Project Assignment 4

David Abeiku Saah

26/04/2023

### Research Question

**Is there an association between levels of trust in government information sources and levels of trust in private media sources?** Explanatory variable: level of trust in government sources. Response variable: level of trust in private media sources.

**Is there an association between levels of trust in public media and levels of trust in private media sources?** Explanatory variable: level of trust in public media sources. Response variable: level of trust in private media sources.

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

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

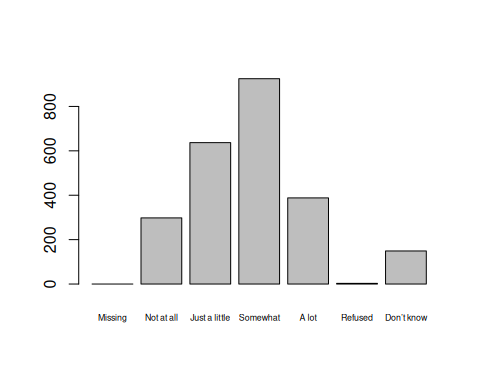
# load dataset and library packages  
load("Ghana\_r8");  
library(descr);  
library(stats);

## 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]);

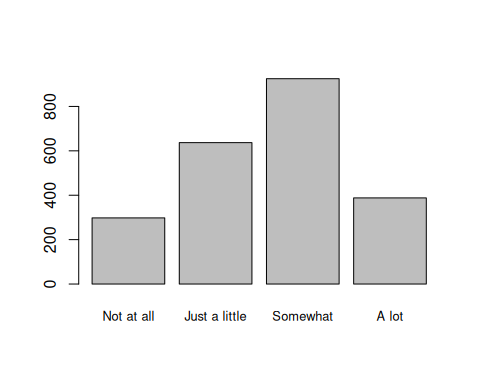
## 3. Data management I: check for and recode errors and NAs

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



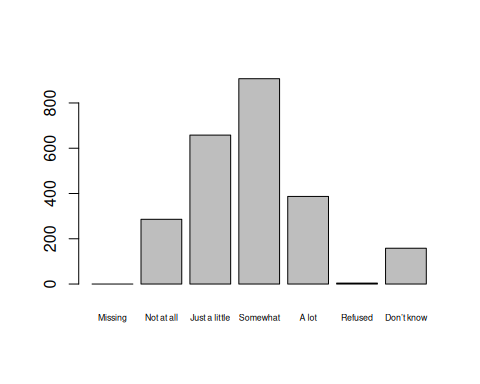
## my\_data$Q41N\_GHA   
## Frequency Percent  
## Missing 0 0.000  
## Not at all 298 12.417  
## Just a little 637 26.542  
## Somewhat 925 38.542  
## A lot 388 16.167  
## Refused 3 0.125  
## Don’t know 149 6.208  
## Total 2400 100.000

# "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);



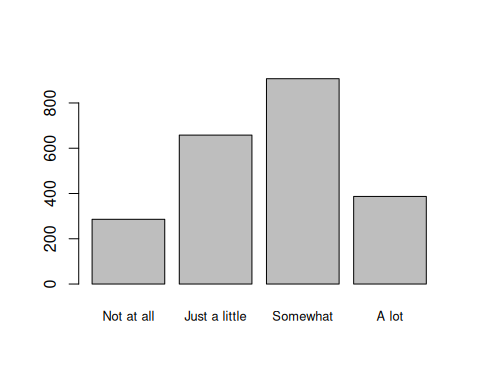
## my\_data$Q41N\_GHA   
## Frequency Percent Valid Percent  
## Not at all 298 12.417 13.26  
## Just a little 637 26.542 28.34  
## Somewhat 925 38.542 41.15  
## A lot 388 16.167 17.26  
## NA's 152 6.333   
## Total 2400 100.000 100.00

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



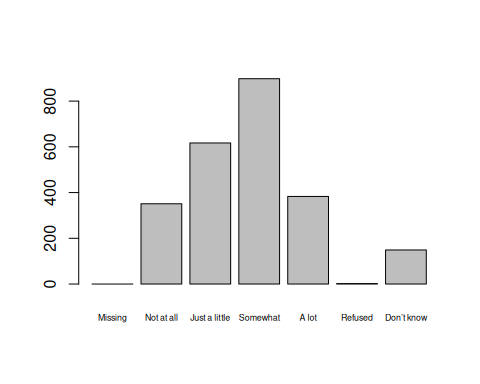
## my\_data$Q41O\_GHA   
## Frequency Percent  
## Missing 0 0.0000  
## Not at all 286 11.9167  
## Just a little 658 27.4167  
## Somewhat 907 37.7917  
## A lot 387 16.1250  
## Refused 4 0.1667  
## Don’t know 158 6.5833  
## Total 2400 100.0000

