CLUA\_AP DATA ANALYSIS

Eva Wanjiru

2022-09-02

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.1.3

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.6 v dplyr 1.0.8  
## v tidyr 1.2.0 v stringr 1.4.0  
## v readr 2.1.2 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(summarytools)

##   
## Attaching package: 'summarytools'

## The following object is masked from 'package:tibble':  
##   
## view

library(ggplot2)  
library(scales)

##   
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':  
##   
## discard

## The following object is masked from 'package:readr':  
##   
## col\_factor

library(ggeasy)

## Warning: package 'ggeasy' was built under R version 4.1.3

library(tidytext)

## Warning: package 'tidytext' was built under R version 4.1.3

#load the two datasets to be used for analysis  
CLUA\_reg<-read.csv("CLUA\_AP\_farmer\_registration\_form.csv")  
CLUA\_farm<-read.csv("CLUA\_AP\_farmer\_registration\_form-farm\_details.csv")  
  
##rename column key  
CLUA\_reg <- CLUA\_reg %>%  
 rename\_at("KEY", ~"PARENT\_KEY")  
  
#merge the two datasets using the parent key  
CLUA\_df<-merge(CLUA\_reg,CLUA\_farm,by="PARENT\_KEY")

##Subset the dataset to be used  
df\_sub<-CLUA\_df[,1:63]  
  
#check the names  
names(df\_sub)

## [1] "PARENT\_KEY"   
## [2] "SubmissionDate"   
## [3] "name\_date.start"   
## [4] "name\_date.today"   
## [5] "name\_date.deviceid"   
## [6] "name\_date.DF\_name"   
## [7] "name\_date.date"   
## [8] "intro\_1"   
## [9] "consent"   
## [10] "District"   
## [11] "village"   
## [12] "gps.Latitude"   
## [13] "gps.Longitude"   
## [14] "gps.Altitude"   
## [15] "gps.Accuracy"   
## [16] "farmer\_details.hh\_id"   
## [17] "farmer\_details.farmer\_name"   
## [18] "farmer\_details.farmer\_gender"   
## [19] "farmer\_details.farmer\_age"   
## [20] "farmer\_details.marital\_status"   
## [21] "farmer\_details.hh\_head"   
## [22] "farmer\_details.relationship"   
## [23] "farmer\_details.other\_relationship"   
## [24] "farmer\_details.contact\_consent"   
## [25] "farmer\_details.contact\_number"   
## [26] "hh\_income.main\_income\_source"   
## [27] "hh\_income.main\_income\_other"   
## [28] "hh\_income.agricultural\_produce"   
## [29] "hh\_income.secondary\_income\_source"   
## [30] "hh\_income.secondary\_income\_other"   
## [31] "group\_house.education\_level"   
## [32] "group\_house.other\_edu\_head"   
## [33] "group\_house.hh\_pop"   
## [34] "group\_house.num\_child"   
## [35] "group\_house.active\_farm"   
## [36] "hh\_farms"   
## [37] "num\_farms"   
## [38] "SET.OF.farm\_details.x"   
## [39] "group\_tree.trees\_onfarm"   
## [40] "group\_tree.tree\_species"   
## [41] "group\_tree.tree\_species\_other"   
## [42] "group\_tree.unavailable\_tree"   
## [43] "group\_tree.unavailable\_tree\_type"   
## [44] "group\_tree.tree\_products"   
## [45] "group\_tree.tree\_products\_other"   
## [46] "group\_tree.tree\_product\_use"   
## [47] "group\_tree.tree\_use"   
## [48] "group\_livestock.livestock\_keeping"   
## [49] "group\_livestock.livestock\_type"   
## [50] "group\_livestock.livestock\_type\_other"   
## [51] "group\_livestock.livestock\_products"   
## [52] "group\_livestock.livestock\_products\_other"   
## [53] "group\_livestock.livestock\_products\_use"   
## [54] "group\_distance.distance\_road"   
## [55] "group\_distance.distance\_market"   
## [56] "group\_distance.distance\_water"   
## [57] "group\_climatechange.climate\_change"   
## [58] "group\_climatechange.cc\_aspects"   
## [59] "group\_climatechange.cc\_experience"   
## [60] "group\_climatechange.cchange\_experience"   
## [61] "group\_climatechange.cchange\_experience.other"  
## [62] "group\_climatechange.cc\_impact"   
## [63] "group\_climatechange.cc\_impact\_other"

#look at the first 6 rows  
head(df\_sub)

## PARENT\_KEY SubmissionDate  
## 1 uuid:00ee1a97-86ea-4f77-a7dc-4bf12628349b Jul 18, 2022 1:56:50 AM  
## 2 uuid:00ee1a97-86ea-4f77-a7dc-4bf12628349b Jul 18, 2022 1:56:50 AM  
## 3 uuid:00ee1a97-86ea-4f77-a7dc-4bf12628349b Jul 18, 2022 1:56:50 AM  
## 4 uuid:01158a38-aac0-4969-ad0a-9c40ace38dbe Aug 7, 2022 3:45:58 AM  
## 5 uuid:024ca5a8-4b20-4cae-995a-277c1ffc2355 Aug 8, 2022 2:49:45 PM  
## 6 uuid:02d4d90c-6986-46d9-8589-12ff6042c3be Jan 12, 2022 11:11:14 AM  
## name\_date.start name\_date.today name\_date.deviceid  
## 1 Jul 15, 2022 6:14:16 AM Jul 15, 2022 collect:jcLH1RWOC8cLbOLl  
## 2 Jul 15, 2022 6:14:16 AM Jul 15, 2022 collect:jcLH1RWOC8cLbOLl  
## 3 Jul 15, 2022 6:14:16 AM Jul 15, 2022 collect:jcLH1RWOC8cLbOLl  
## 4 Jul 3, 2022 9:13:57 AM Jul 3, 2022 collect:Cs3jC2Red4uCu4EP  
## 5 Aug 6, 2022 12:40:48 PM Aug 6, 2022 collect:Cs3jC2Red4uCu4EP  
## 6 Jan 12, 2022 11:07:08 AM Jan 12, 2022 collect:5Q01BMwxh9CVrLdJ  
## name\_date.DF\_name name\_date.date intro\_1 consent District village  
## 1 Rini Caroline Polumuri Jul 15, 2022 NA 1 3 Kapavaram  
## 2 Rini Caroline Polumuri Jul 15, 2022 NA 1 3 Kapavaram  
## 3 Rini Caroline Polumuri Jul 15, 2022 NA 1 3 Kapavaram  
## 4 Godehalu Sandhya Jul 3, 2022 NA 1 1 Korakodu  
## 5 Godehalu Sandhya Aug 6, 2022 NA 1 1 Korrakodu  
## 6 D.Tejaswini Jan 12, 2022 NA 1 3 Aratlakatta  
## gps.Latitude gps.Longitude gps.Altitude gps.Accuracy farmer\_details.hh\_id  
## 1 16.54381 81.69133 35.6 3.566 WG0064  
## 2 16.54381 81.69133 35.6 3.566 WG0064  
## 3 16.54381 81.69133 35.6 3.566 WG0064  
## 4 14.79227 77.31381 435.8 9.000 ATP0111  
## 5 14.77342 77.33145 460.6 23.346 ATP163  
## 6 16.56798 81.66512 0.0 1674.000 WG0003  
## farmer\_details.farmer\_name farmer\_details.farmer\_gender  
## 1 Toleti Eswari 1  
## 2 Toleti Eswari 1  
## 3 Toleti Eswari 1  
## 4 Padmavathi 1  
## 5 E.Vanajakshi 1  
## 6 Digamarthi abbulu 2  
## farmer\_details.farmer\_age farmer\_details.marital\_status  
## 1 47 2  
## 2 47 2  
## 3 47 2  
## 4 42 2  
## 5 32 2  
## 6 55 2  
## farmer\_details.hh\_head farmer\_details.relationship  
## 1 1 NA  
## 2 1 NA  
## 3 1 NA  
## 4 0 1  
## 5 0 1  
## 6 1 NA  
## farmer\_details.other\_relationship farmer\_details.contact\_consent  
## 1 NA 1  
## 2 NA 1  
## 3 NA 1  
## 4 NA 1  
## 5 NA 1  
## 6 NA 1  
## farmer\_details.contact\_number hh\_income.main\_income\_source  
## 1 8121435573 1  
## 2 8121435573 1  
## 3 8121435573 1  
## 4 8688200560 1  
## 5 8688428316 1  
## 6 9395539298 1  
## hh\_income.main\_income\_other hh\_income.agricultural\_produce  
## 1 cereals  
## 2 cereals  
## 3 cereals  
## 4 vegetables  
## 5 vegetables  
## 6 cereals  
## hh\_income.secondary\_income\_source hh\_income.secondary\_income\_other  
## 1 1   
## 2 1   
## 3 1   
## 4 1   
## 5 1   
## 6 1   
## group\_house.education\_level group\_house.other\_edu\_head group\_house.hh\_pop  
## 1 0 NA 3  
## 2 0 NA 3  
## 3 0 NA 3  
## 4 3 NA 3  
## 5 1 NA 4  
## 6 0 NA 6  
## group\_house.num\_child group\_house.active\_farm hh\_farms num\_farms  
## 1 0 3 NA 2  
## 2 0 3 NA 2  
## 3 0 3 NA 2  
## 4 1 2 NA 1  
## 5 2 2 NA 1  
## 6 1 3 NA 1  
## SET.OF.farm\_details.x  
## 1 uuid:00ee1a97-86ea-4f77-a7dc-4bf12628349b/farm\_details  
## 2 uuid:00ee1a97-86ea-4f77-a7dc-4bf12628349b/farm\_details  
## 3 uuid:00ee1a97-86ea-4f77-a7dc-4bf12628349b/farm\_details  
## 4 uuid:01158a38-aac0-4969-ad0a-9c40ace38dbe/farm\_details  
## 5 uuid:024ca5a8-4b20-4cae-995a-277c1ffc2355/farm\_details  
## 6 uuid:02d4d90c-6986-46d9-8589-12ff6042c3be/farm\_details  
## group\_tree.trees\_onfarm group\_tree.tree\_species group\_tree.tree\_species\_other  
## 1 1 10   
## 2 1 10   
## 3 1 10   
## 4 1 10 13 Pomegranate  
## 5 1 10   
## 6 0   
## group\_tree.unavailable\_tree group\_tree.unavailable\_tree\_type  
## 1 0   
## 2 0   
## 3 0   
## 4 0   
## 5 0   
## 6 NA   
## group\_tree.tree\_products group\_tree.tree\_products\_other  
## 1 3 2 NA  
## 2 3 2 NA  
## 3 3 2 NA  
## 4 3 NA  
## 5 3 NA  
## 6 NA  
## group\_tree.tree\_product\_use group\_tree.tree\_use  
## 1 3   
## 2 3   
## 3 3   
## 4 1 <NA>  
## 5 1 <NA>  
## 6 NA   
## group\_livestock.livestock\_keeping group\_livestock.livestock\_type  
## 1 1 6  
## 2 1 6  
## 3 1 6  
## 4 0   
## 5 0   
## 6 1 2  
## group\_livestock.livestock\_type\_other group\_livestock.livestock\_products  
## 1 NA 1 3  
## 2 NA 1 3  
## 3 NA 1 3  
## 4 NA   
## 5 NA   
## 6 NA 2  
## group\_livestock.livestock\_products\_other  
## 1   
## 2   
## 3   
## 4   
## 5   
## 6   
## group\_livestock.livestock\_products\_use group\_distance.distance\_road  
## 1 3 0.0  
## 2 3 0.0  
## 3 3 0.0  
## 4 NA 3.0  
## 5 NA 1.5  
## 6 1 1.0  
## group\_distance.distance\_market group\_distance.distance\_water  
## 1 6.0 0.0  
## 2 6.0 0.0  
## 3 6.0 0.0  
## 4 40.0 0.0  
## 5 85.0 0.0  
## 6 1.4 0.1  
## group\_climatechange.climate\_change group\_climatechange.cc\_aspects  
## 1 1 2 3 4 5  
## 2 1 2 3 4 5  
## 3 1 2 3 4 5  
## 4 1 2 3  
## 5 1 2 3  
## 6 1 1 5  
## group\_climatechange.cc\_experience group\_climatechange.cchange\_experience  
## 1 1 3 1  
## 2 1 3 1  
## 3 1 3 1  
## 4 1 2 3 4  
## 5 1 3 4  
## 6 1 2 3  
## group\_climatechange.cchange\_experience.other group\_climatechange.cc\_impact  
## 1 1 5  
## 2 1 5  
## 3 1 5  
## 4 1 2 3 4 5  
## 5 3 4 5 1  
## 6 1 3 5  
## group\_climatechange.cc\_impact\_other  
## 1   
## 2   
## 3   
## 4   
## 5   
## 6

