Analysis on Afrobarometer round 9

Victor Asum

2024-11-21

**Is the perception of media freedom associated with citizens’ satisfaction with democracy in Ghana?**

*Explainatory Variable*: Perception of Media Freedom (Q33H)

*Response Variable*: Satisfaction of democracy (Q31)

**Does primary langauge used in interview influence the length of the interview?**

*Explanatory Variable*: Language of interview (Q102)

*Response Variable*: Length of interview (LENGTH)

# SECTION 1:

library(descr)

## Warning: package 'descr' was built under R version 4.4.2

library(haven)

## Warning: package 'haven' was built under R version 4.4.2

GH\_AFRO\_9 <- read\_sav("C:/Users/AiAsum/Downloads/afrobarometer\_release-dataset\_gha\_r9\_en\_2023-04-01.sav")

# Section 2:

Sub

vars <- c("Q33H", "Q31", "Q102", "LENGTH")   
cleanData <- GH\_AFRO\_9[vars]  
save(cleanData, file='PA2\_Research.RDATA')

# Section 2 b: Error handling and Labeling variable response categories

# first research questions  
  
# Perception of Media Freedom (Q33H)  
cleanData$Q33H <- as.character(cleanData$Q33H)  
  
cleanData$Q33H[cleanData$Q33H == "8"] <- NA  
cleanData$Q33H[cleanData$Q33H == "9"] <- NA  
cleanData$Q33H[cleanData$Q33H == "-1"] <- NA  
  
  
# Satisfaction of democracy (Q31)  
cleanData$Q31 <- as.character(cleanData$Q31)  
cleanData$Q31[cleanData$Q31 == "8"] <- NA  
cleanData$Q31[cleanData$Q31 == "9"] <- NA  
cleanData$Q31[cleanData$Q31 == "-1"] <- NA  
  
  
  
# Second research question  
# Language of interview (Q102)  
cleanData$Q102 <- as.character(cleanData$Q102)  
  
  
summary(cleanData$LENGTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 30.00 56.00 66.00 71.47 80.00 394.00

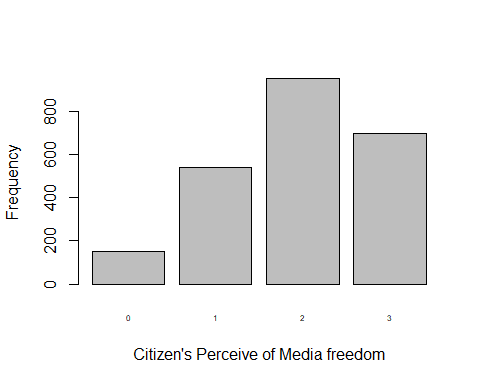
mean(cleanData$LENGTH, na.rm = TRUE)

## [1] 71.46644

sd(cleanData$LENGTH, na.rm = TRUE)

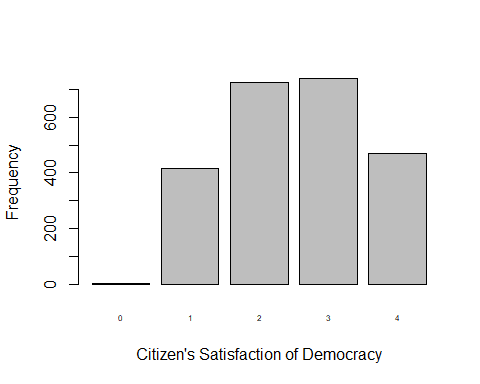
## [1] 28.4942

freq(cleanData$Q33H, cex.names=0.5, ylab = "Frequency", xlab = "Citizen's Perceive of Media freedom")



