PROJECT 1 CODING CLUB

Author: Joan Weru

1.Import,clean and view data

```
library(tidyverse)
library(knitr)
setwd("~/R- CLASS DATASETS")
mobile<-read_csv("mobilemoney_data.csv")
head(mobile)
```

```
## # A tibble: 6 x 28
##    hhid weight account_num account_type district urban gender   age
##   <dbl>  <dbl>       <dbl> <chr>        <chr>    <chr> <chr>  <dbl>
## 1  1001   146.           1 SACCO Accou~ Distric~ Urban male      32
## 2  1001   146.           2 VSLA Account Distric~ Urban male      32
## 3  1002   123.           1 Mobile Money Distric~ Rural male      32
## 4  1002   123.           2 Bank Account Distric~ Rural male      32
## 5  1002   123.           3 VSLA Account Distric~ Rural male      32
## 6  1003   760.           1 Mobile Money Distric~ Urban male      30
## # ... with 20 more variables: hh_members <dbl>,
## #   highest_grade_completed <chr>, mm_account_cancelled <chr>,
## #   prefer_cash <chr>, mm_trust <chr>, mm_account_telco <chr>,
## #   mm_account_telco_main <chr>, v234 <chr>, agent_trust <chr>,
## #   v236 <chr>, v237 <chr>, v238 <chr>, v240 <chr>, v241 <chr>,
## #   v242 <chr>, v243 <chr>, v244 <chr>, v245 <chr>, v246 <chr>,
## #   mm_account <chr>
```

2.Select the variables "age", "gender", "education level" and "number of household members" in R and write them in a new and seperate file. Save the data set as "demographics.csv"

```
demographics<-transmute(mobile,"age"=age,"gender"=gender,"education level"=highest_grade_completed,"numl
head(demographics)%>%knitr::kable()
```

age	gender	education level	numberof household members
32	male	primary 6	1
32	male	primary 6	1
32	male	primary 3	4
32	male	primary 3	4
32	male	primary 3	4
30	male	secondary 6	8

```
write.csv(demographics,"~/R- CLASS DATASETS/demographics.csv")
```

3.Conduct exploratory analysis of the data and write a few bullet points on any descriptive statistics (summary statistics, tables and graphs) you find interesting and why you find them interesting.

- We first explore the different types of account types quantitatively and visually and see if we find any useful insights

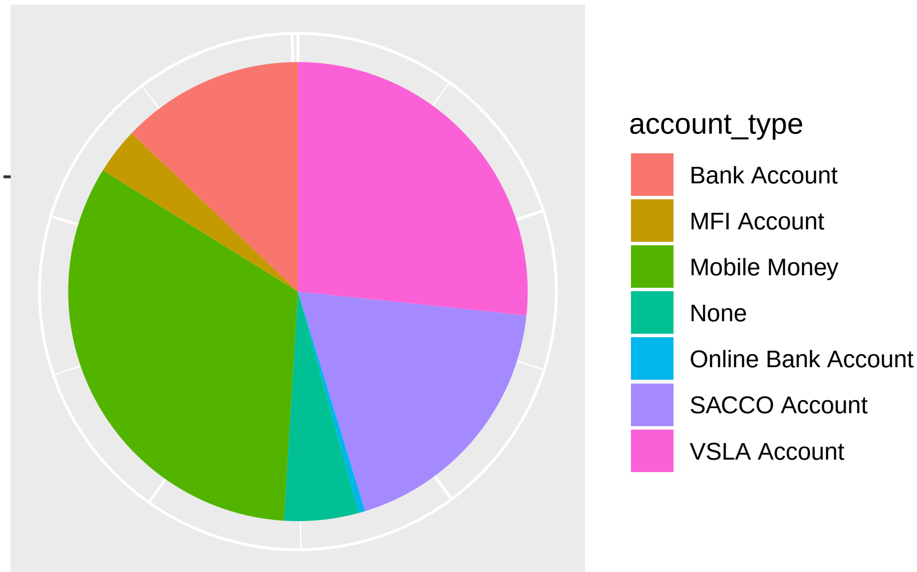
```r
library(ggplot2)
library(knitr)
table(mobile$account_type)%>%knitr::kable()
```