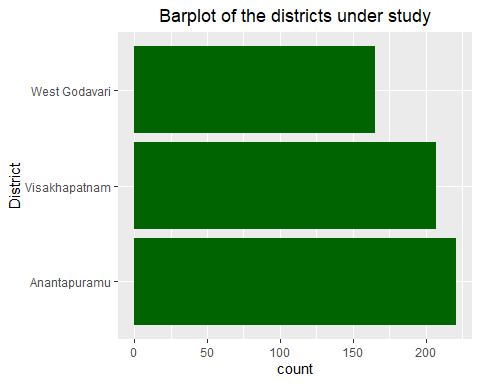
##check for missing values  
colSums(is.na(df\_sub))

## PARENT\_KEY   
## 0   
## SubmissionDate   
## 0   
## name\_date.start   
## 0   
## name\_date.today   
## 0   
## name\_date.deviceid   
## 0   
## name\_date.DF\_name   
## 0   
## name\_date.date   
## 0   
## intro\_1   
## 593   
## consent   
## 0   
## District   
## 0   
## village   
## 0   
## gps.Latitude   
## 0   
## gps.Longitude   
## 0   
## gps.Altitude   
## 0   
## gps.Accuracy   
## 0   
## farmer\_details.hh\_id   
## 0   
## farmer\_details.farmer\_name   
## 0   
## farmer\_details.farmer\_gender   
## 0   
## farmer\_details.farmer\_age   
## 0   
## farmer\_details.marital\_status   
## 0   
## farmer\_details.hh\_head   
## 0   
## farmer\_details.relationship   
## 445   
## farmer\_details.other\_relationship   
## 593   
## farmer\_details.contact\_consent   
## 0   
## farmer\_details.contact\_number   
## 1   
## hh\_income.main\_income\_source   
## 0   
## hh\_income.main\_income\_other   
## 0   
## hh\_income.agricultural\_produce   
## 0   
## hh\_income.secondary\_income\_source   
## 0   
## hh\_income.secondary\_income\_other   
## 0   
## group\_house.education\_level   
## 1   
## group\_house.other\_edu\_head   
## 593   
## group\_house.hh\_pop   
## 0   
## group\_house.num\_child   
## 0   
## group\_house.active\_farm   
## 0   
## hh\_farms   
## 593   
## num\_farms   
## 34   
## SET.OF.farm\_details.x   
## 0   
## group\_tree.trees\_onfarm   
## 2   
## group\_tree.tree\_species   
## 0   
## group\_tree.tree\_species\_other   
## 0   
## group\_tree.unavailable\_tree   
## 221   
## group\_tree.unavailable\_tree\_type   
## 0   
## group\_tree.tree\_products   
## 0   
## group\_tree.tree\_products\_other   
## 593   
## group\_tree.tree\_product\_use   
## 221   
## group\_tree.tree\_use   
## 107   
## group\_livestock.livestock\_keeping   
## 0   
## group\_livestock.livestock\_type   
## 0   
## group\_livestock.livestock\_type\_other   
## 593   
## group\_livestock.livestock\_products   
## 0   
## group\_livestock.livestock\_products\_other   
## 0   
## group\_livestock.livestock\_products\_use   
## 150   
## group\_distance.distance\_road   
## 0   
## group\_distance.distance\_market   
## 0   
## group\_distance.distance\_water   
## 0   
## group\_climatechange.climate\_change   
## 0   
## group\_climatechange.cc\_aspects   
## 0   
## group\_climatechange.cc\_experience   
## 105   
## group\_climatechange.cchange\_experience   
## 0   
## group\_climatechange.cchange\_experience.other   
## 171   
## group\_climatechange.cc\_impact   
## 0   
## group\_climatechange.cc\_impact\_other   
## 0

freq(df\_sub$District)

## Frequencies   
## df\_sub$District   
## Type: Integer   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ----------- ------ --------- -------------- --------- --------------  
## 1 221 37.27 37.27 37.27 37.27  
## 2 207 34.91 72.18 34.91 72.18  
## 3 165 27.82 100.00 27.82 100.00  
## <NA> 0 0.00 100.00  
## Total 593 100.00 100.00 100.00 100.00

##rename the Districts  
df\_sub$District[df\_sub$District=="1"]<-"Anantapuramu"  
df\_sub$District[df\_sub$District=="2"]<-"Visakhapatnam"  
df\_sub$District[df\_sub$District=="3"]<- "West Godavari"  
  
#Barplot of Districts  
df\_sub %>%   
 ggplot(aes(District))+  
 geom\_bar(fill="darkgreen")+  
 ggtitle("Barplot of the districts under study")+  
 ggeasy::easy\_center\_title()+  
 coord\_flip()

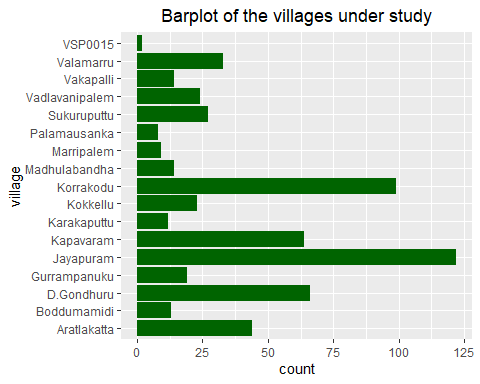


Anantapuramu is the district with the highest count of participants accounting for 46.33% while Visakhapatnam has the least count of participants and accounts for 24.32%.

freq(df\_sub$village)

## Frequencies   
## df\_sub$village   
## Type: Character   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## -------------------- ------ --------- -------------- --------- --------------  
## Aratlakatta 44 7.42 7.42 7.42 7.42  
## Boddumamidi 13 2.19 9.61 2.19 9.61  
## D Gondhuru 55 9.27 18.89 9.27 18.89  
## D.Gondhru 2 0.34 19.22 0.34 19.22  
## D.Gondhuru 9 1.52 20.74 1.52 20.74  
## Gurrampanuku 19 3.20 23.95 3.20 23.95  
## Jayapuram 122 20.57 44.52 20.57 44.52  
## Kapavaram 64 10.79 55.31 10.79 55.31  
## Karakaputtu 12 2.02 57.34 2.02 57.34  
## Kokkellu 23 3.88 61.21 3.88 61.21  
## Korakodu 38 6.41 67.62 6.41 67.62  
## Korrakodu 61 10.29 77.91 10.29 77.91  
## Madhulabandha 14 2.36 80.27 2.36 80.27  
## Marripalem 9 1.52 81.79 1.52 81.79  
## Palamausanka 8 1.35 83.14 1.35 83.14  
## Sukuruputtu 27 4.55 87.69 4.55 87.69  
## Vadlavanipalem 24 4.05 91.74 4.05 91.74  
## Vakapalli 14 2.36 94.10 2.36 94.10  
## Valamarru 33 5.56 99.66 5.56 99.66  
## VSP0015 2 0.34 100.00 0.34 100.00  
## <NA> 0 0.00 100.00  
## Total 593 100.00 100.00 100.00 100.00

#rename the villages  
df\_sub$village[df\_sub$village=="D Gondhuru"]<-"D.Gondhuru"  
df\_sub$village[df\_sub$village=="D.Gondhru"]<-"D.Gondhuru"  
df\_sub$village[df\_sub$village=="D.Gondhuru"]<-"D.Gondhuru"  
df\_sub$village[df\_sub$village=="Korakodu"]<-"Korrakodu"  
  
#visualize  
df\_sub %>%   
 ggplot(aes(village))+  
 geom\_bar(fill="darkgreen")+  
 ggtitle("Barplot of the villages under study")+  
 ggeasy::easy\_center\_title()+  
 coord\_flip()

 Jayapuram is the village that had the highest count of farms under this study and accounts for 25.58%

summary(df\_sub$gps.Latitude)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 14.69 14.77 16.56 16.44 18.06 18.08

summary(df\_sub$gps.Longitude)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 77.31 77.33 81.69 80.34 82.63 82.67

summary(df\_sub$gps.Altitude)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -72.6 5.2 436.6 394.4 474.1 1038.9

summary(df\_sub$gps.Accuracy)

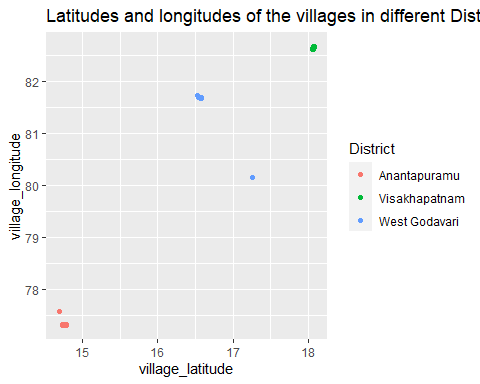
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.500 4.300 5.925 318.526 16.878 2900.000

boxplot.stats(df\_sub$gps.Accuracy)$out

## [1] 1674.000 72.443 1899.999 1899.999 96.448 1674.000 48.176 1000.000  
## [9] 1000.000 1000.000 2601.000 96.460 45.600 63.077 1899.999 1899.999  
## [17] 2200.000 2200.000 41.631 2799.999 2799.999 2500.000 2500.000 2200.000  
## [25] 2200.000 2332.000 2332.000 1674.000 43.809 1979.000 2601.000 2601.000  
## [33] 2200.000 2200.000 64.100 2200.000 2200.000 85.303 96.000 96.000  
## [41] 2299.999 2299.999 58.659 2200.000 2200.000 1674.000 2099.999 2099.999  
## [49] 2200.000 2200.000 1905.000 2299.999 2200.000 2200.000 60.000 1899.999  
## [57] 1899.999 1664.000 1664.000 98.400 87.600 87.600 2601.000 37.244  
## [65] 68.400 2299.999 2299.999 2799.999 2799.999 72.125 1664.000 1664.000  
## [73] 45.847 77.478 97.406 48.161 60.000 2601.000 2601.000 68.464  
## [81] 68.464 68.464 60.000 45.600 45.600 2799.999 2799.999 2299.999  
## [89] 2299.999 2299.999 2900.000 2900.000 2200.000 2200.000 1924.000 2200.000  
## [97] 2200.000 2200.000 2200.000 2500.000 2500.000 1600.000 85.411 40.497  
## [105] 2799.999 2799.999 2299.999 2299.999 2299.999 62.069 96.000 88.163  
## [113] 2799.999 2799.999 2299.999 2299.999 75.591 77.600 77.600 2799.999  
## [121] 2799.999

There are outliers of the gps accuracy and hence further investigation needs to be done.

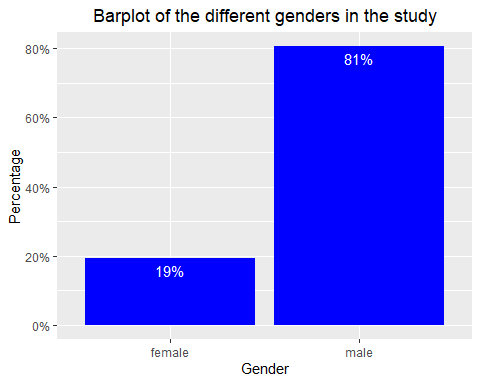
ggplot(df\_sub)+geom\_point(aes(x=gps.Latitude, y=gps.Longitude,colour=District))+  
 xlab("village\_latitude")+  
 ylab("village\_longitude")+  
 ggtitle("Latitudes and longitudes of the villages in different Districts")



df\_sub$farmer\_details.farmer\_gender<-as.factor(df\_sub$farmer\_details.farmer\_gender)  
levels(df\_sub$farmer\_details.farmer\_gender)<-c("female","male")  
freq(df\_sub$farmer\_details.farmer\_gender)

## Frequencies   
## df\_sub$farmer\_details.farmer\_gender   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ------------ ------ --------- -------------- --------- --------------  
## female 115 19.39 19.39 19.39 19.39  
## male 478 80.61 100.00 80.61 100.00  
## <NA> 0 0.00 100.00  
## Total 593 100.00 100.00 100.00 100.00

df\_sub %>%   
 ggplot(aes(farmer\_details.farmer\_gender)) +  
 geom\_bar(fill = "blue", aes(y = (..count..)/sum(..count..)))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust = 1.5, colour = "white")+  
 scale\_y\_continuous(labels =percent) +  
 ggtitle("Barplot of the different genders in the study")+  
 ggeasy::easy\_center\_title()+  
 ylab("Percentage")+  
 xlab("Gender")



Most of the farmers are males and account for 77.57% of the total population under study.

