Sentiment Analysis

Analysis performed using Italian_LIWC2007 Dictionary

Riccardo Ruta

5/2022

Contents


```
# Data
load("data/dfm.Rda")

# Dictionary LWIC Complete
LWIC_ITA <- dictionary(file = "data/large_files/Italian_LIWC2007_Dictionary.dic",
                       format = "LIWC")
```

```
## note: removing empty key: Formale

## note: removing empty key: Passivo
```

```
emotions <- c("Emo_Pos", "Emo_Neg", "Ansia", "Rabbia", "Tristez", "Ottimis" )

# Inspect the words
n.words <- c(
length(LWIC_ITA[["Emo_Pos"]]),
length(LWIC_ITA[["Emo_Neg"]]),
length(LWIC_ITA[["Ansia"]]),
length(LWIC_ITA[["Rabbia"]]),
length(LWIC_ITA[["Tristez"]]),
length(LWIC_ITA[["Ottimis"]])
)

num_words <-  data.frame(emotions,n.words)

# Extracting only the keys we need
myLWIC_ITA <- dictionary(list(positive = LWIC_ITA[["Emo_Pos"]],
                              negative = LWIC_ITA[["Emo_Neg"]],
                              anxiety = LWIC_ITA[["Ansia"]],
                              anger = LWIC_ITA[["Rabbia"]],
                              sadness = LWIC_ITA[["Tristez"]],
```

```
                                optimism = LWIC_ITA[["Ottimis"]]))

myLWIC_ITA_sent <- dictionary(list(positive = LWIC_ITA[["Emo_Pos"]],
                                negative = LWIC_ITA[["Emo_Neg"]]))
```

```
kable(num_words)
```

emotions	n.words
Emo_Pos	200
Emo_Neg	663
Ansia	65
Rabbia	227
Tristez	226
Ottimis	93

Group and weight the dfm

```
# By party & quarter
dfm_weigh_p_quart <- dfm_group(DFM, groups = interaction(party_id, quarter))%>%
  dfm_weight(scheme = "prop")
```

Apply the dictionary

```
# Apply Dictionary to DFM
DFM_emotions <- dfm_lookup(dfm_weigh_p_quart,
                           dictionary = myLWIC_ITA)
DFM_emotions
```

```
## Document-feature matrix of: 110 documents, 6 features (0.76% sparse) and 3 docvars.
##                  features
## docs                 positive   negative     anxiety       anger     sadness
##   CI.1            0.008060854 0.02236603 0.003405995 0.006471390 0.004541326
##   FDI.1           0.006416312 0.02893245 0.002834199 0.011061250 0.006140765
##   FI.1            0.006498830 0.02547256 0.003243474 0.007675035 0.006974064
##   INDIPENDENTE.1 0.005129667 0.01567398 0.001994870 0.005984611 0.003989741
##   IV.1            0.008545455 0.02309091 0.003272727 0.009272727 0.006000000
##   LEGA.1          0.006352373 0.02593448 0.003005565 0.008426081 0.006194876
##                  features
## docs                optimism
##   CI.1            0.01089918
##   FDI.1           0.01487955
##   FI.1            0.01447089
##   INDIPENDENTE.1 0.01025933
##   IV.1            0.01600000
##   LEGA.1          0.01257350
## [ reached max_ndoc ... 104 more documents ]
```

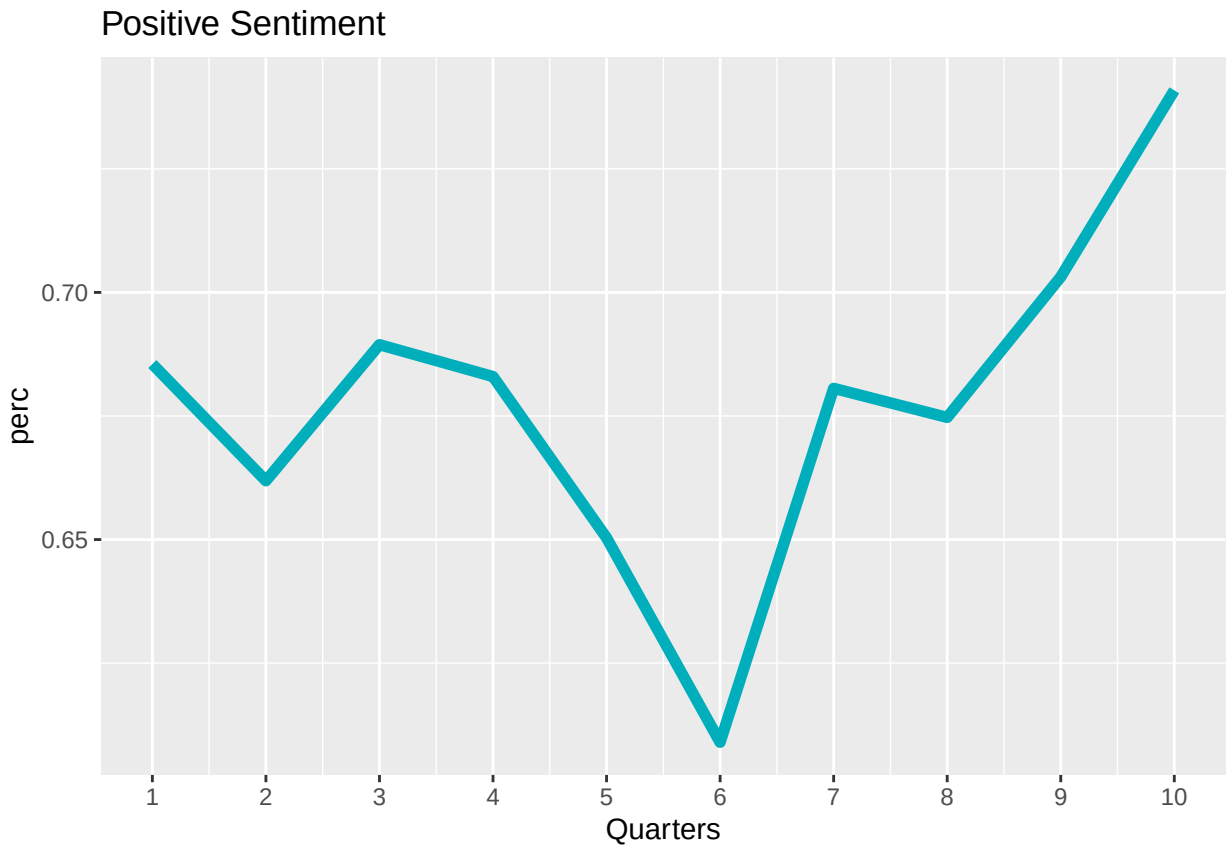
Transform the DFM into an ordinary dataframe