Var1	Freq
Bank Account	323
MFI Account	82
Mobile Money	825
None	131
Online Bank Account	12
SACCO Account	467
VSLA Account	669

```r
account_type.perc=mobile%>%count(account_type)%>%mutate(perc=n/sum(n)*100)%>%arrange(desc(perc))
account_type.perc
```

```
## # A tibble: 7 x 3
##   account_type            n   perc
##   <chr>               <int>  <dbl>
## 1 Mobile Money          825 32.9
## 2 VSLA Account          669 26.7
## 3 SACCO Account         467 18.6
## 4 Bank Account          323 12.9
## 5 None                  131  5.22
## 6 MFI Account            82  3.27
## 7 Online Bank Account    12  0.478
```

```r
ggplot(account_type.perc,aes(x="",y=n,fill=account_type))+geom_bar(width=1,stat="identity")+coord_polar
```

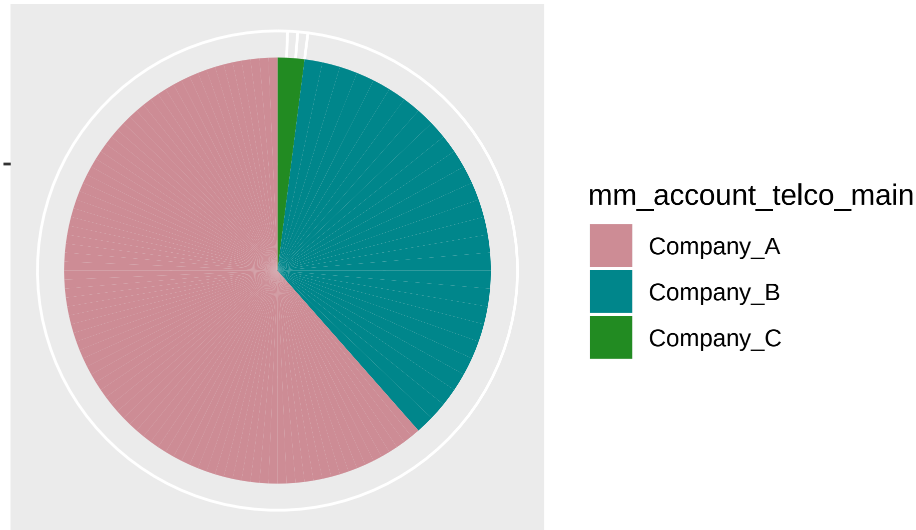


Clearly we see that Mobile Money is leading in terms of popularity

```
library(ggplot2)
library(knitr)
library(tigerstats)
rowPerc(xtabs(~mm_account_telco_main,data=mobile))%>%knitr::kable()
```

Company_A	Company_B	Company_C	Total
59.39	38.73	1.88	100

```
mobile%>%drop_na%>%ggplot(aes(x="",y=mm_account_telco_main,fill=mm_account_telco_main))+geom_bar(width =
```



- Are people who have been victims of fraud less likely to trust mobile money agents

```
mobile%>%drop_na%>%count(v246,agent_trust)%>%rename("victim of fraud"=v246)%>%mutate(perc=n/sum(n)*100)%
```

victim of fraud	agent_trust	n	perc
no	no	60	50.420168
yes	no	29	24.369748
no	yes	23	19.327731
yes	yes	7	5.882353

- We now explore the main three companies in terms of tables, and percent rows

```
library(tigerstats)
table(mobile$mm_account_telco_main)%>%knitr::kable()
```

