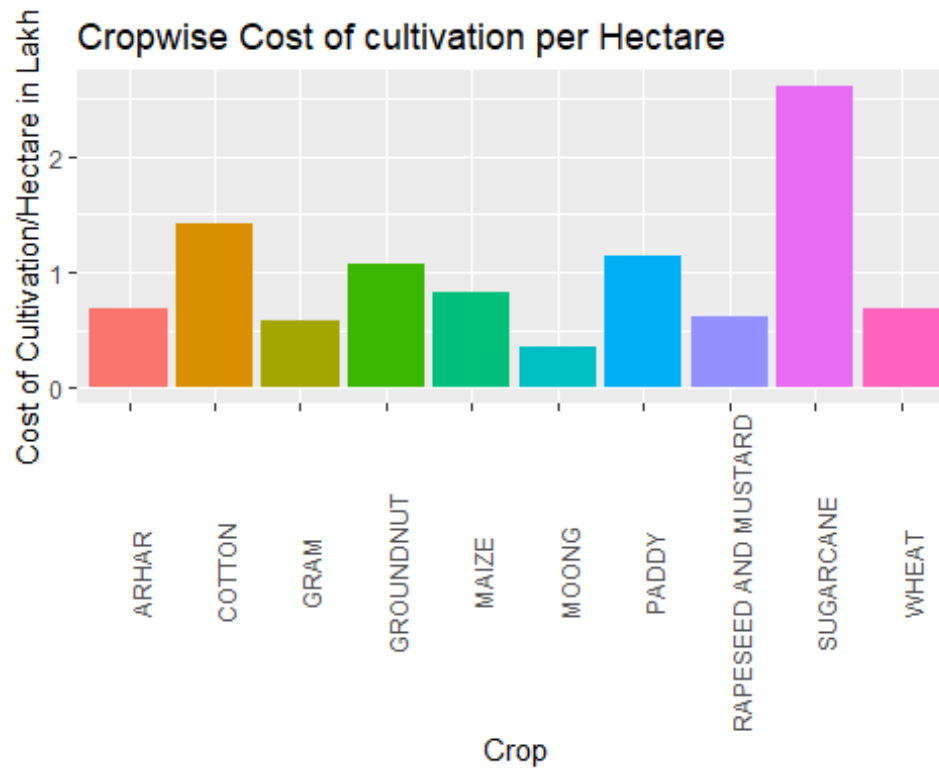
```
ggplot(crop_cost,aes(y=Cost.of.Cultivation....Hectare..C2/100000,x=Crop,fill=Crop))+geom_bar(stat="identity")+ylab("Cost of Cultivation/Hectare in Lakh")+xlab("Crop")+theme(axis.text.x=element_text(angle=90),legend.position="none")+ggtitle("Cropwise Cost of cultivation per Hectare")
```



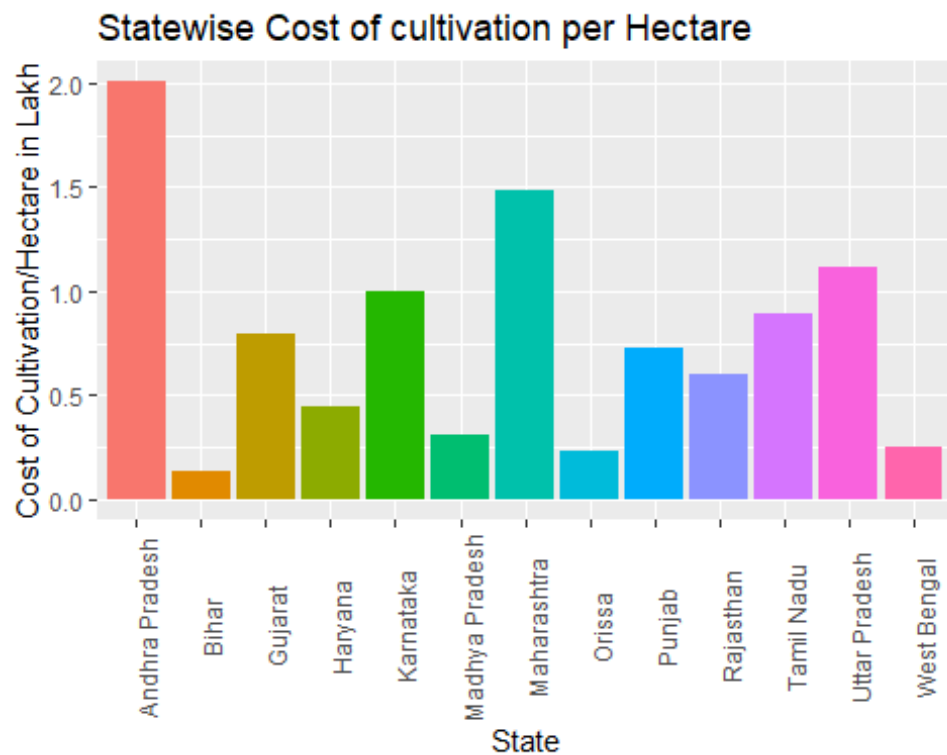
```
ggplot(crop_cost,aes(y=Cost.of.Cultivation....Hectare..C2/100000,x=State))+geom_bar(stat="identity",aes(fill=State))+ylab("Cost of Cultivation/Hectare in Lakh")+xlab("State")+theme(axis.text.x=element_text(angle=90),legend.position="none")+ggtitle("Statewise Cost of cultivation per Hectare")
```



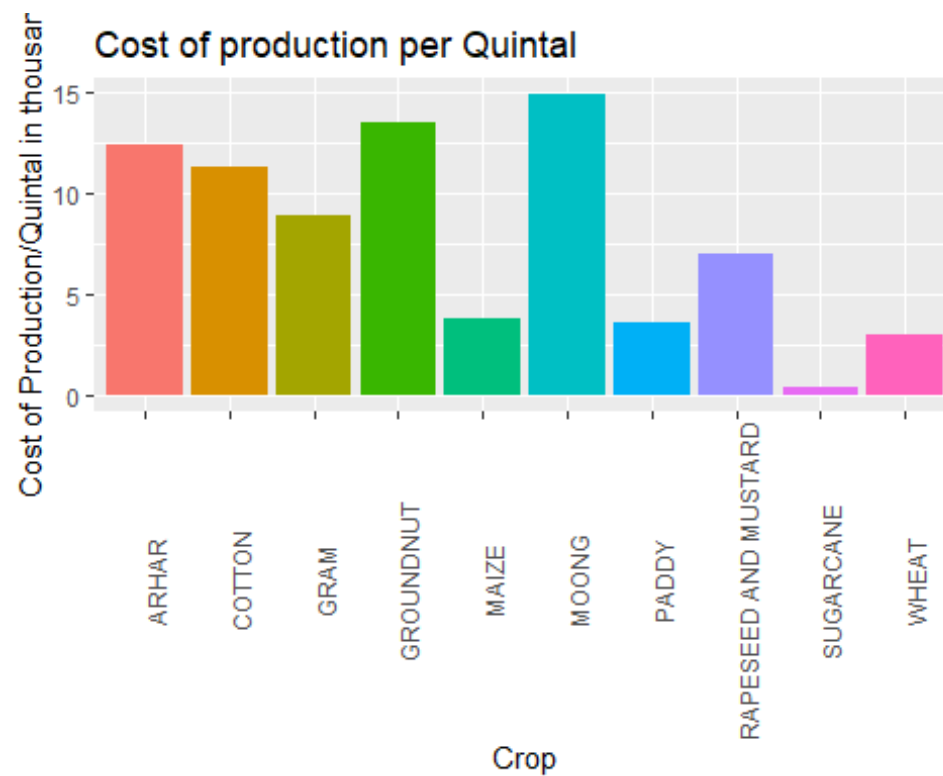
```
ggplot(crop_cost,aes(y=Cost.of.Cultivation....Hectare..A2.FL/100000,x=Crop
,fill=Crop ))+geom_bar(stat="identity")+ylab("Cost of Cultivation/Hectare in
Lakh")+xlab("Crop")+theme(axis.text.x=element_text(angle=90),legend.position=
"none")+ggtitle("Cropwise Cost of cultivation per Hectare")
```