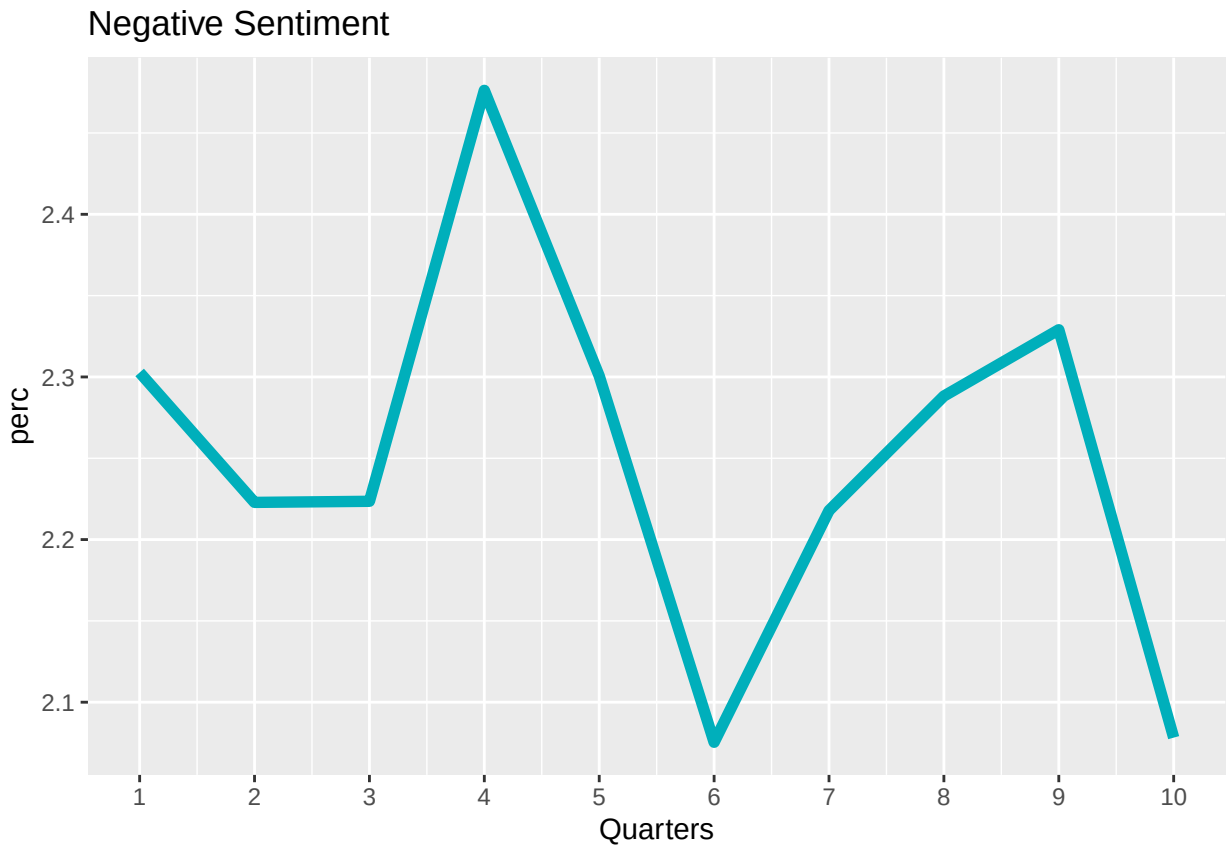
```
data_dict_emo <- DFM_emotions %>%
  quanteda::convert(to = "data.frame") %>%
  cbind(docvars(DFM_emotions))

# Transform the proportion into percentage
data_dict_emo <- data_dict_emo %>% mutate(positive =  positive * 100,
                                          negative = negative * 100,
                                          anxiety = anxiety * 100,
                                          anger = anger *100,
                                          sadness = sadness * 100,
                                          optimism = optimism *100)
```