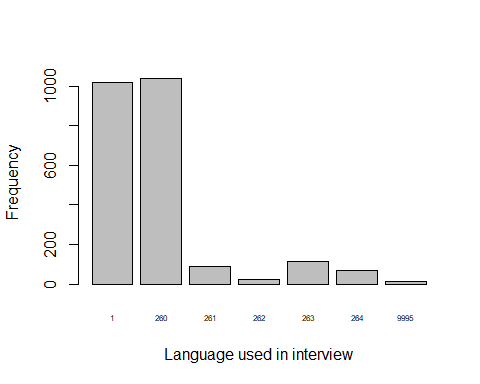
## cleanData$Q33H   
## Frequency Percent Valid Percent  
## 0 150 6.332 6.418  
## 1 540 22.794 23.107  
## 2 951 40.144 40.693  
## 3 696 29.379 29.782  
## NA's 32 1.351   
## Total 2369 100.000 100.000

freq(cleanData$Q31, cex.names=0.5, ylab = "Frequency", xlab = "Citizen's Satisfaction of Democracy")



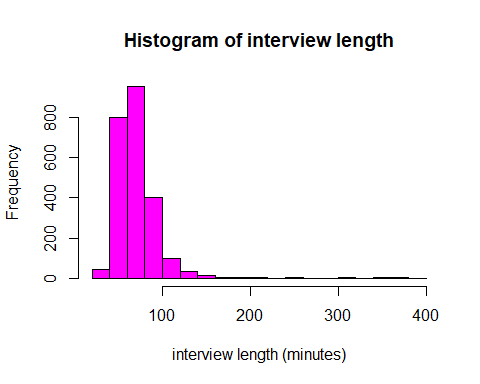
## cleanData$Q31   
## Frequency Percent Valid Percent  
## 0 3 0.1266 0.1274  
## 1 417 17.6024 17.7145  
## 2 725 30.6036 30.7986  
## 3 739 31.1946 31.3934  
## 4 470 19.8396 19.9660  
## NA's 15 0.6332   
## Total 2369 100.0000 100.0000

freq(cleanData$Q102, cex.names=0.5, ylab = "Frequency", xlab = "Language used in interview")



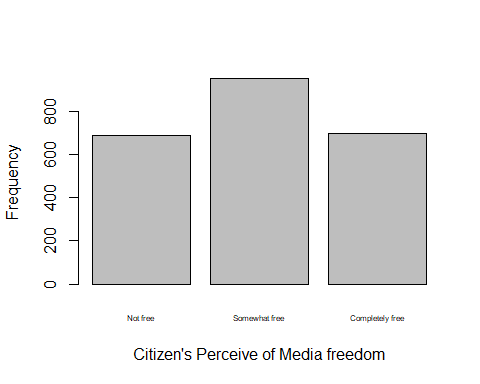
## cleanData$Q102   
## Frequency Percent  
## 1 1022 43.141  
## 260 1039 43.858  
## 261 91 3.841  
## 262 24 1.013  
## 263 112 4.728  
## 264 67 2.828  
## 9995 14 0.591  
## Total 2369 100.000

hist(cleanData$LENGTH, main="Histogram of interview length", xlab="interview length (minutes)", ylab="Frequency", col="#ff00ff")



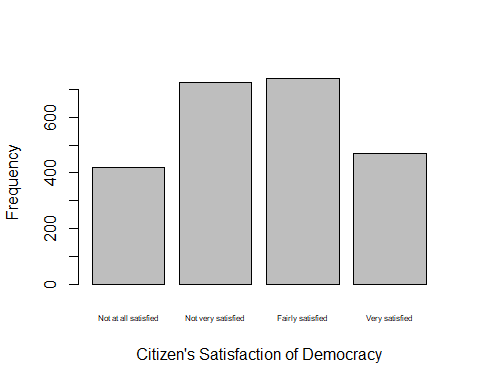
# Section 3 -

# I'm Creating secondary variables for media and democracy  
# I choose to combine Not at all free", "Not very free", as one,"Somewhat free", "Completely free"  
cleanData$Q33H[cleanData$Q33H == "0"] <- "1"  
  
# I choose to combine Not at all free", "Not very free", as one,"Somewhat free", "Completely free"  
cleanData$Q31[cleanData$Q31 == "0"] <- "1"  
  