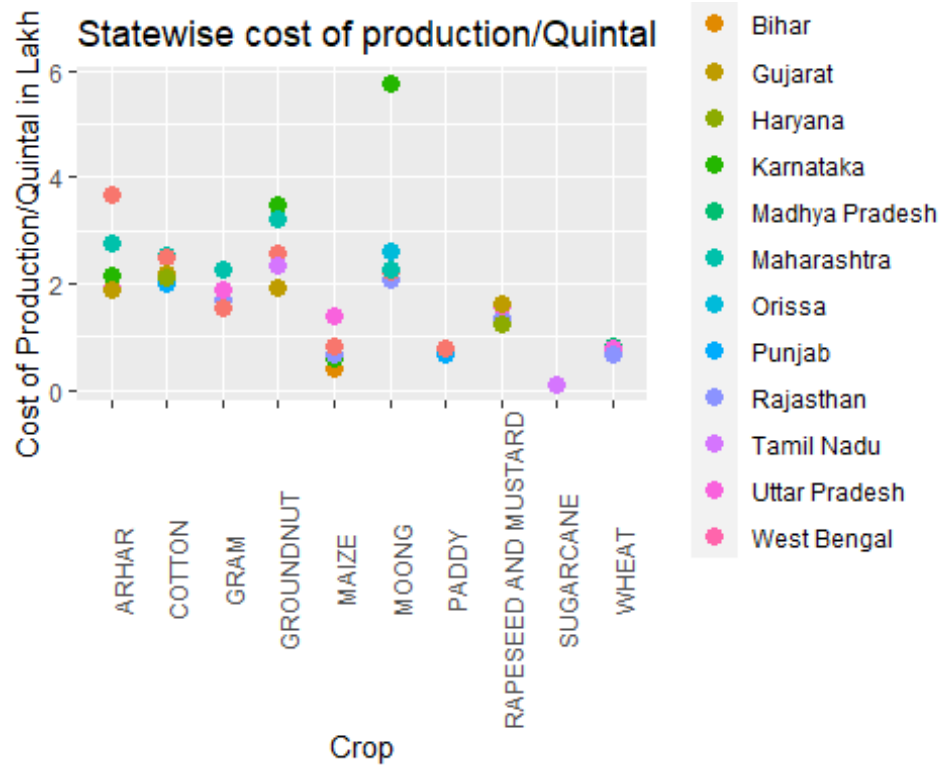
```
ggplot(crop_cost,aes(y=Cost.of.Cultivation...Hectare..A2.FL/100000,x=State))
+geom_bar(stat="identity",aes(fill=State))+ylab("Cost of Cultivation/Hectare
in Lakh")+xlab("State")+theme(axis.text.x=element_text(angle=90),legend.position
="none")+ggtitle("Statewise Cost of cultivation per Hectare")
```



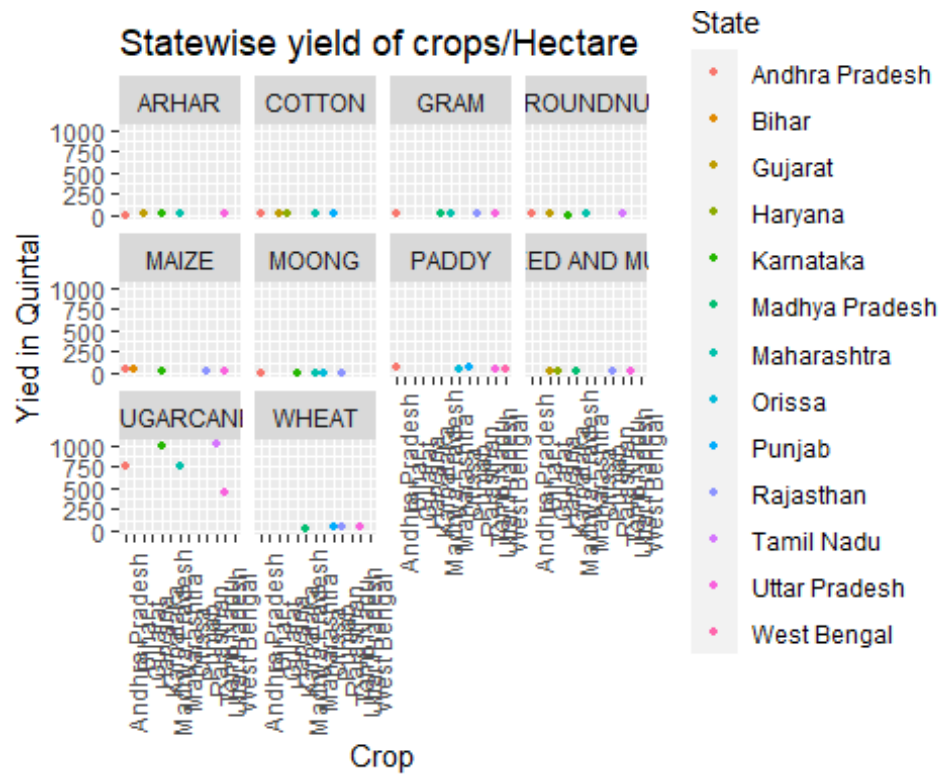
```
ggplot(crop_cost,aes(y=Cost.of.Production....Quintal..C2/1000,x=Crop
,fill=Crop ))+geom_bar(stat="identity")+ylab("Cost of Production/Quintal in
thousand")+xlab("Crop")+theme(axis.text.x=element_text(angle=90),legend.posit
ion="none")+ggtitle("Cost of production per Quintal")
```

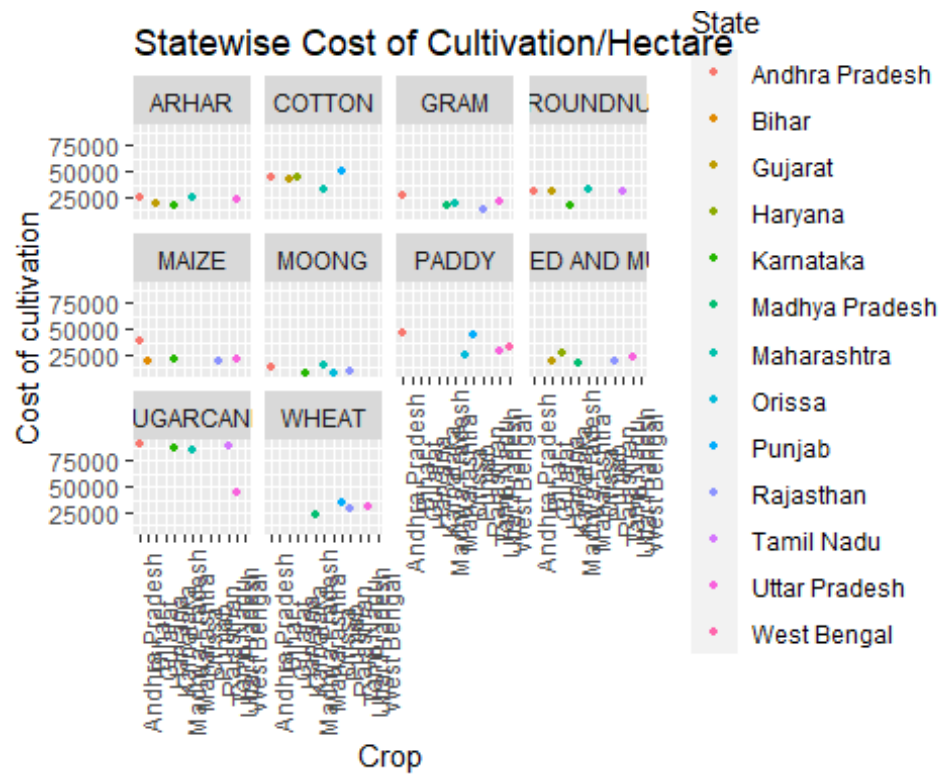
```
ggplot(crop_cost,aes(y=Cost.of.Production....Quintal..C2/1000,x=Crop
,color=State))+geom_point(size=3)+ylab("Cost of Production/Quintal in
Lakh")+xlab("Crop")+theme(axis.text.x=element_text(angle=90))+ggtitle("Statew
ise cost of production/Quintal")
```



```
ggplot(crop_cost,aes(y=Yield..Quintal..Hectare.,x=State,color=State))+geom_point(size=1)+ylab("Yield in Quintal")+xlab("Crop")+theme(axis.text.x=element_text(angle=90))+ggtitle("Statewise yield of crops/Hectare")+facet_wrap(~Crop)
```



```
ggplot(crop_cost,aes(y=Cost.of.Cultivation....Hectare..C2,x=State,color=State
))+geom_point(size=1)+ylab("Cost of
cultivation")+xlab("Crop")+theme(axis.text.x=element_text(angle=90))+ggtitle(
"Statewise Cost of Cultivation/Hectare")+facet_wrap(~Crop )
```



```
ggplot(crop_cost,aes(y=Cost.of.Production....Quintal..C2,x=State,color=State))
+geom_point(size=1)+ylab("Cost of
production")+xlab("Crop")+theme(axis.text.x=element_text(angle=90))+ggtitle("
Statewise Cost of Production/Quintal")+facet_wrap(~Crop )
```