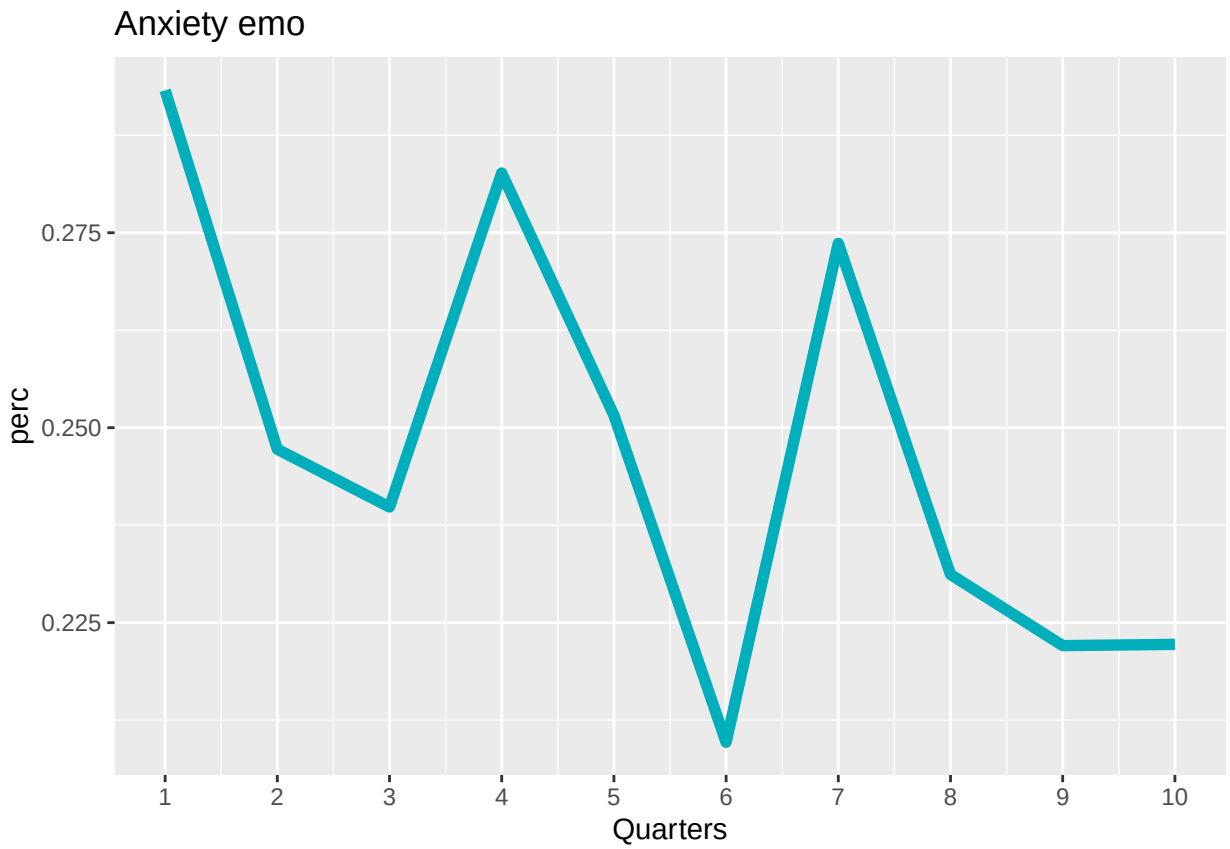
Percentage of the emotions in time

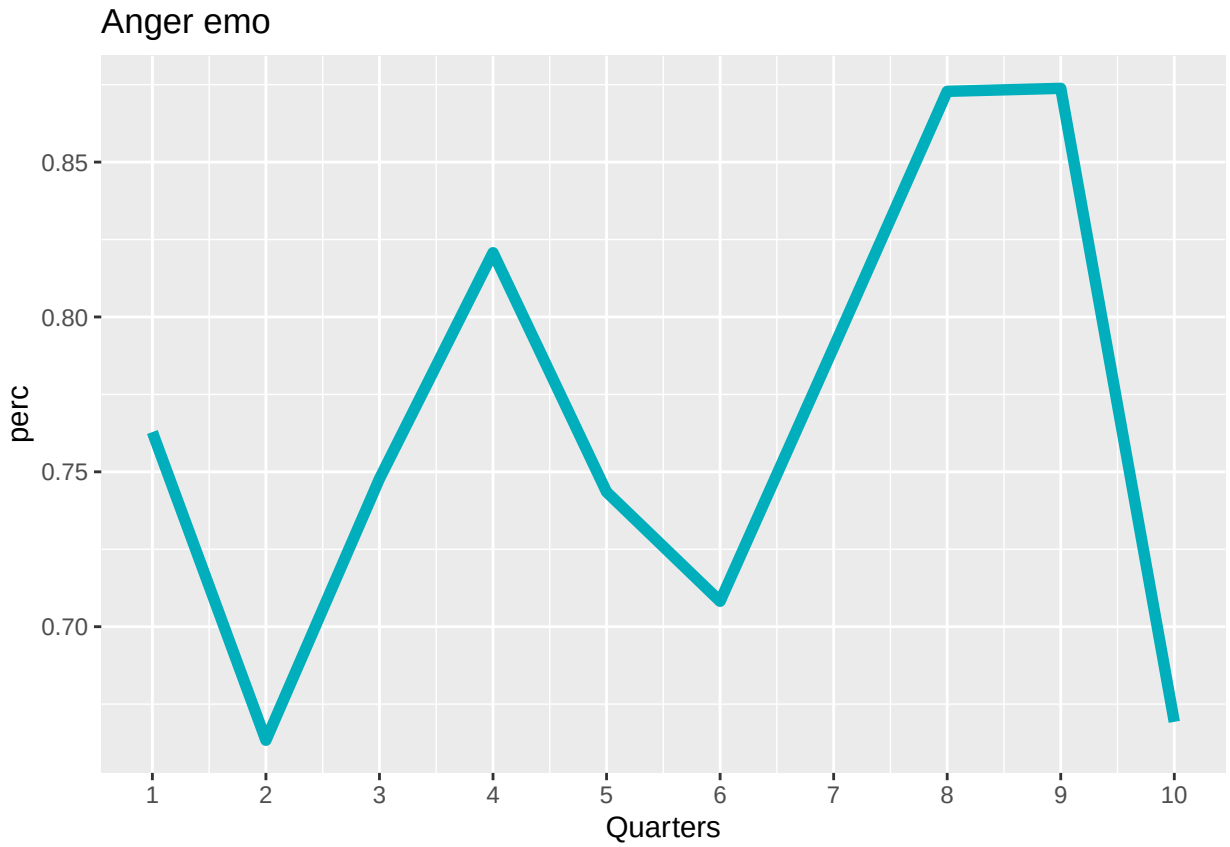
The code is only shown for 'positive' but is identical for all emotions

