PA\_3

Jonathan Kuug

4/3/2022

## Research Questions

1. “Are the country’s current economic conditions associated with how often officials go unpunished?”

* -Explanatory: How often officials go unpunished
* -Response : Current economic conditions

1. “Is people’s refusal to pay taxes associated with the overall direction of the country?”

* -Explanatory: Tax avoidance,
* -Response: Overall direction of the country

1. “Is improving the standard of living of the poor associated with managing the economy?”

* -Explanatory: Management of the economy
* -Response: Standard of living of the poor

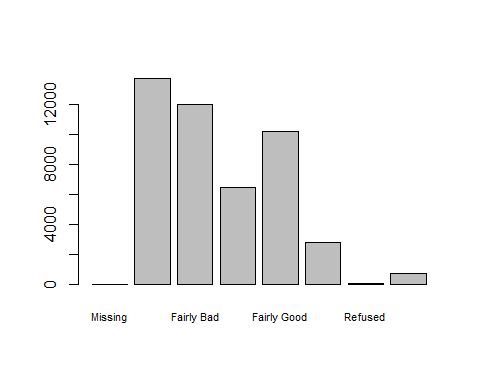
## 1. Load data set(s) and libraries

## 2. Create variable subset

#Interested variables  
vars <- c(  
 "COUNTRY",  
 "Q4A",  
 "Q42E",  
 "Q48D",  
 "Q3",  
 "Q56B",  
 "Q56A"  
   
)  
#Usable Data Objects/Frames  
AfroData <- as.data.frame(AfrobarometerMerged)  
myData <- AfroData[vars]

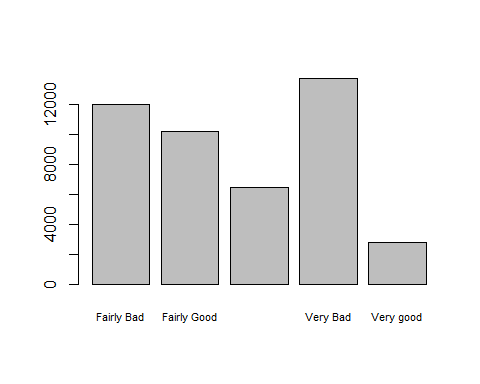
## 3. Data management I: check for and recode errors and NAs

#Data Management Economic condtions vars  
freq(myData$Q4A, cex.names=0.7)



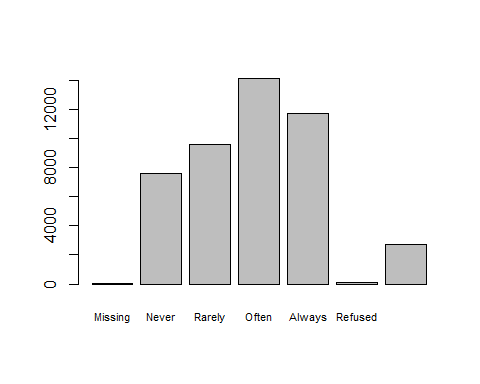
## myData$Q4A   
## Frequency Percent  
## Missing 0 0.00000  
## Very Bad 13706 29.91074  
## Fairly Bad 11984 26.15281  
## Neither good nor bad 6418 14.00607  
## Fairly Good 10178 22.21155  
## Very good 2777 6.06028  
## Refused 27 0.05892  
## Don't know 733 1.59963  
## Total 45823 100.00000

myData$Economic.conditions <- rep(NA, nrow(myData))  
myData$Economic.conditions[myData$Q4A == "Very Bad"] = "Very Bad"  
myData$Economic.conditions[myData$Q4A == "Fairly Bad"] = "Fairly Bad"  
myData$Economic.conditions[myData$Q4A == "Neither good nor bad"] = "Neither good nor bad"  
myData$Economic.conditions[myData$Q4A == "Fairly Good"] = "Fairly Good"  
myData$Economic.conditions[myData$Q4A == "Very good"] = "Very good"  
freq(myData$Economic.conditions, cex.names = 0.7)



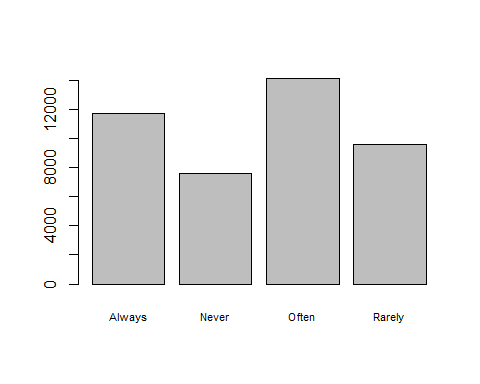
## myData$Economic.conditions   
## Frequency Percent Valid Percent  
## Fairly Bad 11984 26.153 26.594  
## Fairly Good 10178 22.212 22.586  
## Neither good nor bad 6418 14.006 14.242  
## Very Bad 13706 29.911 30.415  
## Very good 2777 6.060 6.162  
## NA's 760 1.659   
## Total 45823 100.000 100.000

#Data Management How Often Officials Go Unpunished vars  
freq(myData$Q42E, cex.names=0.7)



