Project Assignment 3

David Abeiku Saah

02/04/2023

### Research Question

**Are there significant differences in levels of trust in government and private media sources?**

Explanatory variables: Public media, government, private media. Response variable: Level of trust.

All the variables are ***categorical***.

## 1. Load data set(s) and libraries

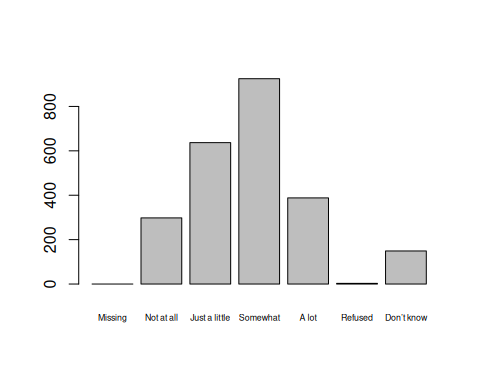
# load dataset and library packages  
load("Ghana\_r8");  
library(descr);  
library(stats);

## 2. Create variable subset

# assign the variables to be used  
vars = c("Q41N\_GHA", "Q41O\_GHA", "Q41M\_GHA");  
  
# create a new data frame with the variables  
my\_data = data.frame(Ghana\_r8[vars]);

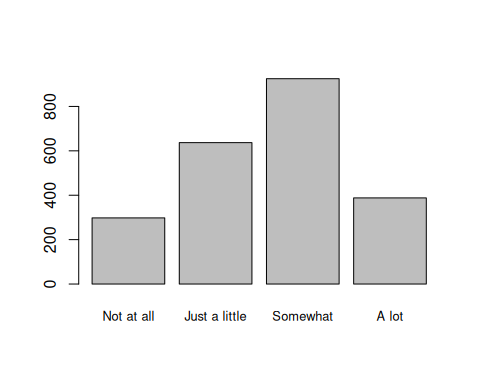
## 3. Data management I: check for and recode errors and NAs

# Check the graph if there are any error codes  
freq(my\_data$Q41N\_GHA, cex.names=0.55);



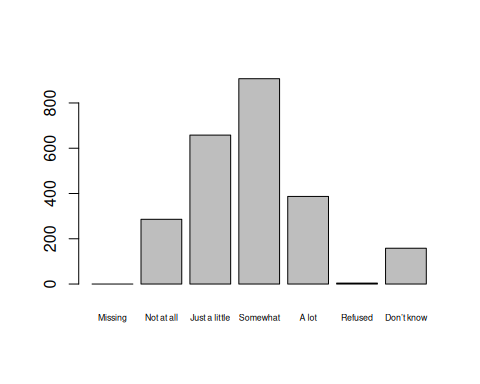
## my\_data$Q41N\_GHA   
## Frequency Percent  
## Missing 0 0.000  
## Not at all 298 12.417  
## Just a little 637 26.542  
## Somewhat 925 38.542  
## A lot 388 16.167  
## Refused 3 0.125  
## Don’t know 149 6.208  
## Total 2400 100.000

# "Missing", "Refused" and "Don't know" are NA values since they identify as missing data  
my\_data$Q41N\_GHA[my\_data$Q41N\_GHA == "Refused"] = NA;  
my\_data$Q41N\_GHA[my\_data$Q41N\_GHA == "Missing"] = NA;  
my\_data$Q41N\_GHA[my\_data$Q41N\_GHA == "Don’t know"] = NA;  
my\_data$Q41N\_GHA = factor(my\_data$Q41N\_GHA);  
freq(my\_data$Q41N\_GHA, cex.names=0.8);



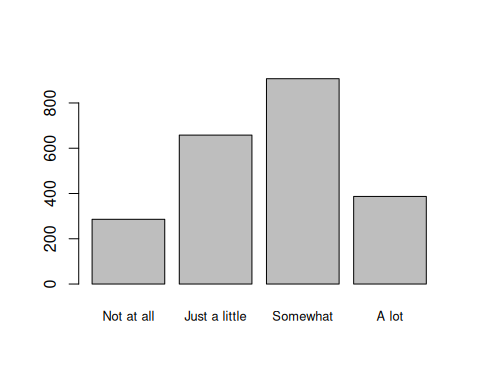
## my\_data$Q41N\_GHA   
## Frequency Percent Valid Percent  
## Not at all 298 12.417 13.26  
## Just a little 637 26.542 28.34  
## Somewhat 925 38.542 41.15  
## A lot 388 16.167 17.26  
## NA's 152 6.333   
## Total 2400 100.000 100.00