Var1	Freq
Company_A	506
Company_B	330
Company_C	16

```
rowPerc(xtabs(~mm_account_telco_main,data=mobile))%>%knitr::kable()
```

Company_A	Company_B	Company_C	Total
59.39	38.73	1.88	100

*Does coming from a rural or urban location influence whether they have one has an mobile money account or not

```
library(tigerstats)
urban.account.relation<-rowPerc(xtabs(~urban+mm_account,data = mobile))%>%knitr::kable()
urban.account.relation
```

	no	yes	Total
Rural	35.71	64.29	100
Urban	14.47	85.53	100

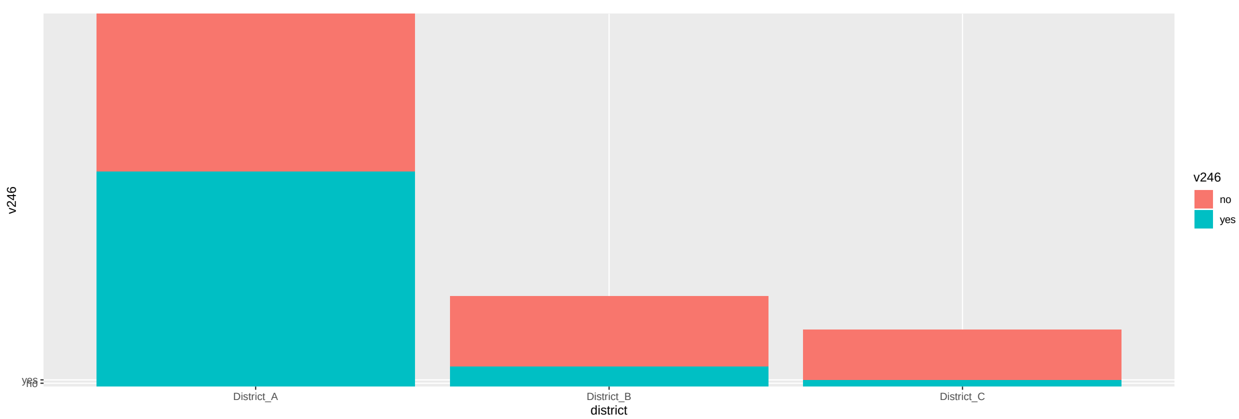
people fromm urban areas have a higher chnace of owning a mobile money account

*We now seek to see areas where fraud cases are prevalent

```
district.fraud<-rowPerc(xtabs(~district+v246,data=mobile))%>%knitr::kable()
district.fraud
```

	no	yes	Total
District_A	84.03	15.97	100
District_B	92.68	7.32	100
District_C	93.80	6.20	100

```
mobile%>%drop_na()%>%ggplot(aes(x=district,y=v246,fill=v246))+geom_col()
```



District A clearly has a high percentage of fraud cases, parties involved should look into the problem and seek solutions