# "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);



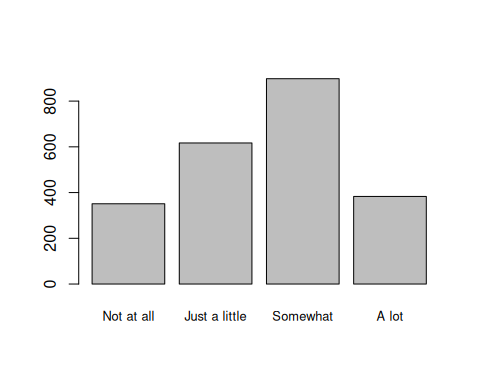
## my\_data$Q41O\_GHA   
## Frequency Percent Valid Percent  
## Not at all 286 11.92 12.78  
## Just a little 658 27.42 29.40  
## Somewhat 907 37.79 40.53  
## A lot 387 16.12 17.29  
## NA's 162 6.75   
## Total 2400 100.00 100.00

# 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

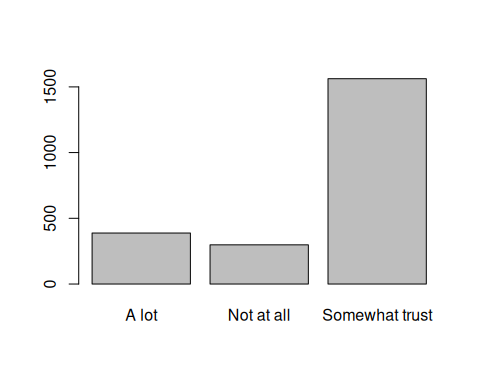
# "Missing", "Refused" and "Don't know" are NA values since they identify as missing data  
my\_data$Q41M\_GHA[my\_data$Q41M\_GHA == "Refused"] = NA;  
my\_data$Q41M\_GHA[my\_data$Q41M\_GHA == "Missing"] = NA;  
my\_data$Q41M\_GHA[my\_data$Q41M\_GHA == "Don’t know"] = NA;  
my\_data$Q41M\_GHA = factor(my\_data$Q41M\_GHA);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

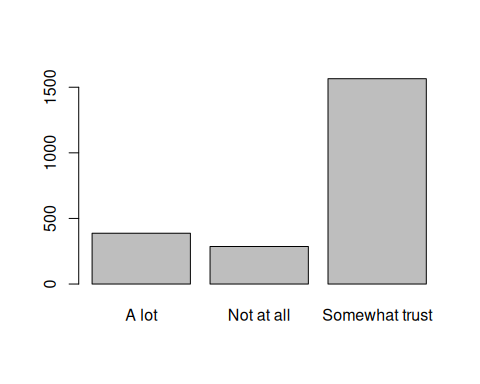
## 4. Data management II: further subset and create secondary variable

# 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);



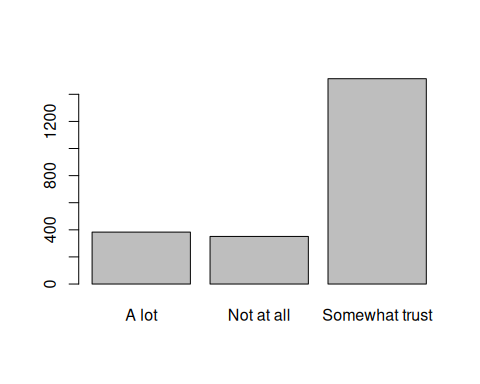
## my\_data$public\_media   
## Frequency Percent Valid Percent  
## A lot 388 16.167 17.26  
## Not at all 298 12.417 13.26  
## Somewhat trust 1562 65.083 69.48  
## NA's 152 6.333   
## Total 2400 100.000 100.00

# 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

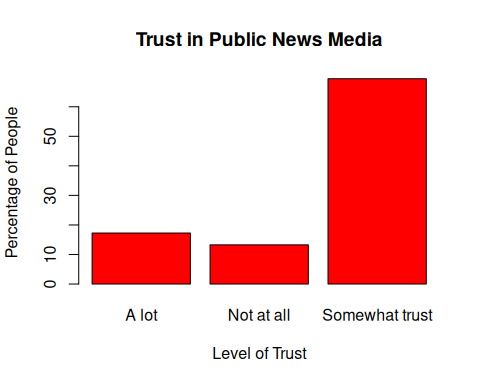
# government sources  
  
my\_data$government[my\_data$Q41M\_GHA == "Not at all"] <- "Not at all";  
my\_data$government[my\_data$Q41M\_GHA == "Just a little"] <- "Somewhat trust";  
my\_data$government[my\_data$Q41M\_GHA == "Somewhat"] <- "Somewhat trust";  
my\_data$government[my\_data$Q41M\_GHA ==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