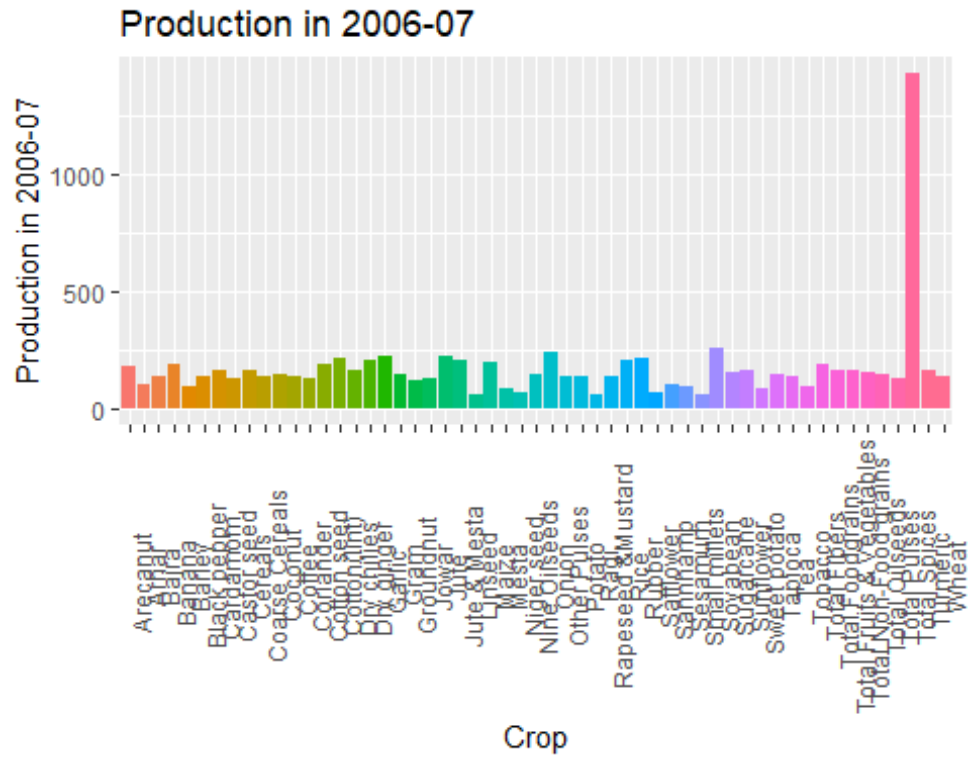
```

cor(crop_cost[, -c(1:2)])

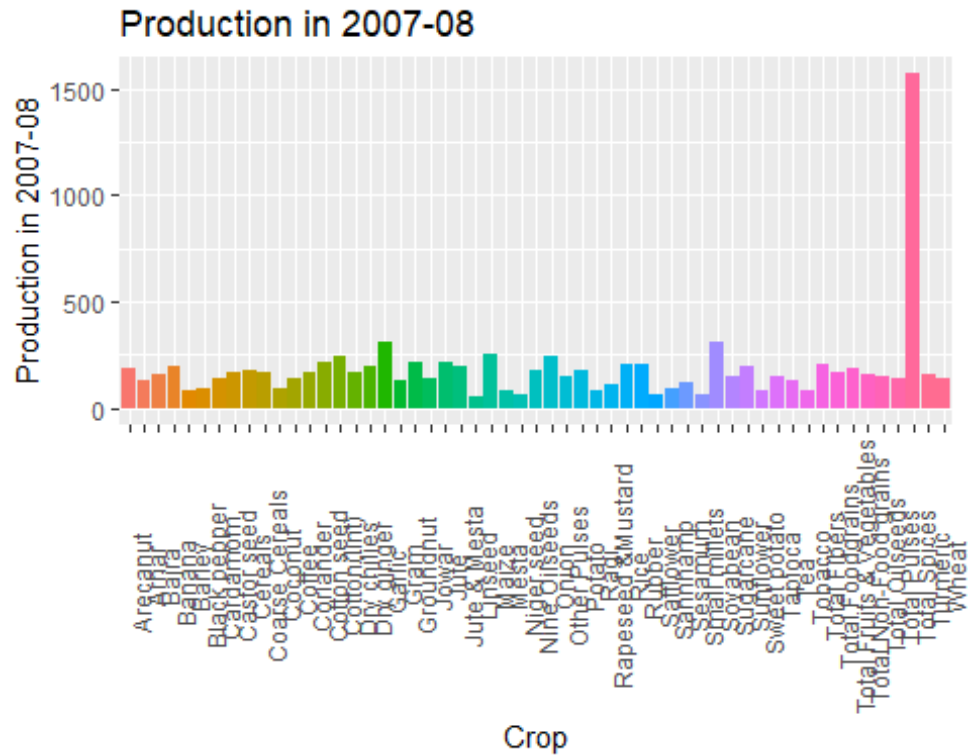
##
Cost.of.Cultivation....Hectare..A2.FL
## Cost.of.Cultivation....Hectare..A2.FL
1.0000000
## Cost.of.Cultivation....Hectare..C2
0.9812253
## Cost.of.Production....Quintal..C2
0.4344223
## Yield..Quintal..Hectare.
0.8634004
##
## Cost.of.Cultivation....Hectare..C2
0.9812253
## Cost.of.Cultivation....Hectare..A2.FL
1.0000000
## Cost.of.Production....Quintal..C2
-0.4970922
## Yield..Quintal..Hectare.
0.8664237
##
## Cost.of.Production....Quintal..C2
-0.4344223
## Cost.of.Cultivation....Hectare..C2
-0.4970922
## Cost.of.Production....Quintal..C2
1.0000000
## Yield..Quintal..Hectare.
-0.4872716
##
## Yield..Quintal..Hectare.
0.8634004
## Cost.of.Cultivation....Hectare..A2.FL
0.8664237
## Cost.of.Cultivation....Hectare..C2
-0.4872716
## Cost.of.Production....Quintal..C2
1.0000000
## Yield..Quintal..Hectare.

ggplot(crop_prod, aes(y=Production.2006.07, x=Crop, fill=Crop
))+geom_bar(stat="identity")+ylab("Production in 2006-
07")+xlab("Crop")+theme(axis.text.x=element_text(angle=90), legend.position="n
one")+ggtitle("Production in 2006-07")

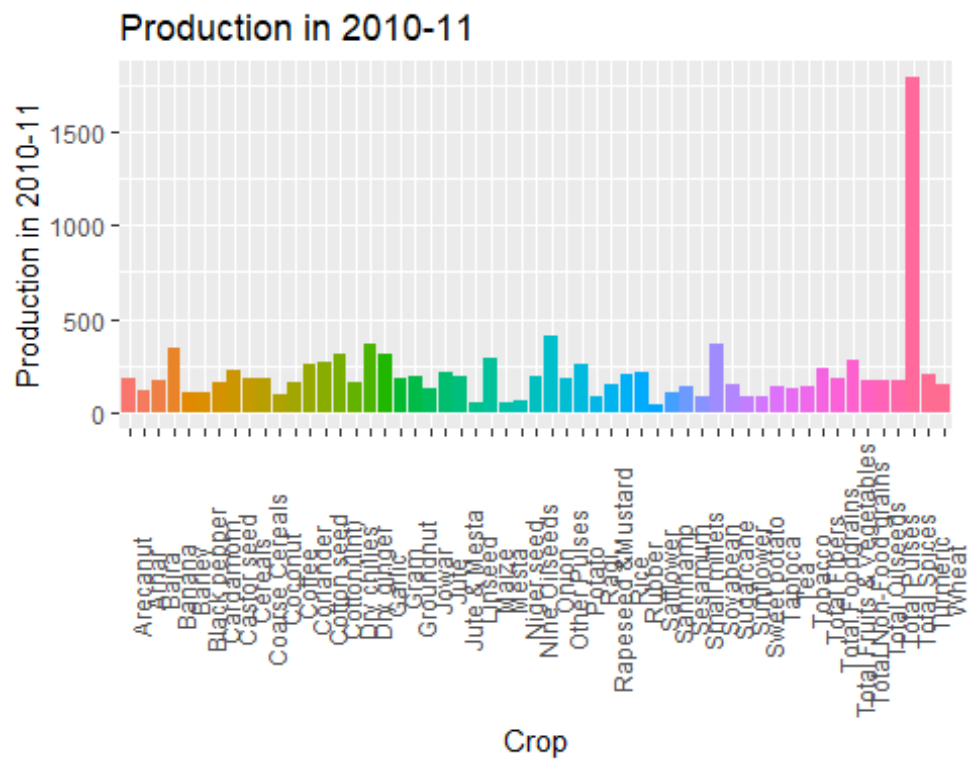
```



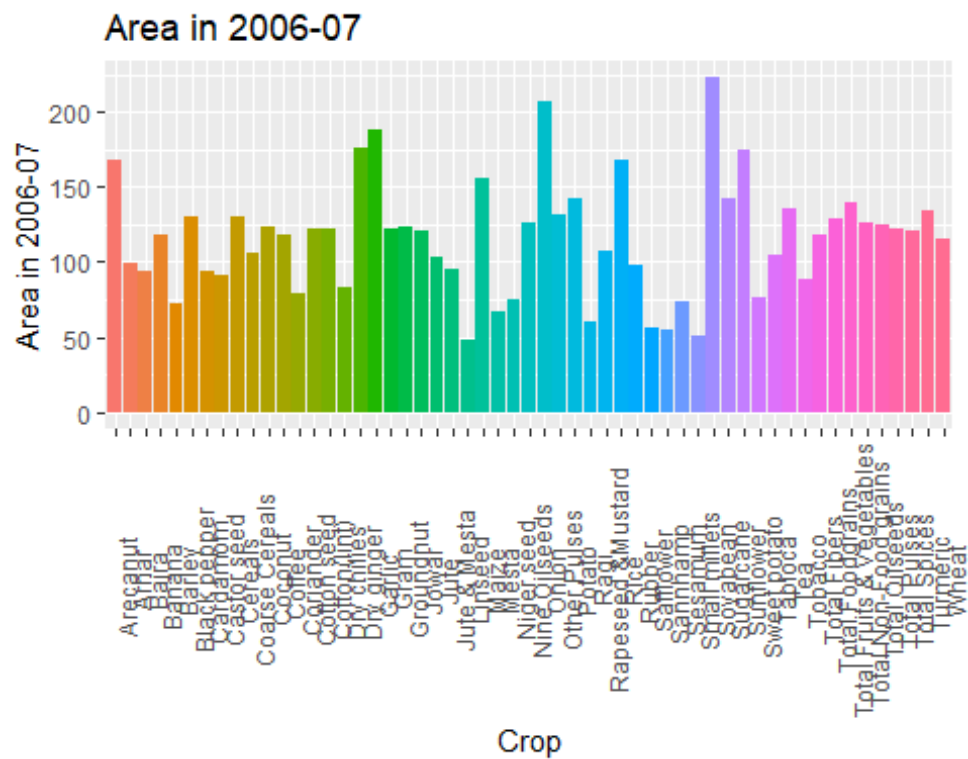
```
ggplot(crop_prod,aes(y=Production.2007.08,x=Crop,fill=Crop))
+geom_bar(stat="identity")+ylab("Production in 2007-08")
+xlab("Crop")+theme(axis.text.x=element_text(angle=90),legend.position="none")
+ggtitle("Production in 2007-08")
```