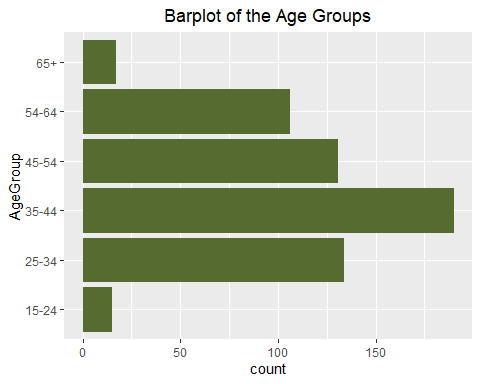
summary(df\_sub$farmer\_details.farmer\_age)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 21.00 35.00 45.00 44.28 52.00 80.00

##create age categories  
labs <- c("below 15","15-24","25-34","35-44","45-54","54-64","65+")  
  
#create a new variable for the age groups  
df\_sub$AgeGroup <- cut(df\_sub$farmer\_details.farmer\_age,breaks=c(0,15,25,35,45,54,65,Inf),labels = labs)  
freq(df\_sub$AgeGroup)

## Frequencies   
## df\_sub$AgeGroup   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## -------------- ------ --------- -------------- --------- --------------  
## below 15 0 0.00 0.00 0.00 0.00  
## 15-24 15 2.53 2.53 2.53 2.53  
## 25-34 134 22.60 25.13 22.60 25.13  
## 35-44 190 32.04 57.17 32.04 57.17  
## 45-54 131 22.09 79.26 22.09 79.26  
## 54-64 106 17.88 97.13 17.88 97.13  
## 65+ 17 2.87 100.00 2.87 100.00  
## <NA> 0 0.00 100.00  
## Total 593 100.00 100.00 100.00 100.00

##visualize the age groups  
ggplot(df\_sub,aes(AgeGroup))+  
 geom\_bar(fill="darkolivegreen")+  
 ggtitle("Barplot of the Age Groups")+  
 ggeasy::easy\_center\_title()+  
 coord\_flip()

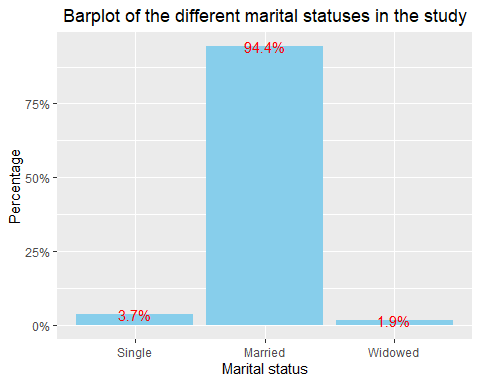


The median age for farmers is 45 but age categories are created to analyse where most farmers lie.Most farmers are aged between 35 and 44 and account for 29.14% of the total population while those aged between 15-24 and above 65 are the minority group and account for 2.94% and 3.14% respectively.

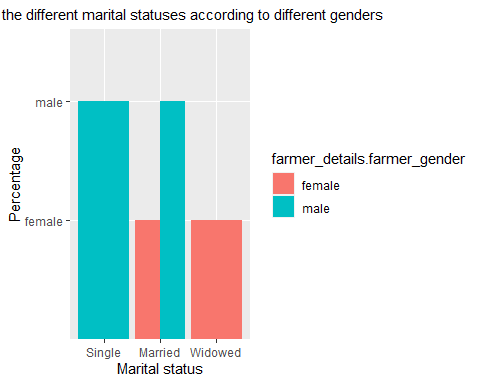
df\_sub$farmer\_details.marital\_status<-as.factor(df\_sub$farmer\_details.marital\_status)  
levels(df\_sub$farmer\_details.marital\_status)<-c("Single","Married","Widowed")  
freq(df\_sub$farmer\_details.marital\_status)

## Frequencies   
## df\_sub$farmer\_details.marital\_status   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ------------- ------ --------- -------------- --------- --------------  
## Single 22 3.71 3.71 3.71 3.71  
## Married 560 94.44 98.15 94.44 98.15  
## Widowed 11 1.85 100.00 1.85 100.00  
## <NA> 0 0.00 100.00  
## Total 593 100.00 100.00 100.00 100.00

##analysis of the marital statuses  
df\_sub %>%   
 ggplot(aes(farmer\_details.marital\_status)) +  
 geom\_bar(fill = "skyblue", aes(y = (..count..)/sum(..count..)))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust = 0.5, colour = "red")+  
 scale\_y\_continuous(labels =percent) +  
 ggtitle("Barplot of the different marital statuses in the study")+  
 ggeasy::easy\_center\_title()+  
 ylab("Percentage")+  
 xlab("Marital status")



##marital analysis according to the different genders  
df\_sub %>%   
 ggplot(aes(farmer\_details.marital\_status,farmer\_details.farmer\_gender,fill=farmer\_details.farmer\_gender)) +  
 geom\_col(position = "dodge")+  
 ggtitle("Barplot of the different marital statuses according to different genders")+  
 ggeasy::easy\_center\_title()+  
 ylab("Percentage")+  
 xlab("Marital status")+  
 theme(plot.title = element\_text(size=11))

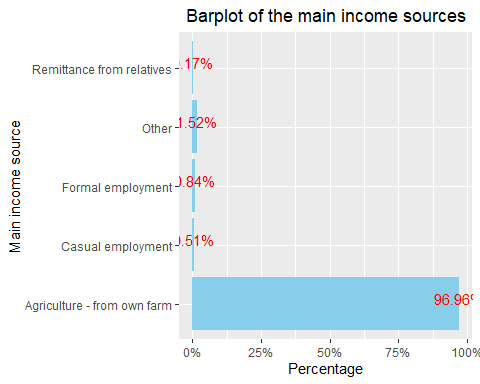


In this study, most of the participants are married and an analysis according to gender is done and the results are shown in the barplot above.

freq(df\_sub$hh\_income.main\_income\_source)

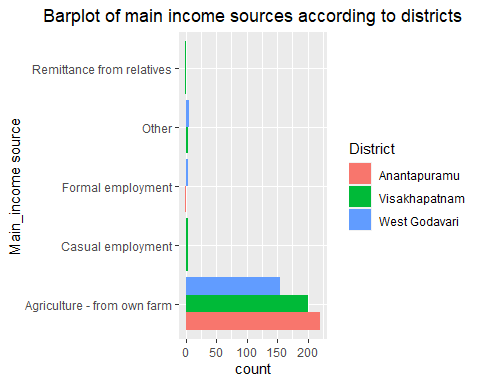
## Frequencies   
## df\_sub$hh\_income.main\_income\_source   
## Type: Integer   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ----------- ------ --------- -------------- --------- --------------  
## 1 575 96.96 96.96 96.96 96.96  
## 2 1 0.17 97.13 0.17 97.13  
## 3 5 0.84 97.98 0.84 97.98  
## 4 3 0.51 98.48 0.51 98.48  
## 7 9 1.52 100.00 1.52 100.00  
## <NA> 0 0.00 100.00  
## Total 593 100.00 100.00 100.00 100.00

##recode the sources   
df\_sub$hh\_income.main\_income\_source[df\_sub$hh\_income.main\_income\_source=="1"]<-"Agriculture - from own farm"  
df\_sub$hh\_income.main\_income\_source[df\_sub$hh\_income.main\_income\_source=="2"]<-"Remittance from relatives"  
df\_sub$hh\_income.main\_income\_source[df\_sub$hh\_income.main\_income\_source=="3"]<-"Formal employment"  
df\_sub$hh\_income.main\_income\_source[df\_sub$hh\_income.main\_income\_source=="4"]<-"Casual employment"  
df\_sub$hh\_income.main\_income\_source[df\_sub$hh\_income.main\_income\_source=="7"]<-"Other"  
  
##visualization  
df\_sub %>%  
 ggplot(aes(hh\_income.main\_income\_source))+  
 geom\_bar(fill = "skyblue", aes(y = (..count..)/sum(..count..)))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust =0.00345, colour = "red")+  
 scale\_y\_continuous(labels =percent) +  
 ggtitle("Barplot of the main income sources")+  
 ggeasy::easy\_center\_title()+  
 ylab("Percentage")+  
 xlab("Main income source")+  
 coord\_flip()



The main source according to most farmers in this study is Agriculture from their own farm which accounts for 96.23%.

df\_sub %>%   
 ggplot(aes(hh\_income.main\_income\_source,fill=District))+  
 geom\_bar(position = "dodge")+  
 ggtitle("Barplot of main income sources according to districts")+  
 ggeasy::easy\_center\_title()+  
 xlab("Main\_income source")+  
 coord\_flip()



freq(df\_sub$hh\_income.main\_income\_other)

## Frequencies   
## df\_sub$hh\_income.main\_income\_other   
## Type: Character   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ---------------------------- ------ --------- -------------- --------- --------------  
## (Empty string) 584 98.48 98.48 98.48 98.48  
## APCNF MT 3 0.51 98.99 0.51 98.99  
## Valamarru village ICRP 6 1.01 100.00 1.01 100.00  
## <NA> 0 0.00 100.00  
## Total 593 100.00 100.00 100.00 100.00

**FUNCTION FOR SPLITTING MESSY COLUMNS**

###SPLITTING THE MESSY COLUMNS####  
##Set the standard function for splitting the columns####  
#use the function below to separate the strings   
split\_into\_multiple <- function(column, pattern = ",", into\_prefix){  
 cols <- str\_split\_fixed(column, pattern, n = Inf)  
 # Sub out the ""'s returned by filling the matrix to the right, with NAs which are useful  
 cols[which(cols == "")] <- NA  
 cols <- as.tibble(cols)  
 # name the 'cols' tibble as 'into\_prefix\_1', 'into\_prefix\_2', ..., 'into\_prefix\_m'   
 # where m = # columns of 'cols'  
 m <- dim(cols)[2]  
  
 names(cols) <- paste(into\_prefix, 1:m, sep = "\_")  
 return(cols)  
}

##split and gather the columns  
df1<-df\_sub%>%   
 bind\_cols(split\_into\_multiple(.$hh\_income.agricultural\_produce," ","agricultural\_produce"))

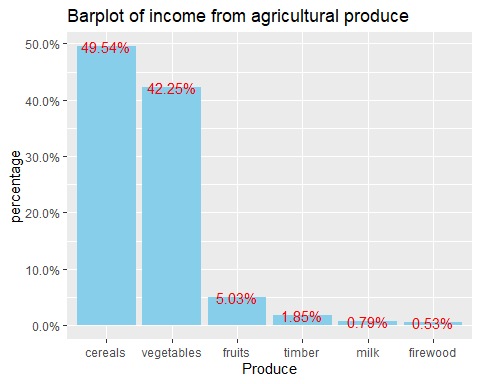
## Warning: `as.tibble()` was deprecated in tibble 2.0.0.  
## Please use `as\_tibble()` instead.  
## The signature and semantics have changed, see `?as\_tibble`.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was generated.

## Warning: The `x` argument of `as\_tibble.matrix()` must have unique column names if `.name\_repair` is omitted as of tibble 2.0.0.  
## Using compatibility `.name\_repair`.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was generated.