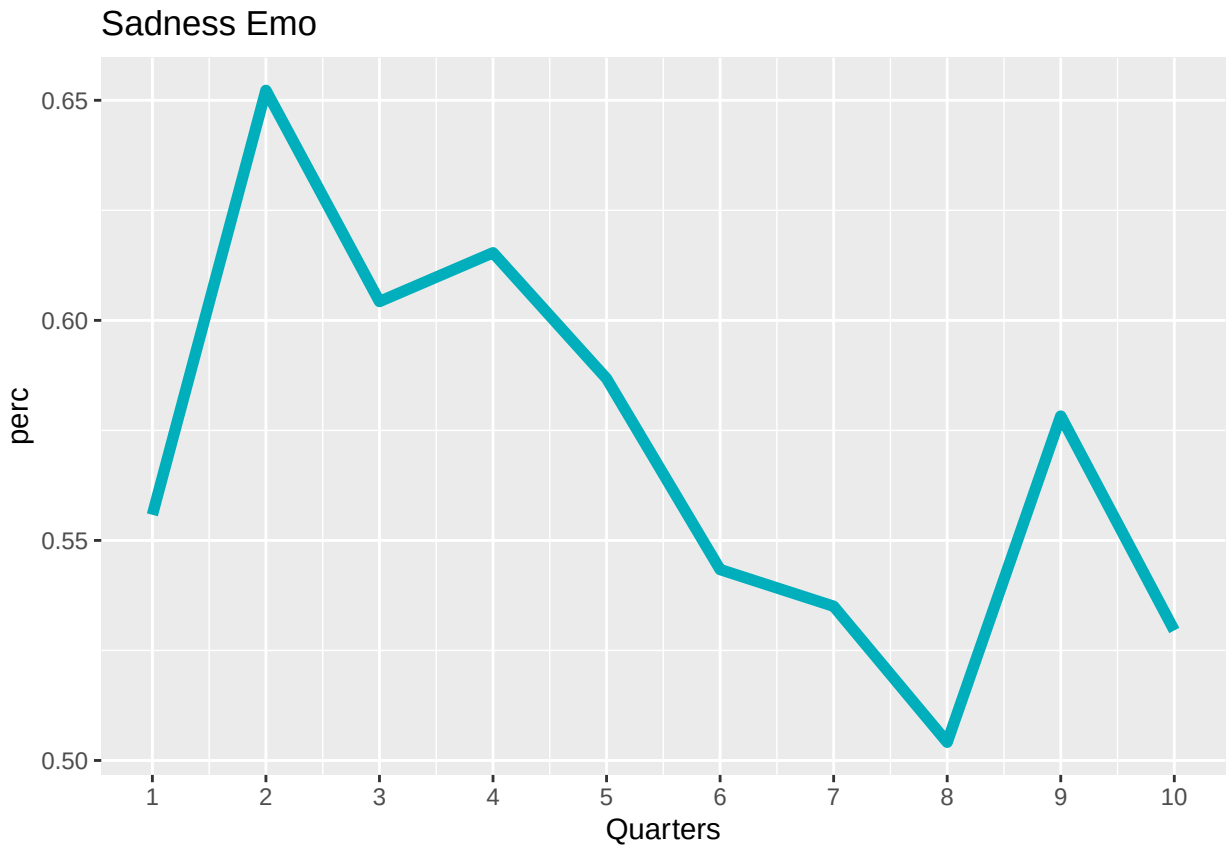
```
#Over time POSITIVE (quarters)
data_quarter_positive <- aggregate(x = data_dict_emo$positive,  # Specify data column
          by = list(data_dict_emo$quarter),  # Specify group indicator
          FUN = mean) # Specify function (i.e. mean)
data_quarter_positive$perc <- data_quarter_positive$x

# plot
plot_positive <- ggplot(data = data_quarter_positive, aes(x = Group.1, y = perc))+
  geom_line(color = "#00AFBB", size = 2)+
  scale_x_continuous("Quarters", labels = as.character(data_quarter_positive$Group.1), breaks = data_qua
  labs(title = "Positive Sentiment")
plot_positive
```