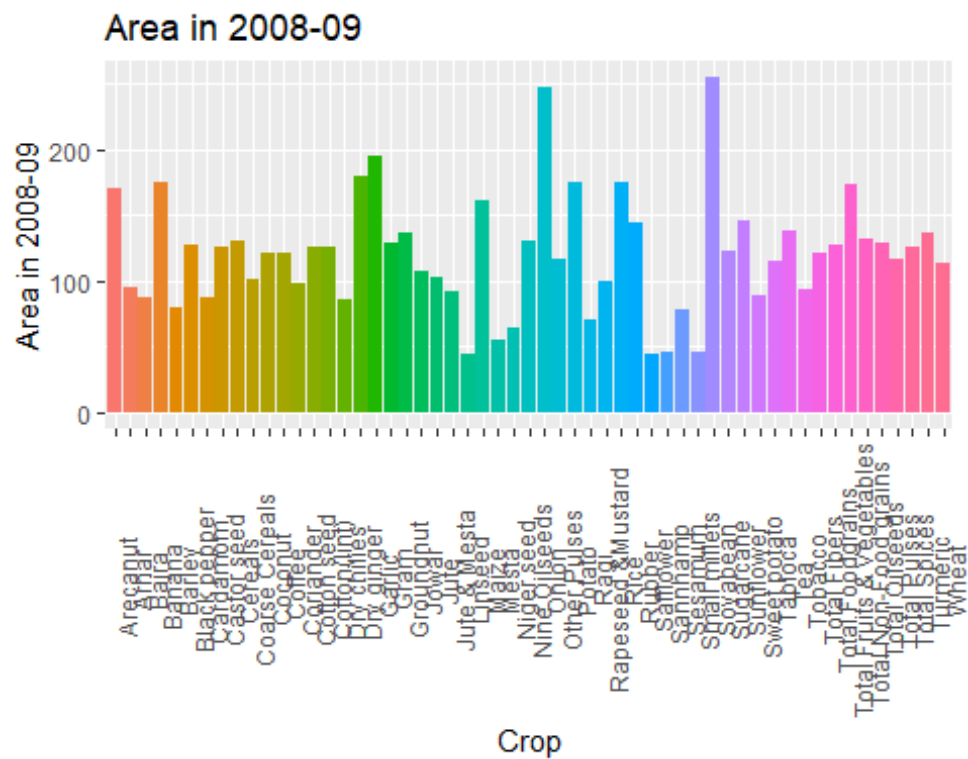
```
ggplot(crop_prod,aes(y=Production.2008.09,x=Crop,fill=Crop))
+geom_bar(stat="identity")+ylab("Production in 2008-09")
+xlab("Crop")+theme(axis.text.x=element_text(angle=90),legend.position="none")
+ggtitle("Production in 2008-09")
```

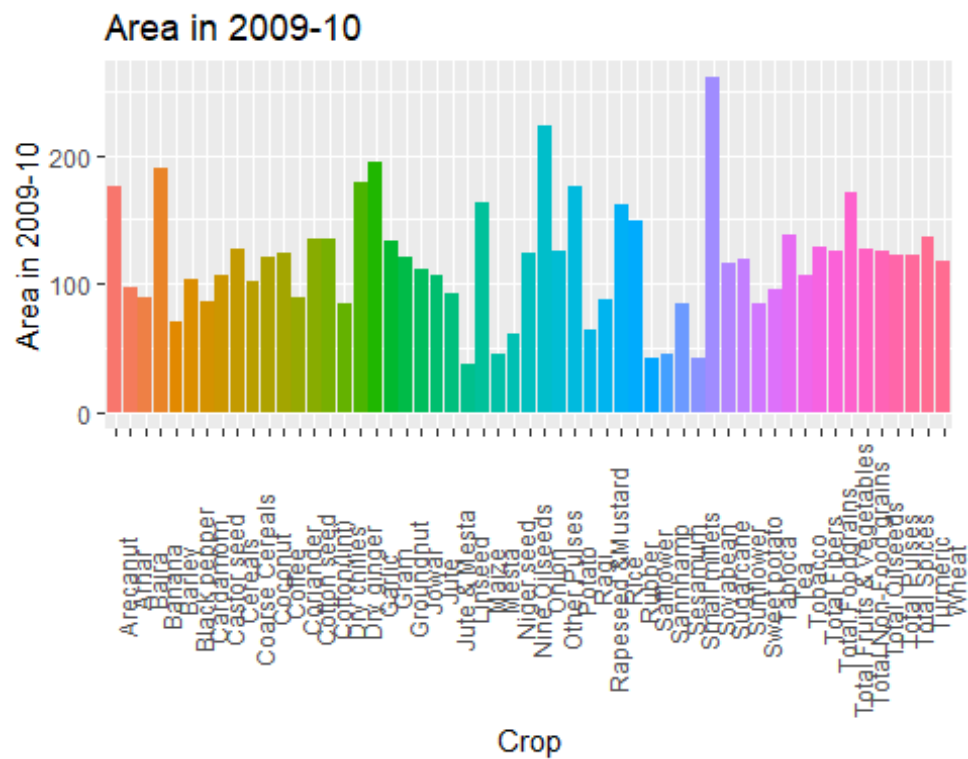
```
ggplot(crop_prod,aes(y=Area.2006.07,x=Crop ,fill=Crop
))+geom_bar(stat="identity")+ylab("Area in 2006-
07")+xlab("Crop")+theme(axis.text.x=element_text(angle=90),legend.position="n
one")+ggtitle("Area in 2006-07")
```



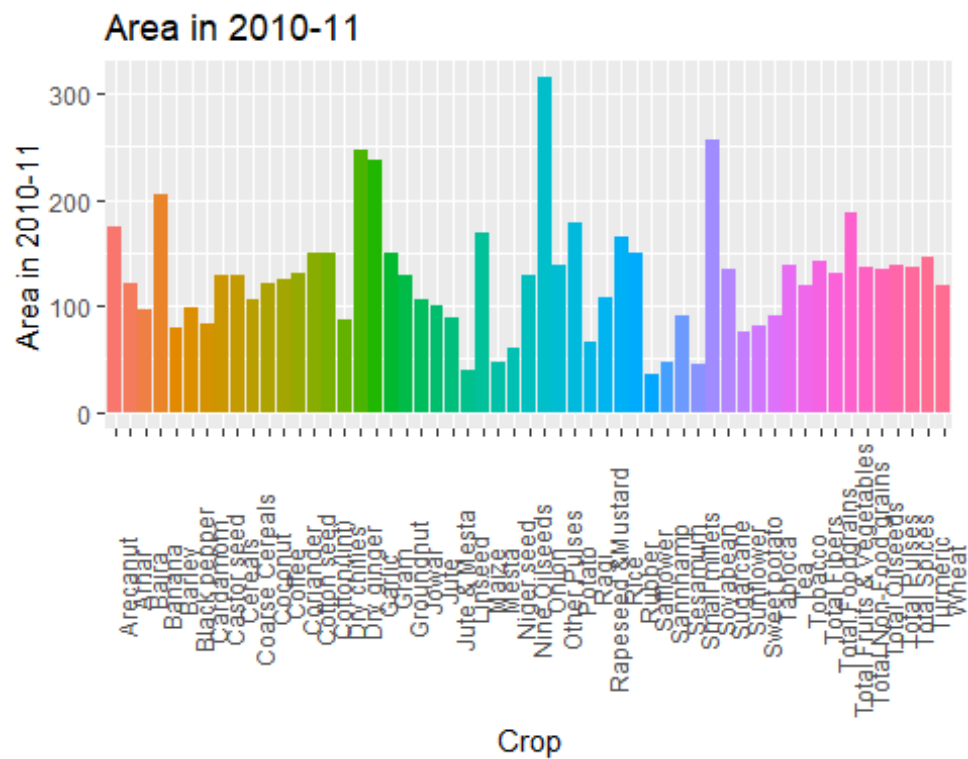
```
ggplot(crop_prod,aes(y=Area.2007.08,x=Crop ,fill=Crop
))+geom_bar(stat="identity")+ylab("Area in 2007-
08")+xlab("Crop")+theme(axis.text.x=element_text(angle=90),legend.position="n
one")+ggtitle("Area in 2007-08")
```