  
# Second Research:  
# Primary Language used for the research  
cleanData$Q102 <- as.character(cleanData$Q102)  
cleanData$Q102[cleanData$Q102 == "261"] <- "9995"  
cleanData$Q102[cleanData$Q102 == "262"] <- "9995"  
cleanData$Q102[cleanData$Q102 == "263"] <- "9995"  
cleanData$Q102[cleanData$Q102 == "264"] <- "9995"  
  
  
# Changing to understanable variables  
 cleanData$Q33H <- factor(cleanData$Q33H,  
 levels = c("1", "2", "3"),  
 labels = c("Not free", "Somewhat free", "Completely free")  
 )  
  
 cleanData$Q31 <- factor(cleanData$Q31,  
 levels = c("1", "2", "3", "4"),  
 labels = c("Not at all satisfied",   
 "Not very satisfied", "Fairly satisfied",   
 "Very satisfied"))  
   
cleanData$Q102 <- factor(cleanData$Q102,  
 levels = c("1", "260", "9995"),  
 labels = c("English", "Twi", "Others"))  
  
freq(cleanData$Q33H, cex.names=0.5, ylab = "Frequency", xlab = "Citizen's Perceive of Media freedom")



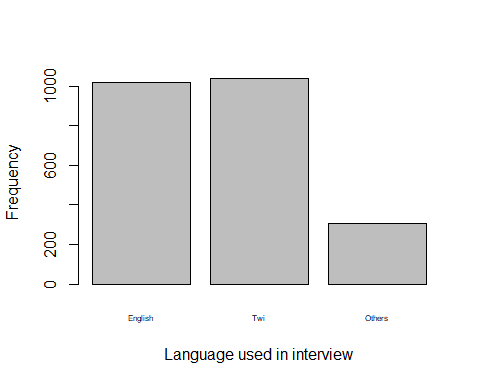
## cleanData$Q33H   
## Frequency Percent Valid Percent  
## Not free 690 29.126 29.53  
## Somewhat free 951 40.144 40.69  
## Completely free 696 29.379 29.78  
## NA's 32 1.351   
## Total 2369 100.000 100.00

freq(cleanData$Q31, cex.names=0.5, ylab = "Frequency", xlab = "Citizen's Satisfaction of Democracy")



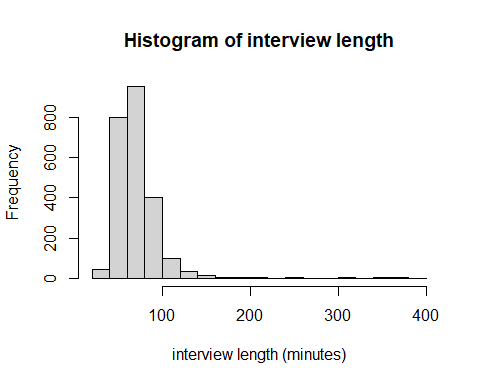
## cleanData$Q31   
## Frequency Percent Valid Percent  
## Not at all satisfied 420 17.7290 17.84  
## Not very satisfied 725 30.6036 30.80  
## Fairly satisfied 739 31.1946 31.39  
## Very satisfied 470 19.8396 19.97  
## NA's 15 0.6332   
## Total 2369 100.0000 100.00

freq(cleanData$Q102, cex.names=0.5, ylab = "Frequency", xlab = "Language used in interview")



## cleanData$Q102   
## Frequency Percent  
## English 1022 43.14  
## Twi 1039 43.86  
## Others 308 13.00  
## Total 2369 100.00

hist(cleanData$LENGTH, main = "Histogram of interview length", xlab = "interview length (minutes)")



# Section 4

# First Research  
  
mediaTable <- table(cleanData$Q31, cleanData$Q33H)  
mediaTable\_colProp <-prop.table(mediaTable, 2)  
round(mediaTable\_colProp, 3)

##   
## Not free Somewhat free Completely free  
## Not at all satisfied 0.267 0.138 0.143  
## Not very satisfied 0.356 0.342 0.216  
## Fairly satisfied 0.249 0.363 0.308  
## Very satisfied 0.128 0.156 0.333

table(cleanData$Q33H, useNA = "no")