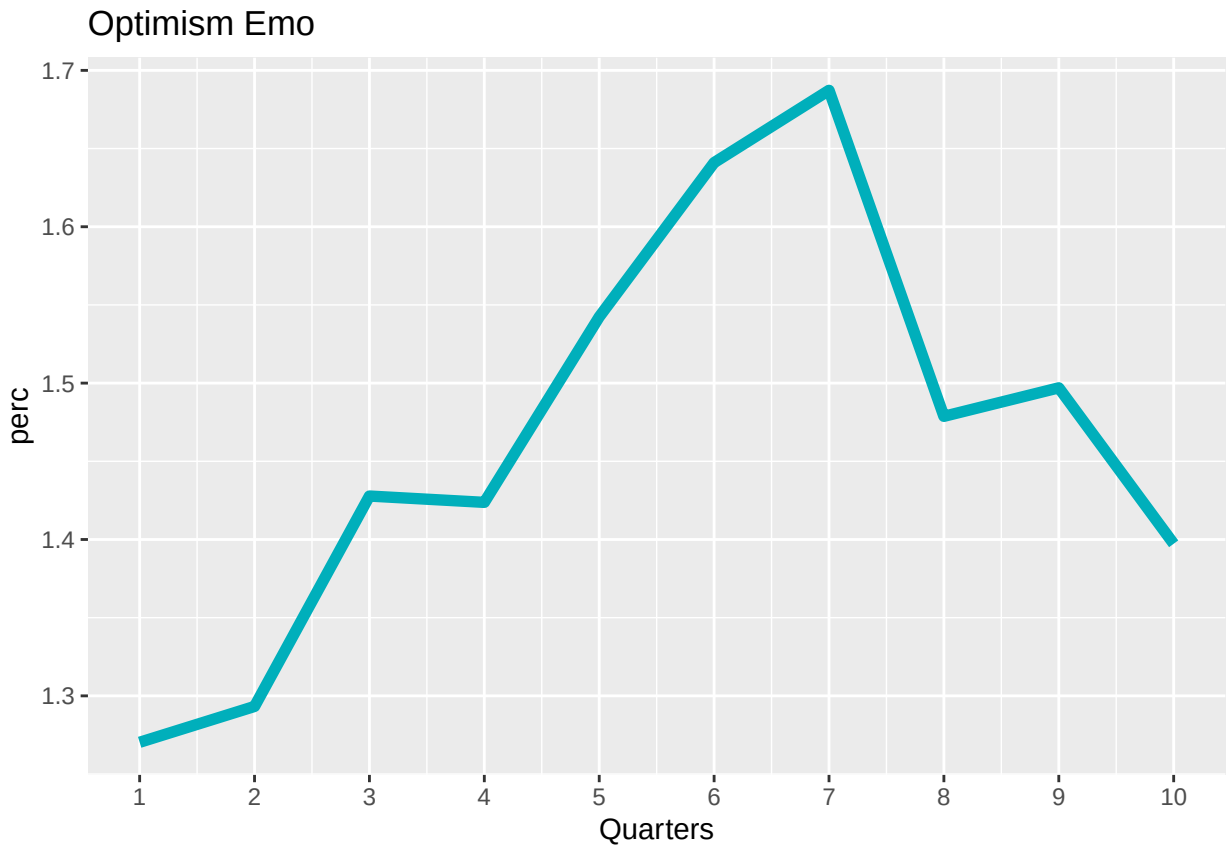

Positive Sentiment
perc
0.70
0.65
1
2
3
4
5
6
7
8
9
10
Quarters

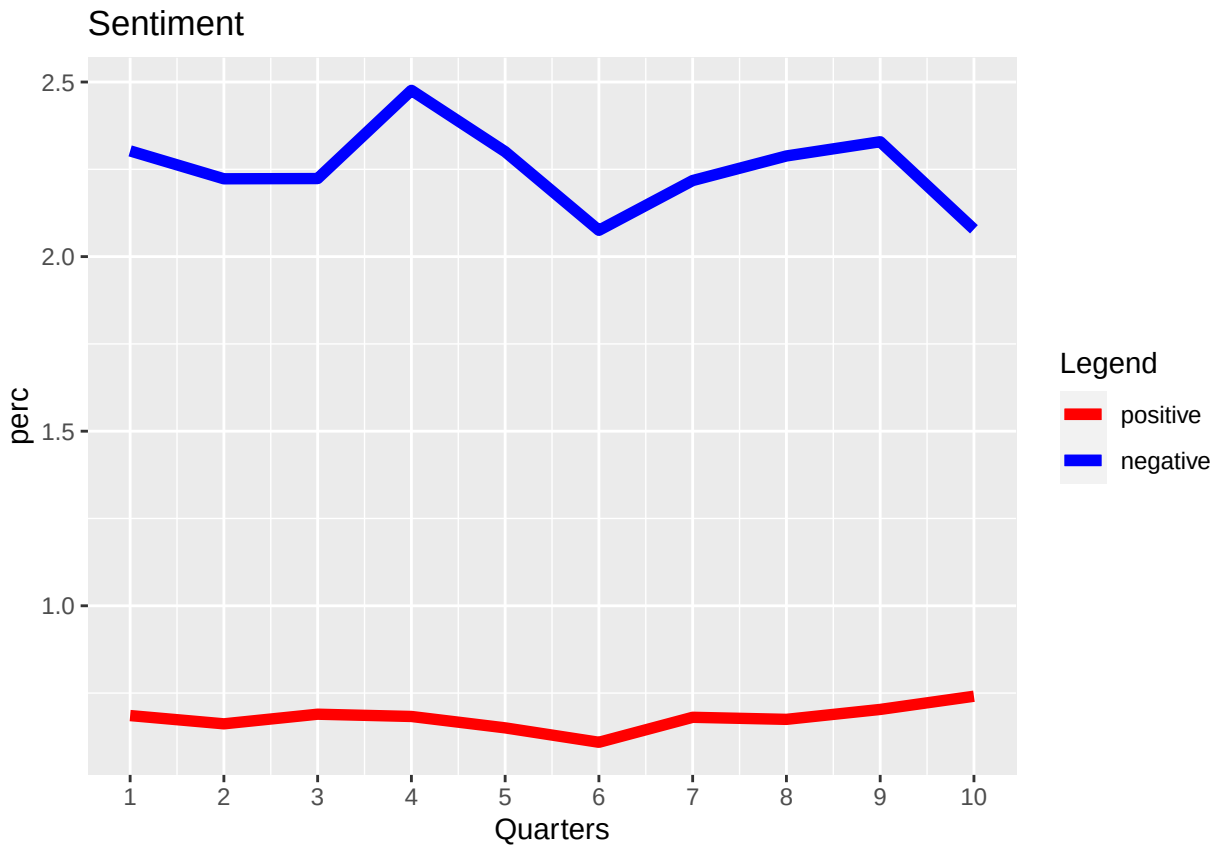

Negative Sentiment
perc
2.4
2.3
2.2
2.1
1
2
3
4
5
6
7
8
9
10
Quarters

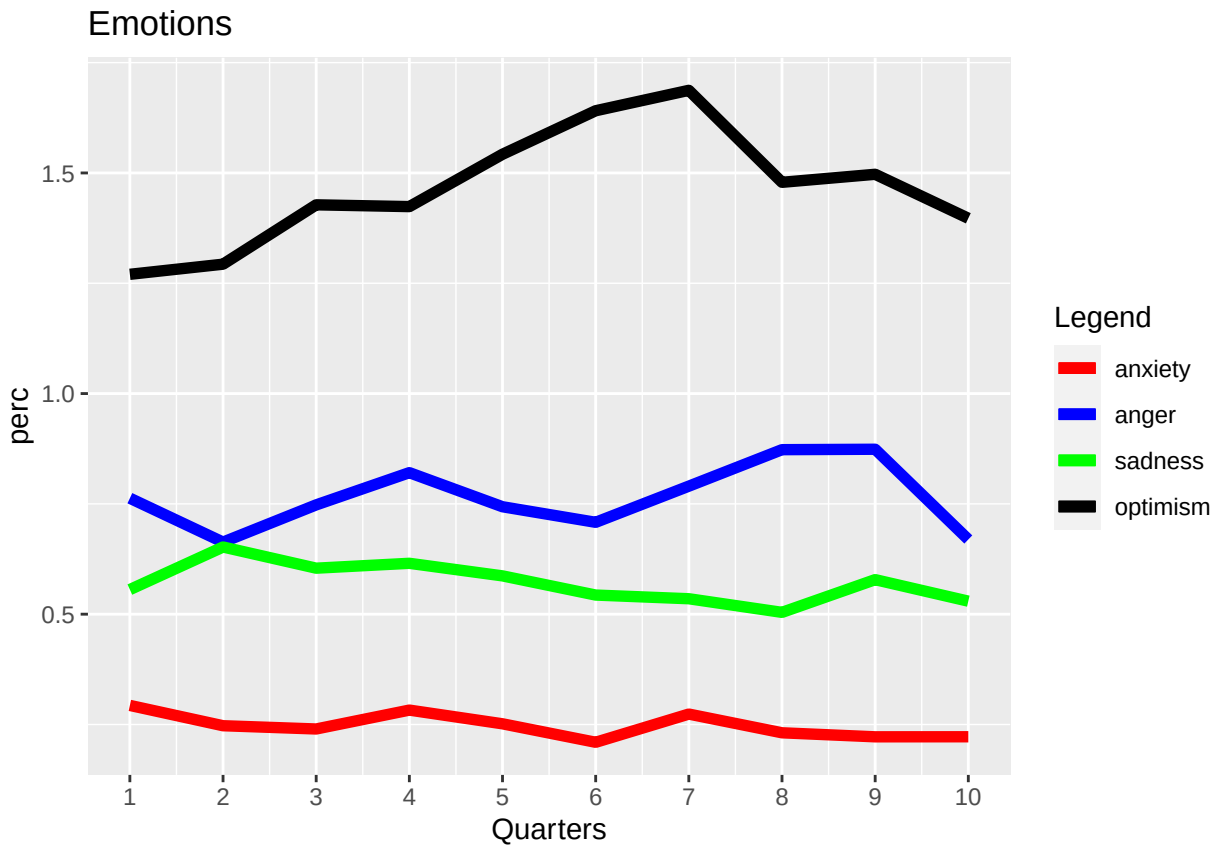

Anxiety emo
perc
0.275
0.250
0.225
1
2
3
4
5
6
7
8
9
10
Quarters


Anger emo
perc
0.85
0.80
0.75
0.70
1
2
3
4
5
6
7
8
9
10
Quarters

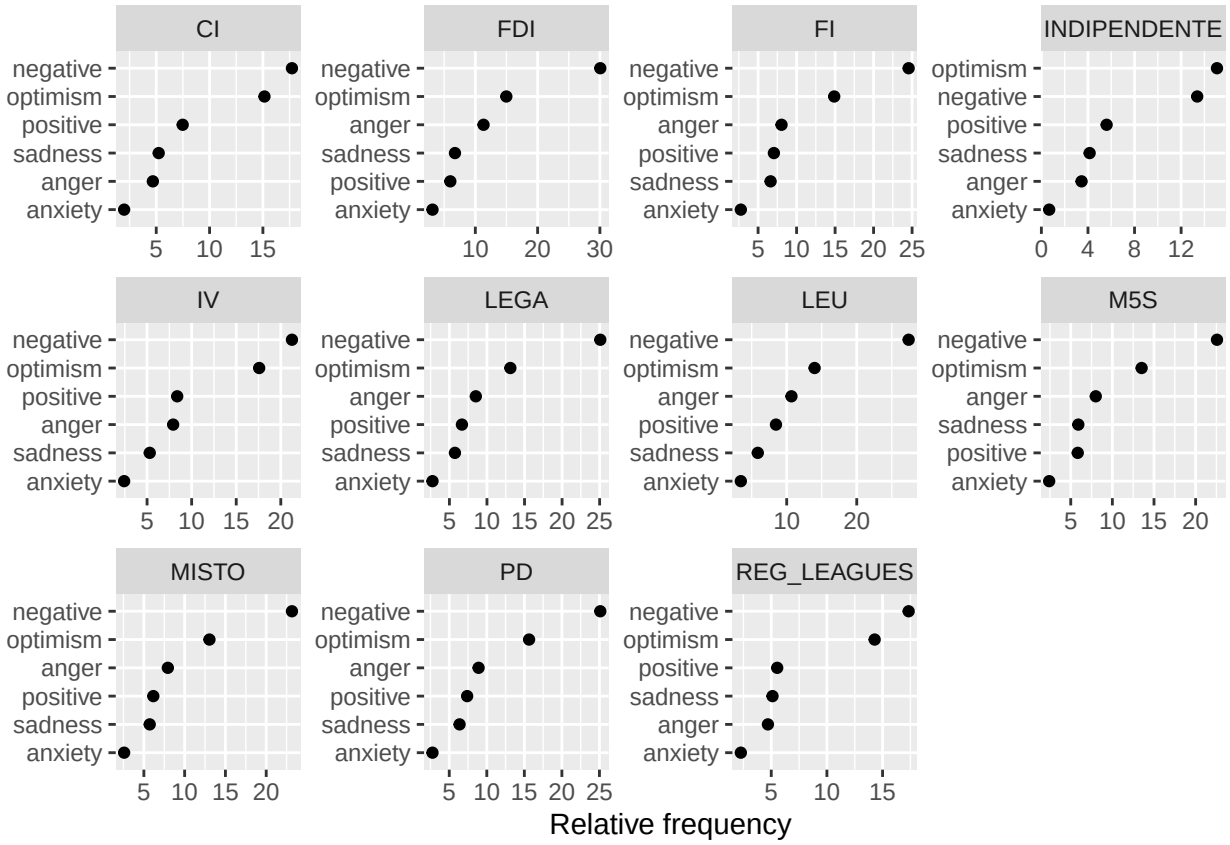

Sadness Emo
0.65
0.60
0.55
0.50
perc
1
2
3
4
5
6
7
8
9
10
Quarters


Optimism Emo
1.7
1.6
1.5
1.4
1.3
perc
1
2
3
4
5
6
7