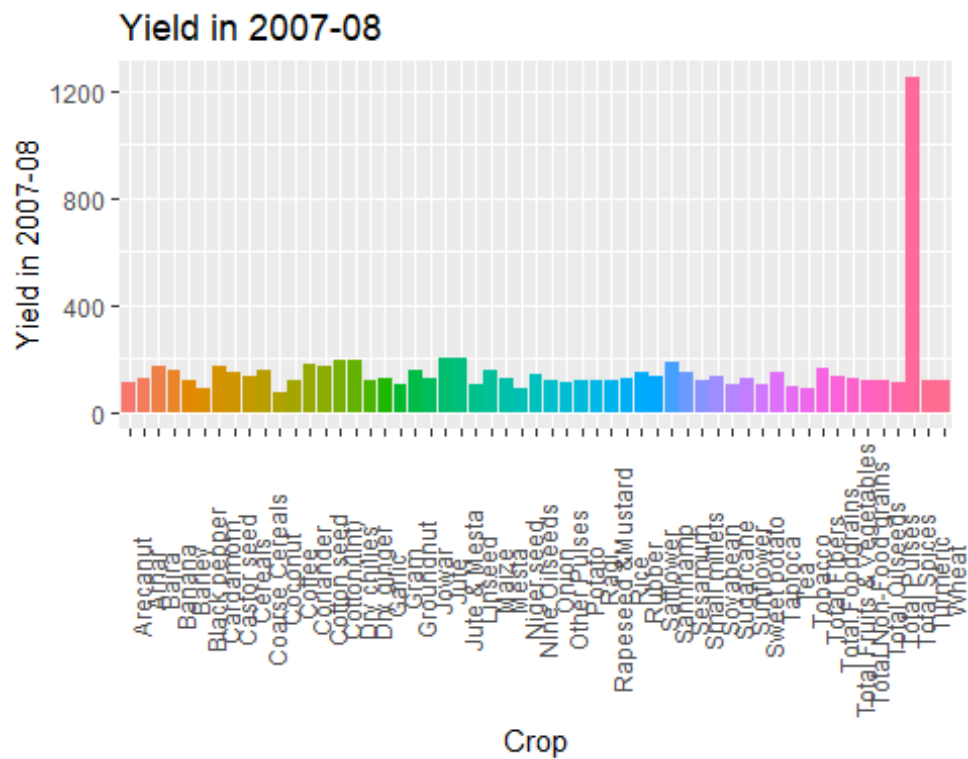
```
ggplot(crop_prod,aes(y=Area.2009.10,x=Crop ,fill=Crop
))+geom_bar(stat="identity")+ylab("Area in 2009-
10")+xlab("Crop")+theme(axis.text.x=element_text(angle=90),legend.position="n
one")+ggtitle("Area in 2009-10")
```



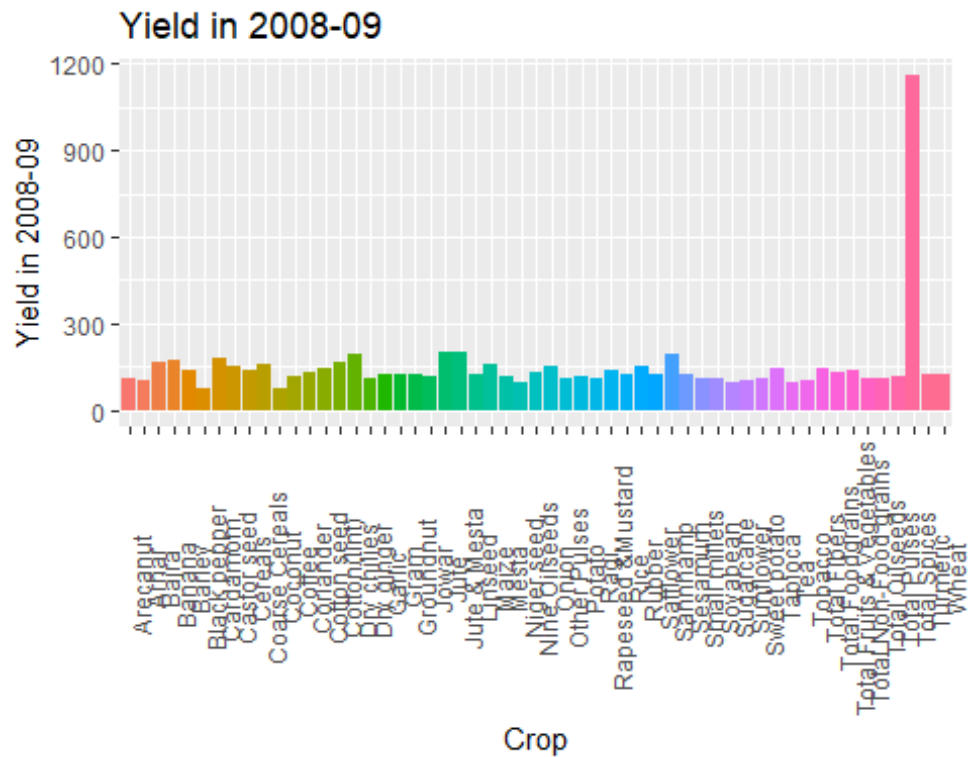
```
ggplot(crop_prod,aes(y=Area.2010.11,x=Crop ,fill=Crop
))+geom_bar(stat="identity")+ylab("Area in 2010-
11")+xlab("Crop")+theme(axis.text.x=element_text(angle=90),legend.position="n
one")+ggtitle("Area in 2010-11")
```



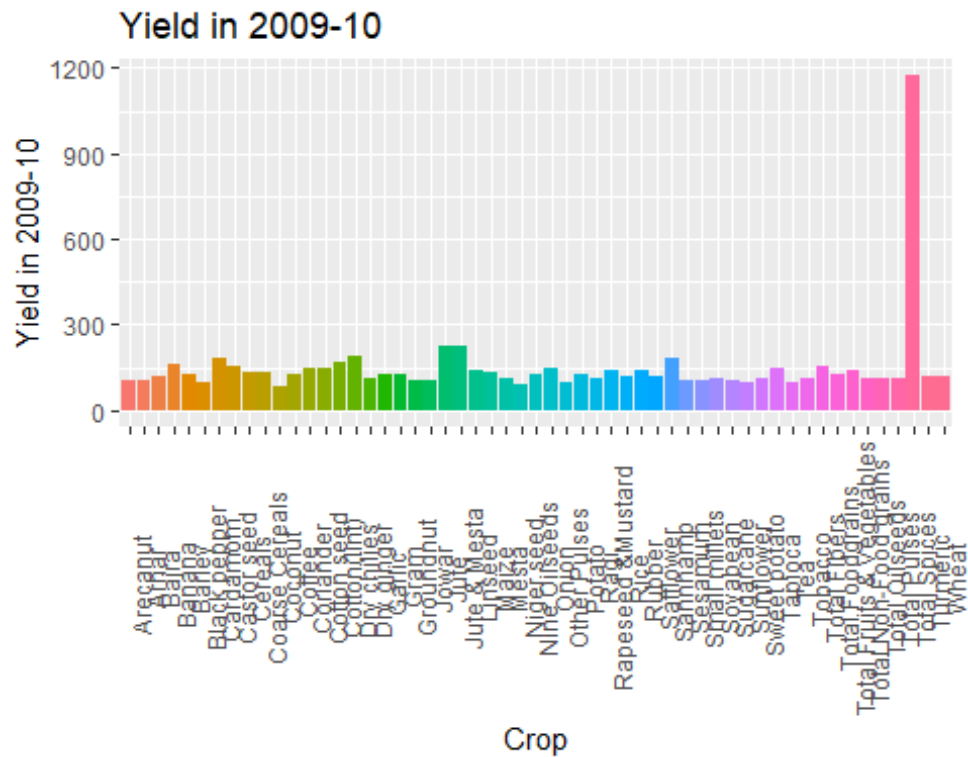
```
ggplot(crop_prod,aes(y=Yield.2006.07,x=Crop ,fill=Crop
))+geom_bar(stat="identity")+ylab("Yield in 2006-
07")+xlab("Crop")+theme(axis.text.x=element_text(angle=90),legend.position="n
one")+ggtitle("Yield in 2006-07")
```

```
ggplot(crop_prod,aes(y=Yield.2008.09,x=Crop ,fill=Crop
))+geom_bar(stat="identity")+ylab("Yield in 2008-
09")+xlab("Crop")+theme(axis.text.x=element_text(angle=90),legend.position="n
one")+ggtitle("Yield in 2008-09")
```