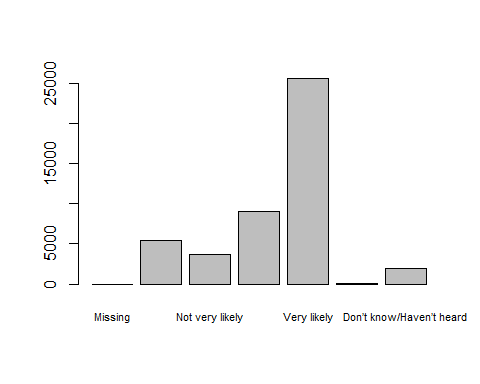
## myData$Q42E   
## Frequency Percent  
## Missing 2 4.365e-03  
## Never 7620 1.663e+01  
## Rarely 9557 2.086e+01  
## Often 14112 3.080e+01  
## Always 11709 2.555e+01  
## Refused 82 1.789e-01  
## Don’t know 2741 5.982e+00  
## Total 45823 1.000e+02

myData$Officials.unpunished <- rep(NA, nrow(myData))  
myData$Officials.unpunished[myData$Q42E == "Never"] = "Never"  
myData$Officials.unpunished[myData$Q42E == "Rarely"] = "Rarely"  
myData$Officials.unpunished[myData$Q42E == "Often"] = "Often"  
myData$Officials.unpunished[myData$Q42E == "Always"] = "Always"  
freq(myData$Officials.unpunished, cex.names=0.7)



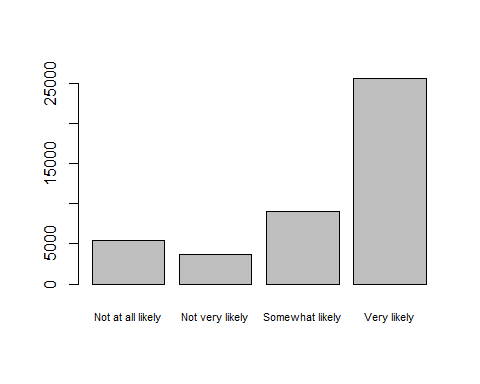
## myData$Officials.unpunished   
## Frequency Percent Valid Percent  
## Always 11709 25.553 27.23  
## Never 7620 16.629 17.72  
## Often 14112 30.797 32.82  
## Rarely 9557 20.856 22.23  
## NA's 2825 6.165   
## Total 45823 100.000 100.00

#Data Management Tax Avoidance vars  
freq(myData$Q48D, cex.names=0.7)



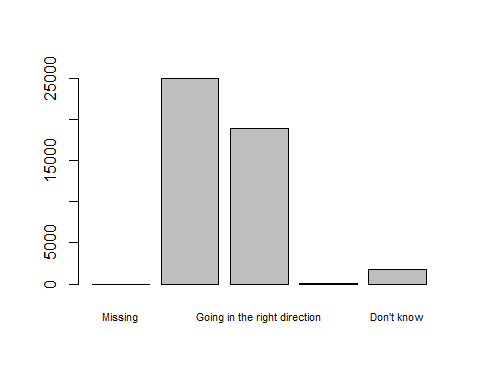
## myData$Q48D   
## Frequency Percent  
## Missing 1 2.182e-03  
## Not at all likely 5440 1.187e+01  
## Not very likely 3729 8.138e+00  
## Somewhat likely 9069 1.979e+01  
## Very likely 25629 5.593e+01  
## Refused 61 1.331e-01  
## Don’t know/Haven’t heard 1894 4.133e+00  
## Total 45823 1.000e+02

myData$Tax.avoidance <- rep(NA, nrow(myData))  
myData$Tax.avoidance[myData$Q48D == "Not at all likely"] = "Not at all likely"  
myData$Tax.avoidance[myData$Q48D == "Not very likely"] = "Not very likely"  
myData$Tax.avoidance[myData$Q48D == "Somewhat likely"] = "Somewhat likely"  
myData$Tax.avoidance[myData$Q48D == "Very likely"] = "Very likely"  
freq(myData$Tax.avoidance, cex.names=0.7)



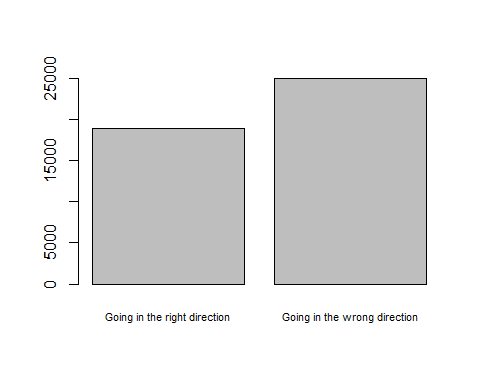
## myData$Tax.avoidance   
## Frequency Percent Valid Percent  
## Not at all likely 5440 11.872 12.401  
## Not very likely 3729 8.138 8.501  
## Somewhat likely 9069 19.791 20.674  
## Very likely 25629 55.930 58.424  
## NA's 1956 4.269   
## Total 45823 100.000 100.000

#Data Management Direction of the Country vars  
freq(myData$Q3, cex.names=0.7)



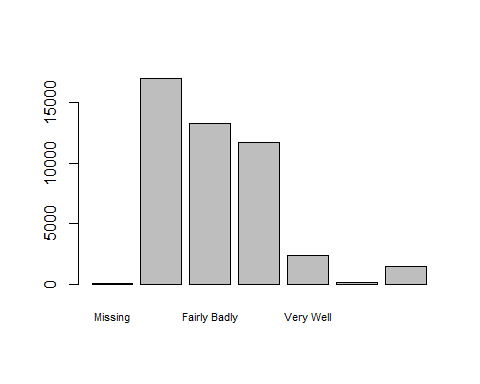
## myData$Q3   
## Frequency Percent  
## Missing 2 4.365e-03  
## Going in the wrong direction 25000 5.456e+01  
## Going in the right direction 18899 4.124e+01  
## Refused 114 2.488e-01  
## Don't know 1808 3.946e+00  
## Total 45823 1.000e+02

myData$Overall.country\_direction <- rep(NA, nrow(myData))  
myData$Overall.country\_direction[myData$Q3 == "Going in the wrong direction"] = "Going in the wrong direction"  
myData$Overall.country\_direction[myData$Q3 == "Going in the right direction"] = "Going in the right direction"  
freq(myData$Overall.country\_direction, cex.names=0.7)