4.During the survey, participants listed all the different types of financial accounts that they have registered. The resulting data has a format where there is one observation per account type. Format the data so that there is now one observation participant. Save data as mobile_new.csv

```
attach(mobile)
acc_details<-select(mobile,hhid,account_type,account_num)

mobile_new<-spread(mobile,account_type,account_num)
#To replace Nas with zero
for (i in 26:33) {
  mobile_new[is.na(mobile_new[,i]),i]<-0
}
str(mobile_new)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    1205 obs. of  33 variables:
##  $ hhid                   : num  1001 1002 1003 1004 1005 ...
##  $ weight                 : num  146 123 760 434 303 ...
##  $ district               : chr  "District_A" "District_B" "District_A" "District_A" ...
##  $ urban                  : chr  "Urban" "Rural" "Urban" "Rural" ...
##  $ gender                 : chr  "male" "male" "male" "male" ...
##  $ age                    : num  32 32 30 68 28 36 66 52 37 45 ...
##  $ hh_members             : num  1 4 8 4 2 7 7 5 5 7 ...
##  $ highest_grade_completed: chr  "primary 6" "primary 3" "secondary 6" "primary 6" ...
##  $ mm_account_cancelled   : chr  "no" "yes" "no" "no" ...
##  $ prefer_cash            : chr  "yes" "yes" "yes" "yes" ...
##  $ mm_trust               : chr  "no" "no" "no" "no" ...
##  $ mm_account_telco       : chr  "Company_A Company_B" NA "Company_A" "Company_A" ...
##  $ mm_account_telco_main  : chr  "Company_A" NA NA NA ...
##  $ v234                   : chr  "yes" NA "yes" "no" ...
##  $ agent_trust            : chr  "no" "no" "no" "no" ...
##  $ v236                   : chr  NA NA NA NA ...
##  $ v237                   : chr  "yes" "yes" "no" "no" ...
##  $ v238                   : chr  "yes" "yes" "yes" "no" ...
##  $ v240                   : chr  "no" "yes" "yes" "no" ...
##  $ v241                   : chr  "yes" "no" "yes" "no" ...
##  $ v242                   : chr  "no" "no" "no" "no" ...
##  $ v243                   : chr  "yes" "no" "yes" "yes" ...
##  $ v244                   : chr  NA NA "yes" NA ...
##  $ v245                   : chr  "yes" "no" "yes" "no" ...
##  $ v246                   : chr  "no" "no" "no" "no" ...
##  $ mm_account             : chr  "yes" "no" "yes" "yes" ...
##  $ Bank Account           : num  0 2 2 0 0 0 0 0 0 0 ...
##  $ MFI Account            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Mobile Money           : num  0 1 1 1 0 1 1 1 0 1 ...
##  $ None                   : num  0 0 0 0 1 0 0 0 1 0 ...
##  $ Online Bank Account    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ SACCO Account          : num  1 0 0 0 0 0 2 2 0 0 ...
##  $ VSLA Account           : num  2 3 0 0 0 0 3 0 0 2 ...
```

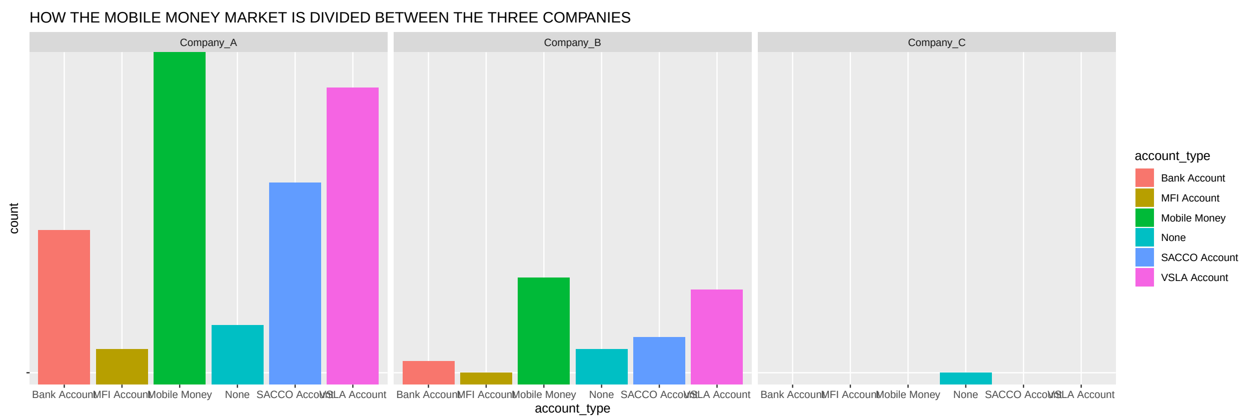
```
write.csv(mobile_new,"~/R- CLASS DATASETS/mobilenew.csv")
```