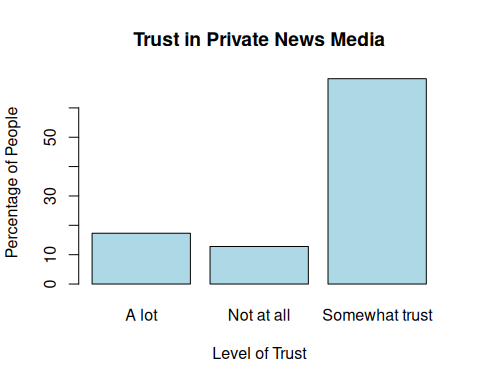
## 5. Descriptive statistics (sample means, standard deviations, proportions) and univariate displays

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



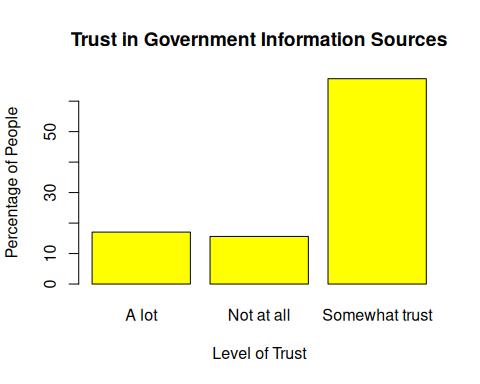
## my\_data$public\_media   
## Frequency Percent Valid Percent  
## A lot 388 16.167 17.26  
## Not at all 298 12.417 13.26  
## Somewhat trust 1562 65.083 69.48  
## NA's 152 6.333   
## Total 2400 100.000 100.00

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



## 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 = "Trust in Government Information Sources", y.axis = "percent", ylab = "Percentage of People", xlab = "Level of Trust", col="yellow");



## 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

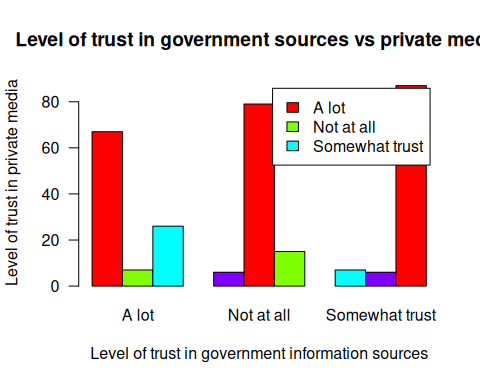
# Is there an association between levels of trust in government and levels of trust in private media sources?  
table1 = table(my\_data$government, my\_data$private\_media)  
table1

##   
## A lot Not at all Somewhat trust  
## A lot 251 17 108  
## Not at all 27 222 93  
## Somewhat trust 98 42 1344

# converting to a table of proportions  
prop\_table1 = round(prop.table(table1, 2), 2) \* 100;prop\_table1

##   
## A lot Not at all Somewhat trust  
## A lot 67 6 7  
## Not at all 7 79 6  
## Somewhat trust 26 15 87

View(prop\_table1)  
  
# bar plot  
barplot(prop\_table1, beside = TRUE, main = "Level of trust in government sources vs private media", xlab = "Level of trust in government information sources", ylab = "Level of trust in private media", legend = row.names(prop\_table1), col = rainbow(4), ylim = c(0, 90), las = 1)

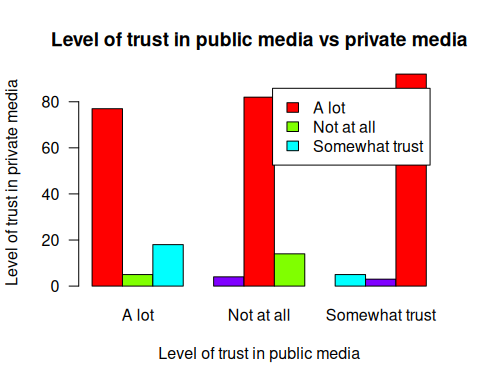


# Is there an association between levels of trust in government and levels of trust in private media sources?  
table2 = table(my\_data$public\_media, my\_data$private\_media)  
View(table2)