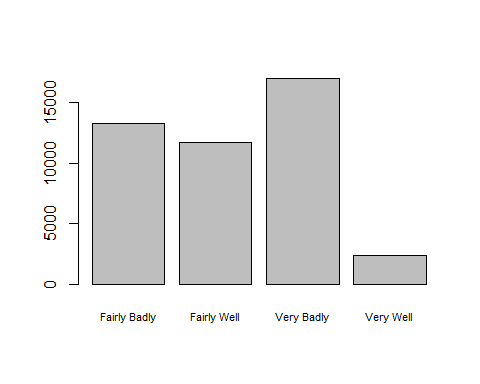
## myData$Overall.country\_direction   
## Frequency Percent Valid Percent  
## Going in the right direction 18899 41.243 43.05  
## Going in the wrong direction 25000 54.558 56.95  
## NA's 1924 4.199   
## Total 45823 100.000 100.00

#Data Management Living Standards vars  
freq(myData$Q56B, cex.names=0.7)



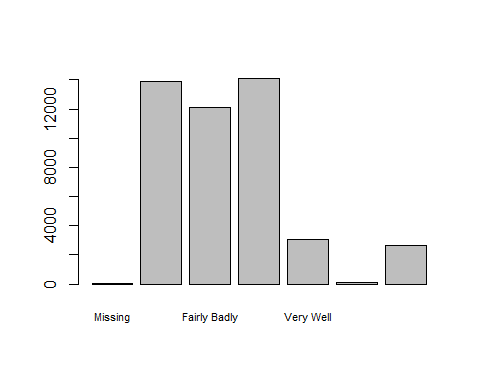
## myData$Q56B   
## Frequency Percent  
## Missing 20 0.04365  
## Very Badly 16972 37.03817  
## Fairly Badly 13249 28.91343  
## Fairly Well 11732 25.60286  
## Very Well 2348 5.12406  
## Refused 85 0.18550  
## Don't know / Haven’t heard enough 1417 3.09233  
## Total 45823 100.00000

myData$Living.standards <- rep(NA, nrow(myData))  
myData$Living.standards[myData$Q56B == "Very Badly"] = "Very Badly"  
myData$Living.standards[myData$Q56B == "Fairly Badly"] = "Fairly Badly"  
myData$Living.standards[myData$Q56B == "Fairly Well"] = "Fairly Well"  
myData$Living.standards[myData$Q56B == "Very Well"] = "Very Well"  
freq(myData$Living.standards, cex.names=0.7)



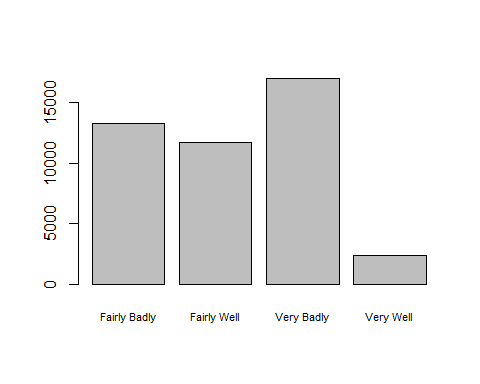
## myData$Living.standards   
## Frequency Percent Valid Percent  
## Fairly Badly 13249 28.913 29.91  
## Fairly Well 11732 25.603 26.48  
## Very Badly 16972 37.038 38.31  
## Very Well 2348 5.124 5.30  
## NA's 1522 3.321   
## Total 45823 100.000 100.00

#Data Management Economy handling/management   
freq(myData$Q56A, cex.names=0.7)



## myData$Q56A   
## Frequency Percent  
## Missing 20 0.04365  
## Very Badly 13863 30.25337  
## Fairly Badly 12078 26.35794  
## Fairly Well 14079 30.72475  
## Very Well 3046 6.64732  
## Refused 95 0.20732  
## Don't know / Haven’t heard enough 2642 5.76566  
## Total 45823 100.00000

myData$Economy.management <- rep(NA, nrow(myData))  
myData$Economy.management[myData$Q56B == "Very Badly"] = "Very Badly"  
myData$Economy.management[myData$Q56B == "Fairly Badly"] = "Fairly Badly"  
myData$Economy.management[myData$Q56B == "Fairly Well"] = "Fairly Well"  
myData$Economy.management[myData$Q56B == "Very Well"] = "Very Well"  
freq(myData$Economy.management, cex.names=0.7)



## myData$Economy.management   
## Frequency Percent Valid Percent  
## Fairly Badly 13249 28.913 29.91  
## Fairly Well 11732 25.603 26.48  
## Very Badly 16972 37.038 38.31  
## Very Well 2348 5.124 5.30  
## NA's 1522 3.321   
## Total 45823 100.000 100.00

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

GhanaData <- myData[myData$COUNTRY == "Ghana",]

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

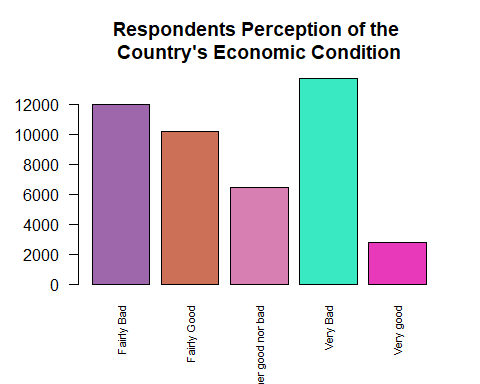
#Sample Proportions  
#How often officials go unpunished against the current economic conditions in Ghana  
table1<- table(GhanaData$Officials.unpunished, GhanaData$Economic.conditions)  
propTable1 <- round(prop.table(table1),3)\* 100  
propTable1

##   
## Fairly Bad Fairly Good Neither good nor bad Very Bad Very good  
## Always 7.7 8.6 2.7 9.7 2.1  
## Never 3.9 4.2 1.2 4.2 1.5  
## Often 10.2 12.2 4.6 9.6 2.0  
## Rarely 4.5 4.7 1.7 3.9 0.8