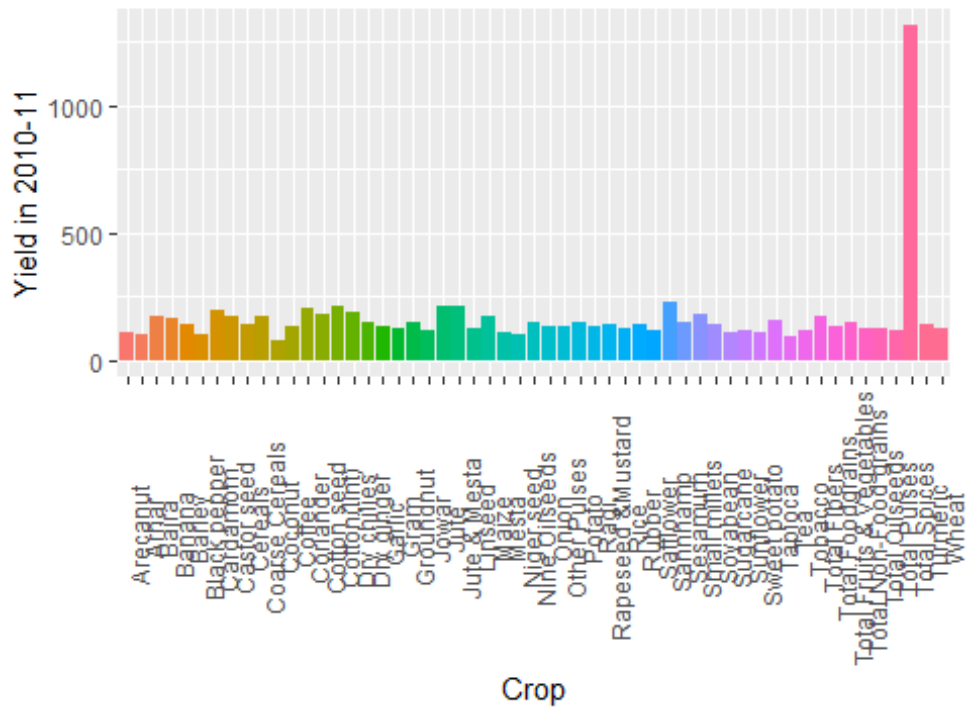


```
ggplot(crop_prod,aes(y=Yield.2009.10,x=Crop ,fill=Crop
))+geom_bar(stat="identity")+ylab("Yield in 2009-
10")+xlab("Crop")+theme(axis.text.x=element_text(angle=90),legend.position="n
one")+ggtitle("Yield in 2009-10")
```



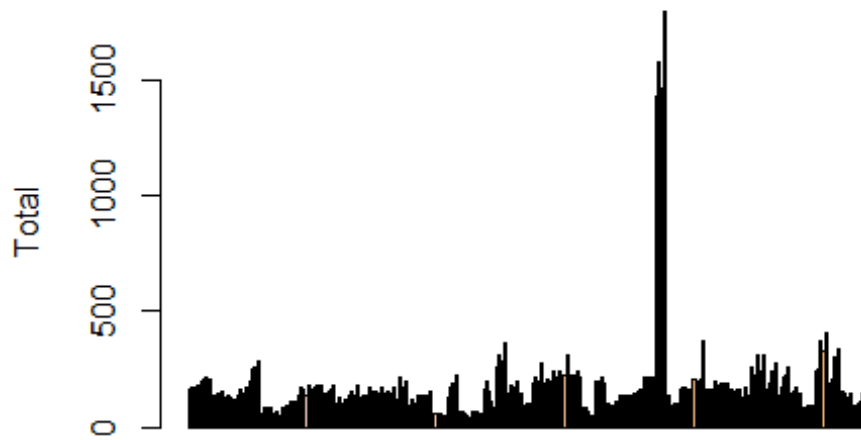
```
ggplot(crop_prod,aes(y=Yield.2010.11,x=Crop ,fill=Crop
))+geom_bar(stat="identity")+ylab("Yield in 2010-
11")+xlab("Crop")+theme(axis.text.x=element_text(angle=90),legend.position="n
one")+ggtitle("Yield in 2010-11")
```

Yield in 2010-11

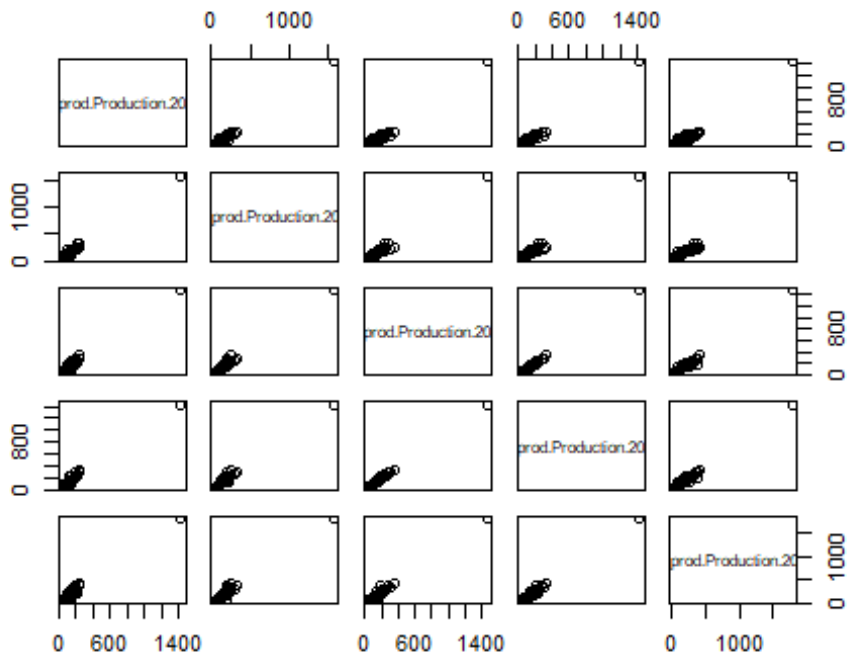


```
Production =  
data.frame(crop_prod$Production.2006.07,crop_prod$Production.2007.08,crop_pro  
d$Production.2008.09,crop_prod$Production.2009.10,crop_prod$Production.2010.1  
1)  
P = as.matrix(Production)  
par(mfrow=c(1,1))  
barplot(t(P), main="Multiple Bar Diagram of Production in five year for  
different crop", ylab="Total", beside=TRUE,  
col=terrain.colors(55))
```

Multiple Bar Diagram of Production in five year for different



```
options(repr.plot.width=15,repr.plot.height=10)
pairs(Production)
```



```

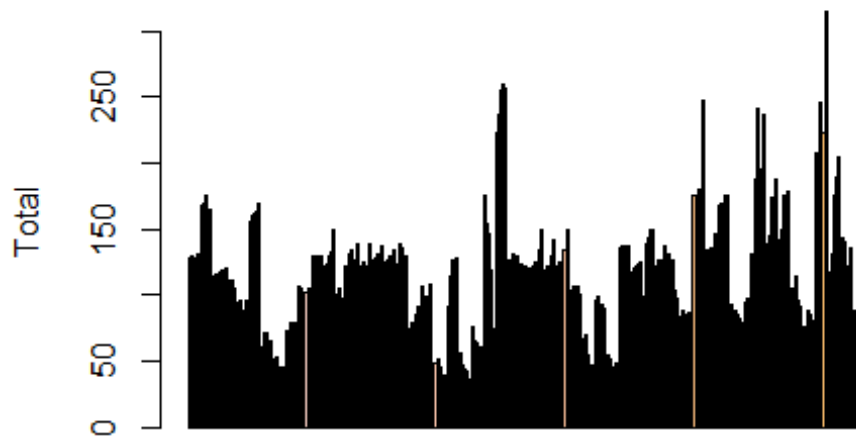
cor(Production)

##                                crop_prod.Production.2006.07
## crop_prod.Production.2006.07                1.0000000
## crop_prod.Production.2007.08                0.9915000
## crop_prod.Production.2008.09                0.9851425
## crop_prod.Production.2009.10                0.9862942
## crop_prod.Production.2010.11                0.9772547
##                                crop_prod.Production.2007.08
## crop_prod.Production.2006.07                0.9915000
## crop_prod.Production.2007.08                1.0000000
## crop_prod.Production.2008.09                0.9865887
## crop_prod.Production.2009.10                0.9839862
## crop_prod.Production.2010.11                0.9817908
##                                crop_prod.Production.2008.09
## crop_prod.Production.2006.07                0.9851425
## crop_prod.Production.2007.08                0.9865887
## crop_prod.Production.2008.09                1.0000000
## crop_prod.Production.2009.10                0.9960675
## crop_prod.Production.2010.11                0.9881706
##                                crop_prod.Production.2009.10
## crop_prod.Production.2006.07                0.9862942
## crop_prod.Production.2007.08                0.9839862
## crop_prod.Production.2008.09                0.9960675
## crop_prod.Production.2009.10                1.0000000
## crop_prod.Production.2010.11                0.9889130
##                                crop_prod.Production.2010.11
## crop_prod.Production.2006.07                0.9772547
## crop_prod.Production.2007.08                0.9817908
## crop_prod.Production.2008.09                0.9881706
## crop_prod.Production.2009.10                0.9889130
## crop_prod.Production.2010.11                1.0000000

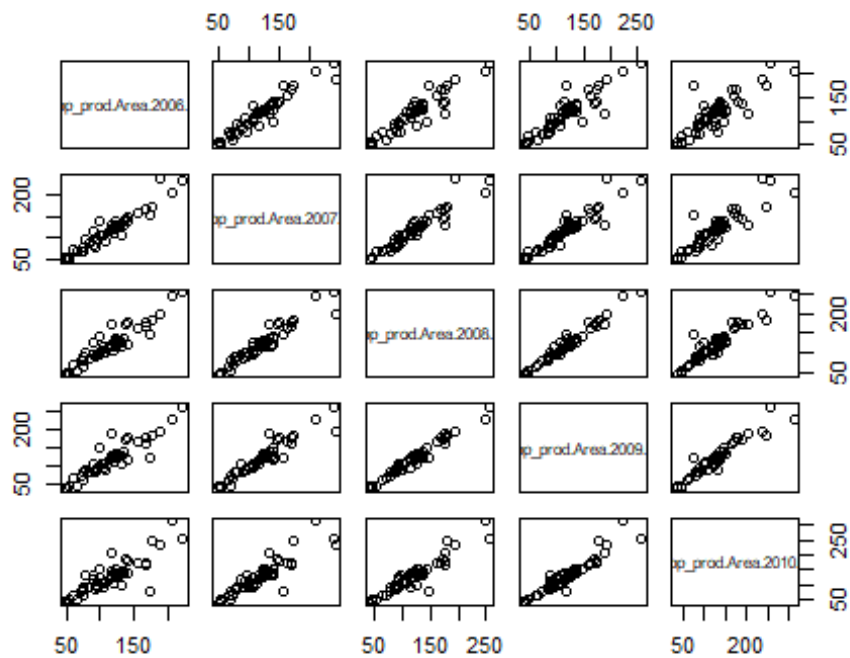
Area =
data.frame(crop_prod$Area.2006.07,crop_prod$Area.2007.08,crop_prod$Area.2008.
09,crop_prod$Area.2009.10,crop_prod$Area.2010.11)
A = as.matrix(Area)
par(mfrow=c(1,1))
barplot(t(A), main="Multiple Bar Diagram of Production in five year for
different crop", ylab="Total", beside=TRUE,
col=terrain.colors(55))