# converting to a table of proportions  
prop\_table2 = round(prop.table(table2, 2), 2) \* 100;prop\_table2

##   
## A lot Not at all Somewhat trust  
## A lot 77 4 5  
## Not at all 5 82 3  
## Somewhat trust 18 14 92

View(prop\_table2)  
  
# bar plot  
barplot(prop\_table2, beside = TRUE, main = "Level of trust in public media vs private media", xlab = "Level of trust in public media", ylab = "Level of trust in private media", legend = row.names(prop\_table2), col = rainbow(4), ylim = c(0, 90), las = 1)



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

**Research Question 1:** Is there an association between levels of trust in government information sources and levels of trust in private media sources?

**Explanatory variable:** level of trust in government sources. **Response variable:** level of trust in private media sources.

### Step 1: Stating The Claim

**Null hypothesis:** There is no association between levels of trust in government information sources and levels of trust in private media sources?

**Alternative hypothesis:** There is an association between levels of trust in government information sources and levels of trust in private media sources?

### Step 2: Collect and Summarise the Sample

# Sample proportion for each category of the explanatory variable  
prop\_alot = 80/sum(table1)  
cat("Sample proportion for A lot:", prop\_alot, "\n")

## Sample proportion for A lot: 0.03633061

prop\_not\_at\_all = 92/sum(table1)  
cat("Sample proportion for Not at all:", prop\_not\_at\_all, "\n")

## Sample proportion for Not at all: 0.0417802

prop\_somewhat = 128/sum(table1)  
cat("Sample proportion for Somewhat:", prop\_somewhat, "\n")

## Sample proportion for Somewhat: 0.05812897

### Step 3: Accessing The Evidence

The sample data satisfies all the conditions for performing a chi-square test.

chisq\_result <- chisq.test(my\_data$private\_media, my\_data$government);chisq\_result

##   
## Pearson's Chi-squared test  
##   
## data: my\_data$private\_media and my\_data$government  
## X-squared = 1789.3, df = 4, p-value < 2.2e-16

chisq\_result$observed

## my\_data$government  
## my\_data$private\_media A lot Not at all Somewhat trust  
## A lot 251 27 98  
## Not at all 17 222 42  
## Somewhat trust 108 93 1344

chisq\_result$expected

## my\_data$government  
## my\_data$private\_media A lot Not at all Somewhat trust  
## A lot 64.20345 58.39782 253.3987  
## Not at all 47.98183 43.64305 189.3751  
## Somewhat trust 263.81471 239.95913 1041.2262

chisq\_result$residuals

## my\_data$government  
## my\_data$private\_media A lot Not at all Somewhat trust  
## A lot 23.312544 -4.108669 -9.762145  
## Not at all -4.472689 26.998055 -10.709340  
## Somewhat trust -9.593100 -9.486979 9.383089

* Test statistic = 1789.3
* Degrees of freedom = 4
* P-value = 2.2e-16

### Step 4: Making Conclusion

The p-value is the probability of getting counts like those observed, assuming the null hypothesis is true. In this case, the p-value (2.2e-16) is less than the alpha value (0.05). This means that there is significant evidence provided by the data to reject the null hypothesis, and accept the alternative hypothesis. We can conclude that there is a relationship between the level of trust people have in government information sources and the level of trust they have in private media sources.

Since we rejected the null hypothesis, we could make a type 1 error. This can occur if there is no relationship between the level of trust people have in government information sources and the level of trust they have in private media sources (null hypothesis).

### Post Hoc Test

rownames(table1) = c("A lot", "Not at all", "Somewhat trust")  
colnames(table1) = c("A lot", "Not at all", "Somewhat trust")  
  
test1 = chisq.test(table1[,1:2]); test1

##   
## Pearson's Chi-squared test  
##   
## data: table1[, 1:2]  
## X-squared = 373.5, df = 2, p-value < 2.2e-16

test2 = chisq.test(table1[,2:3]); test2

##   
## Pearson's Chi-squared test  
##   
## data: table1[, 2:3]  
## X-squared = 897.03, df = 2, p-value < 2.2e-16

The first chi-square test is for the two categories: “A lot” and “Not at all”. The test statistic is 373.5 with 2 degrees of freedom. The p-value is 2.2e-16, which is lower than the alpha value. This means that the relationship between “A lot” and “Not at all” categories are statistically significant. As a result, we reject the null hypothesis and accept the alternative hypothesis.