#Overall direction of the country against tendency to avoid tax.  
table2 <- table(GhanaData$Overall.country\_direction, GhanaData$Tax.avoidance)  
propTable2 <- round(prop.table(table2), 3) \* 100  
propTable2

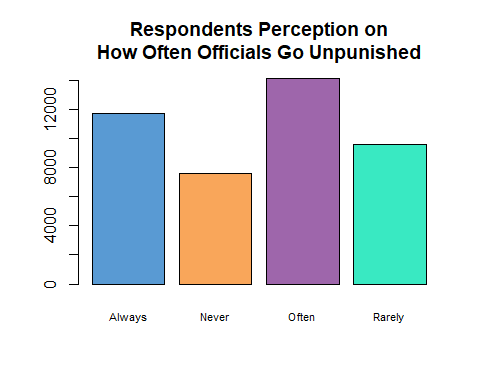
##   
## Not at all likely Not very likely  
## Going in the right direction 4.2 2.9  
## Going in the wrong direction 3.7 3.4  
##   
## Somewhat likely Very likely  
## Going in the right direction 7.6 36.5  
## Going in the wrong direction 9.6 32.3

# Standard of living against handling of the economy  
table3 <- table(GhanaData$Living.standards, GhanaData$Economy.management)  
propTable3 <- round(prop.table(table3), 3) \* 100  
  
#Uni-variate Displays  
freq(myData$Economic.conditions, main="Respondents Perception of the \nCountry's Economic Condition", col = c("#9e66ab","#cd7058","#d77fb3", "#39E9C2", "#E939BB"), cex.names=0.7, las=2)



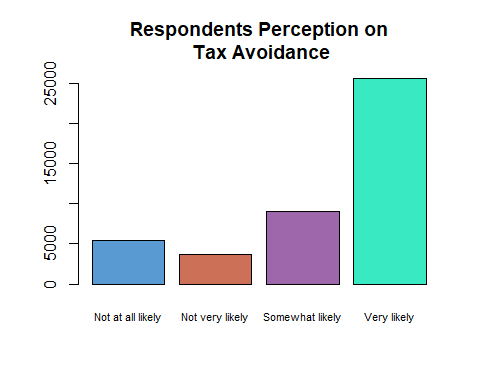
## myData$Economic.conditions   
## Frequency Percent Valid Percent  
## Fairly Bad 11984 26.153 26.594  
## Fairly Good 10178 22.212 22.586  
## Neither good nor bad 6418 14.006 14.242  
## Very Bad 13706 29.911 30.415  
## Very good 2777 6.060 6.162  
## NA's 760 1.659   
## Total 45823 100.000 100.000

freq(myData$Officials.unpunished, main="Respondents Perception on\nHow Often Officials Go Unpunished", col = c("#599ad3", "#f9a65a", "#9e66ab", "#39E9C2", "#E939BB"), cex.names=0.7)



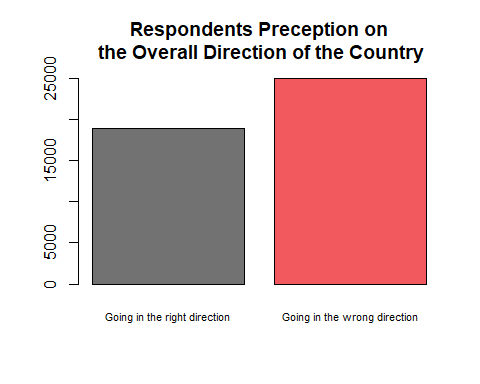
## myData$Officials.unpunished   
## Frequency Percent Valid Percent  
## Always 11709 25.553 27.23  
## Never 7620 16.629 17.72  
## Often 14112 30.797 32.82  
## Rarely 9557 20.856 22.23  
## NA's 2825 6.165   
## Total 45823 100.000 100.00

freq(myData$Tax.avoidance, main="Respondents Perception on\n Tax Avoidance", col = c("#599ad3","#cd7058", "#9e66ab", "#39E9C2", "#E939BB"), cex.names=0.7)



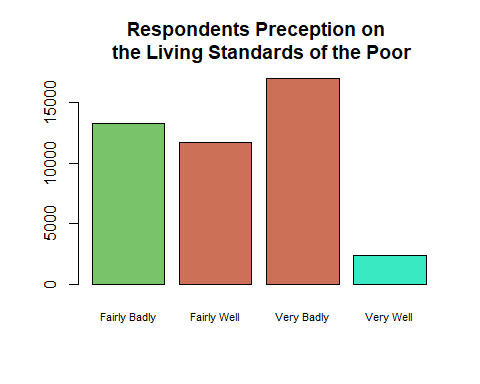
## myData$Tax.avoidance   
## Frequency Percent Valid Percent  
## Not at all likely 5440 11.872 12.401  
## Not very likely 3729 8.138 8.501  
## Somewhat likely 9069 19.791 20.674  
## Very likely 25629 55.930 58.424  
## NA's 1956 4.269   
## Total 45823 100.000 100.000

freq(myData$Overall.country\_direction, main="Respondents Preception on\n the Overall Direction of the Country",col= c("#727272","#f1595f","#79c36a", "#39E9C2", "#E939BB"), cex.names=0.7)



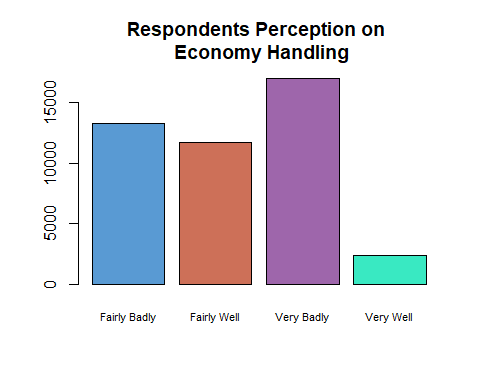
## myData$Overall.country\_direction   
## Frequency Percent Valid Percent  
## Going in the right direction 18899 41.243 43.05  
## Going in the wrong direction 25000 54.558 56.95  
## NA's 1924 4.199   
## Total 45823 100.000 100.00

freq(myData$Living.standards, col= c("#79c36a","#cd7058", "#cd7058", "#39E9C2", "#E939BB"), main="Respondents Preception on \n the Living Standards of the Poor", cex.names=0.7)