```

Multiple Bar Diagram of Production in five year for different



```
options(repr.plot.width=15,repr.plot.height=10)
pairs(Area)
```




```
cor(Area)
```

```
##                crop_prod.Area.2006.07 crop_prod.Area.2007.08
## crop_prod.Area.2006.07                1.0000000                0.9568259
## crop_prod.Area.2007.08                0.9568259                1.0000000
## crop_prod.Area.2008.09                0.9357001                0.9527411
## crop_prod.Area.2009.10                0.9084549                0.9396699
## crop_prod.Area.2010.11                0.8460643                0.8943950
##                crop_prod.Area.2008.09 crop_prod.Area.2009.10
## crop_prod.Area.2006.07                0.9357001                0.9084549
## crop_prod.Area.2007.08                0.9527411                0.9396699
## crop_prod.Area.2008.09                1.0000000                0.9799310
## crop_prod.Area.2009.10                0.9799310                1.0000000
## crop_prod.Area.2010.11                0.9352812                0.9486228
##                crop_prod.Area.2010.11
## crop_prod.Area.2006.07                0.8460643
## crop_prod.Area.2007.08                0.8943950
## crop_prod.Area.2008.09                0.9352812
## crop_prod.Area.2009.10                0.9486228
## crop_prod.Area.2010.11                1.0000000
```

```
Yield =
```

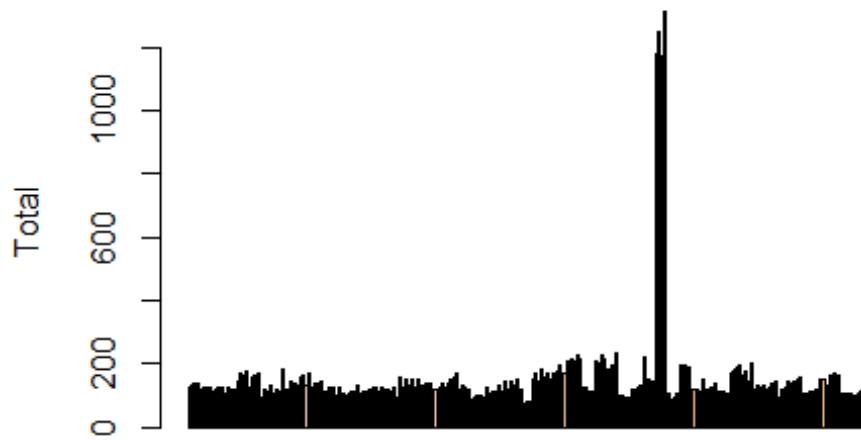
```
data.frame(crop_prod$Yield.2006.07,crop_prod$Yield.2007.08,crop_prod$Yield.2008.09,crop_prod$Yield.2009.10,crop_prod$Yield.2010.11)
```

```
Y = as.matrix(Yield)
```

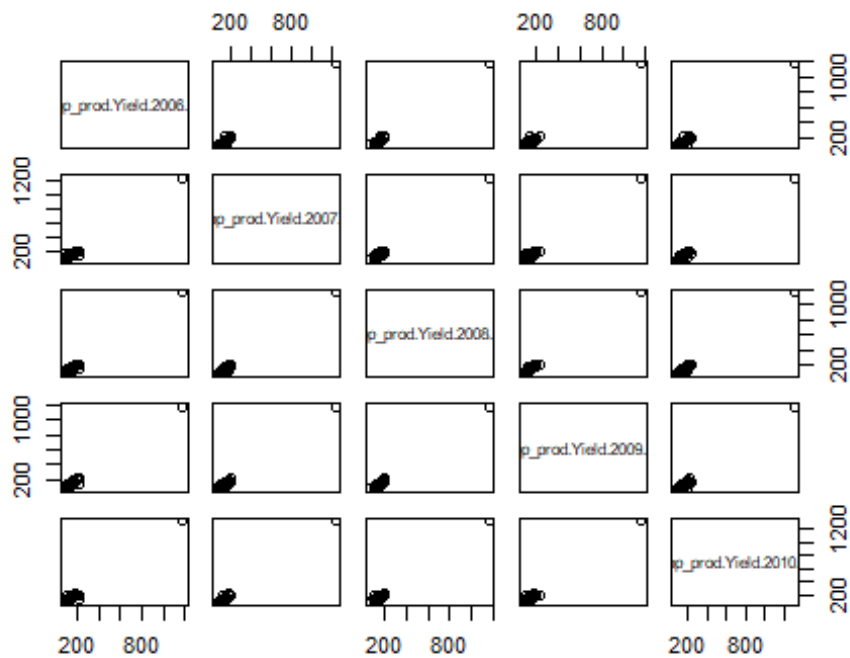
```
par(mfrow=c(1,1))
```

```
barplot(t(Y), main="Multiple Bar Diagram of Production in five year for  
different crop", ylab="Total", beside=TRUE,  
col=terrain.colors(55))
```

Multiple Bar Diagram of Production in five year for different



```
options(repr.plot.width=15,repr.plot.height=10)
pairs(Yield)
```



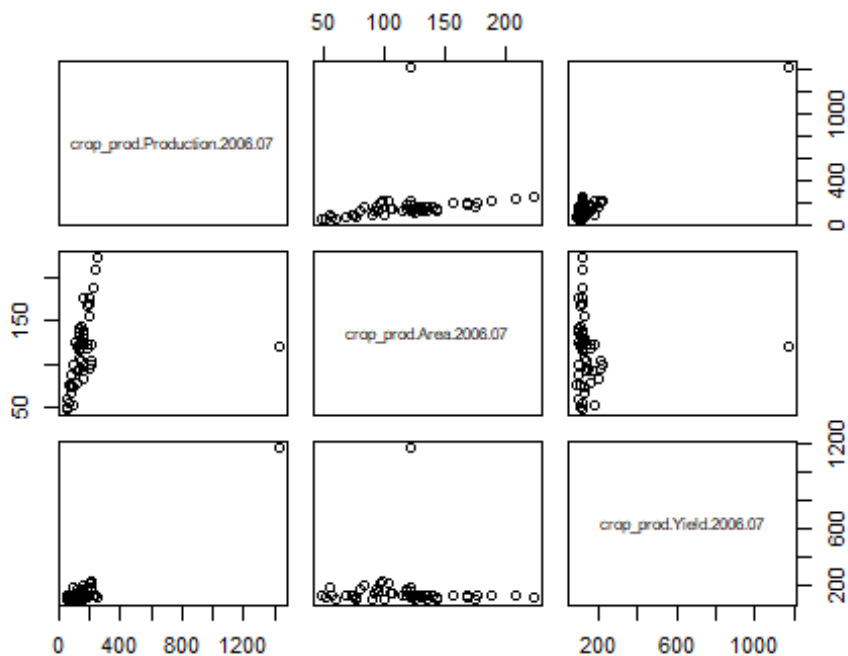
```
cor(Yield)
```

```
##               crop_prod.Yield.2006.07 crop_prod.Yield.2007.08
## crop_prod.Yield.2006.07             1.0000000             0.9922516
## crop_prod.Yield.2007.08             0.9922516             1.0000000
## crop_prod.Yield.2008.09             0.9926255             0.9953347
## crop_prod.Yield.2009.10             0.9935210             0.9924561
## crop_prod.Yield.2010.11             0.9888665             0.9957814
##               crop_prod.Yield.2008.09 crop_prod.Yield.2009.10
## crop_prod.Yield.2006.07             0.9926255             0.9935210
## crop_prod.Yield.2007.08             0.9953347             0.9924561
## crop_prod.Yield.2008.09             1.0000000             0.9962626
## crop_prod.Yield.2009.10             0.9962626             1.0000000
## crop_prod.Yield.2010.11             0.9939387             0.9915548
##               crop_prod.Yield.2010.11
## crop_prod.Yield.2006.07             0.9888665
## crop_prod.Yield.2007.08             0.9957814
## crop_prod.Yield.2008.09             0.9939387
## crop_prod.Yield.2009.10             0.9915548
## crop_prod.Yield.2010.11             1.0000000
```

```
Year_2006_07 =
```

```
data.frame(crop_prod$Production.2006.07,crop_prod$Area.2006.07,crop_prod$Yield.2006.07)
```

```
pairs(Year_2006_07)
```