5.Describe how the mobile money market is divided between the three companies. Include at least one chart or table to illustrate your findings.

```
acc.telco.relation<-rowPerc(xtabs(~account_type+mm_account_telco_main,data = mobile) )%>%knitr::kable()
acc.telco.relation
```

	Company_A	Company_B	Company_C	Total
Bank Account	56.88	39.45	3.67	100
MFI Account	66.67	33.33	0.00	100
Mobile Money	60.42	38.19	1.39	100
None	53.66	43.90	2.44	100
Online Bank Account	0.00	75.00	25.00	100
SACCO Account	59.18	39.46	1.36	100
VSLA Account	60.67	37.66	1.67	100

```
mobile%>%drop_na%>%ggplot(aes(x=account_type,y="",fill=account_type))+geom_col(stat="count")+facet_wrap
```



6.Develop a cross tabulation (in percentages) with gender as the rows and columns with urban, mm_trust and prefer_cash

```
library(gmodels)
library(knitr)
ct1<-CrossTable(mobile_new$gender,mobile_new$urban)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  1205
##
##
```

```
##                    | mobile_new$urban
## mobile_new$gender |     Rural |     Urban | Row Total |
## -------------------|-----------|-----------|-----------|
##             female |       541 |       154 |       695 |
##                    |     0.002 |     0.009 |           |
##                    |     0.778 |     0.222 |     0.577 |
##                    |     0.576 |     0.581 |           |
##                    |     0.449 |     0.128 |           |
## -------------------|-----------|-----------|-----------|
##               male |       399 |       111 |       510 |
##                    |     0.003 |     0.012 |           |
##                    |     0.782 |     0.218 |     0.423 |
##                    |     0.424 |     0.419 |           |
##                    |     0.331 |     0.092 |           |
## -------------------|-----------|-----------|-----------|
##       Column Total |       940 |       265 |      1205 |
##                    |     0.780 |     0.220 |           |
## -------------------|-----------|-----------|-----------|
##
##
```

```
ct2<-CrossTable(mobile_new$gender,mobile_new$mm_trust)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  1113
##
##
##                    | mobile_new$mm_trust
## mobile_new$gender |       -97 |        no |       yes | Row Total |
## -------------------|-----------|-----------|-----------|-----------|
##             female |        12 |       573 |        48 |       633 |
##                    |     1.411 |     0.087 |     0.298 |           |
##                    |     0.019 |     0.905 |     0.076 |     0.569 |
##                    |     0.800 |     0.562 |     0.615 |           |
##                    |     0.011 |     0.515 |     0.043 |           |
## -------------------|-----------|-----------|-----------|-----------|
##               male |         3 |       447 |        30 |       480 |
##                    |     1.860 |     0.115 |     0.394 |           |
##                    |     0.006 |     0.931 |     0.062 |     0.431 |
##                    |     0.200 |     0.438 |     0.385 |           |
##                    |     0.003 |     0.402 |     0.027 |           |
## -------------------|-----------|-----------|-----------|-----------|
##       Column Total |        15 |      1020 |        78 |      1113 |
##                    |     0.013 |     0.916 |     0.070 |           |
```

```
## ------------------|-----------|-----------|-----------|-----------|
##
##
```

```
ct3<-CrossTable(mobile_new$gender,mobile_new$prefer_cash)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  1179
##
##
##                   | mobile_new$prefer_cash
## mobile_new$gender |       -97 |        no |       yes | Row Total |
## ------------------|-----------|-----------|-----------|-----------|
##            female |         0 |        29 |       650 |       679 |
##                   |     0.576 |     0.138 |     0.003 |           |
##                   |     0.000 |     0.043 |     0.957 |     0.576 |
##                   |     0.000 |     0.617 |     0.575 |           |
##                   |     0.000 |     0.025 |     0.551 |           |
## ------------------|-----------|-----------|-----------|-----------|
##              male |         1 |        18 |       481 |       500 |
##                   |     0.782 |     0.187 |     0.004 |           |
##                   |     0.002 |     0.036 |     0.962 |     0.424 |
##                   |     1.000 |     0.383 |     0.425 |           |
##                   |     0.001 |     0.015 |     0.408 |           |
## ------------------|-----------|-----------|-----------|-----------|
##      Column Total |         1 |        47 |      1131 |      1179 |
##                   |     0.001 |     0.040 |     0.959 |           |
## ------------------|-----------|-----------|-----------|-----------|
##
##
```