## myData$Living.standards   
## Frequency Percent Valid Percent  
## Fairly Badly 13249 28.913 29.91  
## Fairly Well 11732 25.603 26.48  
## Very Badly 16972 37.038 38.31  
## Very Well 2348 5.124 5.30  
## NA's 1522 3.321   
## Total 45823 100.000 100.00

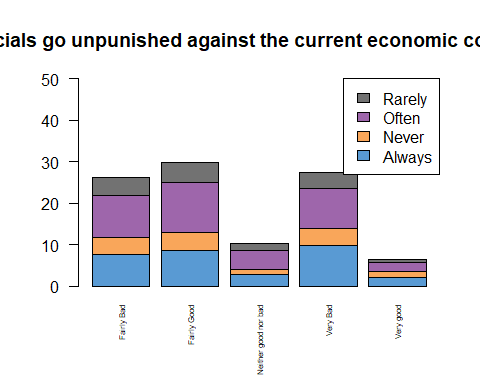
freq(myData$Economy.management, main="Respondents Perception on \n Economy Handling",col = c("#599ad3","#cd7058", "#9e66ab", "#39E9C2", "#E939BB"), cex.names=0.7)



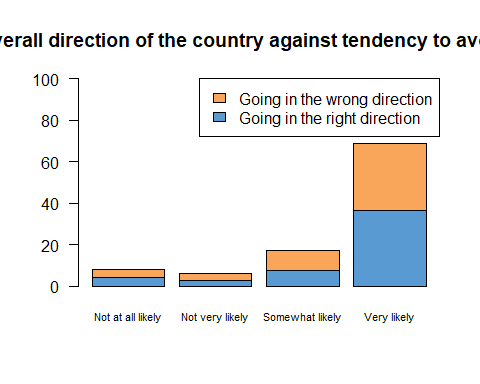
## myData$Economy.management   
## Frequency Percent Valid Percent  
## Fairly Badly 13249 28.913 29.91  
## Fairly Well 11732 25.603 26.48  
## Very Badly 16972 37.038 38.31  
## Very Well 2348 5.124 5.30  
## NA's 1522 3.321   
## Total 45823 100.000 100.00

## 6. Bivariate tables and graphs

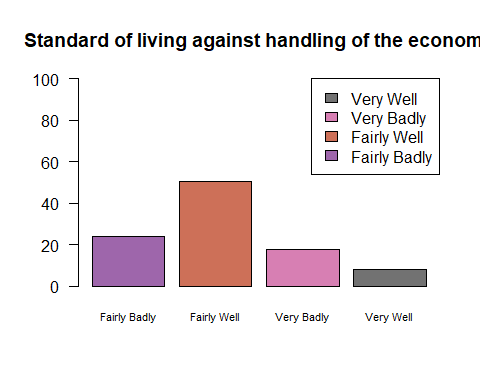
#How often officials go unpunished against the current economic conditions in Ghana(Stack bar plot)  
barplot(propTable1,ylim = c(0, 50),cex.names = 0.47,las=2, col=c("#599ad3", "#f9a65a","#9e66ab", "#727272",  
"#f1595f") ,legend.text=TRUE, args.legend = list(x="topright"), main ="How often officials go unpunished against the current economic conditions in Ghana")



#Overall direction of the country against tendency to avoid tax.  
barplot(propTable2, ylim = c(0, 100), las=1, cex.names=0.7, col = c("#599ad3","#f9a65a"), legend.text=TRUE, args.legend = list(x="topright"), main="Overall direction of the country against tendency to avoid tax")



# Standard of living against handling of the economy  
barplot(propTable3, ylim= c(0, 100), las=1, cex.names= 0.7, col = c("#9e66ab","#cd7058","#d77fb3", "#727272","#f1595f","#79c36a"), legend.text=TRUE, args.legend = list(x="topright"), main="Standard of living against handling of the economy")

 ## Summary In this exercise, i under went a statistical analysis to gain broader understanding in the relationship between three(3) associations involving six(6) different variables presented and recorded in the AfroBarometer dataset. To be more particular, i seeked to gain insights to the relationship between:

* Economic conditions and How often officials go unpunished
* Tax avoidance and the Overall direction of the country(Ghana)
* Standard of living of the poor and handling the economy

After going through the data management exercise, I was able to eliminate redundant responses such as, “I don’t know”, “refused”, “Missing”, allowing focus on more valid responses which were appropriate for proper analysis.

From the Bi-variate distribution of the first association(i.e using the stack bar plot), we notice in general; regardless of their individuals perception there is a common trend most beleive the rate officials go unpunished is between ” always” or “often”. Whereas the minority beleive otherwise (i.e rarely or never).

From the Distribution of the second association, proportions accross all bars(i.e responses) is fairly even.

Moreover, in the third Bi-variate distribution, the data presented allows us to deduce most people sampled Ghana to have a fairly well managed economy and the poor having decent living standards.