# Check the graph if there are any error codes  
freq(my\_data$Q41O\_GHA, cex.names=0.55);



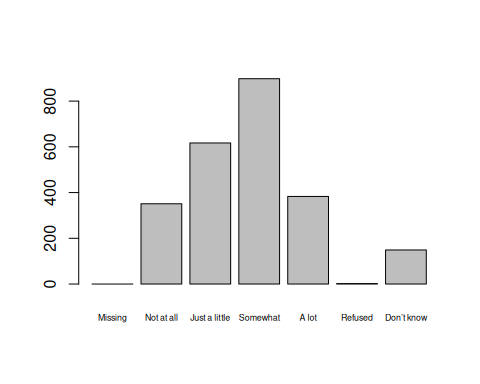
## my\_data$Q41O\_GHA   
## Frequency Percent  
## Missing 0 0.0000  
## Not at all 286 11.9167  
## Just a little 658 27.4167  
## Somewhat 907 37.7917  
## A lot 387 16.1250  
## Refused 4 0.1667  
## Don’t know 158 6.5833  
## Total 2400 100.0000

# "Missing", "Refused" and "Don't know" are NA values since they identify as missing data  
my\_data$Q41O\_GHA[my\_data$Q41O\_GHA == "Refused"] = NA;  
my\_data$Q41O\_GHA[my\_data$Q41O\_GHA == "Missing"] = NA;  
my\_data$Q41O\_GHA[my\_data$Q41O\_GHA == "Don’t know"] = NA;  
my\_data$Q41O\_GHA = factor(my\_data$Q41O\_GHA);  
freq(my\_data$Q41O\_GHA, cex.names=0.8);



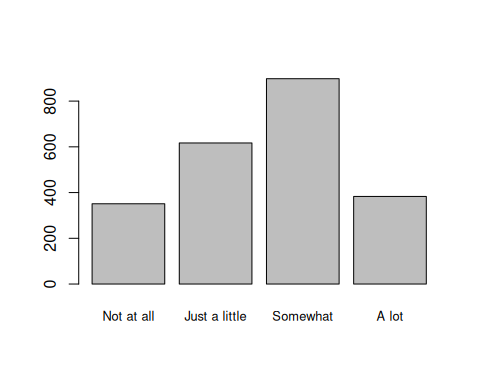
## my\_data$Q41O\_GHA   
## Frequency Percent Valid Percent  
## Not at all 286 11.92 12.78  
## Just a little 658 27.42 29.40  
## Somewhat 907 37.79 40.53  
## A lot 387 16.12 17.29  
## NA's 162 6.75   
## Total 2400 100.00 100.00

# Check the graph if there are any error codes  
freq(my\_data$Q41M\_GHA, cex.names=0.55);



## my\_data$Q41M\_GHA   
## Frequency Percent  
## Missing 0 0.00000  
## Not at all 351 14.62500  
## Just a little 617 25.70833  
## Somewhat 898 37.41667  
## A lot 383 15.95833  
## Refused 2 0.08333  
## Don’t know 149 6.20833  
## Total 2400 100.00000

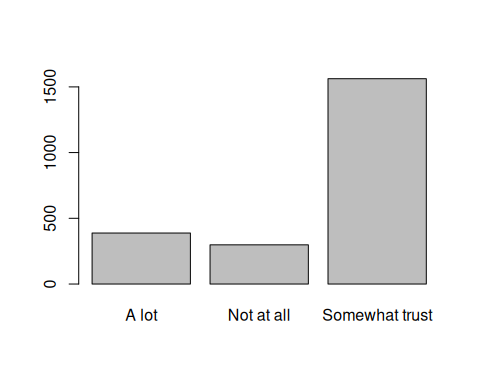
# "Missing", "Refused" and "Don't know" are NA values since they identify as missing data  
my\_data$Q41M\_GHA[my\_data$Q41M\_GHA == "Refused"] = NA;  
my\_data$Q41M\_GHA[my\_data$Q41M\_GHA == "Missing"] = NA;  
my\_data$Q41M\_GHA[my\_data$Q41M\_GHA == "Don’t know"] = NA;  
my\_data$Q41M\_GHA = factor(my\_data$Q41M\_GHA);  
freq(my\_data$Q41M\_GHA, cex.names=0.8);