##   
## Not free Somewhat free Completely free   
## 690 951 696

# Second research  
tapply(cleanData$LENGTH, cleanData$Q102, mean, na.rm = TRUE)

## English Twi Others   
## 71.19472 72.52069 68.81169

tapply(cleanData$LENGTH, cleanData$Q102, sd, na.rm = TRUE)

## English Twi Others   
## 25.61713 25.21517 43.81041

tapply(cleanData$LENGTH, cleanData$Q102, length)

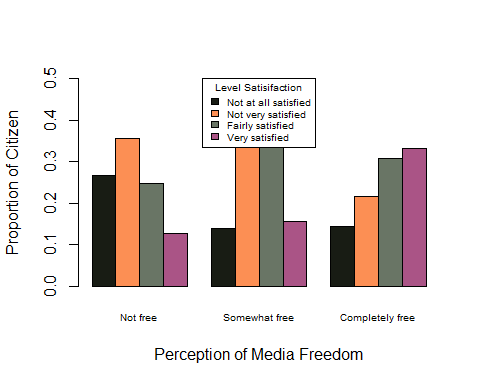
## English Twi Others   
## 1022 1039 308

table(cleanData$Q102, !is.na(cleanData$LENGTH))

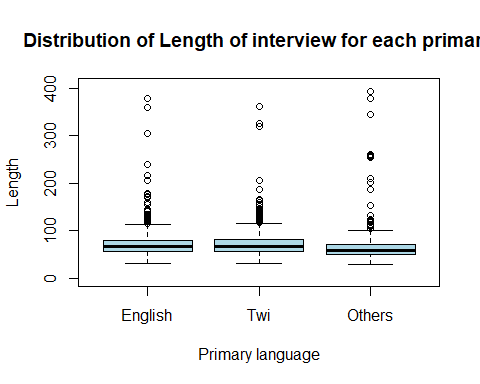
##   
## TRUE  
## English 1022  
## Twi 1039  
## Others 308

# SECTION 4:

colors1 <- c("#AA5486", "#47663B", "#E8ECD7", "#EED3B1")  
colors2 <- c("#181C14", "#FC8F54", "#697565","#AA5486")  
  
barplot(  
 mediaTable\_colProp,   
 beside = TRUE,   
 cex.names = 0.6,  
 ylim = c(0, 0.5),  
 col = colors2,   
 xlab = "Perception of Media Freedom",   
 ylab = "Proportion of Citizen"  
)  
  
legend(  
 "top",   
 legend = rownames(mediaTable\_colProp),  
 fill = colors2,  
 title = "Level Satisifaction",  
 cex = 0.6  
)



# Box and whiaker plot  
boxplot(cleanData$LENGTH ~ factor(cleanData$Q102),  
 names = levels(factor(cleanData$Q102)),  
 main = "Distribution of Length of interview for each primary",  
 xlab = "Primary language",  
 ylab = "Length",  
 ylim = c(0, max(cleanData$LENGTH, na.rm = TRUE) + 10),  
 col = "lightblue"  
)



## Summary

Based on the summary and visualized data, there appears to be a significant relationship between the perception of media freedom and satisfaction with democracy in Ghana. Respondents who perceived the media as “Not free” demonstrated a higher proportion of dissatisfaction (e.g., around 36% reported “Not very satisfied”) and a much lower proportion of satisfaction (e.g., approximately 13% reported being “Very satisfied”).In contrast, as perceptions of media freedom increased (from “Somewhat free” to “Completely free”), satisfaction with democracy also increased. For instance, among those who viewed the media as “Completely free,” about 33% reported being “Very satisfied,” while dissatisfaction decreased to less than 13% for the “Not at all satisfied” category. With this trend I propose that as citizens perceive greater media freedom, their satisfaction with democracy tends to improve.

Based on the summary and visualized data, interview lengths varied slightly across primary languages, with Twi showing the highest average (72.52 minutes), followed by English (71.19 minutes) and Others (68.81 minutes). However, the “Others” category displayed significantly higher variability (SD = 43.81), compared to English (SD = 25.62) and Twi (SD = 25.22).While the majority of interviews were of similar duration, all categories exhibited outliers, particularly with longer interviews in the “Others” group. This variability may reflect differences in linguistic or contextual factors influencing interview length.