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

#Relationship 1  
chisq\_results <- chisq.test(GhanaData$Officials.unpunished, GhanaData$Economic.conditions)  
chisq\_results

##   
## Pearson's Chi-squared test  
##   
## data: GhanaData$Officials.unpunished and GhanaData$Economic.conditions  
## X-squared = 26.725, df = 12, p-value = 0.008463

chisq\_results$observed

## GhanaData$Economic.conditions  
## GhanaData$Officials.unpunished Fairly Bad Fairly Good Neither good nor bad  
## Always 172 193 60  
## Never 88 93 26  
## Often 229 274 102  
## Rarely 101 105 38  
## GhanaData$Economic.conditions  
## GhanaData$Officials.unpunished Very Bad Very good  
## Always 218 46  
## Never 94 34  
## Often 216 44  
## Rarely 88 19

chisq\_results$expected

## GhanaData$Economic.conditions  
## GhanaData$Officials.unpunished Fairly Bad Fairly Good Neither good nor bad  
## Always 181.47768 204.54688 69.51518  
## Never 88.23661 99.45312 33.79911  
## Often 227.83482 256.79688 87.27232  
## Rarely 92.45089 104.20312 35.41339  
## GhanaData$Economic.conditions  
## GhanaData$Officials.unpunished Very Bad Very good  
## Always 189.475 43.98527  
## Never 92.125 21.38616  
## Often 237.875 55.22098  
## Rarely 96.525 22.40759

chisq\_results$residuals

## GhanaData$Economic.conditions  
## GhanaData$Officials.unpunished Fairly Bad Fairly Good Neither good nor bad  
## Always -0.70354255 -0.80736151 -1.14124036  
## Never -0.02518857 -0.64708430 -1.34150493  
## Often 0.07719380 1.07352578 1.57650747  
## Rarely 0.88912988 0.07806383 0.43465703  
## GhanaData$Economic.conditions  
## GhanaData$Officials.unpunished Very Bad Very good  
## Always 2.07228601 0.30378316  
## Never 0.19534960 2.72760125  
## Often -1.41831815 -1.51000647  
## Rarely -0.86770976 -0.71986271

mySubset\_0\_1 <- subset(GhanaData, Economic.conditions == "Fairly Bad" | Economic.conditions == "Fairly Good")  
  
# pair-wise chi-square for explanatory categories 0 -1  
chisq\_0\_1 <- chisq.test(mySubset\_0\_1$Economic.conditions, mySubset\_0\_1$Economic.conditions)  
chisq\_0\_1

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: mySubset\_0\_1$Economic.conditions and mySubset\_0\_1$Economic.conditions  
## X-squared = 1303, df = 1, p-value < 2.2e-16

mySubset\_0\_2 <- subset(GhanaData, Economic.conditions == "Fairly Bad" | Economic.conditions == "Neither good nor bad")  
  
# pair-wise chi-square for explanatory categories 0 - 2  
chisq\_0\_2 <- chisq.test(mySubset\_0\_2$Economic.conditions, mySubset\_0\_2$Economic.conditions)  
chisq\_0\_2

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: mySubset\_0\_2$Economic.conditions and mySubset\_0\_2$Economic.conditions  
## X-squared = 843.05, df = 1, p-value < 2.2e-16

mySubset\_0\_3 <- subset(GhanaData, Economic.conditions == "Fairly Bad" | Economic.conditions == "Very Bad")  
  
# pair-wise chi-square for explanatory categories 0 - 3  
chisq\_0\_3 <- chisq.test(mySubset\_0\_3$Economic.conditions, mySubset\_0\_3$Economic.conditions)  
chisq\_0\_3

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: mySubset\_0\_3$Economic.conditions and mySubset\_0\_3$Economic.conditions  
## X-squared = 1259, df = 1, p-value < 2.2e-16

mySubset\_0\_4 <- subset(GhanaData, Economic.conditions == "Fairly Bad" | Economic.conditions == "Very good")  
  
# pair-wise chi-square for explanatory categories 0 -4  
chisq\_0\_4 <- chisq.test(mySubset\_0\_4$Economic.conditions, mySubset\_0\_4$Economic.conditions)  
chisq\_0\_4

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: mySubset\_0\_4$Economic.conditions and mySubset\_0\_4$Economic.conditions  
## X-squared = 751.65, df = 1, p-value < 2.2e-16

mySubset\_1\_2 <- subset(GhanaData, Economic.conditions == "Fairly Good" | Economic.conditions == "Neither good nor bad")  
  
# pair-wise chi-square for explanatory categories  
chisq\_1\_2 <- chisq.test(mySubset\_1\_2$Economic.conditions, mySubset\_1\_2$Economic.conditions)  
chisq\_1\_2

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: mySubset\_1\_2$Economic.conditions and mySubset\_1\_2$Economic.conditions  
## X-squared = 929.74, df = 1, p-value < 2.2e-16

mySubset\_1\_3 <- subset(GhanaData, Economic.conditions == "Fairly Good" | Economic.conditions == "Very Bad")  
  
# pair-wise chi-square for explanatory categories  
chisq\_1\_3 <- chisq.test(mySubset\_1\_3$Economic.conditions, mySubset\_1\_3$Economic.conditions)  
chisq\_1\_3

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: mySubset\_1\_3$Economic.conditions and mySubset\_1\_3$Economic.conditions  
## X-squared = 1346, df = 1, p-value < 2.2e-16

mySubset\_1\_4 <- subset(GhanaData, Economic.conditions == "Fairly Good" | Economic.conditions == "Very good")  
  
# pair-wise chi-square for explanatory categories  
chisq\_1\_4 <- chisq.test(mySubset\_1\_4$Economic.conditions, mySubset\_1\_4$Economic.conditions)  
chisq\_1\_4

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: mySubset\_1\_4$Economic.conditions and mySubset\_1\_4$Economic.conditions  
## X-squared = 838.09, df = 1, p-value < 2.2e-16

mySubset\_2\_3 <- subset(GhanaData, Economic.conditions == "Neither good nor bad" | Economic.conditions == "Very Bad")  
  
# pair-wise chi-square for explanatory categories  
chisq\_2\_3 <- chisq.test(mySubset\_2\_3$Economic.conditions, mySubset\_2\_3$Economic.conditions)  
chisq\_2\_3