## my\_data$Q41M\_GHA   
## Frequency Percent Valid Percent  
## Not at all 351 14.625 15.61  
## Just a little 617 25.708 27.43  
## Somewhat 898 37.417 39.93  
## A lot 383 15.958 17.03  
## NA's 151 6.292   
## Total 2400 100.000 100.00

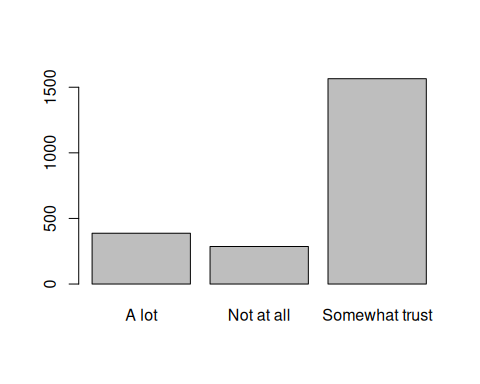
## 4. Data management II: further subset and create secondary variable

# Merge "Just a little" and "Somewhat" into "Somewhat trust"  
  
# create new variables  
my\_data$public\_media <- rep(NA,nrow(my\_data));  
my\_data$private\_media <- rep(NA,nrow(my\_data));  
my\_data$government <- rep(NA,nrow(my\_data));  
  
# public media  
my\_data$public\_media[my\_data$Q41N\_GHA == "Not at all"] <- "Not at all";  
my\_data$public\_media[my\_data$Q41N\_GHA == "Just a little"] <- "Somewhat trust";  
my\_data$public\_media[my\_data$Q41N\_GHA == "Somewhat"] <- "Somewhat trust";  
my\_data$public\_media[my\_data$Q41N\_GHA == "A lot"] <- "A lot";  
freq(my\_data$public\_media);



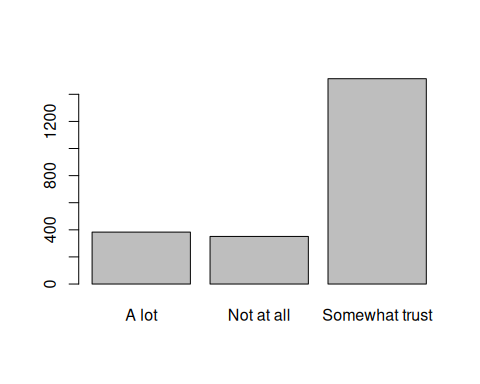
## my\_data$public\_media   
## Frequency Percent Valid Percent  
## A lot 388 16.167 17.26  
## Not at all 298 12.417 13.26  
## Somewhat trust 1562 65.083 69.48  
## NA's 152 6.333   
## Total 2400 100.000 100.00

# private media  
my\_data$private\_media[my\_data$Q41O\_GHA == "Not at all"] <- "Not at all";  
my\_data$private\_media[my\_data$Q41O\_GHA == "Just a little"] <- "Somewhat trust";  
my\_data$private\_media[my\_data$Q41O\_GHA == "Somewhat"] <- "Somewhat trust";  
my\_data$private\_media[my\_data$Q41O\_GHA == "A lot"] <- "A lot";  
freq(my\_data$private\_media);



## my\_data$private\_media   
## Frequency Percent Valid Percent  
## A lot 387 16.12 17.29  
## Not at all 286 11.92 12.78  
## Somewhat trust 1565 65.21 69.93  
## NA's 162 6.75   
## Total 2400 100.00 100.00

