First RMarkdown Project

Malin Chimoyi

7/12/2019

# Coding Club

## Assignment Report

## Introduction

The data was collected on the use of mobile money services and the experiences of mobile money customers in three districts of Rwanda. The dataset has been provided and you have been asked to write a short summary outlining key findings from the project. You should perform all the statistical analysis required and write up the key findings. You should remember that the target audiences do not have a statistical background.

#This covers subsetting of the data using the dplyr package  
demographics <- dta %>% select(age,gender,highest\_grade\_completed,hh\_members)  
#write.csv(demographics,"demographics.csv", row.names = FALSE)  
head(demographics,6)

## # A tibble: 6 x 4  
## age gender highest\_grade\_completed hh\_members  
## <dbl> <fct> <chr> <dbl>  
## 1 32 male primary 6 1  
## 2 32 male primary 6 1  
## 3 32 male primary 3 4  
## 4 32 male primary 3 4  
## 5 32 male primary 3 4  
## 6 30 male secondary 6 8

# Distribution of data

#conducting an EDA of the data using demographics  
table(demographics$age)

##   
## 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35   
## 40 53 60 43 51 67 87 79 91 79 106 69 97 89 89 65 53 70   
## 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53   
## 89 60 90 41 58 41 63 44 38 53 38 44 34 37 41 21 49 21   
## 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71   
## 22 31 23 23 34 15 40 10 13 30 12 12 14 10 11 7 2 6   
## 72 74 75 76 77 78 79 80 81 82 83 84 85 87 88 97   
## 7 4 2 3 3 4 2 3 1 1 1 4 4 1 3 1

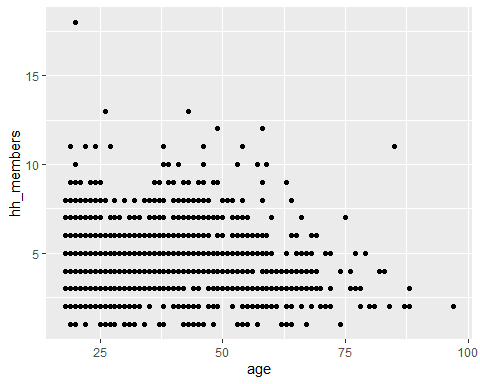
## Plotting

This is the plot of age against the number of household members.

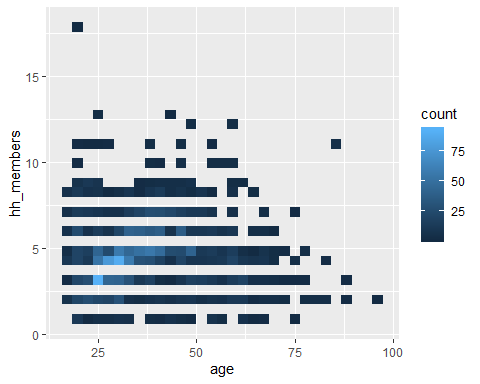
#leading ggplot2  
library(ggplot2)  
library(hexbin)

## Warning: package 'hexbin' was built under R version 3.6.1

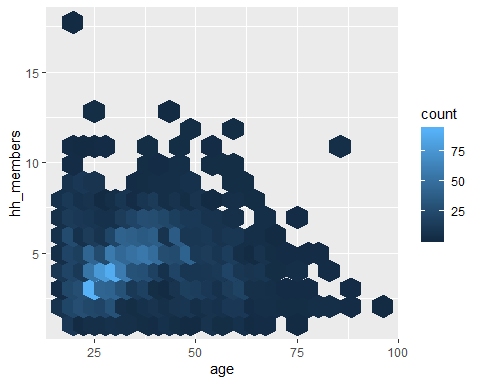
ggplot(data= demographics) + geom\_point(mapping = aes (age,hh\_members))



ggplot(data= demographics) + geom\_bin2d(mapping = aes (age,hh\_members))

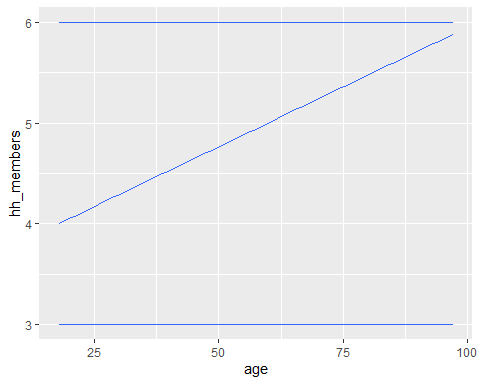


ggplot(data= demographics) + geom\_hex(mapping = aes (age,hh\_members))



ggplot(data= demographics) + geom\_quantile(mapping = aes (age,hh\_members))

## Smoothing formula not specified. Using: y ~ x



## The distribution of account-types

The accounts with the highest percentages are mobilemoney accounts and the VSLA accounts while the one with the lowest is the Online Bank Account.

dta %>% group\_by(account\_type) %>% summarise(number=n()) %>% mutate(percentage=number/sum(number)\*100) %>% arrange(desc(percentage))

## # A tibble: 7 x 3  
## account\_type number percentage  
## <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