8
9
10
Quarters


Sentiment
2.5
2.0
1.5
1.0
perc
1
2
3
4
5
6
7
8
9
10
Quarters
Legend
positive
negative


Emotions
perc
1.5
1.0
0.5
1
2
3
4
5
6
7
8
9
10
Quarters
Legend
anxiety
anger
sadness
optimism

Main emotion for each parliamentary group



The code is only shown for 'positive' but is identical for all emotions

```
# POSITIVE
data_party_positive <- aggregate(x = data_dict_emo$positive,  # Specify data column
          by = list(data_dict_emo$party_id),  # Specify group indicator
          FUN = mean) # Specify function (i.e. mean)
data_party_positive$perc <- round(data_party_positive$x,3)
kable(data_party_positive %>% select(Group.1, perc) %>%  arrange(desc(perc)), caption = "POSITIVE")
```

```
ggplot(data=data_party_positive, aes(x=Group.1, y=perc)) +
  geom_bar(stat="identity", fill="steelblue")+
  geom_text(aes(label=perc), vjust=0, color="black", size=3.5)+
  theme_minimal()+
  labs(title = "Positive Sentiment")+
  coord_flip()
```

Table 1: POSITIVE

Group.1	perc
LEU	0.847
IV	0.838
CI	0.748
PD	0.738
FI	0.706
LEGA	0.667
MISTO	0.616
FDI	0.598
M5S	0.584
INDIPENDENTE	0.560
REG_LEAGUES	0.554