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: mySubset\_2\_3$Economic.conditions and mySubset\_2\_3$Economic.conditions  
## X-squared = 885.9, df = 1, p-value < 2.2e-16

mySubset\_2\_4 <- subset(GhanaData, Economic.conditions == "Neither good nor bad" | Economic.conditions == "Very good")  
  
# pair-wise chi-square for explanatory categories  
chisq\_2\_4 <- chisq.test(mySubset\_2\_4$Economic.conditions, mySubset\_2\_4$Economic.conditions)  
chisq\_2\_4

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: mySubset\_2\_4$Economic.conditions and mySubset\_2\_4$Economic.conditions  
## X-squared = 381.78, df = 1, p-value < 2.2e-16

mySubset\_3\_4 <- subset(GhanaData, Economic.conditions == "Very Bad" | Economic.conditions == "Very good")  
  
# pair-wise chi-square for explanatory categories  
chisq\_3\_4 <- chisq.test(mySubset\_3\_4$Economic.conditions, mySubset\_3\_4$Economic.conditions)  
chisq\_3\_4

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: mySubset\_3\_4$Economic.conditions and mySubset\_3\_4$Economic.conditions  
## X-squared = 794.37, df = 1, p-value < 2.2e-16

#Relationship 2  
chisq\_results2 <- chisq.test(GhanaData$Overall.country\_direction, GhanaData$Tax.avoidance)  
chisq\_results2

##   
## Pearson's Chi-squared test  
##   
## data: GhanaData$Overall.country\_direction and GhanaData$Tax.avoidance  
## X-squared = 11.326, df = 3, p-value = 0.01009

chisq\_results2$observed

## GhanaData$Tax.avoidance  
## Not at all likely Not very likely  
## Going in the right direction 91 63  
## Going in the wrong direction 80 74  
## GhanaData$Tax.avoidance  
## Somewhat likely Very likely  
## Going in the right direction 166 797  
## Going in the wrong direction 210 705

chisq\_results2$expected

## GhanaData$Tax.avoidance  
## Not at all likely Not very likely  
## Going in the right direction 87.3774 70.00412  
## Going in the wrong direction 83.6226 66.99588  
## GhanaData$Tax.avoidance  
## Somewhat likely Very likely  
## Going in the right direction 192.1281 767.4904  
## Going in the wrong direction 183.8719 734.5096

chisq\_results2$residuals

## GhanaData$Tax.avoidance  
## Not at all likely Not very likely  
## Going in the right direction 0.3875436 -0.8371275  
## Going in the wrong direction -0.3961488 0.8557154  
## GhanaData$Tax.avoidance  
## Somewhat likely Very likely  
## Going in the right direction -1.8850037 1.0651897  
## Going in the wrong direction 1.9268590 -1.0888415

mySubset2\_0\_1 <- subset(GhanaData, Tax.avoidance == "Not at all likely" | Tax.avoidance == "Not very likely")  
  
# pair-wise chi-square for explanatory categories 0 -1  
chisq2\_0\_1 <- chisq.test(mySubset2\_0\_1$Tax.avoidance, mySubset2\_0\_1$Tax.avoidance)  
chisq2\_0\_1

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: mySubset2\_0\_1$Tax.avoidance and mySubset2\_0\_1$Tax.avoidance  
## X-squared = 327.96, df = 1, p-value < 2.2e-16

mySubset2\_0\_2 <- subset(GhanaData, Tax.avoidance == "Not at all likely" | Tax.avoidance == "Somewhat likely")  
  
# pair-wise chi-square for explanatory categories 0 -1  
chisq2\_0\_2 <- chisq.test(mySubset2\_0\_2$Tax.avoidance, mySubset2\_0\_2$Tax.avoidance)  
chisq2\_0\_2

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: mySubset2\_0\_2$Tax.avoidance and mySubset2\_0\_2$Tax.avoidance  
## X-squared = 571.42, df = 1, p-value < 2.2e-16

mySubset2\_0\_3 <- subset(GhanaData, Tax.avoidance == "Not at all likely" | Tax.avoidance == "Very likely")  
  
# pair-wise chi-square for explanatory categories 0 -1  
chisq2\_0\_3 <- chisq.test(mySubset2\_0\_3$Tax.avoidance, mySubset2\_0\_3$Tax.avoidance)  
chisq2\_0\_3

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: mySubset2\_0\_3$Tax.avoidance and mySubset2\_0\_3$Tax.avoidance  
## X-squared = 1782.2, df = 1, p-value < 2.2e-16

mySubset2\_1\_2 <- subset(GhanaData, Tax.avoidance == "Not very likely" | Tax.avoidance == "Somewhat likely")  
  
# pair-wise chi-square for explanatory categories 0 -1  
chisq2\_1\_2 <- chisq.test(mySubset2\_1\_2$Tax.avoidance, mySubset2\_1\_2$Tax.avoidance)  
chisq2\_1\_2

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: mySubset2\_1\_2$Tax.avoidance and mySubset2\_1\_2$Tax.avoidance  
## X-squared = 532.98, df = 1, p-value < 2.2e-16

mySubset2\_1\_3 <- subset(GhanaData, Tax.avoidance == "Not very likely" | Tax.avoidance == "Very likely")  
  
# pair-wise chi-square for explanatory categories 0 -1  
chisq2\_1\_3 <- chisq.test(mySubset2\_1\_3$Tax.avoidance, mySubset2\_1\_3$Tax.avoidance)  
chisq2\_1\_3

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: mySubset2\_1\_3$Tax.avoidance and mySubset2\_1\_3$Tax.avoidance  
## X-squared = 1742, df = 1, p-value < 2.2e-16

mySubset2\_2\_3 <- subset(GhanaData, Tax.avoidance == "Somewhat likely" | Tax.avoidance == "Very likely")  
  
# pair-wise chi-square for explanatory categories 0 -1  
chisq2\_2\_3 <- chisq.test(mySubset2\_2\_3$Tax.avoidance, mySubset2\_2\_3$Tax.avoidance)  
chisq2\_2\_3

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: mySubset2\_2\_3$Tax.avoidance and mySubset2\_2\_3$Tax.avoidance  
## X-squared = 1992.6, df = 1, p-value < 2.2e-16

1. “Are the country’s current economic conditions associated with how often officials go unpunished?”