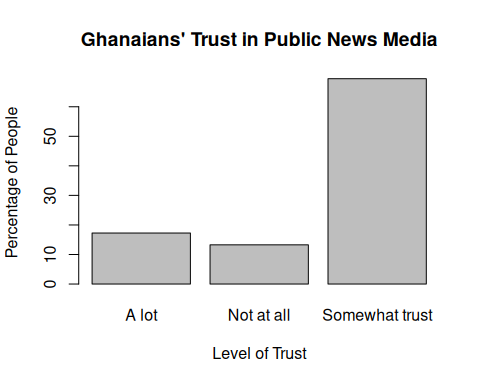
# government sources  
  
my\_data$government[my\_data$Q41M\_GHA == "Not at all"] <- "Not at all";  
my\_data$government[my\_data$Q41M\_GHA == "Just a little"] <- "Somewhat trust";  
my\_data$government[my\_data$Q41M\_GHA == "Somewhat"] <- "Somewhat trust";  
my\_data$government[my\_data$Q41M\_GHA == "A lot"] <- "A lot";  
freq(my\_data$government);



## my\_data$government   
## Frequency Percent Valid Percent  
## A lot 383 15.958 17.03  
## Not at all 351 14.625 15.61  
## Somewhat trust 1515 63.125 67.36  
## NA's 151 6.292   
## Total 2400 100.000 100.00

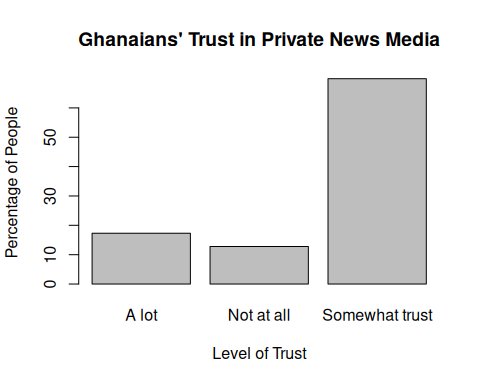
## 5. Descriptive statistics (sample means, standard deviations, proportions) and univariate displays

# Univariate display for public media information source  
freq(my\_data$public\_media, main = "Ghanaians' Trust in Public News Media", y.axis = "percent", ylab = "Percentage of People", xlab = "Level of Trust");



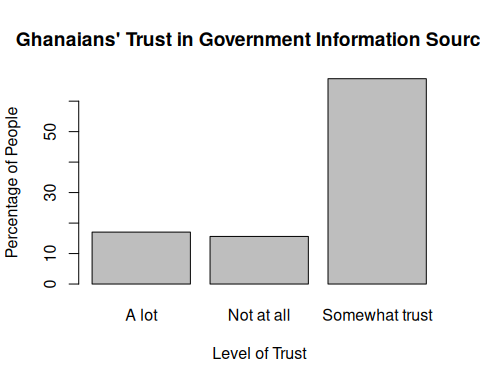
## my\_data$public\_media   
## Frequency Percent Valid Percent  
## A lot 388 16.167 17.26  
## Not at all 298 12.417 13.26  
## Somewhat trust 1562 65.083 69.48  
## NA's 152 6.333   
## Total 2400 100.000 100.00

# Univariate display for private media information source  
freq(my\_data$private\_media, main = "Ghanaians' Trust in Private News Media", y.axis = "percent", ylab = "Percentage of People", xlab = "Level of Trust");



## my\_data$private\_media   
## Frequency Percent Valid Percent  
## A lot 387 16.12 17.29  
## Not at all 286 11.92 12.78  
## Somewhat trust 1565 65.21 69.93  
## NA's 162 6.75   
## Total 2400 100.00 100.00

# Univariate display for government information source  
freq(my\_data$government, main = "Ghanaians' Trust in Government Information Sources", y.axis = "percent", ylab = "Percentage of People", xlab = "Level of Trust");



## my\_data$government   
## Frequency Percent Valid Percent  
## A lot 383 15.958 17.03  
## Not at all 351 14.625 15.61  
## Somewhat trust 1515 63.125 67.36  
## NA's 151 6.292   
## Total 2400 100.000 100.00

## 6. Bivariate tables and graphs

I have learnt that most Ghanaians are indecisive of what type of information they believe or trust. The proportion of Ghanaians from the sample that are indecisive are 69.48%, 69.93% and 67.36% for public media, private media and government information sources. A few percentage of people made a choice to accept or reject information as truth depending on its source.

It is possible that there is no significant difference between the level of trust in government and private media sources.

## 7. Bivariate analysis (hypothesis tests and post-hoc tests)

## 8. Moderation

## 9. Save

# save the data  
save(my\_data, file = "./my\_data.RData");