Agric\_produce<-df1%>%select(c(agricultural\_produce\_1,agricultural\_produce\_2,agricultural\_produce\_3,agricultural\_produce\_4)) %>% gather(agricultural\_produce,type)  
#3get the frequency  
freq(Agric\_produce$type)

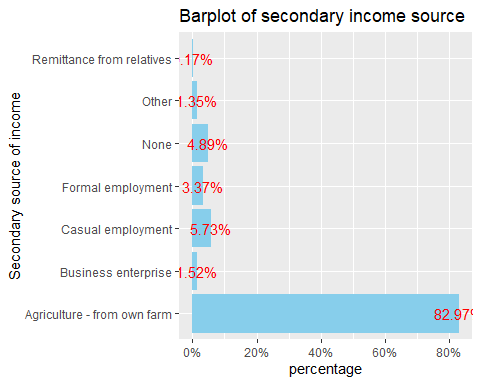
## Frequencies   
## Agric\_produce$type   
## Type: Character   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ---------------- ------ --------- -------------- --------- --------------  
## cereals 374 49.54 49.54 15.77 15.77  
## firewood 4 0.53 50.07 0.17 15.94  
## fruits 38 5.03 55.10 1.60 17.54  
## milk 6 0.79 55.89 0.25 17.79  
## timber 14 1.85 57.75 0.59 18.38  
## vegetables 319 42.25 100.00 13.45 31.83  
## <NA> 1617 68.17 100.00  
## Total 2372 100.00 100.00 100.00 100.00

##barplot  
Agric\_produce %>%drop\_na(type) %>%   
 ggplot(aes(factor(fct\_infreq(type))))+  
 geom\_bar(fill="skyblue",aes(y=(..count..)/sum(..count..)))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust =0.5, colour = "red")+  
 ylab("percentage")+  
 xlab("Produce")+  
 scale\_y\_continuous(labels=percent)+  
 ggtitle("Barplot of income from agricultural produce")



Most farmers(49.21%) produce vegetables in their farms for cash income and the least group(0.63%) produce firewood.

##recode the levels  
df1$hh\_income.secondary\_income\_source[df1$hh\_income.secondary\_income\_source=="1"]<-"Agriculture - from own farm"  
df1$hh\_income.secondary\_income\_source[df1$hh\_income.secondary\_income\_source=="2"]<-"Remittance from relatives"  
df1$hh\_income.secondary\_income\_source[df1$hh\_income.secondary\_income\_source=="3"]<-"Formal employment"  
df1$hh\_income.secondary\_income\_source[df1$hh\_income.secondary\_income\_source=="4"]<-"Casual employment"  
df1$hh\_income.secondary\_income\_source[df1$hh\_income.secondary\_income\_source=="5"]<-"Business enterprise"  
df1$hh\_income.secondary\_income\_source[df1$hh\_income.secondary\_income\_source=="6"]<-"None"  
df1$hh\_income.secondary\_income\_source[df1$hh\_income.secondary\_income\_source=="7"]<-"Other"  
  
#get the frequency   
df1 %>%  
 ggplot(aes(hh\_income.secondary\_income\_source))+  
 geom\_bar(fill="skyblue",aes(y=(..count..)/sum(..count..)))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust =0.5, colour = "red")+  
 ylab("percentage")+  
 xlab("Secondary source of income")+  
 scale\_y\_continuous(labels=percent)+  
 coord\_flip()+  
 ggtitle("Barplot of secondary income source")



From the graph above, it can be seen that most of the secondary income is from Agriculture in the farmer’s own farms

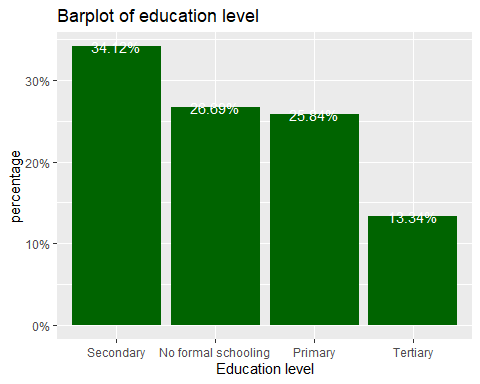
freq(df1$hh\_income.secondary\_income\_other)

## Frequencies   
## df1$hh\_income.secondary\_income\_other   
## Type: Character   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## -------------------------------- ------ --------- -------------- --------- --------------  
## (Empty string) 585 98.65 98.65 98.65 98.65  
## L2 of poduru 1 0.17 98.82 0.17 98.82  
## MPTC 2 0.34 99.16 0.34 99.16  
## Politician ( Ward member) 3 0.51 99.66 0.51 99.66  
## Sarpanch of the panchayath 2 0.34 100.00 0.34 100.00  
## <NA> 0 0.00 100.00  
## Total 593 100.00 100.00 100.00 100.00

df1$group\_house.education\_level<-as.factor(df1$group\_house.education\_level)  
levels(df1$group\_house.education\_level)<-c("No formal schooling","Primary","Secondary","Tertiary","Other")  
freq(df1$group\_house.education\_level)

## Frequencies   
## df1$group\_house.education\_level   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ------------------------- ------ --------- -------------- --------- --------------  
## No formal schooling 158 26.69 26.69 26.64 26.64  
## Primary 153 25.84 52.53 25.80 52.45  
## Secondary 202 34.12 86.66 34.06 86.51  
## Tertiary 79 13.34 100.00 13.32 99.83  
## Other 0 0.00 100.00 0.00 99.83  
## <NA> 1 0.17 100.00  
## Total 593 100.00 100.00 100.00 100.00

df1 %>% drop\_na(group\_house.education\_level) %>%   
 ggplot(aes(factor(fct\_infreq(group\_house.education\_level))))+  
 geom\_bar(fill="darkgreen",aes(y=((..count..)/sum(..count..))))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust =0.5, colour = "white")+  
 ylab("percentage")+  
 scale\_y\_continuous(labels=percent)+  
 ggtitle("Barplot of education level")+  
 xlab("Education level")



From the output above,most farmers have secondary level information.

summary(df1$group\_house.hh\_pop)

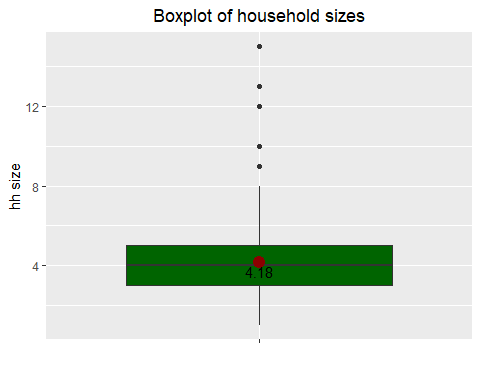
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 3.000 4.000 4.184 5.000 15.000

boxplot.stats(df1$group\_house.hh\_pop)$out

## [1] 12 12 12 12 9 9 10 9 15 15 9 10 12 12 12 13 13 13 13 10 10

##boxplot  
meanFunction <- function(x){  
return(data.frame(y=round(mean(x),2),label=round(mean(x,na.rm=T),2)))}  
  
df1 %>% drop\_na(group\_house.hh\_pop) %>%   
 ggplot(aes(x="",y=group\_house.hh\_pop))+  
 geom\_boxplot(fill="darkgreen")+  
 stat\_summary(fun.y = mean, geom="point",colour="darkred", size=4) +  
stat\_summary(fun.data = meanFunction, geom="text", size = 4, vjust=1.3) +  
 ggtitle("Boxplot of household sizes")+  
 xlab("")+  
 ylab("hh size")+  
 ggeasy::easy\_center\_title()

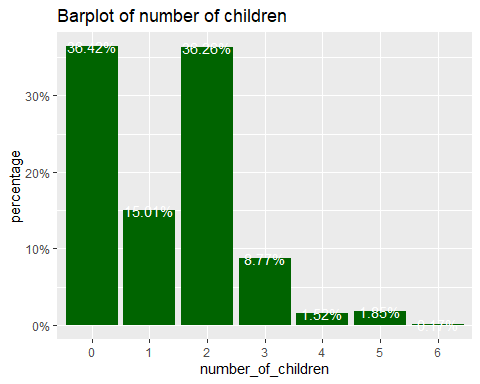
## Warning: `fun.y` is deprecated. Use `fun` instead.

 The summary statistics for the household population are given and there are some outliers that is families having a population above 8 members. The boxplot is however shown for easy visualization and it can be seen that the average number of members in a household is 4 members.

freq(df1$group\_house.num\_child)

## Frequencies   
## df1$group\_house.num\_child   
## Type: Integer   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ----------- ------ --------- -------------- --------- --------------  
## 0 216 36.42 36.42 36.42 36.42  
## 1 89 15.01 51.43 15.01 51.43  
## 2 215 36.26 87.69 36.26 87.69  
## 3 52 8.77 96.46 8.77 96.46  
## 4 9 1.52 97.98 1.52 97.98  
## 5 11 1.85 99.83 1.85 99.83  
## 6 1 0.17 100.00 0.17 100.00  
## <NA> 0 0.00 100.00  
## Total 593 100.00 100.00 100.00 100.00

df1$group\_house.num\_child<-as.character(df1$group\_house.num\_child)  
  
  
df1 %>% ggplot(aes(group\_house.num\_child))+  
 geom\_bar(fill="darkgreen",aes(y=((..count..)/sum(..count..))))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust =0.5, colour = "white")+  
 ylab("percentage")+  
 xlab("number\_of\_children")+  
 scale\_y\_continuous(labels=percent)+  
 ggtitle("Barplot of number of children")

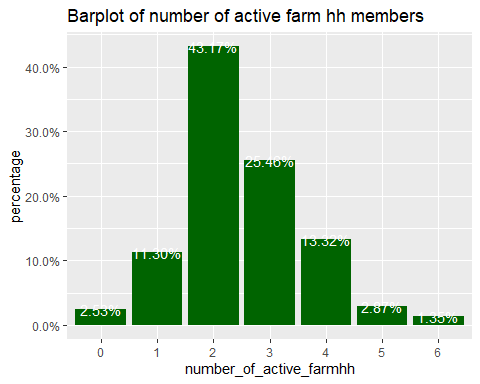


In this study,most families had 2 children accounting for 39.83% of the total population and 32.70% for families with no children while the least 0.21% accounts for families with 6 children.

##How many of the household members are active on the farm?  
df1$group\_house.active\_farm<-as.character(df1$group\_house.active\_farm)  
freq(df1$group\_house.active\_farm)

## Frequencies   
## df1$group\_house.active\_farm   
## Type: Character   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ----------- ------ --------- -------------- --------- --------------  
## 0 15 2.53 2.53 2.53 2.53  
## 1 67 11.30 13.83 11.30 13.83  
## 2 256 43.17 57.00 43.17 57.00  
## 3 151 25.46 82.46 25.46 82.46  
## 4 79 13.32 95.78 13.32 95.78  
## 5 17 2.87 98.65 2.87 98.65  
## 6 8 1.35 100.00 1.35 100.00  
## <NA> 0 0.00 100.00  
## Total 593 100.00 100.00 100.00 100.00

df1 %>% ggplot(aes(group\_house.active\_farm))+  
 geom\_bar(fill="darkgreen",aes(y=((..count..)/sum(..count..))))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust =0.5, colour = "white")+  
 ylab("percentage")+  
 xlab("number\_of\_active\_farmhh")+  
 scale\_y\_continuous(labels=percent)+  
 ggtitle("Barplot of number of active farm hh members")

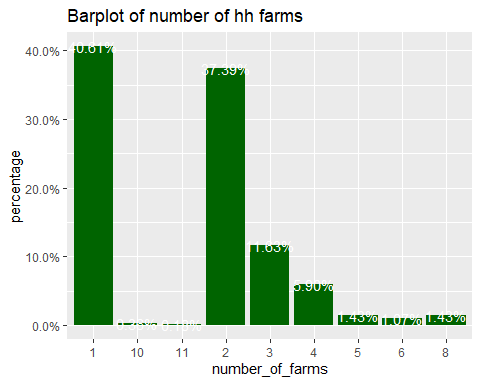


Most families have 2 household members who are active on the farm.

df1$num\_farms<-as.character(df1$num\_farms)  
freq(df1$num\_farms)

## Frequencies   
## df1$num\_farms   
## Type: Character   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ----------- ------ --------- -------------- --------- --------------  
## 1 227 40.61 40.61 38.28 38.28  
## 10 2 0.36 40.97 0.34 38.62  
## 11 1 0.18 41.14 0.17 38.79  
## 2 209 37.39 78.53 35.24 74.03  
## 3 65 11.63 90.16 10.96 84.99  
## 4 33 5.90 96.06 5.56 90.56  
## 5 8 1.43 97.50 1.35 91.91  
## 6 6 1.07 98.57 1.01 92.92  
## 8 8 1.43 100.00 1.35 94.27  
## <NA> 34 5.73 100.00  
## Total 593 100.00 100.00 100.00 100.00

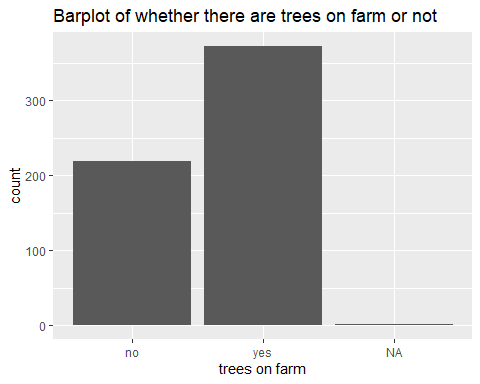
df1 %>% drop\_na(num\_farms) %>%   
 ggplot(aes(num\_farms))+  
 geom\_bar(fill="darkgreen",aes(y=((..count..)/sum(..count..))))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust =0.5, colour = "white")+  
 ylab("percentage")+  
 xlab("number\_of\_farms")+  
 scale\_y\_continuous(labels=percent)+  
 ggtitle("Barplot of number of hh farms")



df1$group\_tree.trees\_onfarm<-as.factor(df1$group\_tree.trees\_onfarm)  
freq(df1$group\_tree.trees\_onfarm)

## Frequencies   
## df1$group\_tree.trees\_onfarm   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ----------- ------ --------- -------------- --------- --------------  
## 0 219 37.06 37.06 36.93 36.93  
## 1 372 62.94 100.00 62.73 99.66  
## <NA> 2 0.34 100.00  
## Total 593 100.00 100.00 100.00 100.00

levels(df1$group\_tree.trees\_onfarm)<-c("no","yes")  
  
df1 %>% ggplot(aes(group\_tree.trees\_onfarm))+  
 geom\_bar()+  
 ggtitle("Barplot of whether there are trees on farm or not")+  
 xlab("trees on farm")

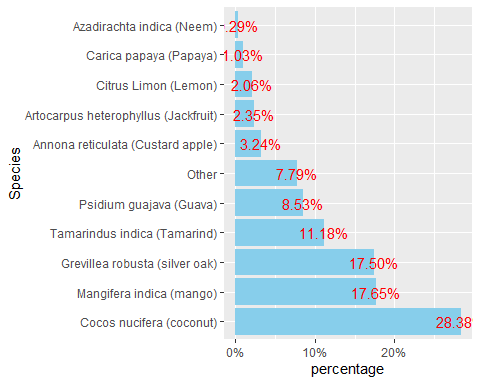


The larger percentage 61.64% of the farmers during this period of study had trees on their farms.

###Which tree species do you have on your farm?  
df2<-df1%>%   
 bind\_cols(split\_into\_multiple(.$group\_tree.tree\_species," ","tree\_species"))  
Tree\_species<-df2 %>% select(tree\_species\_1,tree\_species\_2,tree\_species\_3,tree\_species\_4,tree\_species\_5) %>%   
 gather(tree\_species,type)  
#reorder the levels  
Tree\_species$type<-as.factor(Tree\_species$type)  
levels(Tree\_species$type)<-c("Mangifera indica (mango)","Cocos nucifera (coconut)","Grevillea robusta (silver oak)","Other","Annona reticulata (Custard apple)","Psidium guajava (Guava)","Tamarindus indica (Tamarind)","Artocarpus heterophyllus (Jackfruit)","Azadirachta indica (Neem)","Carica papaya (Papaya)","Citrus Limon (Lemon)")  
freq(Tree\_species$type)

## Frequencies   
## Tree\_species$type   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ------------------------------------------ ------ --------- -------------- --------- --------------  
## Mangifera indica (mango) 120 17.647 17.647 4.047 4.047  
## Cocos nucifera (coconut) 193 28.382 46.029 6.509 10.556  
## Grevillea robusta (silver oak) 119 17.500 63.529 4.013 14.570  
## Other 53 7.794 71.324 1.788 16.358  
## Annona reticulata (Custard apple) 22 3.235 74.559 0.742 17.099  
## Psidium guajava (Guava) 58 8.529 83.088 1.956 19.056  
## Tamarindus indica (Tamarind) 76 11.176 94.265 2.563 21.619  
## Artocarpus heterophyllus (Jackfruit) 16 2.353 96.618 0.540 22.159  
## Azadirachta indica (Neem) 2 0.294 96.912 0.067 22.226  
## Carica papaya (Papaya) 7 1.029 97.941 0.236 22.462  
## Citrus Limon (Lemon) 14 2.059 100.000 0.472 22.934  
## <NA> 2285 77.066 100.000  
## Total 2965 100.000 100.000 100.000 100.000

Tree\_species %>% drop\_na(type) %>%   
 ggplot(aes(factor(fct\_infreq(type))))+  
 geom\_bar(fill="skyblue",aes(y=(..count..)/sum(..count..)))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust =0.5, colour = "red")+  
 coord\_flip()+  
 scale\_y\_continuous(labels = percent)+  
 ylab("percentage")+  
 xlab("Species")



Coconut is the most prevalent tree species and accounts for 38.03% of the total trees available.

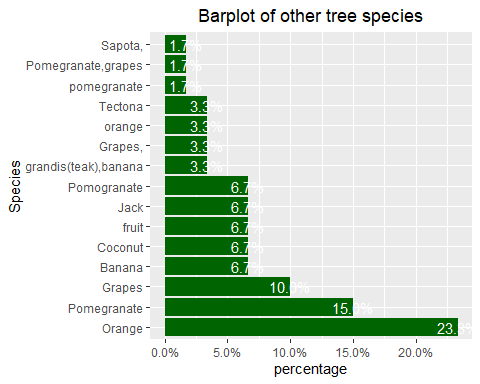
##What other tree species do you have on your farm  
##split the column  
df3<-df2%>%   
 bind\_cols(split\_into\_multiple(.$group\_tree.tree\_species\_other," ","other\_tree\_species"))  
Other\_species<-df3 %>% select(other\_tree\_species\_1,other\_tree\_species\_2) %>%   
 gather(other\_tree\_species,type)  
##change into factor and rename the levels  
Other\_species$type<-factor(Other\_species$type)  
levels(Other\_species$type)

## [1] "Banana" "Coconut" "fruit"   
## [4] "grandis(teak),banana" "Grapes" "Grapes,"   
## [7] "Jack" "orange" "Orange"   
## [10] "pomegranate" "Pomegranate" "Pomegranate,grapes"   
## [13] "Pomogranate" "Sapota," "Tectona"

#3check the frequency  
freq(Other\_species$type)

## Frequencies   
## Other\_species$type   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## -------------------------- ------ --------- -------------- --------- --------------  
## Banana 4 6.667 6.667 0.337 0.337  
## Coconut 4 6.667 13.333 0.337 0.675  
## fruit 4 6.667 20.000 0.337 1.012  
## grandis(teak),banana 2 3.333 23.333 0.169 1.180  
## Grapes 6 10.000 33.333 0.506 1.686  
## Grapes, 2 3.333 36.667 0.169 1.855  
## Jack 4 6.667 43.333 0.337 2.192  
## orange 2 3.333 46.667 0.169 2.361  
## Orange 14 23.333 70.000 1.180 3.541  
## pomegranate 1 1.667 71.667 0.084 3.626  
## Pomegranate 9 15.000 86.667 0.759 4.384  
## Pomegranate,grapes 1 1.667 88.333 0.084 4.469  
## Pomogranate 4 6.667 95.000 0.337 4.806  
## Sapota, 1 1.667 96.667 0.084 4.890  
## Tectona 2 3.333 100.000 0.169 5.059  
## <NA> 1126 94.941 100.000  
## Total 1186 100.000 100.000 100.000 100.000

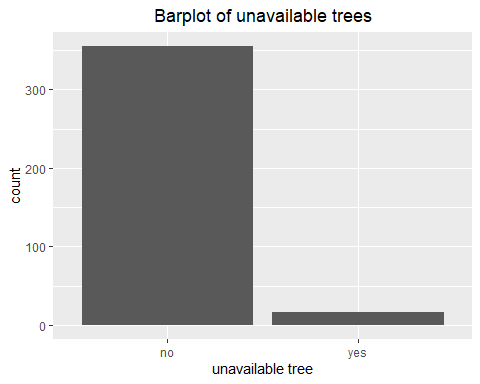
##barplot  
Other\_species %>% drop\_na(type) %>%   
 ggplot(aes(factor(fct\_infreq(type))))+  
 geom\_bar(fill="darkgreen",aes(y=(..count..)/sum(..count..)))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust =0.5, colour = "white")+  
 scale\_y\_continuous(labels = percent)+  
 ylab("percentage")+  
 xlab("Species")+  
 ggtitle("Barplot of other tree species")+  
 ggeasy::easy\_center\_title()+  
 coord\_flip()



##Is there a tree(s) that has disappeared in your farm which was there before?  
df3$group\_tree.unavailable\_tree<-as.factor(df3$group\_tree.unavailable\_tree)  
levels(df3$group\_tree.unavailable\_tree)<-c("no","yes")  
freq(df3$group\_tree.unavailable\_tree)

## Frequencies   
## df3$group\_tree.unavailable\_tree   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ----------- ------ --------- -------------- --------- --------------  
## no 355 95.43 95.43 59.87 59.87  
## yes 17 4.57 100.00 2.87 62.73  
## <NA> 221 37.27 100.00  
## Total 593 100.00 100.00 100.00 100.00

##barplot  
df3 %>% drop\_na(group\_tree.unavailable\_tree) %>%   
 ggplot(aes(group\_tree.unavailable\_tree))+  
 geom\_bar()+  
 ggtitle("Barplot of unavailable trees")+  
 xlab("unavailable tree")+  
 ggeasy::easy\_center\_title()



The larger percentage(97.96%) of the farmers during this period of study did not have trees that disappeared from their farms that were there before.

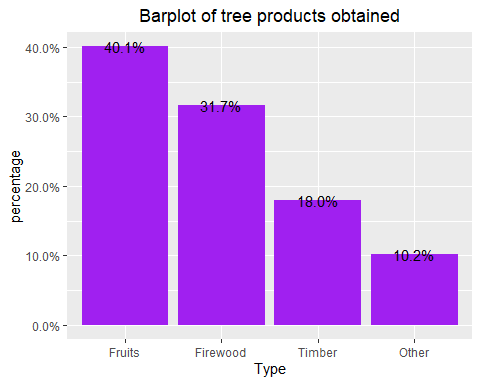
freq(df3$group\_tree.unavailable\_tree\_type)

## Frequencies   
## df3$group\_tree.unavailable\_tree\_type   
## Type: Character   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## -------------------- ------ --------- -------------- --------- --------------  
## (Empty string) 591 99.66 99.66 99.66 99.66  
## Sapota 2 0.34 100.00 0.34 100.00  
## <NA> 0 0.00 100.00  
## Total 593 100.00 100.00 100.00 100.00

###What products do you currently obtain from the trees grown on your farm?  
df4<-df3%>%   
 bind\_cols(split\_into\_multiple(.$group\_tree.tree\_products," ","tree\_products"))  
Tree\_products<-df4 %>% select(tree\_products\_1,tree\_products\_2,tree\_products\_3,tree\_products\_4) %>%   
 gather(tree\_products,type)  
  
##change into a factor variable  
Tree\_products$type<-as.factor(Tree\_products$type)  
levels(Tree\_products$type)<-c("Timber","Firewood","Fruits","Other")  
freq(Tree\_products$type)

## Frequencies   
## Tree\_products$type   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## -------------- ------ --------- -------------- --------- --------------  
## Timber 113 17.99 17.99 4.76 4.76  
## Firewood 199 31.69 49.68 8.39 13.15  
## Fruits 252 40.13 89.81 10.62 23.78  
## Other 64 10.19 100.00 2.70 26.48  
## <NA> 1744 73.52 100.00  
## Total 2372 100.00 100.00 100.00 100.00

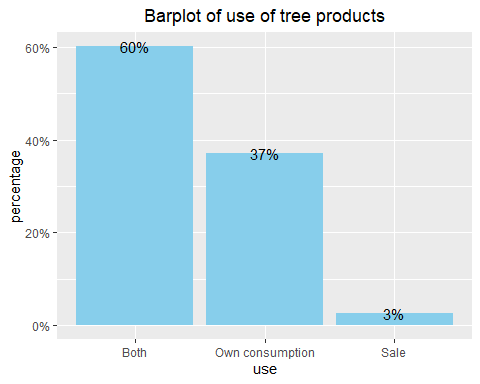
##visualize  
Tree\_products %>% drop\_na(type) %>%   
 ggplot(aes(factor(fct\_infreq(type))))+  
 geom\_bar(fill="purple",aes(y=(..count..)/sum(..count..)))+  
 geom\_text(aes(y=((..count..)/sum(..count..)),label=scales::percent((..count..)/sum(..count..))),stat = "count",vjust=0.5,colour="black")+  
 scale\_y\_continuous(labels=percent)+  
 ylab("percentage")+  
 xlab("Type")+  
 ggtitle("Barplot of tree products obtained")+  
 ggeasy::easy\_center\_title()



df4$group\_tree.tree\_product\_use<-factor(df4$group\_tree.tree\_product\_use)  
levels(df4$group\_tree.tree\_product\_use)<-c("Own consumption","Sale","Both")  
  
freq(df4$group\_tree.tree\_product\_use)

## Frequencies   
## df4$group\_tree.tree\_product\_use   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## --------------------- ------ --------- -------------- --------- --------------  
## Own consumption 138 37.10 37.10 23.27 23.27  
## Sale 10 2.69 39.78 1.69 24.96  
## Both 224 60.22 100.00 37.77 62.73  
## <NA> 221 37.27 100.00  
## Total 593 100.00 100.00 100.00 100.00

##visualization  
df4 %>% drop\_na(group\_tree.tree\_product\_use) %>%   
 ggplot(aes(factor(fct\_infreq(group\_tree.tree\_product\_use))))+  
 geom\_bar(fill="skyblue",aes(y=(..count..)/sum(..count..)))+  
 geom\_text(aes(y=((..count..)/sum(..count..)),label=scales::percent((..count..)/sum(..count..))),stat = "count",vjust=0.5,colour="black")+  
 scale\_y\_continuous(labels=percent)+  
 ylab("percentage")+  
 xlab("use")+  
 ggtitle("Barplot of use of tree products")+  
 ggeasy::easy\_center\_title()



60% of the farmers use the tree products for both own consumption and sale and the least percentage(3%) uses tree products for sale.

##How else do you use the trees on your farm?  
#trigram  
df4<-df4 %>% drop\_na(group\_tree.tree\_use)  
trigrams\_separated\_uncommon <- df4 %>%   
 select(group\_tree.tree\_use) %>%   
 unnest\_tokens(trigram, group\_tree.tree\_use, token = "ngrams", n = 2) %>%   
 separate(trigram, c("word1", "word2","word3"), sep = " ") %>%  
 filter(!word1 %in% stop\_words$word,  
 !word2 %in% stop\_words$word) %>%  
 count(word1, word2, sort = TRUE)

## Warning: Expected 3 pieces. Missing pieces filled with `NA` in 297 rows [23, 24,  
## 25, 42, 43, 59, 60, 64, 65, 68, 69, 70, 71, 82, 83, 84, 85, 86, 87, 88, ...].

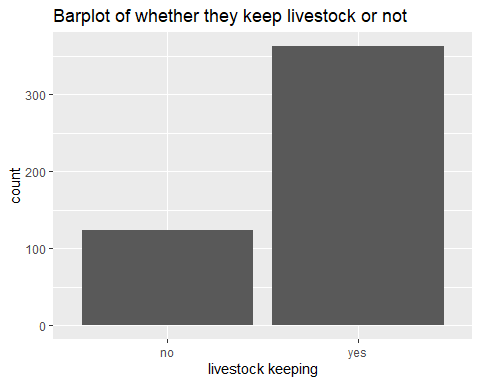
### unite the trigrams  
trigrams\_united\_uncommon <- trigrams\_separated\_uncommon %>%  
 unite(trigram, word1, word2, sep = " ")  
trigrams\_united\_uncommon %>% View()

There is not much found out about how trees are used but it can be observed that they are used for housing purposes and as wind breakers.

##Do you keep livestock?  
df4$group\_livestock.livestock\_keeping<-factor(df4$group\_livestock.livestock\_keeping)  
levels(df4$group\_livestock.livestock\_keeping)<-c("no","yes")  
freq(df4$group\_livestock.livestock\_keeping)

## Frequencies   
## df4$group\_livestock.livestock\_keeping   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ----------- ------ --------- -------------- --------- --------------  
## no 123 25.31 25.31 25.31 25.31  
## yes 363 74.69 100.00 74.69 100.00  
## <NA> 0 0.00 100.00  
## Total 486 100.00 100.00 100.00 100.00

##visualize  
df4 %>% ggplot(aes(group\_livestock.livestock\_keeping))+  
 geom\_bar()+  
 xlab("livestock keeping")+  
 ggtitle("Barplot of whether they keep livestock or not")

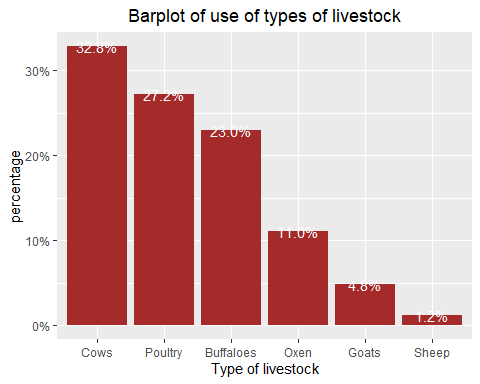


The larger percentage of farmers(74.09%) keeps livestock in their farms.

##check the frequency  
##freq(df4$group\_livestock.livestock\_type)  
  
df5<-df4 %>%   
 bind\_cols(split\_into\_multiple(.$group\_livestock.livestock\_type," ","livestock\_type"))  
  
Livestock\_Type<-df5 %>% select(livestock\_type\_1,livestock\_type\_2,livestock\_type\_3,livestock\_type\_4) %>%   
 gather(livestock\_type,type)  
  
  
#investigate the individual livestock types  
Livestock\_Type$type<-factor(Livestock\_Type$type)  
levels(Livestock\_Type$type)<-c("Cows","Buffaloes","Goats","Sheep","Oxen","Poultry")  
##check the frequency  
freq(Livestock\_Type$type)

## Frequencies   
## Livestock\_Type$type   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## --------------- ------ --------- -------------- --------- --------------  
## Cows 226 32.85 32.85 11.63 11.63  
## Buffaloes 158 22.97 55.81 8.13 19.75  
## Goats 33 4.80 60.61 1.70 21.45  
## Sheep 8 1.16 61.77 0.41 21.86  
## Oxen 76 11.05 72.82 3.91 25.77  
## Poultry 187 27.18 100.00 9.62 35.39  
## <NA> 1256 64.61 100.00  
## Total 1944 100.00 100.00 100.00 100.00

##visualize  
Livestock\_Type %>% drop\_na(type) %>%   
 ggplot(aes(fct\_infreq(type)))+  
 geom\_bar(fill="brown",aes(y=(..count..)/sum(..count..)))+  
 geom\_text(aes(y=((..count..)/sum(..count..)),label=scales::percent((..count..)/sum(..count..))),stat = "count",vjust=0.5,colour="white")+  
 scale\_y\_continuous(labels=percent)+  
 ylab("percentage")+  
 xlab("Type of livestock")+  
 ggtitle("Barplot of use of types of livestock")+  
 ggeasy::easy\_center\_title()



The highest percentage(35.5%) of the farmers keep cows and the least percentage (1.2%) rears sheep.

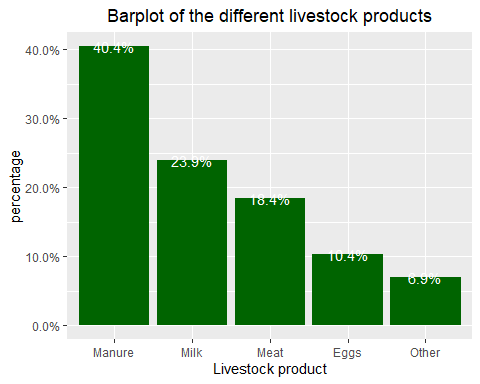
###What products do you currently obtain from the livestock kept on your farm?  
freq(df5$group\_livestock.livestock\_products)

## Frequencies   
## df5$group\_livestock.livestock\_products   
## Type: Character   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## -------------------- ------ --------- -------------- --------- --------------  
## (Empty string) 123 25.31 25.31 25.31 25.31  
## 1 4 0.82 26.13 0.82 26.13  
## 1 2 3 4 34 7.00 33.13 7.00 33.13  
## 1 2 4 3 7 1.44 34.57 1.44 34.57  
## 1 3 6 1.23 35.80 1.23 35.80  
## 1 3 4 7 1.44 37.24 1.44 37.24  
## 1 4 41 8.44 45.68 8.44 45.68  
## 1 4 3 8 1.65 47.33 1.65 47.33  
## 1 4 5 12 2.47 49.79 2.47 49.79  
## 1 5 3 0.62 50.41 0.62 50.41  
## 2 8 1.65 52.06 1.65 52.06  
## 2 3 4 1 5 1.03 53.09 1.03 53.09  
## 2 4 120 24.69 77.78 24.69 77.78  
## 2 4 1 2 0.41 78.19 0.41 78.19  
## 3 1 5 1.03 79.22 1.03 79.22  
## 3 2 4 1 6 1.23 80.45 1.23 80.45  
## 3 4 5 5 1.03 81.48 1.03 81.48  
## 3 5 2 0.41 81.89 0.41 81.89  
## 4 31 6.38 88.27 6.38 88.27  
## 4 1 8 1.65 89.92 1.65 89.92  
## 4 1 5 2 0.41 90.33 0.41 90.33  
## 4 2 14 2.88 93.21 2.88 93.21  
## 4 5 29 5.97 99.18 5.97 99.18  
## 4 5 1 1 0.21 99.38 0.21 99.38  
## 5 3 0.62 100.00 0.62 100.00  
## <NA> 0 0.00 100.00  
## Total 486 100.00 100.00 100.00 100.00

df6<-df5 %>%   
 bind\_cols(split\_into\_multiple(.$group\_livestock.livestock\_products,"","livestock\_products"))  
  
##gather the split columns  
Livestock\_products<-df6 %>%   
 select(livestock\_products\_1,livestock\_products\_2,livestock\_products\_3,livestock\_products\_4,livestock\_products\_5,livestock\_products\_6,livestock\_products\_7) %>%   
 gather(livestock\_products,type)  
  
##check for the frequency  
freq(Livestock\_products$type)

## Frequencies   
## Livestock\_products$type   
## Type: Character   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ----------- ------ --------- -------------- --------- --------------  
## · 458 35.81 35.81 13.46 13.46  
## 1 151 11.81 47.62 4.44 17.90  
## 2 196 15.32 62.94 5.76 23.66  
## 3 85 6.65 69.59 2.50 26.16  
## 4 332 25.96 95.54 9.76 35.92  
## 5 57 4.46 100.00 1.68 37.60  
## <NA> 2123 62.40 100.00  
## Total 3402 100.00 100.00 100.00 100.00

#convert into factor  
Livestock\_products$type<-factor(Livestock\_products$type)  
  
##rename  
levels(Livestock\_products$type)<-c("","Meat","Milk","Eggs","Manure","Other")  
  
##barplot  
Livestock\_products %>% filter(type %in% c("Meat","Milk","Eggs","Manure","Other")) %>%   
 ggplot(aes(fct\_infreq(type)))+  
 geom\_bar(fill="darkgreen",aes(y=(..count..)/sum(..count..)))+  
 geom\_text(aes(y=((..count..)/sum(..count..)),label=scales::percent((..count..)/sum(..count..))),stat = "count",vjust=0.5,colour="white")+ scale\_y\_continuous(labels = percent)+  
 ylab("percentage")+  
 xlab("Livestock product")+  
 ggtitle("Barplot of the different livestock products")+  
 ggeasy::easy\_center\_title()

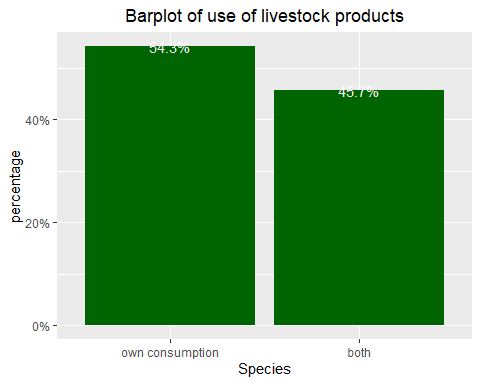


The highest percentage of farmers mainly obtain manure from the livestock kept in the farm and this accounts for 39.6% of the total livestock products.

##Are these products for home consumption or for sale?  
df6$group\_livestock.livestock\_products\_use<-factor(df6$group\_livestock.livestock\_products\_use)  
levels(df6$group\_livestock.livestock\_products\_use)<-c("own consumption","both")  
freq(df6$group\_livestock.livestock\_products\_use)

## Frequencies   
## df6$group\_livestock.livestock\_products\_use   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## --------------------- ------ --------- -------------- --------- --------------  
## own consumption 197 54.27 54.27 40.53 40.53  
## both 166 45.73 100.00 34.16 74.69  
## <NA> 123 25.31 100.00  
## Total 486 100.00 100.00 100.00 100.00

##barplot  
df6 %>% drop\_na(group\_livestock.livestock\_products\_use) %>%   
 ggplot(aes(group\_livestock.livestock\_products\_use))+  
 geom\_bar(fill="darkgreen",aes(y=(..count..)/sum(..count..)))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust =0.5, colour = "white")+  
 scale\_y\_continuous(labels = percent)+  
 ylab("percentage")+  
 xlab("Species")+  
 ggtitle("Barplot of use of livestock products")+  
 ggeasy::easy\_center\_title()



55.94% of farmers which is the higher percentage use livestock products for their own consumption and 44.06% use it for both own consumption and sale.

##What is the distance from your farm to nearest main road?  
summary(df6$group\_distance.distance\_road)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 0.500 1.800 1.974 3.000 15.000

#check for outliers  
boxplot.stats(df6$group\_distance.distance\_road)$out

## [1] 12 12 15 12 7 12 12 12 12 10 10 8 8 8

There are outliers in the distance from the farm to the nearest main road.

summary(df6$group\_distance.distance\_market)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.60 6.00 8.00 17.35 32.00 86.00

#check for outliers  
boxplot.stats(df6$group\_distance.distance\_market)$out

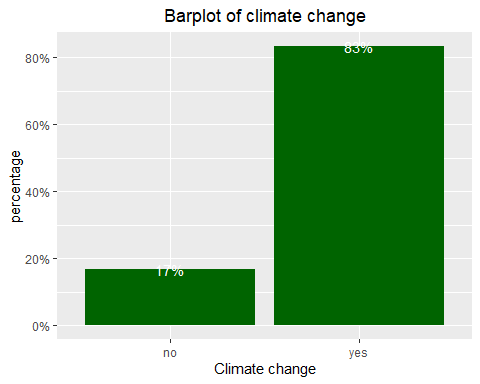
## [1] 85 86 85 85 80 85 85 85 85 84 80 85 85 85 85 85

There are outliers in the distance to the market.

##Have you heard of climate change?  
df6$group\_climatechange.climate\_change<-factor(df6$group\_climatechange.climate\_change)  
levels(df6$group\_climatechange.climate\_change)<-c("no","yes")  
freq(df6$group\_climatechange.climate\_change)

## Frequencies   
## df6$group\_climatechange.climate\_change   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ----------- ------ --------- -------------- --------- --------------  
## no 81 16.67 16.67 16.67 16.67  
## yes 405 83.33 100.00 83.33 100.00  
## <NA> 0 0.00 100.00  
## Total 486 100.00 100.00 100.00 100.00

##visualize  
df6 %>%   
 ggplot(aes(group\_climatechange.climate\_change))+  
 geom\_bar(fill="darkgreen",aes(y=(..count..)/sum(..count..)))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust =0.5, colour = "white")+  
 scale\_y\_continuous(labels = percent)+  
 ylab("percentage")+  
 xlab("Climate change")+  
 ggtitle("Barplot of climate change")+  
 ggeasy::easy\_center\_title()



79% of farmers in this study have heard avout climate change.

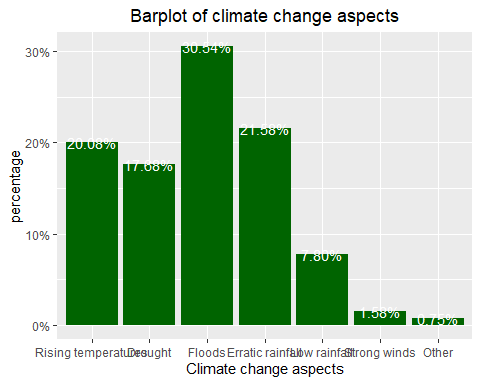
###What aspects of climate change have you heard of?  
##split the column for further analysis.  
df7<-df6 %>%   
 bind\_cols(split\_into\_multiple(.$group\_climatechange.cc\_aspects," ","climate\_changeaspects"))  
##gather the columns  
Climate\_change\_aspects<-df7 %>% select(climate\_changeaspects\_1,climate\_changeaspects\_2,climate\_changeaspects\_3,climate\_changeaspects\_4,climate\_changeaspects\_5,climate\_changeaspects\_6) %>%   
 gather(Climate\_change\_aspects,type)  
##check the frequency  
freq(Climate\_change\_aspects$type)

## Frequencies   
## Climate\_change\_aspects$type   
## Type: Character   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ----------- ------ --------- -------------- --------- --------------  
## 1 242 20.08 20.08 8.30 8.30  
## 2 213 17.68 37.76 7.30 15.60  
## 3 368 30.54 68.30 12.62 28.22  
## 4 260 21.58 89.88 8.92 37.14  
## 5 94 7.80 97.68 3.22 40.36  
## 6 19 1.58 99.25 0.65 41.02  
## 8 9 0.75 100.00 0.31 41.32  
## <NA> 1711 58.68 100.00  
## Total 2916 100.00 100.00 100.00 100.00

##convert into factor  
Climate\_change\_aspects$type<-factor(Climate\_change\_aspects$type)  
levels(Climate\_change\_aspects$type)<-c("Rising temperatures","Drought","Floods","Erratic rainfall","Low rainfall","Strong winds","Other")  
  
##frequency  
freq(Climate\_change\_aspects$type)

## Frequencies   
## Climate\_change\_aspects$type   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ------------------------- ------ --------- -------------- --------- --------------  
## Rising temperatures 242 20.08 20.08 8.30 8.30  
## Drought 213 17.68 37.76 7.30 15.60  
## Floods 368 30.54 68.30 12.62 28.22  
## Erratic rainfall 260 21.58 89.88 8.92 37.14  
## Low rainfall 94 7.80 97.68 3.22 40.36  
## Strong winds 19 1.58 99.25 0.65 41.02  
## Other 9 0.75 100.00 0.31 41.32  
## <NA> 1711 58.68 100.00  
## Total 2916 100.00 100.00 100.00 100.00

##barplot  
 Climate\_change\_aspects%>% drop\_na(type) %>%   
 ggplot(aes(type))+  
 geom\_bar(fill="darkgreen",aes(y=(..count..)/sum(..count..)))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust =0.5, colour = "white")+  
 scale\_y\_continuous(labels = percent)+  
 ylab("percentage")+  
 xlab("Climate change aspects")+  
 ggtitle("Barplot of climate change aspects")+  
 ggeasy::easy\_center\_title()

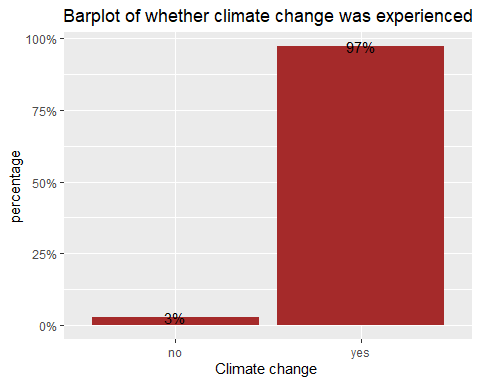


Most farmers 30.13% of the total number of farmers interviewed noted that they have mostly heard of floods followed by drought(21.94%).

##convert into factor with levels  
df7$group\_climatechange.cc\_experience<-factor(df7$group\_climatechange.cc\_experience)  
levels(df7$group\_climatechange.cc\_experience)<-c("no","yes")  
freq(df7$group\_climatechange.cc\_experience)

## Frequencies   
## df7$group\_climatechange.cc\_experience   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ----------- ------ --------- -------------- --------- --------------  
## no 11 2.72 2.72 2.26 2.26  
## yes 394 97.28 100.00 81.07 83.33  
## <NA> 81 16.67 100.00  
## Total 486 100.00 100.00 100.00 100.00

##visualization  
df7 %>% drop\_na(group\_climatechange.cc\_experience) %>%   
 ggplot(aes(group\_climatechange.cc\_experience))+  
 geom\_bar(fill="brown",aes(y=(..count..)/sum(..count..)))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust =0.5, colour = "black")+  
 scale\_y\_continuous(labels = percent)+  
 ylab("percentage")+  
 xlab("Climate change ")+  
 ggtitle("Barplot of whether climate change was experienced")+  
 ggeasy::easy\_center\_title()



96% of the farmers noted that they experienced climate change and the next code chunk investigates the different types of climate cjhange aspects that were experienced.

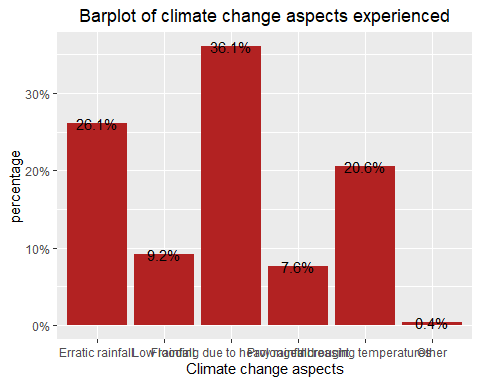
##check the frequency  
freq(df7$group\_climatechange.cchange\_experience)

## Frequencies   
## df7$group\_climatechange.cchange\_experience   
## Type: Character   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## -------------------- ------ --------- -------------- --------- --------------  
## (Empty string) 92 18.93 18.93 18.93 18.93  
## 1 4 0.82 19.75 0.82 19.75  
## 1 2 3 5 40 8.23 27.98 8.23 27.98  
## 1 2 4 5 2 0.41 28.40 0.41 28.40  
## 1 2 5 4 0.82 29.22 0.82 29.22  
## 1 3 45 9.26 38.48 9.26 38.48  
## 1 3 2 5 2 0.41 38.89 0.41 38.89  
## 1 3 4 4 0.82 39.71 0.82 39.71  
## 1 3 4 5 1 0.21 39.92 0.21 39.92  
## 1 3 5 103 21.19 61.11 21.19 61.11  
## 1 3 5 2 1 0.21 61.32 0.21 61.32  
## 1 3 5 6 2 0.41 61.73 0.41 61.73  
## 1 4 3 1 0.21 61.93 0.21 61.93  
## 1 5 7 1.44 63.37 1.44 63.37  
## 1 5 2 2 0.41 63.79 0.41 63.79  
## 1 5 3 1 0.21 63.99 0.21 63.99  
## 1 5 6 2 0.41 64.40 0.41 64.40  
## 2 1 3 5 2 0.41 64.81 0.41 64.81  
## 2 3 5 1.03 65.84 1.03 65.84  
## 2 3 4 6 1.23 67.08 1.23 67.08  
## 2 3 4 1 1 0.21 67.28 0.21 67.28  
## 2 5 4 0.82 68.11 0.82 68.11  
## 3 34 7.00 75.10 7.00 75.10  
## 3 1 12 2.47 77.57 2.47 77.57  
## 3 1 2 12 2.47 80.04 2.47 80.04  
## 3 1 5 2 0.41 80.45 0.41 80.45  
## 3 2 5 1.03 81.48 1.03 81.48  
## 3 2 4 1 0.21 81.69 0.21 81.69  
## 3 2 5 1 0.21 81.89 0.21 81.89  
## 3 4 47 9.67 91.56 9.67 91.56  
## 3 5 9 1.85 93.42 1.85 93.42  
## 3 5 1 2 0.41 93.83 0.41 93.83  
## 3 5 4 1 0.21 94.03 0.21 94.03  
## 4 3 5 1.03 95.06 1.03 95.06  
## 4 3 1 4 0.82 95.88 0.82 95.88  
## 4 3 2 1 0.21 96.09 0.21 96.09  
## 4 5 1 0.21 96.30 0.21 96.30  
## 5 3 0.62 96.91 0.62 96.91  
## 5 1 3 0.62 97.53 0.62 97.53  
## 5 2 3 3 0.62 98.15 0.62 98.15  
## 5 3 5 1.03 99.18 1.03 99.18  
## 5 3 1 3 0.62 99.79 0.62 99.79  
## 5 4 3 1 0.21 100.00 0.21 100.00  
## <NA> 0 0.00 100.00  
## Total 486 100.00 100.00 100.00 100.00

##split the column  
df8<-df7 %>%   
 bind\_cols(split\_into\_multiple(.$group\_climatechange.cchange\_experience," ", "Climate\_experience"))  
##gather the columns  
Climate\_aspects\_experienced<-df8 %>%   
 select(Climate\_experience\_1,Climate\_experience\_2,Climate\_experience\_3,Climate\_experience\_4) %>%   
 gather(Climate\_experience,type)  
  
##convert to factor  
Climate\_aspects\_experienced$type<-factor(Climate\_aspects\_experienced$type)  
levels(Climate\_aspects\_experienced$type)<-c("Erratic rainfall","Low rainfall","Flooding due to heavy rainfall","Prolonged drought","Increasing temperatures","Other")  
##frequency  
freq(Climate\_aspects\_experienced$type)

## Frequencies   
## Climate\_aspects\_experienced$type   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ------------------------------------ ------ --------- -------------- --------- --------------  
## Erratic rainfall 262 26.12 26.12 13.48 13.48  
## Low rainfall 92 9.17 35.29 4.73 18.21  
## Flooding due to heavy rainfall 362 36.09 71.39 18.62 36.83  
## Prolonged drought 76 7.58 78.96 3.91 40.74  
## Increasing temperatures 207 20.64 99.60 10.65 51.39  
## Other 4 0.40 100.00 0.21 51.59  
## <NA> 941 48.41 100.00  
## Total 1944 100.00 100.00 100.00 100.00

##visualize  
 Climate\_aspects\_experienced%>% drop\_na(type) %>%   
 ggplot(aes(type))+  
 geom\_bar(fill="firebrick",aes(y=(..count..)/sum(..count..)))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust =0.5, colour = "black")+  
 scale\_y\_continuous(labels = percent)+  
 ylab("percentage")+  
 xlab("Climate change aspects")+  
 ggtitle("Barplot of climate change aspects experienced")+  
 ggeasy::easy\_center\_title()



The highest percentage of farmers have experienced floods which accounts for 36.57% of the total climate aspects experienced and the least percentage 0.28% accounts for those that have experienced other climate change aspects.

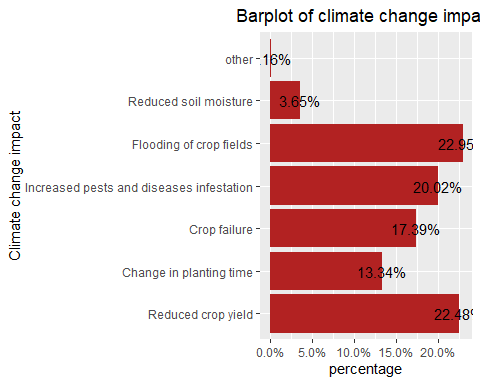
freq(df8$group\_climatechange.cc\_impact)

## Frequencies   
## df8$group\_climatechange.cc\_impact   
## Type: Character   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## -------------------- ------ --------- -------------- --------- --------------  
## (Empty string) 92 18.93 18.93 18.93 18.93  
## 1 16 3.29 22.22 3.29 22.22  
## 1 2 10 2.06 24.28 2.06 24.28  
## 1 2 3 2 0.41 24.69 0.41 24.69  
## 1 2 3 4 2 0.41 25.10 0.41 25.10  
## 1 2 3 4 5 33 6.79 31.89 6.79 31.89  
## 1 2 3 4 5 6 12 2.47 34.36 2.47 34.36  
## 1 2 3 4 6 3 0.62 34.98 0.62 34.98  
## 1 2 3 4 6 5 2 0.41 35.39 0.41 35.39  
## 1 2 3 5 3 0.62 36.01 0.62 36.01  
## 1 2 3 5 4 4 0.82 36.83 0.82 36.83  
## 1 2 3 5 6 4 2 0.41 37.24 0.41 37.24  
## 1 2 3 6 4 0.82 38.07 0.82 38.07  
## 1 2 4 3 0.62 38.68 0.62 38.68  
## 1 2 4 3 5 2 0.41 39.09 0.41 39.09  
## 1 2 4 5 3 0.62 39.71 0.62 39.71  
## 1 2 4 5 3 3 0.62 40.33 0.62 40.33  
## 1 2 4 6 2 0.41 40.74 0.41 40.74  
## 1 2 4 6 5 2 0.41 41.15 0.41 41.15  
## 1 3 3 0.62 41.77 0.62 41.77  
## 1 3 2 4 5 6 2 0.41 42.18 0.41 42.18  
## 1 3 4 2 0.41 42.59 0.41 42.59  
## 1 3 4 5 44 9.05 51.65 9.05 51.65  
## 1 3 4 5 6 1 0.21 51.85 0.21 51.85  
## 1 3 5 10 2.06 53.91 2.06 53.91  
## 1 4 18 3.70 57.61 3.70 57.61  
## 1 4 2 1 0.21 57.82 0.21 57.82  
## 1 4 3 5 2 0.41 58.23 0.41 58.23  
## 1 4 5 16 3.29 61.52 3.29 61.52  
## 1 4 5 2 1 0.21 61.73 0.21 61.73  
## 1 4 5 3 8 1.65 63.37 1.65 63.37  
## 1 4 5 6 1 0.21 63.58 0.21 63.58  
## 1 5 38 7.82 71.40 7.82 71.40  
## 1 5 3 4 1 0.21 71.60 0.21 71.60  
## 1 5 4 2 0.41 72.02 0.41 72.02  
## 1 5 4 3 3 0.62 72.63 0.62 72.63  
## 1 7 2 0.41 73.05 0.41 73.05  
## 2 7 1.44 74.49 1.44 74.49  
## 2 1 2 0.41 74.90 0.41 74.90  
## 2 1 3 5 2 0.41 75.31 0.41 75.31  
## 2 3 4 1 5 1 0.21 75.51 0.21 75.51  
## 2 3 4 5 7 1.44 76.95 1.44 76.95  
## 2 3 4 5 6 7 1.44 78.40 1.44 78.40  
## 2 3 5 6 1.23 79.63 1.23 79.63  
## 2 3 5 6 8 1.65 81.28 1.65 81.28  
## 2 4 9 1.85 83.13 1.85 83.13  
## 2 4 5 6 1.23 84.36 1.23 84.36  
## 2 5 10 2.06 86.42 2.06 86.42  
## 3 6 1.23 87.65 1.23 87.65  
## 3 1 2 0.41 88.07 0.41 88.07  
## 3 1 5 4 1 0.21 88.27 0.21 88.27  
## 3 2 4 5 2 0.41 88.68 0.41 88.68  
## 3 2 5 1 0.21 88.89 0.21 88.89  
## 3 4 3 0.62 89.51 0.62 89.51  
## 3 4 1 1 0.21 89.71 0.21 89.71  
## 3 4 5 3 0.62 90.33 0.62 90.33  
## 3 4 5 2 1 1 0.21 90.53 0.21 90.53  
## 3 5 3 0.62 91.15 0.62 91.15  
## 3 5 4 1 0.21 91.36 0.21 91.36  
## 4 7 1.44 92.80 1.44 92.80  
## 4 1 3 5 3 0.62 93.42 0.62 93.42  
## 4 3 2 5 2 0.41 93.83 0.41 93.83  
## 4 3 5 2 0.41 94.24 0.41 94.24  
## 4 3 5 1 1 0.21 94.44 0.21 94.44  
## 4 5 10 2.06 96.50 2.06 96.50  
## 4 5 1 2 0.41 96.91 0.41 96.91  
## 4 5 3 2 0.41 97.33 0.41 97.33  
## 4 5 3 1 4 0.82 98.15 0.82 98.15  
## 5 6 1.23 99.38 1.23 99.38  
## 5 2 1 0.21 99.59 0.21 99.59  
## 5 3 4 1 0.21 99.79 0.21 99.79  
## 5 4 3 1 0.21 100.00 0.21 100.00  
## <NA> 0 0.00 100.00  
## Total 486 100.00 100.00 100.00 100.00

##split the column  
df9<-df8 %>%   
 bind\_cols(split\_into\_multiple(.$group\_climatechange.cc\_impact," ", "Climate\_impact"))  
##gather the columns  
Climate\_impact<-df9 %>%   
 select(Climate\_impact\_1,Climate\_impact\_2,Climate\_impact\_3,Climate\_impact\_4,Climate\_impact\_5,Climate\_impact\_6) %>%   
 gather(Climate\_impact,type)  
  
##recode the columns  
Climate\_impact$type<-factor(Climate\_impact$type)  
levels(Climate\_impact$type)<-c("Reduced crop yield","Change in planting time","Crop failure","Increased pests and diseases infestation","Flooding of crop fields","Reduced soil moisture","other")  
##frequency  
freq(Climate\_impact$type)

## Frequencies   
## Climate\_impact$type   
## Type: Factor   
##   
## Freq % Valid % Valid Cum. % Total % Total Cum.  
## ---------------------------------------------- ------ --------- -------------- --------- --------------  
## Reduced crop yield 283 22.478 22.478 9.705 9.705  
## Change in planting time 168 13.344 35.822 5.761 15.466  
## Crop failure 219 17.395 53.217 7.510 22.977  
## Increased pests and diseases infestation 252 20.016 73.233 8.642 31.619  
## Flooding of crop fields 289 22.955 96.187 9.911 41.529  
## Reduced soil moisture 46 3.654 99.841 1.578 43.107  
## other 2 0.159 100.000 0.069 43.176  
## <NA> 1657 56.824 100.000  
## Total 2916 100.000 100.000 100.000 100.000

##visualize  
 Climate\_impact%>% drop\_na(type) %>%   
 ggplot(aes(type))+  
 geom\_bar(fill="firebrick",aes(y=(..count..)/sum(..count..)))+  
 geom\_text(aes(y = ((..count..)/sum(..count..)), label = scales::percent((..count..)/sum(..count..))), stat = "count", vjust =0.5, colour = "black")+  
 scale\_y\_continuous(labels = percent)+  
 ylab("percentage")+  
 xlab("Climate change impact")+  
 ggtitle("Barplot of climate change impact")+  
 ggeasy::easy\_center\_title()+  
 coord\_flip()



The most prevalent impact of climate change is reduced crop yield which accounts for 22.61% of the total impacts and is closely followed by flooding of crop fields which accounts for 22.27%.