### Stating hypothesis

* -H0: There is no linear relationship between present day economic conditions and and how frequently corrupt officials elude persecution.
* -Ha: There is a linear relationship between present day economic conditions and and how frequently corrupt officials elude persecution.

### Collecting and summarizing sample (Explanatory proportions)

* -Fairly Bad : 7.7%, 3.8%, 10.2%, 4.5% (sample size: 590)
* -Farily Good: 8.6%, 4.2%, 12.2%, 4.7% (sample size: 665)
* -Neither good nor bad: 2.7%, 1.2%, 4.6%, 1.7% (sample size: 226)
* -Very Bad: 9.7%, 4.2%, 9.6%, 3.9% (sample size: 616 )
* -Very good: 2.1%, 1.5%, 2.0%, 0.8% (sample size: 143)

### Conditions for safe use of chi-square test

* + The sample was random
  + All expected counts were greater than 1 and more than 80%(almost all) of the expected counts were greater than 5

### Access the evidence ( Test statistic and p-value report)

* X-squared = 26.725  
    
  df = 12  
    
  p-value = 0.008463

### Making conclusions

* From the reports of the test statitic, we can conclude by applying the decision rule based of 𝛼= 0.05 ( i.e level of significance of the test)  
    
  The p-value obtained (0.008463) is less than 𝛼 = 0.05 hence we reject the null hypothesis H0: (There is no linear relationship between present day economic conditions and and how frequently corrupt officials elude persecution.). (𝑝―𝑣𝑎𝑙𝑢𝑒< 𝛼, reject 𝐻0), finally arriving at the conclusion there is a relationship between economic conditions and how often corrupt officials elude punishment.  
    
  Since there is a rejection of the null hypothesis, there is the possiblity for a Type 1 error.

1. “Is people’s refusal to pay taxes associated with the overall direction of the country?”

* -H0: There is no linear relationship between the payment of taxes and the overall direction of the state.
* -Ha: There is a linear relationship between the payment of taxes and the overall direction of the state.

### Explanatory proportion

* -Not at all likely: 4.2%, 3.7% (sample size: 171)
* -Not very likely: 2.8%, 3.4% (sample size: 137)
* -Somewhat likely: 7.6%, 9.6% (sample size: 376)
* -Very likely: 36.5%, 32.3% (sample size: 1502)

### Conditions for safe use of chi-square test

* + The sample was random
  + All expected counts were greater than 1 and more than 80%(almost all) of the expected counts were greater than 5

### Access the evidence ( Test statistic and p-value report)

* X-squared = 11.326
* df = 3
* p-value = 0.01009

### Making conclusions

* From the reports of the test statitic, we can conclude by applying the decision rule based of 𝛼= 0.05 ( i.e level of significance of the test)  
    
  The p-value obtained (0.01009) is less than 𝛼 = 0.05 hence we reject the null hypothesis H0: (There is no linear relationship between the payment of taxes and the overall direction of the state.). (𝑝―𝑣𝑎𝑙𝑢𝑒< 𝛼, reject 𝐻0), finally arriving at the conclusion there is a linear relationship between the payment of taxes and the overall direction of the state.  
    
  Since there is a rejection of the null hypothesis, there is the possiblity for a Type 1 error.

### Summary of Overall results

My analysis exercise tackled two C -> C relationships which I performed hypothesis tests on using the chiq-square test. For both relationships, the end results provided by the test statistic and p-value suggested a rejection of the null hypothesis since the p-values of both were well below that of the level of significance ( 𝛼 = 0.05 ). Hence, in both the alternate hypothesis Ha was the final conclusion on for both relationships ( i.e a linear relationship between economic conditions and how often corrupt officials elude punishment and a linear relationship between the payment of taxes and the overall direction of the state). However in both since there is a rejection of the null hypothesis there is the possiblity of Type 1 errors. Furthermore, both relationships had explanatory variables with more than 2 levels hence there was a need for a post hoc test to avoid type 1 error. Results from the post hoc test. To see if there was a significant difference between the groups, I utilized a post-hoc test. All of the matching categories had an exact p-value of less than 0.005 when 𝛼 = 0.05 was used. As a result, the p-values of the pairings did not differ significantly. “There is a linear relationship between the payment of taxes and the overall direction of the state,” and “There is a relationship between economic conditions and how often corrupt officials elude punishment.” I can conclude both statements as a consequence of my original research.

There was adequate evidence to reject the null hypothesis for my first research question because the p-value was less than 0.005 using 𝛼 = 0.05. I suspected that I had made a type 1 error, so I paired the levels and ran a post-hoc test because there was more than one level in my variable. Although there was a difference in the p-value after the test, it was not significant since the p-value was still less than alpha; thus, 0.005. As a result, I must continue to reject the null hypothesis. However, based on my initial study, I can infer that “there is a linear relationship between the extent of democracy and the current economic conditions”.

Because my P-value was less than 0.005 utilizing the 𝛼 = 0.05 method, I had to reject the null hypothesis for my second research question. I used a post-hoc test to avoid committing a type 1 error, which would mean rejecting the null hypothesis when it could be true. I used a post-hoc test to examine if there was any significant difference between the categories. Using the alpha = 0.05, all of the matched categories had the exact p-value, which was less than 0.005. The p-values of the pairings did not differ significantly as a result. As a result of my original research, I can draw the following conclusion: “There is a linear relationship between current economic conditions and current living conditions”.