The second chi-square test is for the two categories: “Not at all” and “Somewhat trust”. The test statistic is 897.03 with 2 degrees of freedom. The p-value is 2.2e-16, which is lower than the alpha value. This means that the relationship between “Not at all” and “Somewhat trust” categories are statistically significant. As a result, we reject the null hypothesis and accept the alternative hypothesis.

**Research Question 2:** Is there an association between levels of trust in public media sources and levels of trust in private media sources?

**Explanatory variable:** level of trust in public media sources. **Response variable:** level of trust in private media sources.

### Step 1: Stating The Claim

**Null hypothesis:** There is no association between levels of trust in public media sources and levels of trust in private media sources.

**Alternative hypothesis:** There is an association between levels of trust in public media sources and levels of trust in private media sources.

### Step 2: Collect and Summarise the Sample

# Sample proportion for each category of the explanatory variable  
prop\_alot = 86/sum(table2)  
cat("Sample proportion for A lot:", prop\_alot, "\n")

## Sample proportion for A lot: 0.03859964

prop\_not\_at\_all = 90/sum(table2)  
cat("Sample proportion for Not at all:", prop\_not\_at\_all, "\n")

## Sample proportion for Not at all: 0.04039497

prop\_somewhat = 124/sum(table2)  
cat("Sample proportion for Somewhat:", prop\_somewhat, "\n")

## Sample proportion for Somewhat: 0.0556553

### Step 3: Accessing The Evidence

The sample data satisfies all the conditions for performing a chi-square test.

chisq\_result <- chisq.test(my\_data$private\_media, my\_data$public\_media);chisq\_result

##   
## Pearson's Chi-squared test  
##   
## data: my\_data$private\_media and my\_data$public\_media  
## X-squared = 2519.9, df = 4, p-value < 2.2e-16

chisq\_result$observed

## my\_data$public\_media  
## my\_data$private\_media A lot Not at all Somewhat trust  
## A lot 296 18 69  
## Not at all 12 234 39  
## Somewhat trust 77 44 1439

chisq\_result$expected

## my\_data$public\_media  
## my\_data$private\_media A lot Not at all Somewhat trust  
## A lot 66.18268 50.88330 265.9340  
## Not at all 49.24820 37.86355 197.8882  
## Somewhat trust 269.56912 207.25314 1083.1777

chisq\_result$residuals

## my\_data$public\_media  
## my\_data$private\_media A lot Not at all Somewhat trust  
## A lot 28.249490 -4.609861 -12.076297  
## Not at all -5.307746 31.874810 -11.294884  
## Somewhat trust -11.728746 -11.339946 10.811433

* Test statistic = 2519.9
* Degrees of freedom = 4
* P-value = 2.2e-16

### Step 4: Making Conclusion

The p-value is the probability of getting counts like those observed, assuming the null hypothesis is true. In this case, the p-value (2.2e-16) is less than the alpha value (0.05). This means that there is significant evidence provided by the data to reject the null hypothesis, and accept the alternative hypothesis. We can conclude that there is a relationship between the level of trust people have in public media sources and the level of trust they have in private media sources.

Since we rejected the null hypothesis, we could make a type 1 error. This can occur if there is no relationship between the level of trust people have in public media sources and the level of trust they have in private media sources (null hypothesis).

### Post Hoc Test

rownames(table2) = c("A lot", "Not at all", "Somewhat trust")  
colnames(table2) = c("A lot", "Not at all", "Somewhat trust")  
  
test1 = chisq.test(table2[,1:2]); test1

##   
## Pearson's Chi-squared test  
##   
## data: table2[, 1:2]  
## X-squared = 450.67, df = 2, p-value < 2.2e-16

test2 = chisq.test(table2[,2:3]); test2

##   
## Pearson's Chi-squared test  
##   
## data: table2[, 2:3]  
## X-squared = 1191.2, df = 2, p-value < 2.2e-16

The first chi-square test is for the two categories: “A lot” and “Not at all”. The test statistic is 450.67 with 2 degrees of freedom. The p-value is 2.2e-16, which is lower than the alpha value. This means that the relationship between “A lot” and “Not at all” categories are statistically significant. As a result, we reject the null hypothesis and accept the alternative hypothesis.

The second chi-square test is for the two categories: “Not at all” and “Somewhat trust”. The test statistic is 1191.2 with 2 degrees of freedom. The p-value is 2.2e-16, which is lower than the alpha value. This means that the relationship between “Not at all” and “Somewhat trust” categories are statistically significant. As a result, we reject the null hypothesis and accept the alternative hypothesis.

## 8. Moderation

## 9. Save

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