# Section 7

# Testing for first research  
  
# Null Hypothesis: There is no association between the perception of media freedom and citizens' satisfaction with democracy in Ghana.  
  
# Alternative Hypothesis: There is an association between the perception of media freedom and citizens' satisfaction with democracy in Ghana.  
  
# Sample proportion  
sample\_media <- table(cleanData$Q31, cleanData$Q33H)  
sample\_prop\_media <-prop.table(mediaTable, 2)  
round(sample\_prop\_media, 3)

##   
## Not free Somewhat free Completely free  
## Not at all satisfied 0.267 0.138 0.143  
## Not very satisfied 0.356 0.342 0.216  
## Fairly satisfied 0.249 0.363 0.308  
## Very satisfied 0.128 0.156 0.333

media\_chisq\_results <- chisq.test(cleanData$Q31, cleanData$Q33H)  
media\_chisq\_results

##   
## Pearson's Chi-squared test  
##   
## data: cleanData$Q31 and cleanData$Q33H  
## X-squared = 176.38, df = 6, p-value < 2.2e-16

media\_chisq\_results$observed

## cleanData$Q33H  
## cleanData$Q31 Not free Somewhat free Completely free  
## Not at all satisfied 184 131 99  
## Not very satisfied 245 324 149  
## Fairly satisfied 171 344 213  
## Very satisfied 88 148 230

media\_chisq\_results$expected

## cleanData$Q33H  
## cleanData$Q31 Not free Somewhat free Completely free  
## Not at all satisfied 122.4557 168.5546 122.9897  
## Not very satisfied 212.3749 292.3242 213.3009  
## Fairly satisfied 215.3328 296.3955 216.2717  
## Very satisfied 137.8366 189.7257 138.4377

media\_chisq\_results$residuals

## cleanData$Q33H  
## cleanData$Q31 Not free Somewhat free Completely free  
## Not at all satisfied 5.5615798 -2.8926297 -2.1631675  
## Not very satisfied 2.2387230 1.8526602 -4.4027190  
## Fairly satisfied -3.0211321 2.7651069 -0.2224719  
## Very satisfied -4.2448890 -3.0292885 7.7819736

# I think since the p value is < 0.05, I need to conduct a Post Hoc test  
  
# 1. Subset for Not Free and Somewhat Free  
mySubset\_0\_1 <- subset(cleanData, Q33H == "Not free" | Q33H == "Somewhat free")  
  
chisq\_0\_1 <- chisq.test(mySubset\_0\_1$Q31, mySubset\_0\_1$Q33H)  
print(chisq\_0\_1)

##   
## Pearson's Chi-squared test  
##   
## data: mySubset\_0\_1$Q31 and mySubset\_0\_1$Q33H  
## X-squared = 53.571, df = 3, p-value = 1.385e-11

# Subset for Not Free and Completely Free  
mySubset\_0\_2 <- subset(cleanData, Q33H == "Not free" | Q33H == "Completely free")  
chisq\_0\_2 <- chisq.test(mySubset\_0\_2$Q31, mySubset\_0\_2$Q33H)  
print(chisq\_0\_2)

##   
## Pearson's Chi-squared test  
##   
## data: mySubset\_0\_2$Q31 and mySubset\_0\_2$Q33H  
## X-squared = 116.92, df = 3, p-value < 2.2e-16

# Subset for Somewhat Free and Completely Free  
mySubset\_1\_2 <- subset(cleanData, Q33H == "Somewhat free" | Q33H == "Completely free")  
chisq\_1\_2 <- chisq.test(mySubset\_1\_2$Q31, mySubset\_1\_2$Q33H)  
print(chisq\_1\_2)

##   
## Pearson's Chi-squared test  
##   
## data: mySubset\_1\_2$Q31 and mySubset\_1\_2$Q33H  
## X-squared = 79.734, df = 3, p-value < 2.2e-16

# o avoid a Type I error (false positive)  
  
# END OF FIRST RESEARCH  
  
# SECOND RESEARCH  
  
# Null Hypothesis: The primary language used in the interview has no effect on the length of the interview.  
# H0: μ(English) =μ2(Twi) =μ3 (Others)   
  
  
# Alternative hypothesis: The primary language used in the interview does influence the length of the interview.  
# H1 :At least one language group has a different mean interview length.   
  
# Assumptions  
 # Independence  
 # Normality  
 # Homogeneity of Variance  
  
# Data summary:  
tapply(cleanData$LENGTH, cleanData$Q102, mean, na.rm = TRUE)

## English Twi Others   
## 71.19472 72.52069 68.81169

tapply(cleanData$LENGTH, cleanData$Q102, sd, na.rm = TRUE)

## English Twi Others   
## 25.61713 25.21517 43.81041

tapply(cleanData$LENGTH, cleanData$Q102, length)

## English Twi Others   
## 1022 1039 308

table(cleanData$Q102, !is.na(cleanData$LENGTH))

##   
## TRUE  
## English 1022  
## Twi 1039  
## Others 308

#Conduct ANOVA test  
anova\_results <- aov(cleanData$LENGTH ~ cleanData$Q102, data = cleanData)  
summary(anova\_results)

## Df Sum Sq Mean Sq F value Pr(>F)  
## cleanData$Q102 2 3401 1700.5 2.096 0.123  
## Residuals 2366 1919225 811.2

#For p-value < 0.05 conduct a Post Hoc test  
TukeyHSD(anova\_results)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = cleanData$LENGTH ~ cleanData$Q102, data = cleanData)  
##   
## $`cleanData$Q102`  
## diff lwr upr p adj  
## Twi-English 1.325977 -1.616665 4.2686184 0.5411308  
## Others-English -2.383028 -6.724696 1.9586402 0.4025177  
## Others-Twi -3.709005 -8.042440 0.6244302 0.1106531

## STEPS FOR FIRST RESEARCH

#### State Claim:

First, I started by making a claim about the relationship between media freedom and satisfaction with democracy. I needed to state my hypotheses for this test formally.

* My null hypothesis (H₀) was that there is no relationship between how people feel about media freedom and how satisfied they are with democracy.
* My alternative hypothesis (H₁) was that there is a relationship between perceptions of media freedom and satisfaction of democracy.

#### Step 2: Collect and Summarize the Sample

Next, I used R build functions, and then I calculated the proportions of people in each combination of satisfaction and media freedom categories and displayed them in a contingency table. For example:

Not Free (media perception) and Not at All Satisfied (democracy satisfaction): 26.7% Somewhat Free and Fairly Satisfied: 36.3% Completely Free and Very Satisfied: 33.3% These proportions helped me understand how the groups were distributed.

Counts for Each Category: I created a table to show how many people fell into each group: For example, 184 people said the media was “Not Free” and they were “Not at All Satisfied.” 230 people said the media was “Completely Free” and they were “Very Satisfied.”

Expected Counts: I calculated the expected counts for each cell in the table to compare them with the observed counts.

Sample Size: The total sample size was 2366 respondents, and there were no missing values, so I didn’t need to remove anything.

#### Step 3: Assess the Evidence

After summarizing the data, I ran a Chi-Square test to see if there was a relationship between media freedom and satisfaction with democracy. The test gave me the following results:

Overall Test: X² = 176.38, df = 6, p-value < 2.2e-16 This means the result is highly significant because the p-value is way smaller than 0.05.

Post-Hoc Analysis: Because there were three categories for media freedom (Not Free, Somewhat Free, and Completely Free), I did additional tests (pairwise comparisons) to find out which groups had the most significant differences.

Not Free vs. Somewhat Free: X² = 53.57, p-value = 1.39e-11 Not Free vs. Completely Free: X² = 116.92, p-value < 2.2e-16 Somewhat Free vs. Completely Free: X² = 79.73, p-value < 2.2e-16 All these results showed that the differences between these groups were statistically significant.