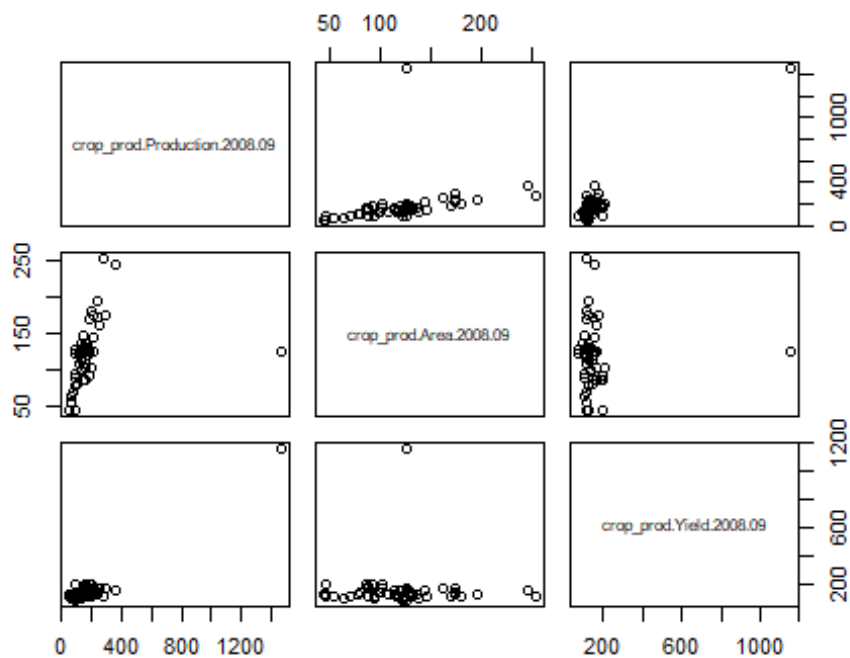



```

crop_prod.Yield.2007.08
## crop_prod.Production.2007.08          0.272526752
0.962377457
## crop_prod.Area.2007.08              1.000000000
0.006301568
## crop_prod.Yield.2007.08            0.006301568
1.000000000

Year_2008_09 =
data.frame(crop_prod$Production.2008.09,crop_prod$Area.2008.09,crop_prod$Yield.2008.09)
pairs(Year_2008_09)

```



```

cor(Year_2008_09)

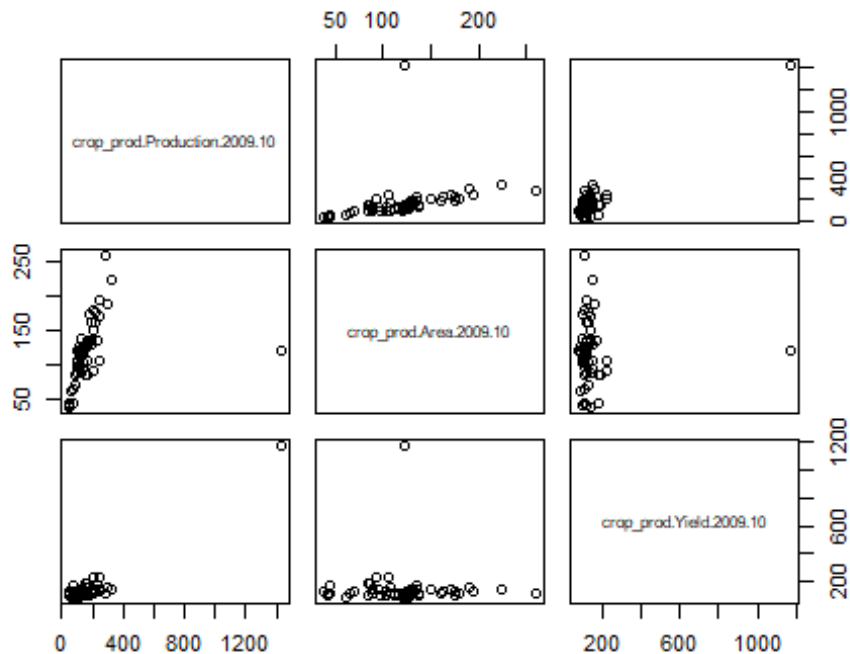
##                                crop_prod.Production.2008.09
## crop_prod.Production.2008.09          1.0000000
## crop_prod.Area.2008.09              0.3093606
## crop_prod.Yield.2008.09            0.9479256
##                                crop_prod.Area.2008.09
crop_prod.Yield.2008.09
## crop_prod.Production.2008.09          0.309360647
0.947925551
## crop_prod.Area.2008.09              1.000000000      -
0.002569112
## crop_prod.Yield.2008.09            -0.002569112
1.000000000

```

```

Year_2009_10 =
data.frame(crop_prod$Production.2009.10,crop_prod$Area.2009.10,crop_prod$Yield.2009.10)
pairs(Year_2009_10)

```



```

cor(Year_2009_10)

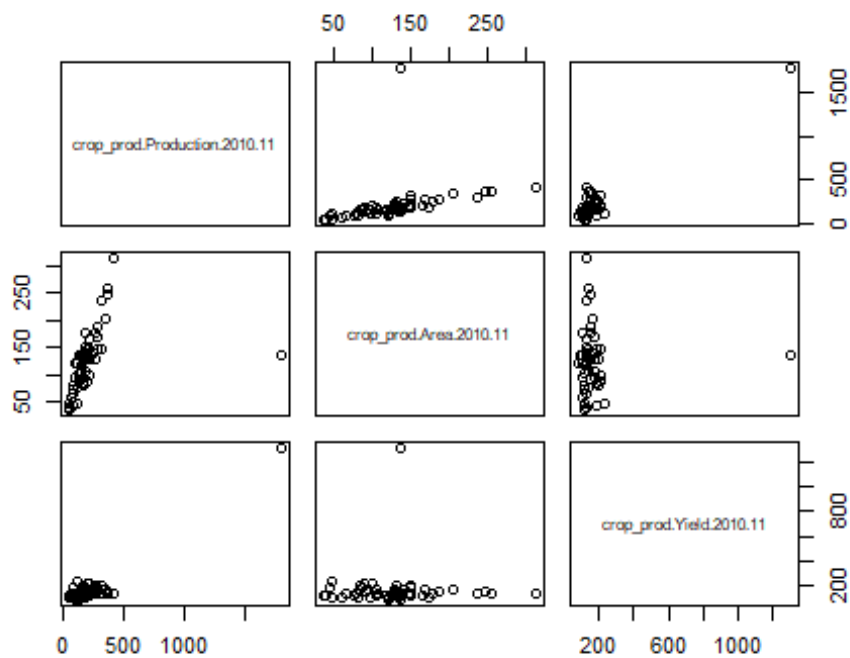
##                                crop_prod.Production.2009.10
## crop_prod.Production.2009.10                        1.0000000
## crop_prod.Area.2009.10                               0.3123837
## crop_prod.Yield.2009.10                              0.9472567
##                                crop_prod.Area.2009.10
crop_prod.Yield.2009.10
## crop_prod.Production.2009.10                        0.312383744
0.947256744
## crop_prod.Area.2009.10                               1.000000000
0.001207165
## crop_prod.Yield.2009.10                             -0.001207165
1.000000000

```

```

Year_2010_11 =
data.frame(crop_prod$Production.2010.11,crop_prod$Area.2010.11,crop_prod$Yield.2010.11)
pairs(Year_2010_11)

```



```
cor(Year_2010_11)
```

```
##                                crop_prod.Production.2010.11
## crop_prod.Production.2010.11                1.0000000
## crop_prod.Area.2010.11                      0.3434309
## crop_prod.Yield.2010.11                     0.9394721
##                                crop_prod.Area.2010.11
crop_prod.Yield.2010.11
## crop_prod.Production.2010.11                0.343430872
0.939472134
## crop_prod.Area.2010.11                      1.000000000
0.009508415
## crop_prod.Yield.2010.11                     0.009508415
1.000000000
```

Conclusion: