PAD-ATA Call Log Analysis

Libraries used in the Project and uploading data

```r
library(tidyverse)

library(dplyr)
library(stringr)
library(svMisc)

library(ggplot2)
library(gtable)
library("ggpubr")

log_data <- read_csv("~/Desktop/PAE/github/data/logData.csv")

#MAKE IT ALL UPPER!!!

log_data$logInfo <- toupper(log_data$logInfo)

### GETTING THE DATE MONTH AND YEAR IF I WANT!!!
extractdate <- function(date) {
    day <- format(date, format="%d")
    month <- format(date, format="%m")
    year <- format(date, format="%Y")

    cbind(day, month, year)
}

#making first calls
first_call<-extractdate(log_data$callTime)


#making last call
last_day <-substr(log_data$lastCallTime, 9, 10)
last_month <- substr(log_data$lastCallTime, 6, 7)
last_year <-substr(log_data$lastCallTime, 1, 4)

log_data<-cbind(log_data, first_call,last_day,last_month,last_year)

#Converting columns to numeric
log_data$day <-as.numeric(log_data$day)
log_data$last_day <- as.numeric(log_data$last_day)

## Warning: NAs introduced by coercion

log_data$month <-as.numeric(log_data$month)
log_data$last_month <- as.numeric(log_data$last_month)
```

```
## Warning: NAs introduced by coercion

log_data$year <- as.numeric(log_data$year)
log_data$last_year <- as.numeric(log_data$last_year )

## Warning: NAs introduced by coercion

log_data$year <- log_data$year +log_data$month/12
log_data$last_year <- log_data$last_year + log_data$last_month/12

log_data$caller_lifetime <- log_data$last_year - log_data$year
```

PRELIMINARY DISTRIBUTIONS

In this section, we display some of the basic distributions of the data including languages, total lifeetime of a given caller, colls made, content accessed, and the ratio of content listened to calls made.

Based on the graphs below, we are nervous that some of the data is not random given the incredibly high ratio for content listened to calls made, but given that the data is distributed fairly evenly across thee different graphs, we are unsure

```
lang_spoken <- log_data %>%
  group_by(callerId, langId) %>%
  count()

#Ditribution of languages spoken
ggplot(data=lang_spoken, aes(x=langId)) +
  geom_bar()+
  labs(
    title = 'Distribution of Languages
    1 = Amharic, 2=Oromiffa, 3=Tigrigna, 4=Wolayitta, and 5=Sidamigna',
    x = 'Language',
    y= 'count'
  )
```

Distribution of Languages
1 = Amharic, 2=Oromiffa, 3=Tigrigna, 4=Wolayitta, a[...]
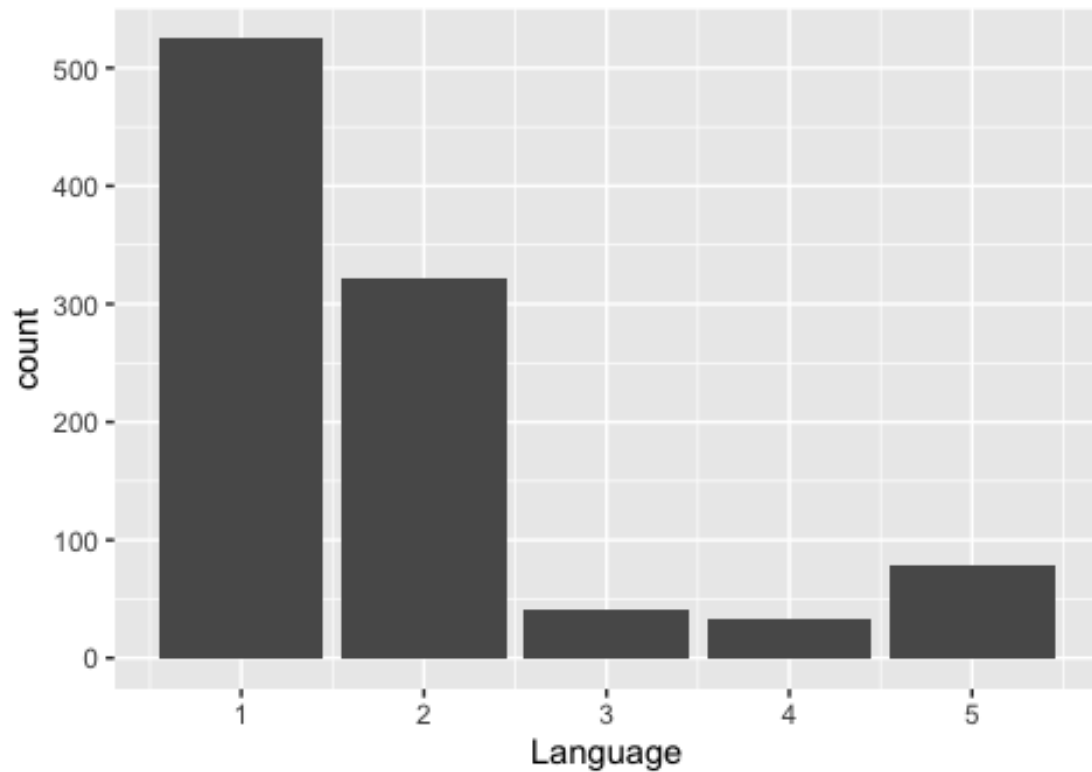

```r
#Ditribution otiem difference beetween firs tand last call

lifetime<- log_data %>%
  group_by(callerId, caller_lifetime) %>%
  count()

ggplot(data=lifetime, aes(x=caller_lifetime)) +
  geom_histogram(fill='brown', color='black')+
  labs(
    title = 'Distribution of Call Lifetimes',
    x = 'Years between First and Last Call',
    y= 'count'
  )
```

Distribution of Call Lifetimes
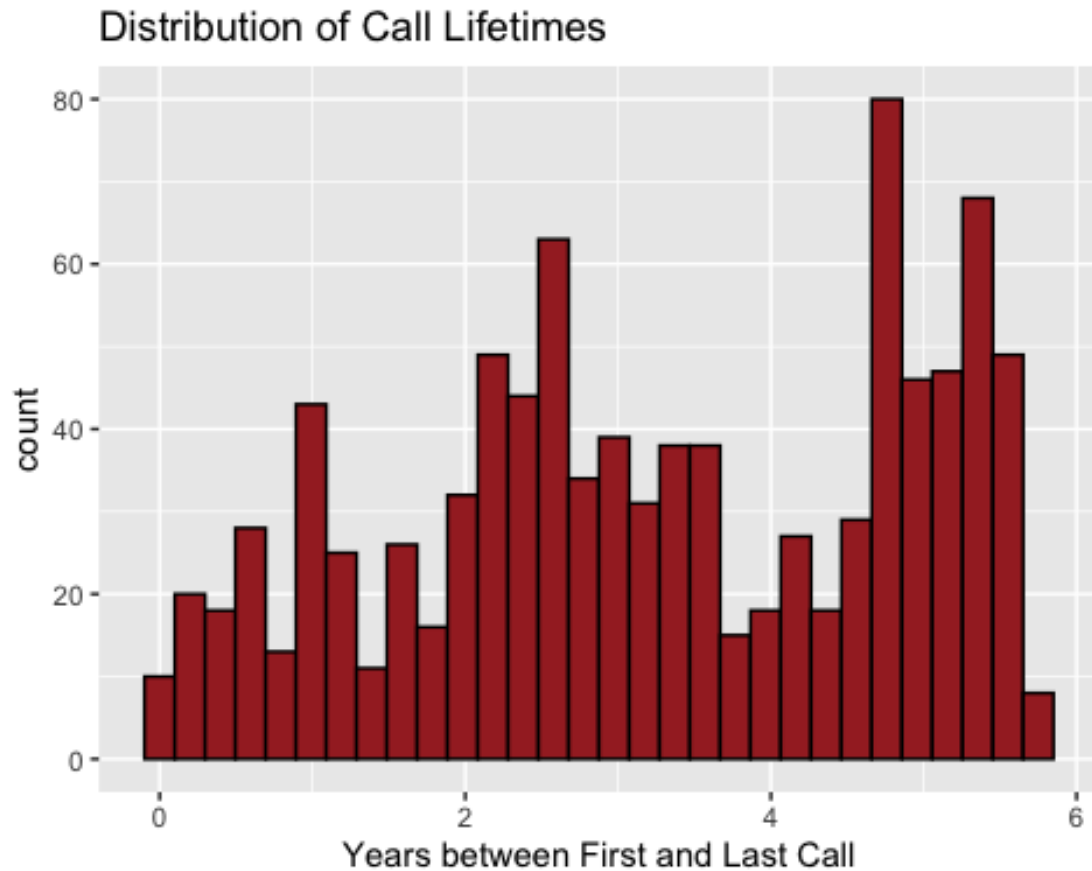

```
#distribution of calls made

calls_made <- log_data %>%
  group_by(callerId, noCallsMade) %>%
  count()

ggplot(calls_made, aes(x=noCallsMade)) +
  geom_histogram(color="black", fill="blue") +
  xlim(0,350) +
  labs(
    title = 'Distribution of calls made per individual (Outliers Excluded) \n
Min. 1st Qu.  Median  Mean   3rd Qu.    Max.
    3.0    45.0    69.0    99.9    106.0    7110.0 ',
    x = 'Total number of calls'
  )
```

Distribution of calls made per individual (Outliers Excluc

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3.0	45.0	69.0	99.9	106.0	7110.0

Total number of calls
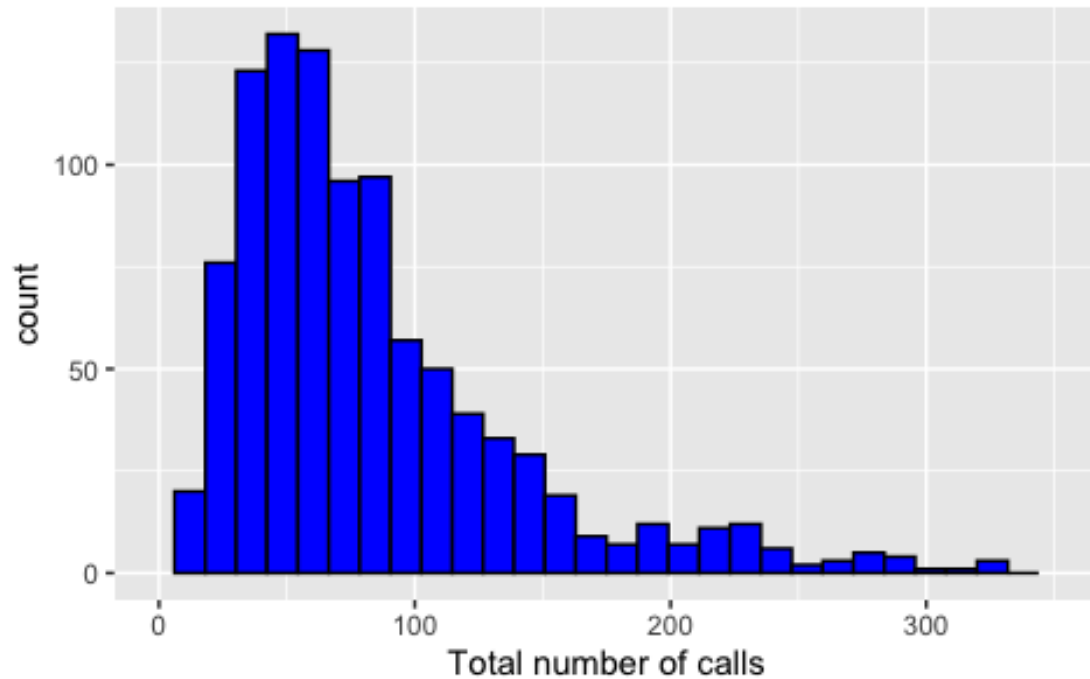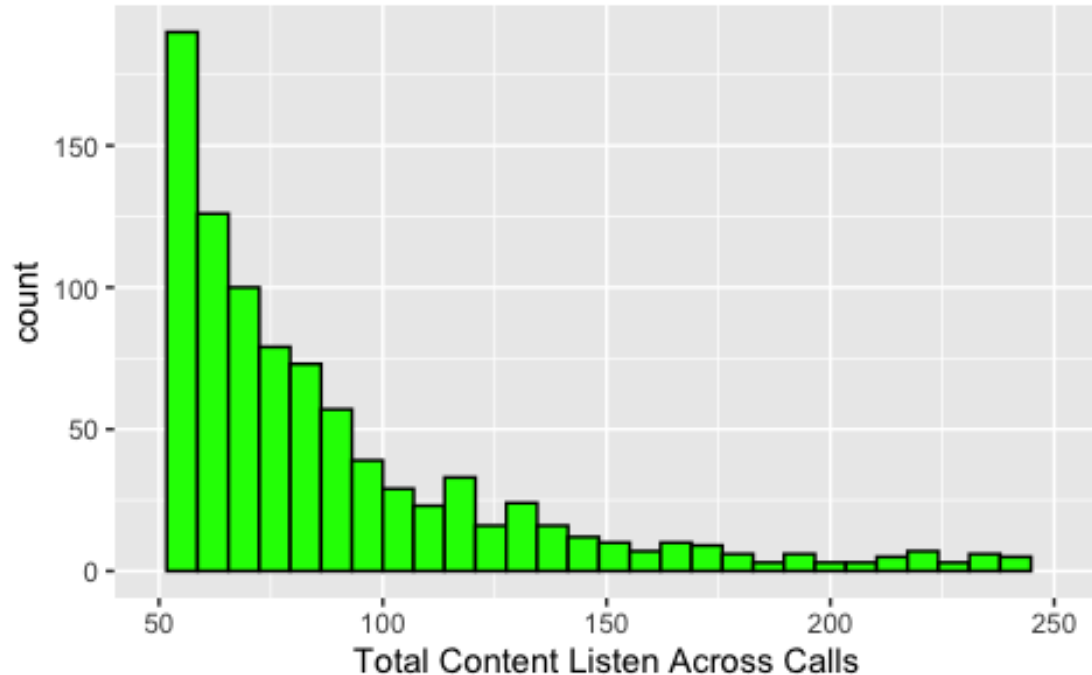
```r
#Distribution of contenet listened to

content_listened <- log_data %>%
  group_by(callerId, noContentListened) %>%
  count()

#summary(content_listened$noContentListened)

ggplot(content_listened, aes(x=noContentListened)) +
  geom_histogram(color="black", fill="green") +
  xlim(50, 250) +
  labs(
    title = 'Distribution of total Content listened to all Across all Calls (
Outliers Excluded) \n
  Min. 1st Qu.  Median  Mean  3rd Qu.    Max. n\
  50.00   58.00   74.50   99.97  105.00   2367.00 ',
    x = 'Total Content Listen Across Calls'
  )
```

Distribution of total Content listened to all Across all Ca

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	n
50.00	58.00	74.50	99.97	105.00	2367.00	


```r
# Ratio of content per call - creating the ratio and graphing the results
content_call_ratio_df <- inner_join(calls_made,content_listened,by="callerId"
)
content_call_ratio_df$ratio <- content_call_ratio_df$noContentListened/conten
t_call_ratio_df$noCallsMade

#summary(content_call_ratio_df$ratio)

ggplot(content_call_ratio_df, aes(x=ratio)) +
  geom_histogram(color="black", fill="red") +
  xlim(0, 5) +
  labs(
    title = 'Distribution of ratio of content listened to Calls (Outliers Exc
luded) \n
    Min.       1st Qu.    Median     Mean     3rd Qu.      Max.
 0.05963    0.84598  1.18586  1.40733  1.69773    22.33333 ',
    x = 'Content to Call Ratio'
  )

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
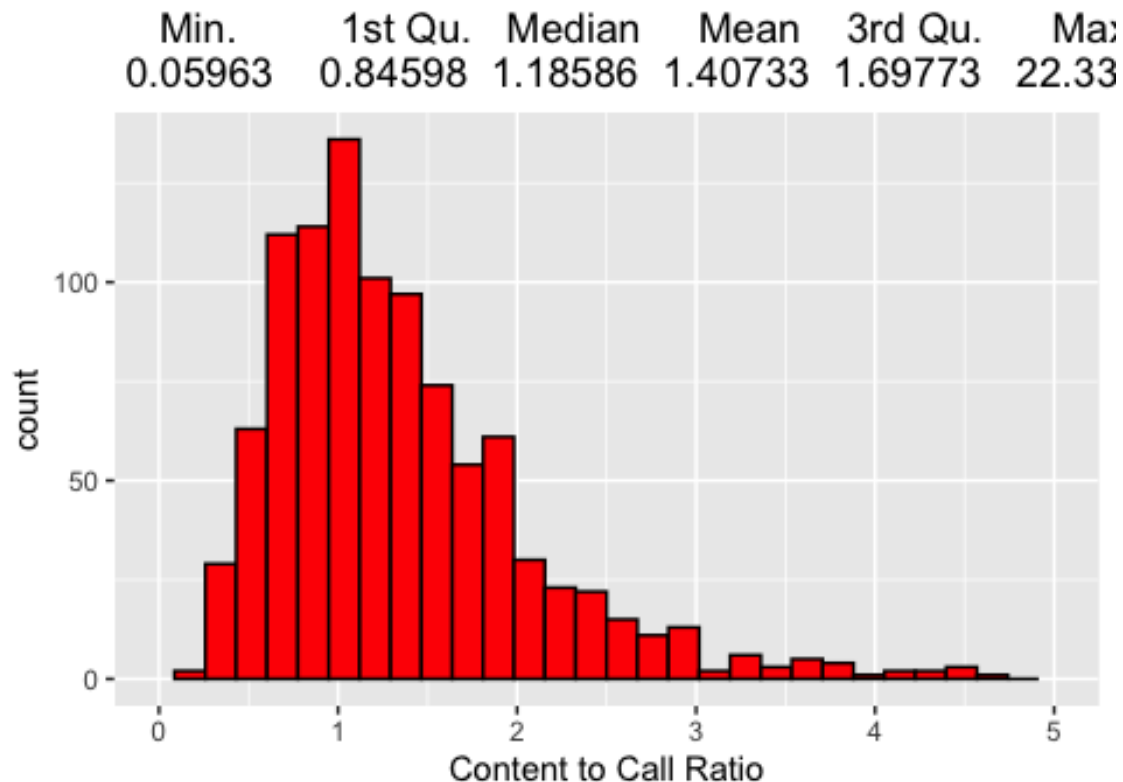
```
## Warning: Removed 11 rows containing non-finite values (stat_bin).

## Warning: Removed 2 rows containing missing values (geom_bar).
```


Distribution of ratio of content listened to Calls (Outliers

Min.	1st Qu.	Median	Mean	3rd Qu.	Ma:
0.05963	0.84598	1.18586	1.40733	1.69773	22.33

Next, we want to Breakd down the initial Menu options and delve into the distributions of the selections that the farmer can make at a different state in the IVR System

For the first analysis, we looked at the distribution of the TOP MENU in the data. This comes in as follows:

INCOMING CALL STARTED –> WELCOME MESSAGE PLAYED –> ASSINGED TO EXPERIMENT (OPTIONAL) –> TOP MENU

TOP MENU

- RAIN OPTION
- HHI OPTION (household Irrigation)
- RESET PROFILE OPTION
- TOP MENU REPLAY
- LIVESTOCK OPTION

- COVID OPTION

Please note, we are unsure of the order these are presented. For other menus, we do have the order

Below, we have visualized the number of farmers, out of our sample, that in the life time of their calling history have accessed a given TOP MENU item at least once

Additionally, we also visualized this same selection divided by language to see if the patterns of seleciont are similar or different across langaugesv

```r
# filtering out top Menu keys

top_menu <-filter(log_data, grepl("TOP MENU",logInfo))

#cleaning the axis
top_menu$logInfo<- str_replace(top_menu$logInfo, "TOP MENU - ", "")
top_menu$logInfo<- str_replace(top_menu$logInfo, " PRESSED", "")
top_menu$logInfo<- str_replace(top_menu$logInfo, " PRESSED", "")
top_menu$logInfo<- str_replace(top_menu$logInfo, " SELECTED", "")

top_menu <- top_menu %>%
  group_by(callerId, logInfo, langId) %>%
  count()


ggplot(top_menu, aes(logInfo)) +
  theme(axis.text.x = element_text(angle = 30, hjust =1))+
  geom_bar()+
  labs(
    title = 'Distirution of TOP MENU items selected at least once by a unique caller',
    x = 'TOP MENU options'
  )
```
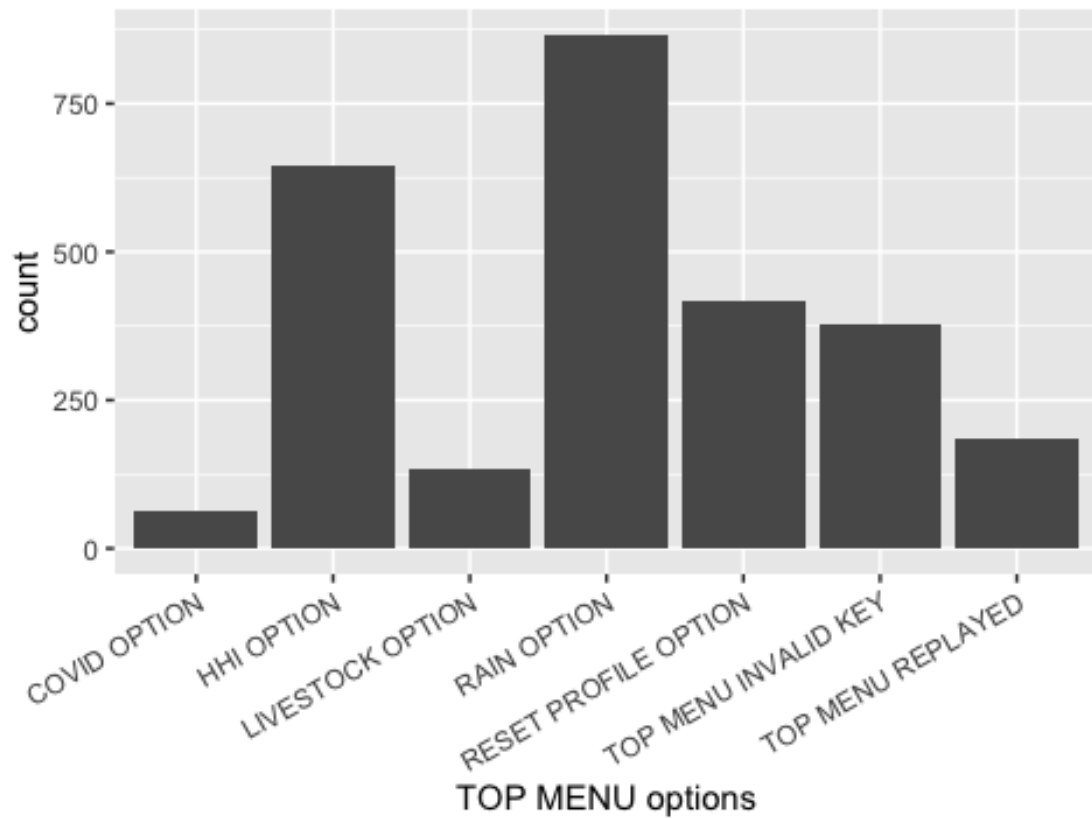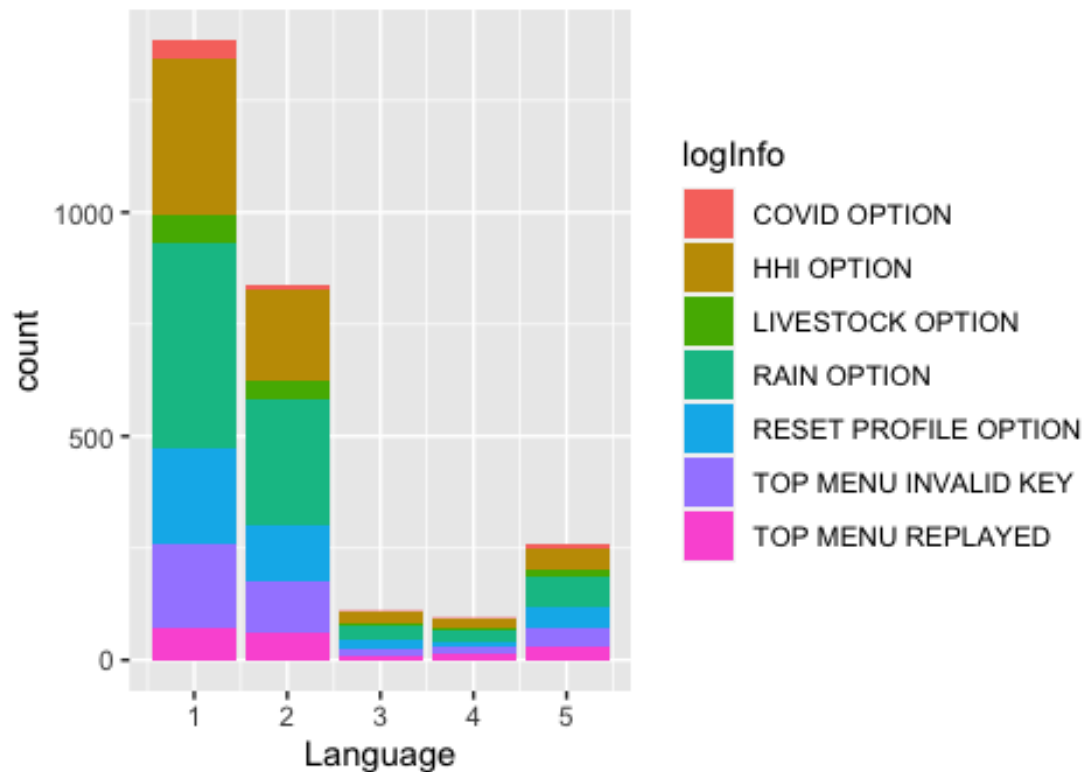
Distirution of TOP MENU items selected at least once b


```
ggplot(top_menu, aes(langId, fill = logInfo)) +
  geom_bar()+
  labs(
    title = 'Distirution of Languages divided by TOP MENU option
    1 = Amharic, 2=Oromiffa, 3=Tigrigna, 4=Wolayitta, and 5=Sidamigna',
    x = 'Language'
  )
```

Distirution of Languages divided by TOP MENU option
1 = Amharic, 2=Oromiffa, 3=Tigrigna, 4=Wolayitta, a

As we can easily see, the RAIN options and the HHI option dominate the choices from the top Menu. So, we decided to dive deeper into each of these menus to see if there are any outlier selections that are deep within the system. Additionally, ths distribution across languages seems to be fairly consistent at face value. We can run tests in the future, but for EDA this will be adequate. For 3/4 langauge options - having more data would be apprecaited

RAIN MENU

INCOMING CALL STARTED –> WELCOME MESSAGE PLAYED –> RANDOM EXPERIMENT –> TOP MENU –> RAIN MENU

Here, we look at the breakdown of individuals who have selected a given option in the rain menu at least once. In the data, MAIN MENU == RAIN MENU.

RAIN MENU

- PRE PLANTING OPTION
- PLANTING OPTION
- CROP PROTECTION OPTION
- FERTILIZGER SIDE DREESSING
- POST HARVEST AND PROCESSING

- REPEAT RAIN MENU
- RETURN TO TOP MENU
- INVALID KEY PRESSED (ADDITIONAL OUTCOME)

```r
# Pulling out the rain Menu

#RAIN == HHI_RAIN
rain_menu <- filter(log_data, grepl("MAIN MENU - ",logInfo))
rain_menu <- filter(rain_menu, !grepl("LIVESTOCK MAIN MENU",logInfo))
rain_menu <- filter(rain_menu, !grepl("HHI MAIN MENU",logInfo))
rain_menu <- filter(rain_menu, !grepl("COVID-19 MAIN MENU",logInfo))

#grouping the observations
rain_menu <-rain_menu %>%
  group_by(callerId, logInfo, langId) %>%
  count()

rain_menu$logInfo<- str_replace(rain_menu$logInfo, "MAIN MENU - ", "")
rain_menu$logInfo<- str_replace(rain_menu$logInfo, " SELECTED", "")
rain_menu$logInfo<- str_replace(rain_menu$logInfo, " PRESSED", "")
rain_menu$logInfo<- str_replace(rain_menu$logInfo, " OPTION", "")

#plotting rain menu breakdown
ggplot(data=rain_menu, aes(x=logInfo)) +
  theme(axis.text.x = element_text(angle = 30, hjust =1))+
  geom_bar(fill='blue')+
  labs(
    title = 'Breakdown for Individuals By Content Selected at least Once from
the rain menu',
    x = 'Content Selected',
    y= 'count'
  )
```
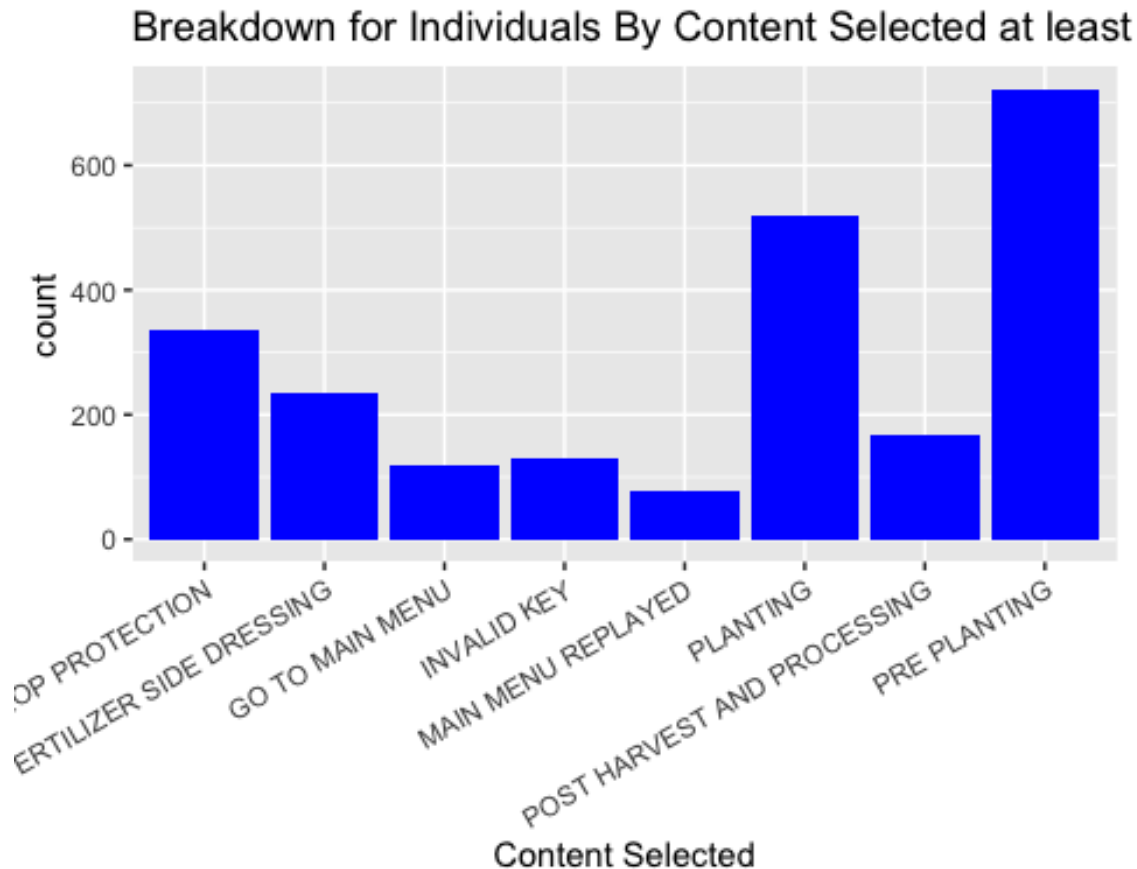
Breakdown for Individuals By Content Selected at least

We can see that the pre-planting menu, which still has over 75% of unique users accesss it at least once expereinces incredibly high usage. So, we now dive into this menu. We will breakdown conent selection by language and overall

RAIN MENU/PRE PLANTING MENU

INCOMING CALL STARTED –> WELCOME MESSAGE PLAYED –> RANDOM EXPERIMENT –>

TOP MENU –> RAIN MENU –> PRE PLANTING MENU

Please note, the pre planting menu under the rain menu is referred to as MENU 1 in the data

PRE PLANTING MENU

• LAND PREPARATION
• SEED VARIETY
• REPEAT MENU
• GO TO MAIN MENU (RAIN MENU)
• INVALID KEY (ADDITIONAL OUTCOME)

```
#filtering out MENU 1 options as the pre planting
MENU_1 <- filter(log_data, grepl("MENU 1",logInfo))
```

```r
MENU_1 <- filter(MENU_1, !grepl("HHIMENU 1",logInfo))
MENU_1 <- filter(MENU_1, !grepl("APICULTURE MENU",logInfo))
MENU_1 <- filter(MENU_1, !grepl("APICULTURE SUB4 MENU",logInfo))
MENU_1 <- filter(MENU_1, !grepl("DAIRY MENU",logInfo))
MENU_1 <- filter(MENU_1, !grepl("SMALL-SCALE SUB5 MENU",logInfo))
MENU_1 <- filter(MENU_1, !grepl("DAIRY SUB2 MENU",logInfo))
MENU_1 <- filter(MENU_1, !grepl("FATTENING MANU",logInfo))

MENU_1 <- filter(MENU_1, !grepl("FATTENING SUB1 MENU",logInfo))
MENU_1 <- filter(MENU_1, !grepl("FATTENING SUB2 MENU",logInfo))
MENU_1 <- filter(MENU_1, !grepl("HOUSEHOLD MENU",logInfo))
MENU_1 <- filter(MENU_1, !grepl("HOUSEHOLD SUB1 MENU ",logInfo))


MENU_1 <- MENU_1 %>%
  group_by(callerId, logInfo, langId) %>%
  count()

MENU_1$logInfo<- str_replace(MENU_1$logInfo, "MENU 1 - ", "")
MENU_1$logInfo<- str_replace(MENU_1$logInfo, " PRESSED", "")
MENU_1$logInfo<- str_replace(MENU_1$logInfo, " SELECTED", "")

# plotting
ggplot(data=MENU_1, aes(x=logInfo)) +
  theme(axis.text.x = element_text(angle = 25, hjust= 1))+
  geom_bar(fill='cyan')+
  labs(
    title = 'Individuals By Content Selected at least Once from
    RAIN MENU/PRE PLANTING OPTION',
    x = 'Content Selected',
    y= 'count'
  )
```
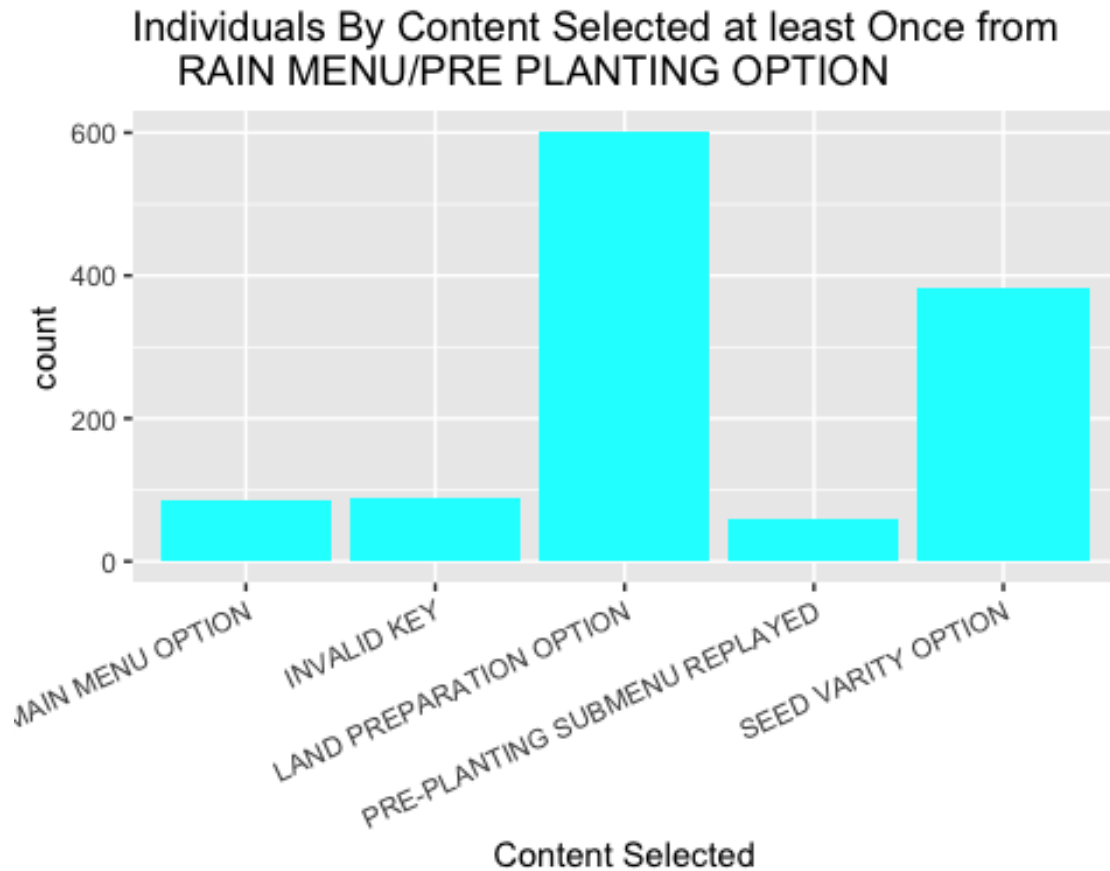
Individuals By Content Selected at least Once from RAIN MENU/PRE PLANTING OPTION

RAIN MENU/PRE PLANTING MENU/LAND PREPARATION

We continue our pattern of following the most accessed submenu from the previous subment - land preparation! This is also the bottom of a tree. We will breakdown conent selection by language and overall

INCOMING CALL STARTED –> WELCOME MESSAGE PLAYED –> RANDOM EXPERIMENT –>

TOP MENU –> RAIN MENU –> PRE PLANTING MENU –> LAND PREPARATION

LAND PREPARATION

- BARLEY
- MAIZE
- SORGHUM
- TEF
- WHEAT
- SESAME
- FAVA BEAN
- CHICKPEA
- COMMON BEAN

- COTTON

ORDER UNKOWN:

- RICE
- DURAME WHEAT
- LENTIL
- MALT BARLEY

```r
#GETTING THE CROP MENU FROM MENU 1 (rain/preplanting/land preparation sub menu)
land_prep <- filter(log_data, grepl("CONTENT PLAYED - LAND PREPARATION - ",logInfo))

land_prep_final <- land_prep %>%
  group_by(callerId, logInfo, langId) %>%
  count()

land_prep_final$logInfo<- str_replace(land_prep_final$logInfo, "CONTENT PLAYED - LAND PREPARATION -", "")

ggplot(data=land_prep_final, aes(x=logInfo)) +
  theme(axis.text.x = element_text(angle = 45, hjust=1))+
  geom_bar()+
  labs(
    title = 'RAIN MENU/PRE PLANTING/LAND PREPARATION
    content actually played by language',
    x = 'Content Played',
    y= 'count'
  )
```
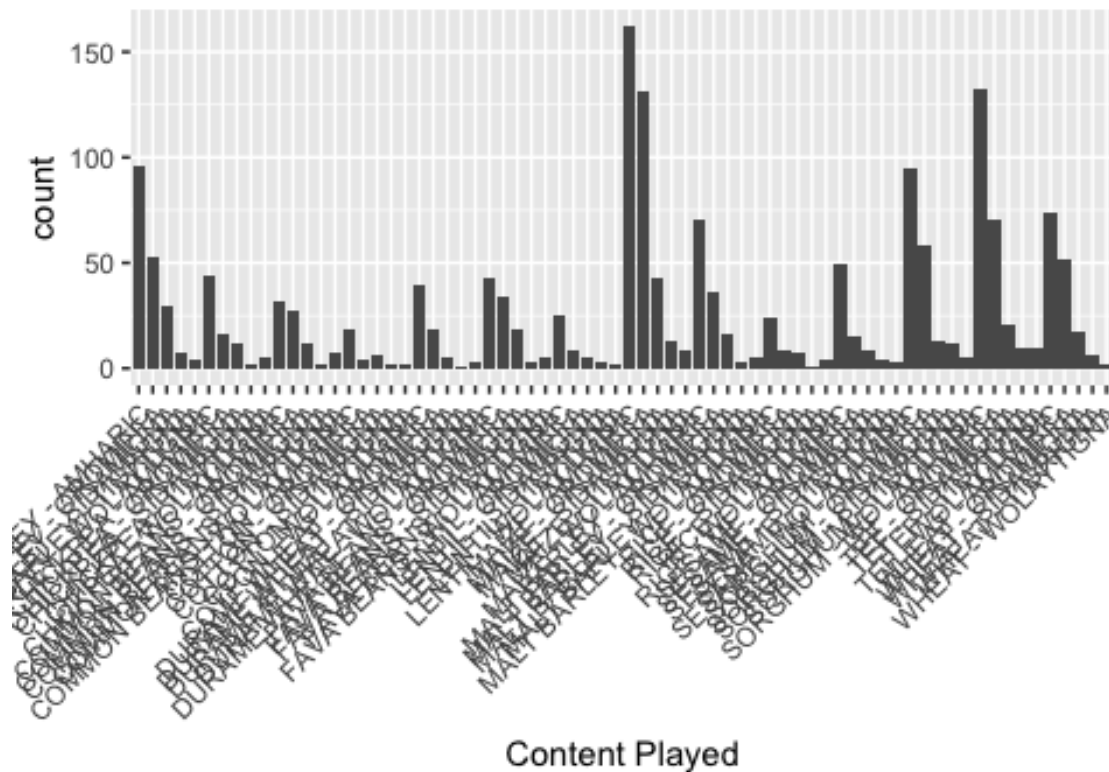
RAIN MENU/PRE PLANTING/LAND PREPARATION
content actually played by language
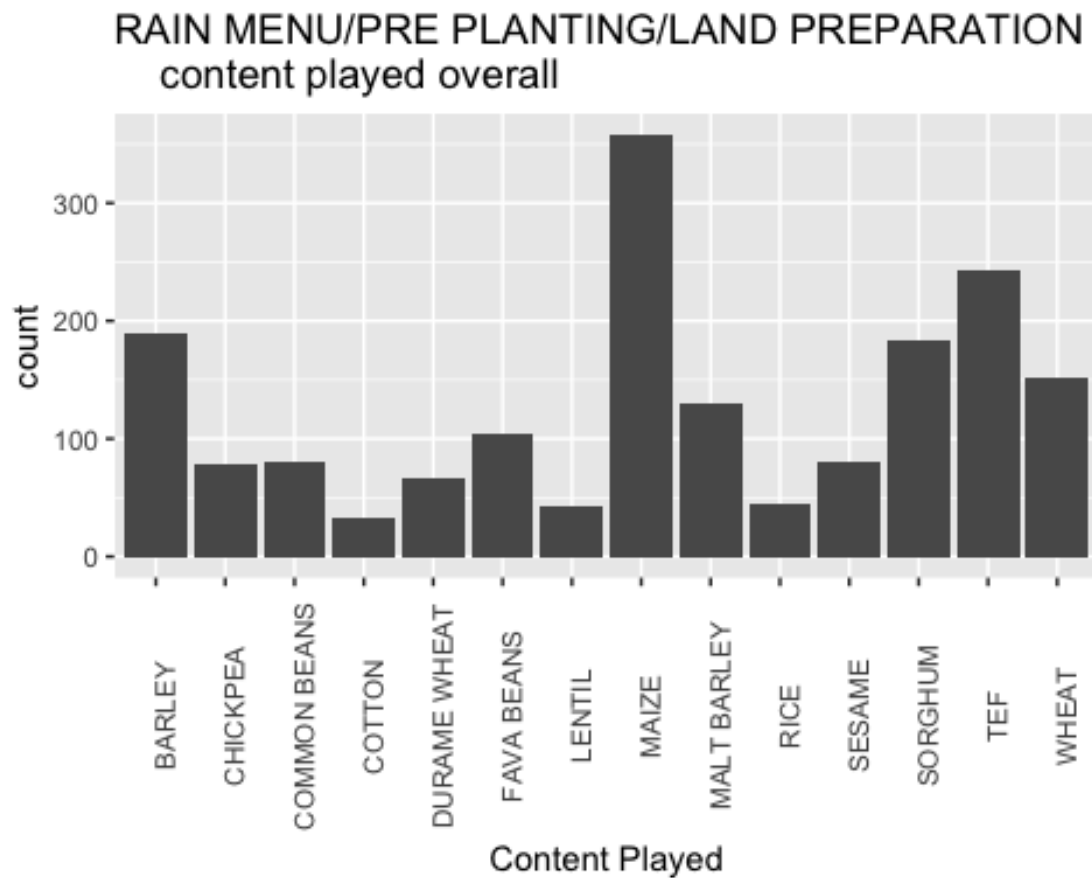
Content Played

```
#get rid of the language breakdown
land_prep_no_lang <- land_prep_final

land_prep_no_lang$logInfo<- str_replace(land_prep_no_lang$logInfo, "- AMHARIC
", "")
land_prep_no_lang$logInfo<- str_replace(land_prep_no_lang$logInfo, "- OROMIFF
A", "")
land_prep_no_lang$logInfo<- str_replace(land_prep_no_lang$logInfo, "- TIGRIGN
A", "")
land_prep_no_lang$logInfo<- str_replace(land_prep_no_lang$logInfo, "- WOLAYIT
TA", "")
land_prep_no_lang$logInfo<- str_replace(land_prep_no_lang$logInfo, "- WOLAYTI
GNA", "")
land_prep_no_lang$logInfo<- str_replace(land_prep_no_lang$logInfo, "- SIDAMIG
NA", "")

ggplot(land_prep_no_lang, aes(x=logInfo)) +
  theme(axis.text.x = element_text(angle = 90))+
  geom_bar()+
  labs(
    title = 'RAIN MENU/PRE PLANTING/LAND PREPARATION
    content played overall',
    x = 'Content Played',
```

```
    y= 'count'
)
```

RAIN MENU/PRE PLANTING/LAND PREPARATION
content played overall


```
ggplot(land_prep_no_lang, aes(langId, fill = logInfo)) +
  geom_bar()+
  labs(
    title = 'ALTERNATIVE RAIN MENU/PRE PLANTING/LAND PREPARATION
    content actually played by language
    1 = Amharic, 2=Oromiffa, 3=Tigrigna, 4=Wolayitta, and 5=Sidamigna
    ',
    x = 'Language'
)
```

ALTERNATIVE RAIN MENU/PRE PLANTING/LAND PI
content actually played by language
1 = Amharic, 2=Oromiffa, 3=Tigrigna, 4=Wolayitta, a
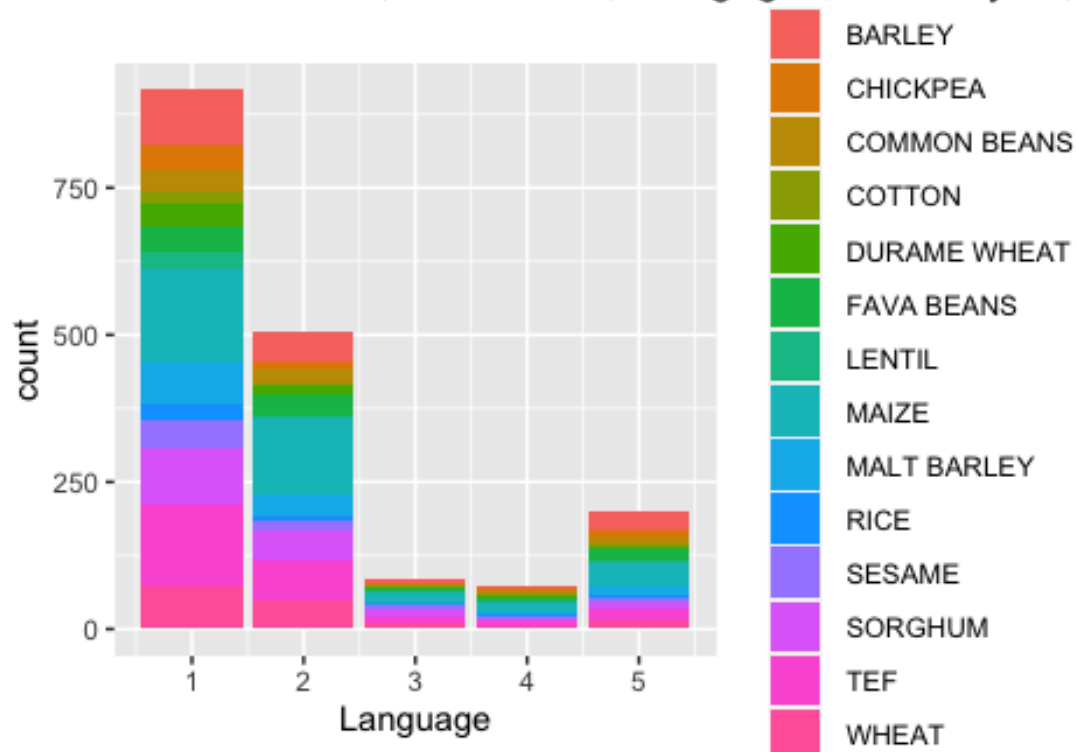
BIG TAKEAWAY: 35% of users end up accessing the maize menu at some point! Wheat, Tef, Sorghum, and Barley also appear to be qualitatively significant

There is no menu to go deeper in, so let's go back up a level and check to see if other branches have a lot of access

RAIN MENU/PRE PLANTING MENU/SEED VARIETY

INCOMING CALL STARTED –> WELCOME MESSAGE PLAYED –> RANDOM EXPERIMENT –>

TOP MENU –> RAIN MENU –> PRE PLANTING MENU –> SEED VARIETY
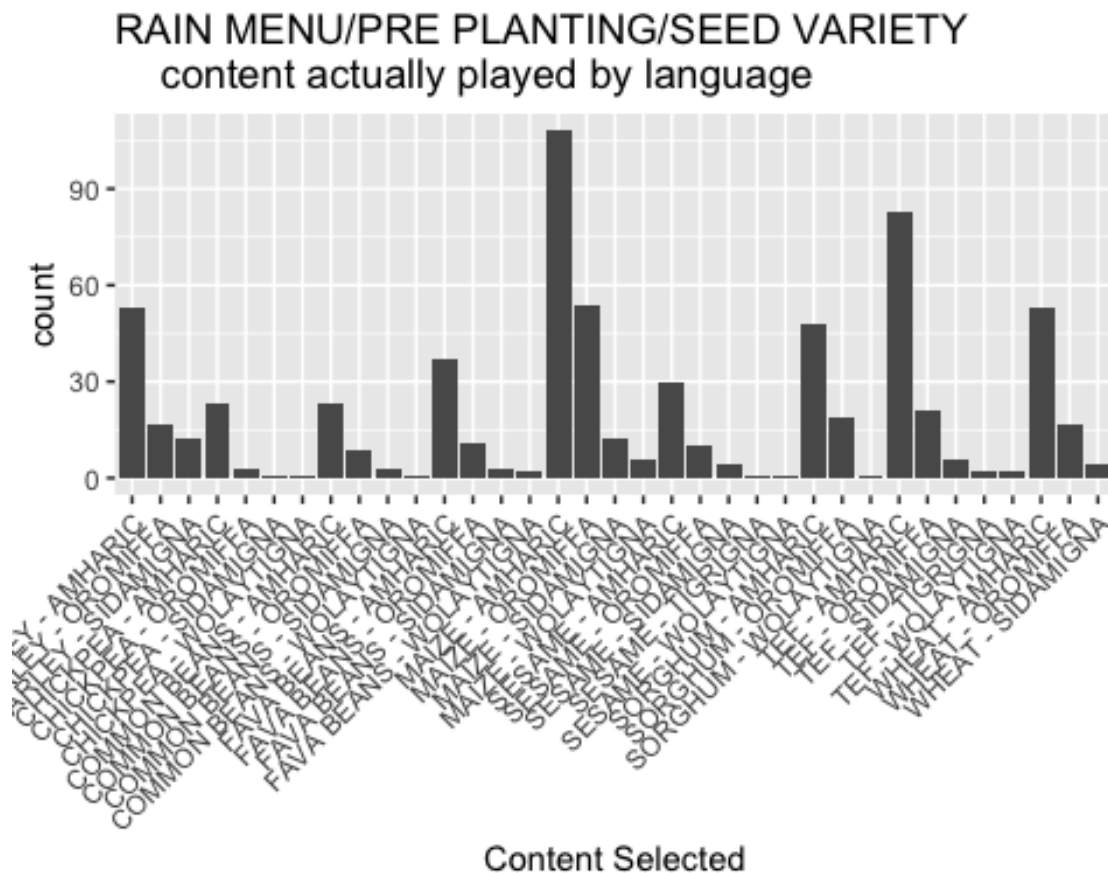
SEED VARIETY

- BARLEY
- MAIZE
- SORGHUM
- TEF
- WHEAT
- SESAME
- FAVA BEAN
- CHICKPEA

- COMMON BEAN

```r
#GETTING THE CROP MENU FROM MENU 1 (rain/preplanting/seed variety sub menu)
seed_variety <- filter(log_data, grepl("CONTENT PLAYED - SEED VARITY - ",logInfo))

seed_variety_final <- seed_variety %>%
  group_by(callerId, logInfo, langId) %>%
  count()

seed_variety_final$logInfo<- str_replace(seed_variety_final$logInfo, "CONTENT PLAYED - SEED VARITY - ", "")

ggplot(data=seed_variety_final, aes(x=logInfo)) +
  theme(axis.text.x = element_text(angle = 45, hjust=1))+
  geom_bar()+
  labs(
    title = 'RAIN MENU/PRE PLANTING/SEED VARIETY
    content actually played by language',
    x = 'Content Selected',
    y= 'count'
  )
```

```r
#get rid of the language breakdown
seed_variety_no_lang <- seed_variety_final

seed_variety_no_lang$logInfo<- str_replace(seed_variety_no_lang$logInfo, "- A
MHARIC", "")
seed_variety_no_lang$logInfo<- str_replace(seed_variety_no_lang$logInfo, "- O
ROMIFFA", "")
seed_variety_no_lang$logInfo<- str_replace(seed_variety_no_lang$logInfo, "- T
IGRIGNA", "")
seed_variety_no_lang$logInfo<- str_replace(seed_variety_no_lang$logInfo, "- W
OLAYITTA", "")
seed_variety_no_lang$logInfo<- str_replace(seed_variety_no_lang$logInfo, "- W
OLAYTIGNA", "")
seed_variety_no_lang$logInfo<- str_replace(seed_variety_no_lang$logInfo, "- S
IDAMIGNA", "")

#alternative with langauge view
ggplot(seed_variety_no_lang, aes(langId, fill = logInfo)) +
  geom_bar()+
  labs(
    title = 'ALTERNATIVE RAIN MENU/PRE PLANTING/SEE VARIETY
    content actually played by language
    1 = Amharic, 2=Oromiffa, 3=Tigrigna, 4=Wolayitta, and 5=Sidamigna
    ',
    x = 'Language'
  )
```
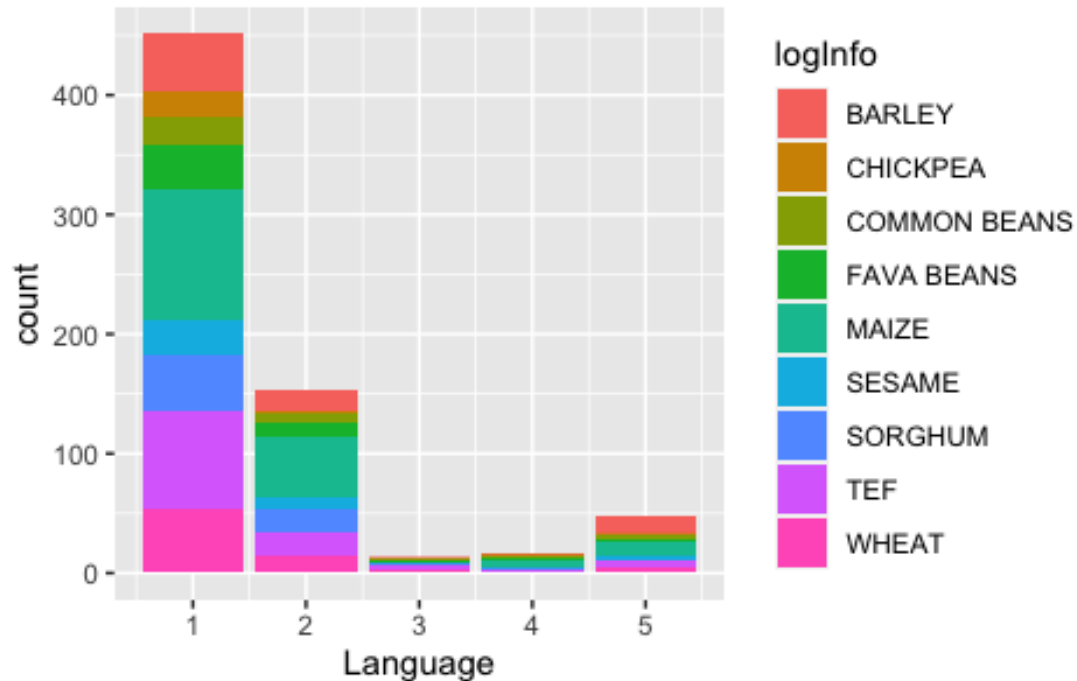
ALTERNATIVE RAIN MENU/PRE PLANTING/SEE VAI
content actually played by language
1 = Amharic, 2=Oromiffa, 3=Tigrigna, 4=Wolayitta, al


```
#no language - overall
ggplot(seed_variety_no_lang, aes(x=logInfo)) +
  theme(axis.text.x = element_text(angle = 30, hjust=1))+
  geom_bar()+
  labs(
    title = 'RAIN MENU/PRE PLANTING/seed variety
    content actually played',
    x = 'Content Selected',
    y= 'count'
  )
```

RAIN MENU/PRE PLANTING/seed variety
content actually played
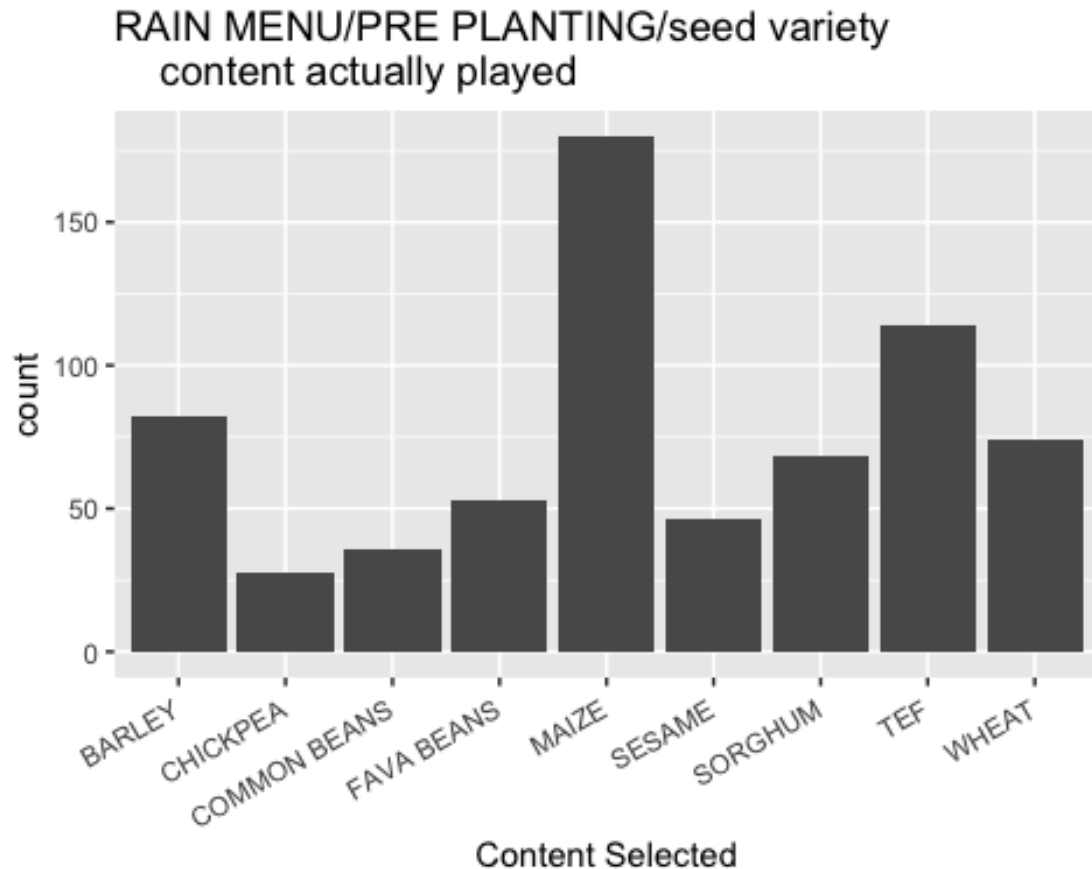
We have mostly exhausted the pre-planting option menu, and if we go back up to the original rain menu (aka MAIN MENU), we can see that planting also has a large porportion of users - over half

RAIN MENU/PLANTING MENU

INCOMING CALL STARTED –> WELCOME MESSAGE PLAYED –> RANDOM EXPERIMENT –>

TOP MENU –> RAIN MENU –> PLANTING MENU

The planting option as a subsection of the rain option is known as MENU 2 in the data

PLANTING MENU

• SEED RATE
• SOIL DEPTH PLANTING
• TRANSPLANTING
• MOISTURE CONSERVATION
• FERTILIZER APPLICATION
• REPLAY MENY
• RETURN TO RAIN MENY
• INVALID KEY (ADDITIONAL OUTCOME)

```r
#PLANTING OPTION - second most accessed MENU 2
MENU_2 <- filter(log_data, grepl("MENU 2",logInfo))
MENU_2 <- filter(MENU_2, !grepl("HHIMENU 2",logInfo))
MENU_2 <- filter(MENU_2, !grepl("APICULTURE MENU",logInfo))
MENU_2 <- filter(MENU_2, !grepl("APICULTURE SUB2 MENU",logInfo))
MENU_2 <- filter(MENU_2, !grepl("DAIRY MENU",logInfo))
MENU_2 <- filter(MENU_2, !grepl("FATTENING MANU",logInfo))

MENU_2 <- MENU_2 %>%
  group_by(callerId, logInfo, langId) %>%
  count()

#cleaning
MENU_2$logInfo<- str_replace(MENU_2$logInfo, "MENU 2 - ", "")
MENU_2$logInfo<- str_replace(MENU_2$logInfo, " SELECTED", "")
MENU_2$logInfo<- str_replace(MENU_2$logInfo, " PRESSED", "")
MENU_2$logInfo<- str_replace(MENU_2$logInfo, " OPTION", "")


#plotting
ggplot(data=MENU_2, aes(x=fct_infreq(logInfo))) +
  theme(axis.text.x = element_text(angle = 45, hjust=1))+
  geom_bar(fill='olivedrab')+
  labs(
    title = 'Individuals By Content Selected at least Once from
    RAIN MENU/PLANTING OPTION',
    x = 'Content Selected',
    y= 'count'
  )
```

Individuals By Content Selected at least Once from RAIN MENU/PLANTING OPTION
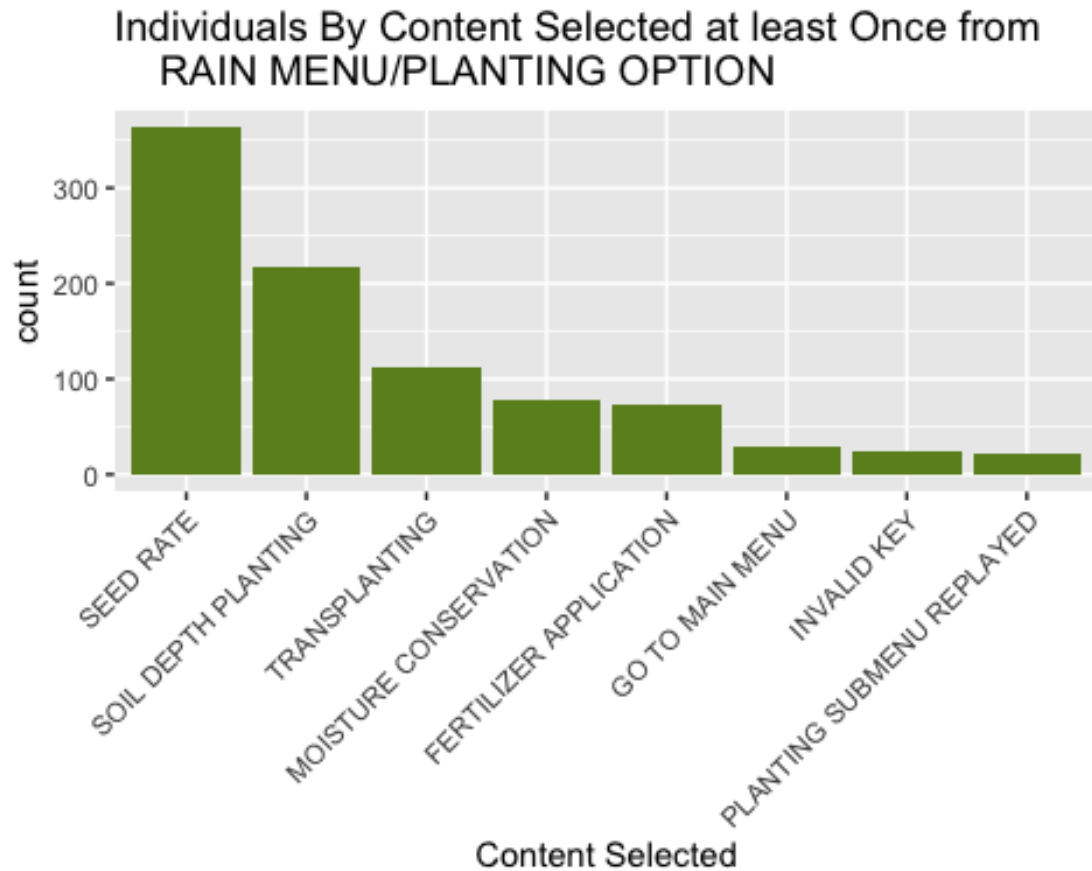
As per usual, we will go into the most common selection: seed rate!

RAIN MENU/PLANTING MENU/SEED RATE

INCOMING CALL STARTED –> WELCOME MESSAGE PLAYED –> RANDOM EXPERIMENT –>

TOP MENU –> RAIN MENU –> PLANTING MENU –> SEED RATE

SEED RATE

- BARLEY
- MAIZE
- SORGHUM
- TEF
- WHEAT
- SESAME
- FAVA BEAN
- CHICKPEA
- COMMON BEAN
- COTTON

ORDER UNKOWN:

- RICE
- DURAME WHEAT
- LENTIL
- MALT BARLEY

```r
seed_rate <- filter(log_data, grepl("CONTENT PLAYED - SEED RATE",logInfo))

seed_rate_final <- seed_rate %>%
  group_by(callerId, logInfo, langId) %>%
  count()

seed_rate_final$logInfo<- str_replace(seed_rate_final$logInfo, "CONTENT PLAYE
D - SEED RATE -", "")

ggplot(data=seed_rate_final, aes(x=logInfo)) +
  theme(axis.text.x = element_text(angle = 90))+
  geom_bar()+
  labs(
    title = 'RAIN MENU/PLANTING/SEED RATE
    content actually played by language',
    x = 'Content Selected',
    y= 'count'
  )
```


RAIN MENU/PLANTING/SEED RATE
content actually played by language
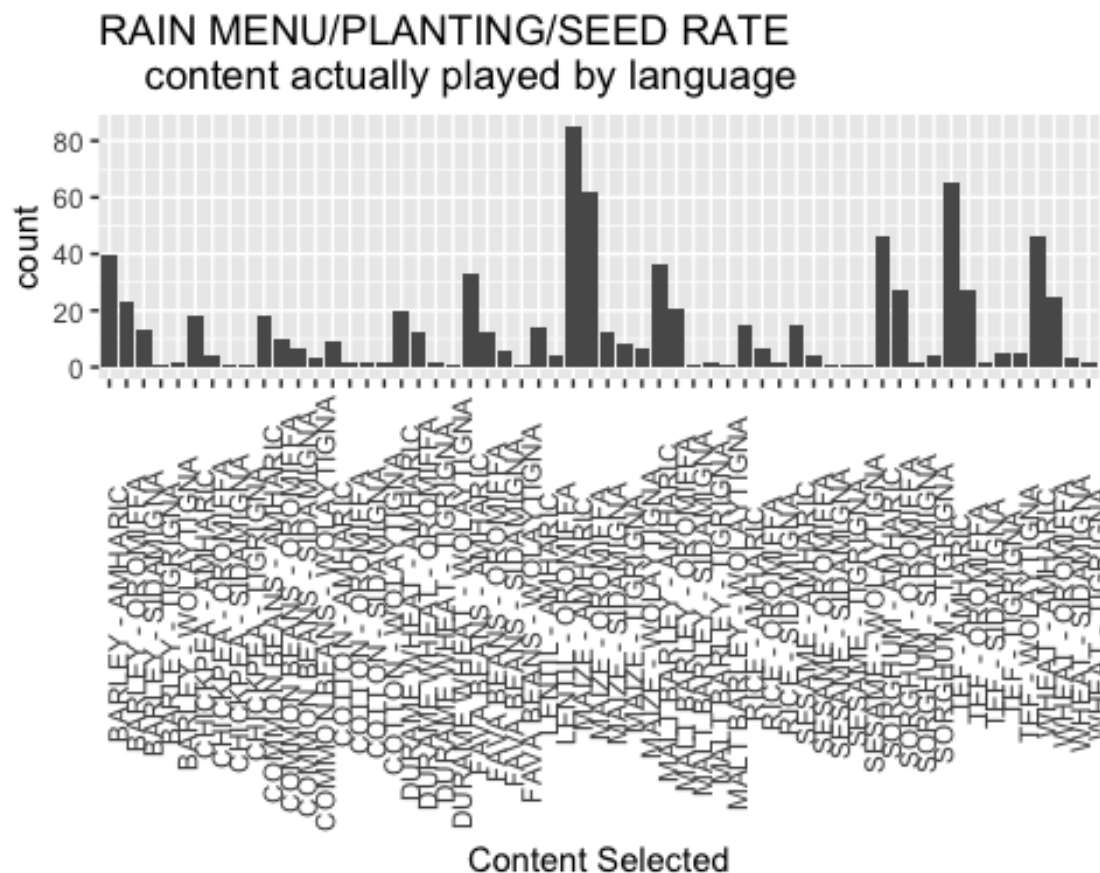
```
#get rid of the language breakdown
seed_rate_no_lang <- seed_rate_final

seed_rate_no_lang$logInfo<- str_replace(seed_rate_no_lang$logInfo, "- AMHARIC
", "")
seed_rate_no_lang$logInfo<- str_replace(seed_rate_no_lang$logInfo, "- OROMIFF
A", "")
seed_rate_no_lang$logInfo<- str_replace(seed_rate_no_lang$logInfo, "- TIGRIGN
A", "")
seed_rate_no_lang$logInfo<- str_replace(seed_rate_no_lang$logInfo, "- WOLAYIT
TA", "")
seed_rate_no_lang$logInfo<- str_replace(seed_rate_no_lang$logInfo, "- WOLAYTI
GNA", "")
seed_rate_no_lang$logInfo<- str_replace(seed_rate_no_lang$logInfo, "- SIDAMIG
NA", "")

#alternative with langauge view
ggplot(seed_rate_no_lang, aes(langId, fill = logInfo)) +
  geom_bar()+
  labs(
    title = 'ALTERNATIVE RAIN MENU/PLANTING/SEED RATE
    content actually played by language
    1 = Amharic, 2=Oromiffa, 3=Tigrigna, 4=Wolayitta, and 5=Sidamigna
    ',
    x = 'Language'
  )
```
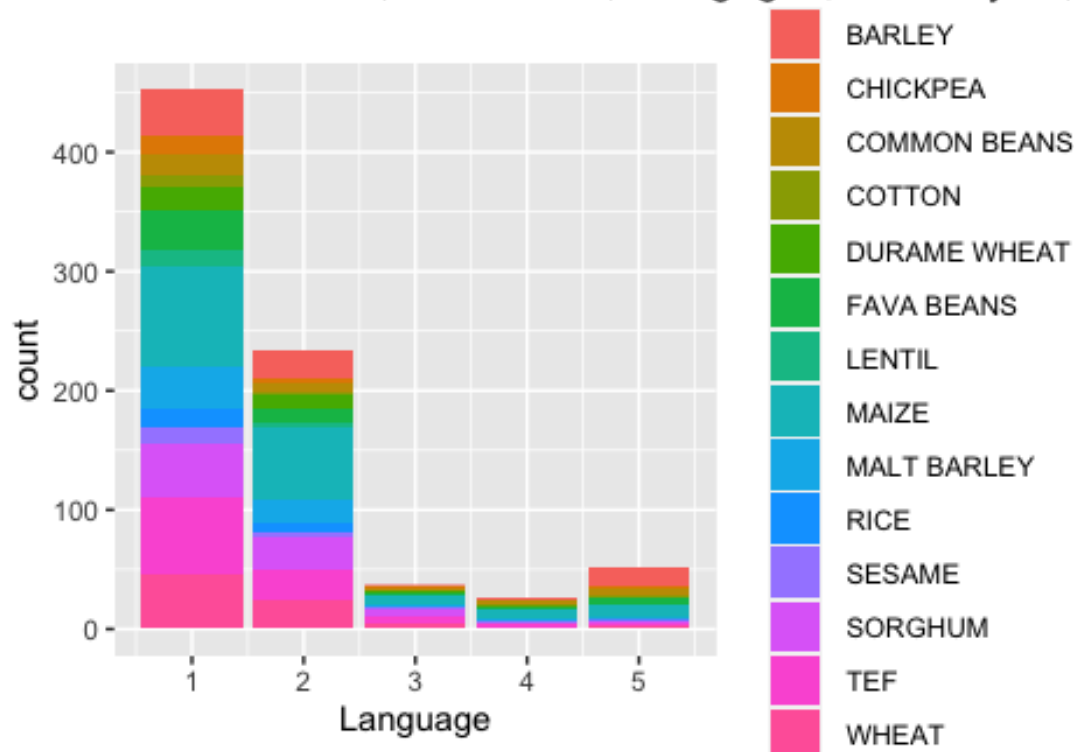
ALTERNATIVE RAIN MENU/PLANTING/SEED RATE
content actually played by language
1 = Amharic, 2=Oromiffa, 3=Tigrigna, 4=Wolayitta, ai


```
#no Language
ggplot(seed_rate_no_lang, aes(x=logInfo)) +
  theme(axis.text.x = element_text(angle = 90))+
  geom_bar()+
  labs(
    title = 'RAIN MENU/PLANTING/SEED RATE
    content actually played',
    x = 'Content Selected',
    y= 'count'
  )
```

RAIN MENU/PLANTING/SEED RATE
### content actually played
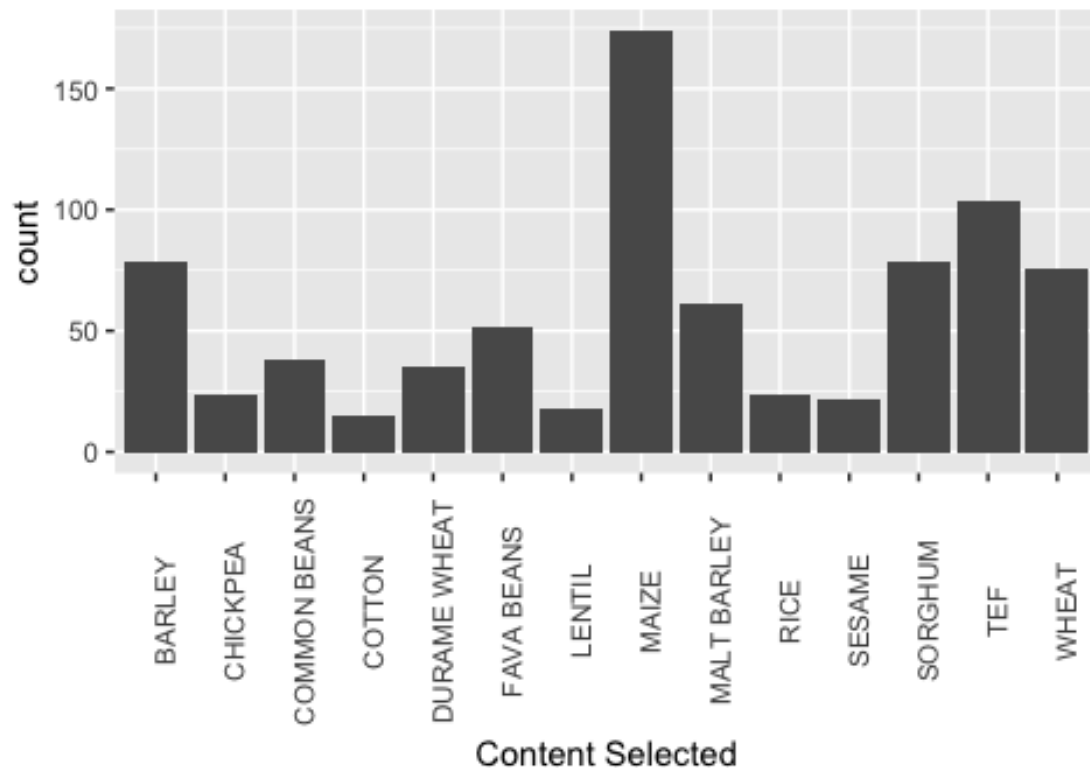

```r
crop_menu <- log_data %>%
  filter(grepl("CROP MENU - ",logInfo)) %>%
  filter(!grepl("HHI",logInfo)) %>%
  filter(!grepl("REPLAYED",logInfo))

crop_menu <- crop_menu %>%
  group_by(callerId, logInfo, langId) %>%
  count()

crop_menu$logInfo<- str_replace(crop_menu$logInfo, "CROP MENU - ", "")
crop_menu$logInfo<- str_replace(crop_menu$logInfo, " SELECTED", "")
crop_menu$logInfo<- str_replace(crop_menu$logInfo, "GO TO", "")


ggplot(crop_menu, aes(x=logInfo)) +
  theme(axis.text.x = element_text(angle = 60, hjust =1, size = 7))+
  geom_bar()+
  labs(
    title = 'Crop menu breakdown',
    x = 'Crops',
    y= 'count'
  )
```
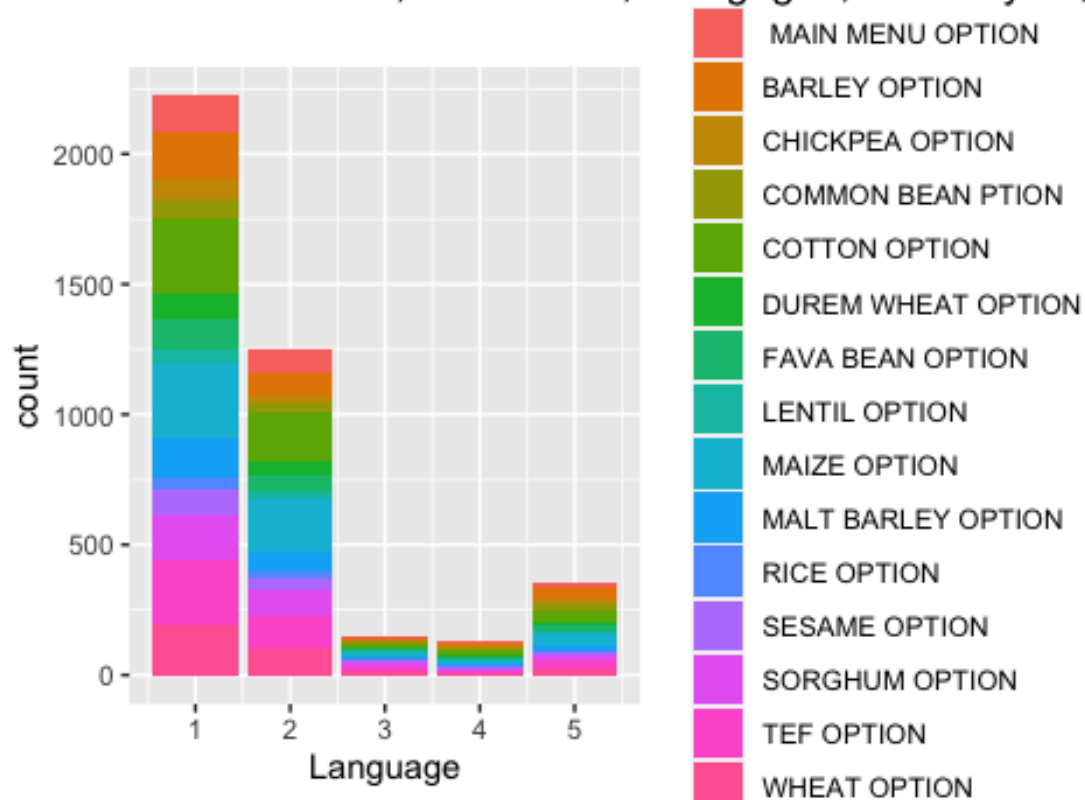
Crop menu breakdown
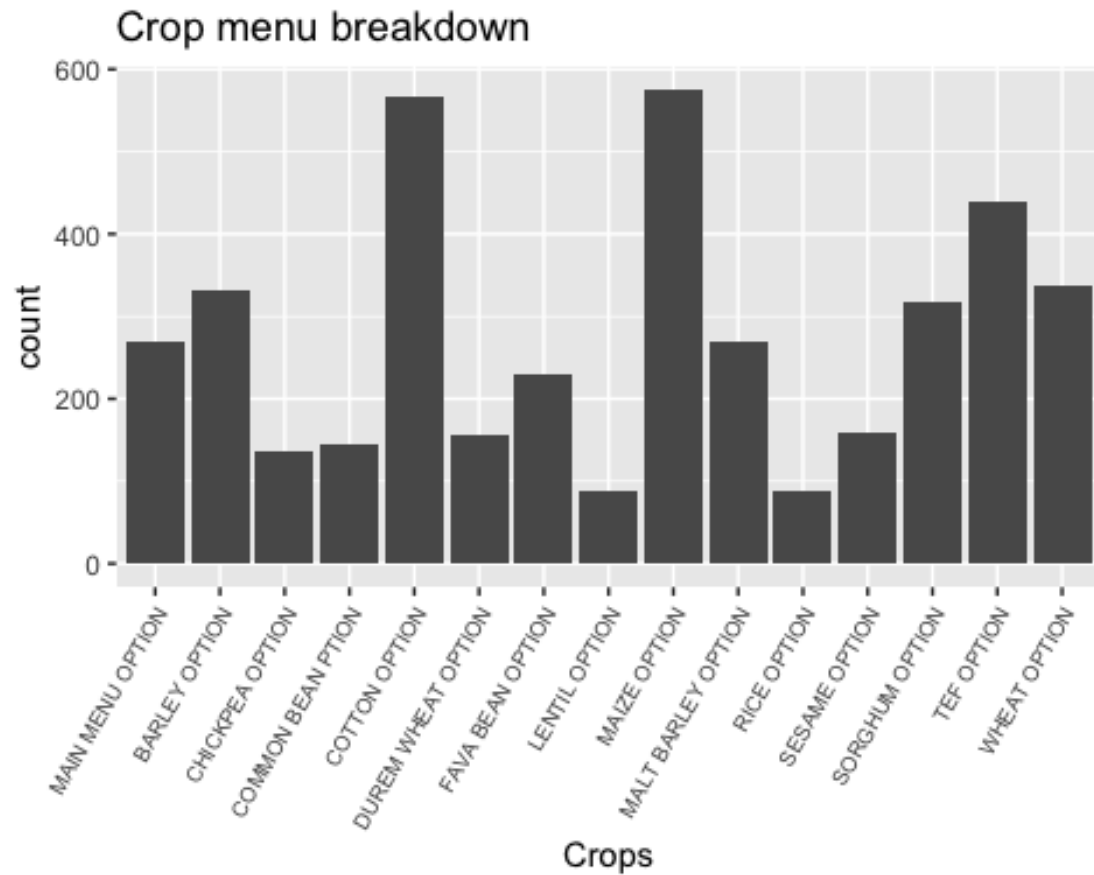

```
ggplot(crop_menu, aes(langId, fill = logInfo)) +
  geom_bar()+
  labs(
    title = 'CROP MENU BY LANGUAGE
    1 = Amharic, 2=Oromiffa, 3=Tigrigna, 4=Wolayitta, and 5=Sidamigna
    ',
    x = 'Language'
  )
```

CROP MENU BY LANGUAGE
1 = Amharic, 2=Oromiffa, 3=Tigrigna, 4=Wolayitta, a ... tplnfe

Legend:
MAIN MENU OPTION
BARLEY OPTION
CHICKPEA OPTION
COMMON BEAN PTION
COTTON OPTION
DUREM WHEAT OPTION
FAVA BEAN OPTION
LENTIL OPTION
MAIZE OPTION
MALT BARLEY OPTION
RICE OPTION
SESAME OPTION
SORGHUM OPTION
TEF OPTION
WHEAT OPTION

```
num_crop_types <- data.frame(as.numeric(table(crop_menu$callerId)))
num_crop_types$frequencies <- num_crop_types$as.numeric.table.crop_menu.calle
rId.

summary(num_crop_types$frequencies)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   4.000   4.915   7.000  15.000

ggplot(num_crop_types, aes(frequencies)) +
  geom_histogram(color="black", fill="red") +
  labs(
    title = 'Distribution of number crops ever selected by individual \n
      Min.  1st Qu.  Median    Mean   3rd Qu.    Max.
   1.000    2.000    4.000    4.915    7.000    15.000 ',
    x = 'Number of crops selected '
  )

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
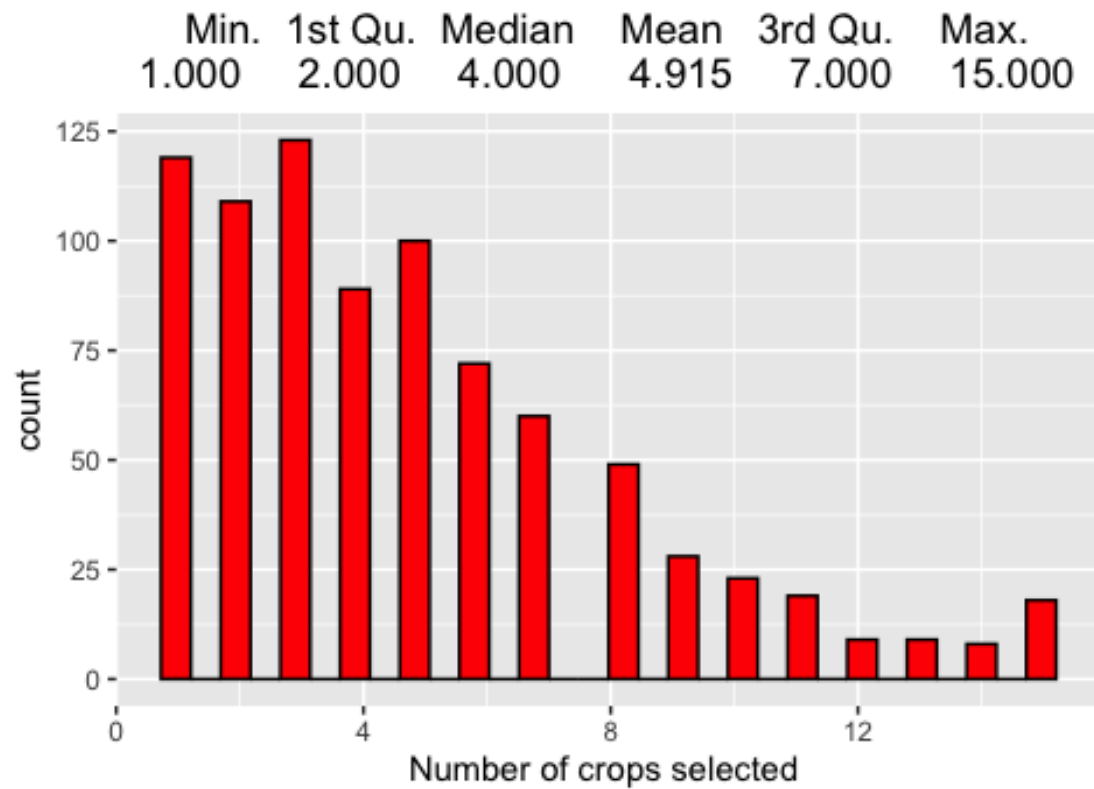
Distribution of number crops ever selected by individual

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	2.000	4.000	4.915	7.000	15.000

HOW MANY CROPS PER USER ARE LOOKED AT??????????