#### Step 4: Make a Conclusion

From the test results, I could confidently say that the perception of media freedom is associated with people’s satisfaction with democracy in Ghana because the p-values were statistically significant.

Since the p-value was less than 0.05, I rejected the null hypothesis. This means that people’s views on whether the media is free or not are likely related to how satisfied they feel with democracy.

Type of Error: If I made a mistake here, it would be a Type I error. This means I might have said there’s a relationship when there really isn’t one. That is why I needed to conduct the post-hoc analysis further.

## STEPS FOR SECOND RESEARCH

#### Step 1: State the Claim

My null hypothesis (H₀) was that the primary language used in the interview does not influence the length of the interview. In other words, the mean interview length is the same for all language groups (e.g., English, Twi, Others). (H₀): u1 = u2 = u3,where u1 = mean of English language, u2= mean of Twi and u3= mean of Other languages

My alternative hypothesis (H₁) was that the primary language used in the interview does influence the length of the interview. This means that at least one language group would have a different mean interview length compared to the others.

#### Step 2: Collect and Summarize the Sample

Next, I looked at the data to summarize the important details. Here’s what I found:

Groups and Data: There were three language groups in my data: English, Twi, and Others. The dependent variable was the length of the interview, measured in minutes.

Sample Sizes: The total number of interviews (sample size) was 2366. Each group had enough data to analyze properly.

Summary Statistics: I calculated the mean interview lengths for each language group. I found that the mean of the times for English, Twi, and Others were *71.19472*, *72.52069*, and *68.81169*, respectively. From this, I could tell that the differences between the group means were minor (e.g., Twi was about 1.3 minutes longer than English on average). However, to validate, I needed to conduct an ANOVA test.

Assumptions for ANOVA: Before running the ANOVA test, I checked the assumptions: Independence: Each interview was conducted separately, so this was satisfied. Normality: The interview lengths for each language group were approximately normally distributed from the box plot I did above. Homogeneity of Variance: I assumed the variances across the groups were similar enough for the test to work properly.

#### Step 3: Assess the Evidence

I used an ANOVA test to compare the average interview lengths across the three language groups. Here’s what the test results showed:

Degrees of Freedom (Df): The between-groups degrees of freedom (for the language variable) was 2 because there were three groups (3-1 = 2).

The residual degrees of freedom (for the within-group variation) was 2366, based on the total number of data points minus the number of groups.

The sum of Squares (Sum Sq): The variation between the groups (language groups) was 3401. The variation within the groups (individual differences in interview lengths) was 1919225.

Mean Squares (Mean Sq): The mean square for the groups was 1700.5 (3401 / 2). The mean square for the residuals was 811.2 (1919225 / 2366).

F-Statistic and p-Value: The F-statistic was 2.096, and the p-value was 0.123.

*Since the p-value was more significant than 0.05, I failed to reject the null hypothesis. This means there was no strong evidence that the primary language used in the interview affected the length of the interview*.

Post-Hoc Analysis (Tukey’s Test): Even though the ANOVA test wasn’t significant, I ran a Tukey HSD test to look at the pairwise comparisons between the language groups to avoid type II errors.

Twi vs. English: The mean difference in interview length was 1.326 minutes, with a confidence interval of -1.617 to 4.269. The p-value was 0.5411, meaning the difference was not significant.

Others vs. English: The mean difference was -2.383 minutes, with a confidence interval of -6.725 to 1.959. The p-value was 0.4025, meaning the difference was not significant.

Others vs. Twi: The mean difference was -3.709 minutes, with a confidence interval of -8.042 to 0.624. The p-value was 0.1107, meaning the difference was also not significant.

Type of Error: If I made a mistake here, it would be a Type II error. This means I might have missed a real difference between the language groups, even though the test didn’t find one.

#### Step 4: Make a Conclusion

Based on the results of the ANOVA test, I concluded that the primary language used in the interview does not significantly affect the length of the interview. In other words, the average interview lengths for English, Twi, and Others were not different enough to be considered statistically significant.