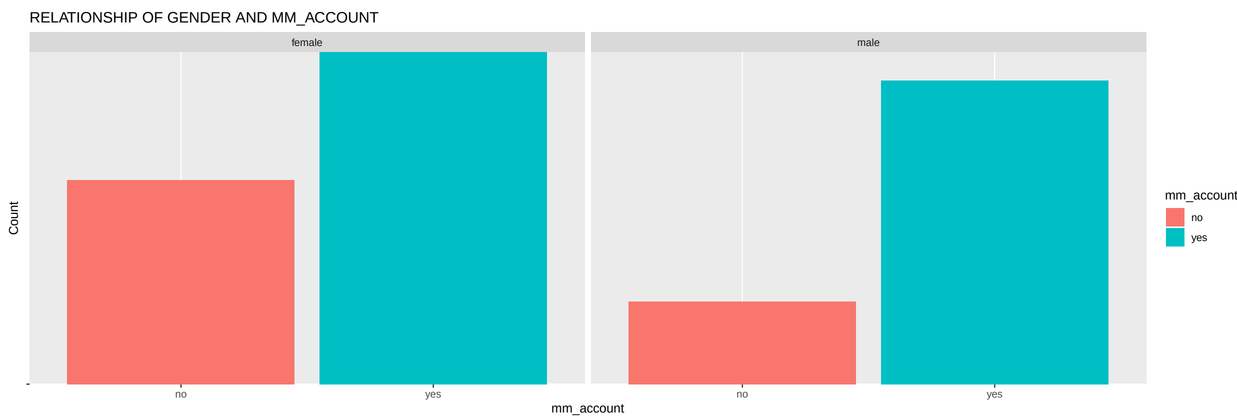
```
ct4<-cbind(c(ct1$prop.row,ct2$prop.row,ct3$prop.row))
ct5<-ct4*100
ct5
```

```
##             [,1]
##  [1,] 77.841727
##  [2,] 78.235294
##  [3,] 22.158273
##  [4,] 21.764706
##  [5,]  1.895735
##  [6,]  0.625000
##  [7,] 90.521327
##  [8,] 93.125000
##  [9,]  7.582938
```

```
## [10,]  6.250000
## [11,]  0.000000
## [12,]  0.200000
## [13,]  4.270987
## [14,]  3.600000
## [15,] 95.729013
## [16,] 96.200000
```

7.Plot the graph of gender and mm_account in the same bar graph and clearly label your axis

```r
mobile%>%ggplot(aes(x=mm_account,y="",fill=mm_account))+geom_col(stat="count")+facet_wrap(~gender)+labs
```



8.Is there a difference in the share of customers who have experienced failed mobile money transactions in rural and urban villages? If so, is it statistically significant? Explain your findings including any assumptions and limitations.(Hint: Chi square test)

Chi square test is a statistical method which is used to determine if two categorical variables have a significant correlation between them. In this case our two categorical variables are failed mobile money transaction and urban

- H_0-he two variables are independent
- H_1-The two variabes relate to each other
- in this cae our level of significance is 0.05

```r
cq<-chisq.test(mobile$v240,mobile$urban)
cq
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  mobile$v240 and mobile$urban
## X-squared = 63.19, df = 1, p-value = 1.877e-15
```

our p value is less than 0.05(level of significance),we reject the null hypothesis and conclude that there is a significant relationship between failed mobile money transactions and whether one is from a rural or urban area