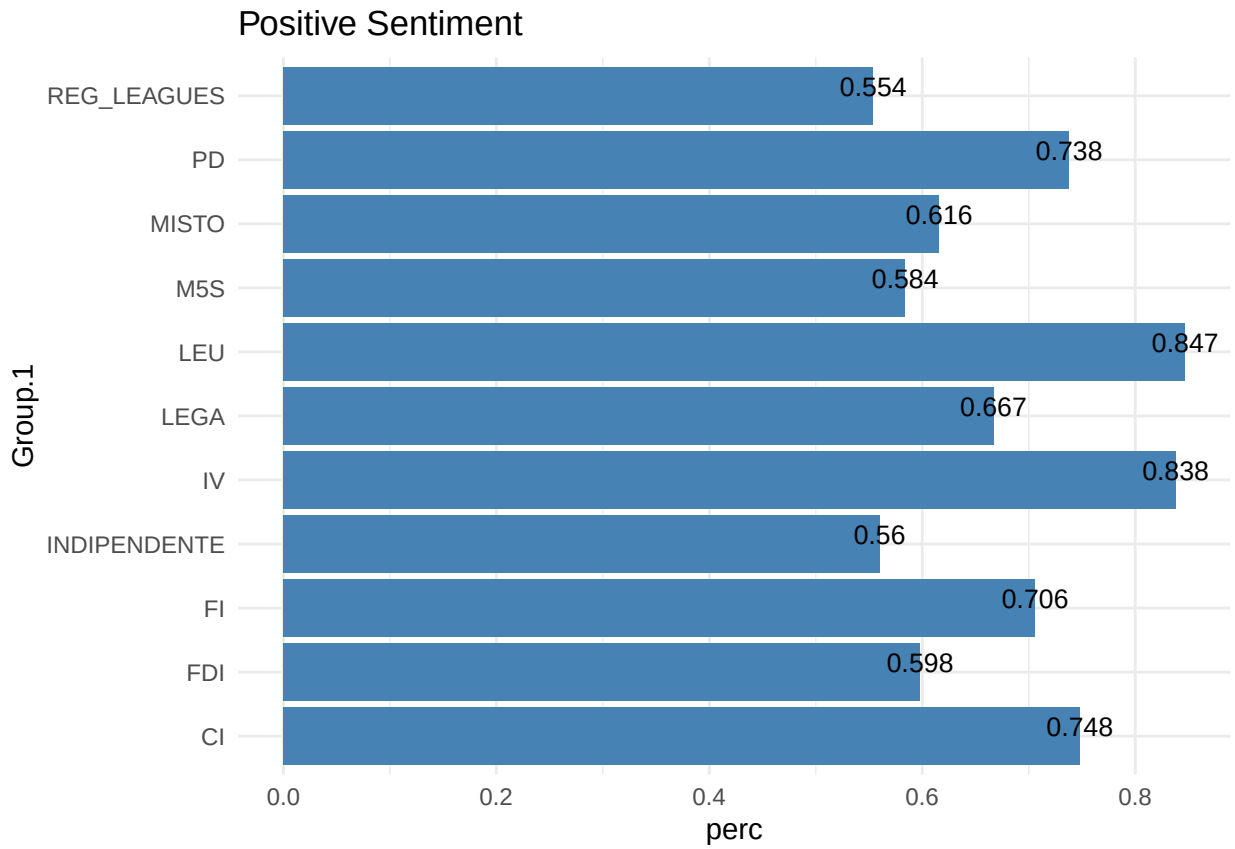


Table 2: NEGATIVE

Group.1	perc
FDI	3.006
LEU	2.741
PD	2.512
LEGA	2.509
FI	2.455
MISTO	2.316
M5S	2.257
IV	2.125
CI	1.772
REG_LEAGUES	1.734
INDIPENDENTE	1.338

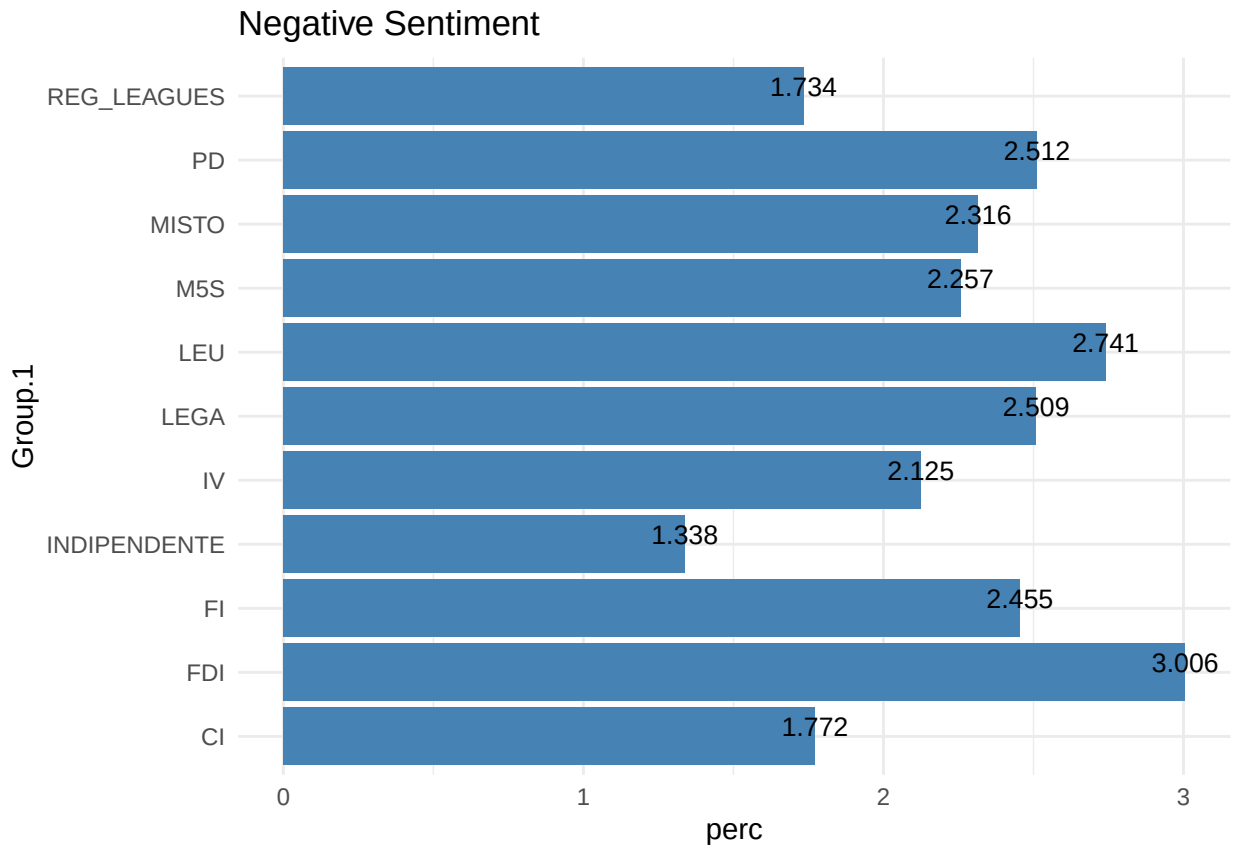


Table 3: ANXIETY

Group.1	perc
LEU	0.345
FDI	0.312
PD	0.277
FI	0.276
LEGA	0.275
MISTO	0.258
IV	0.243
M5S	0.241
REG_LEAGUES	0.227
CI	0.199
INDIPENDENTE	0.067

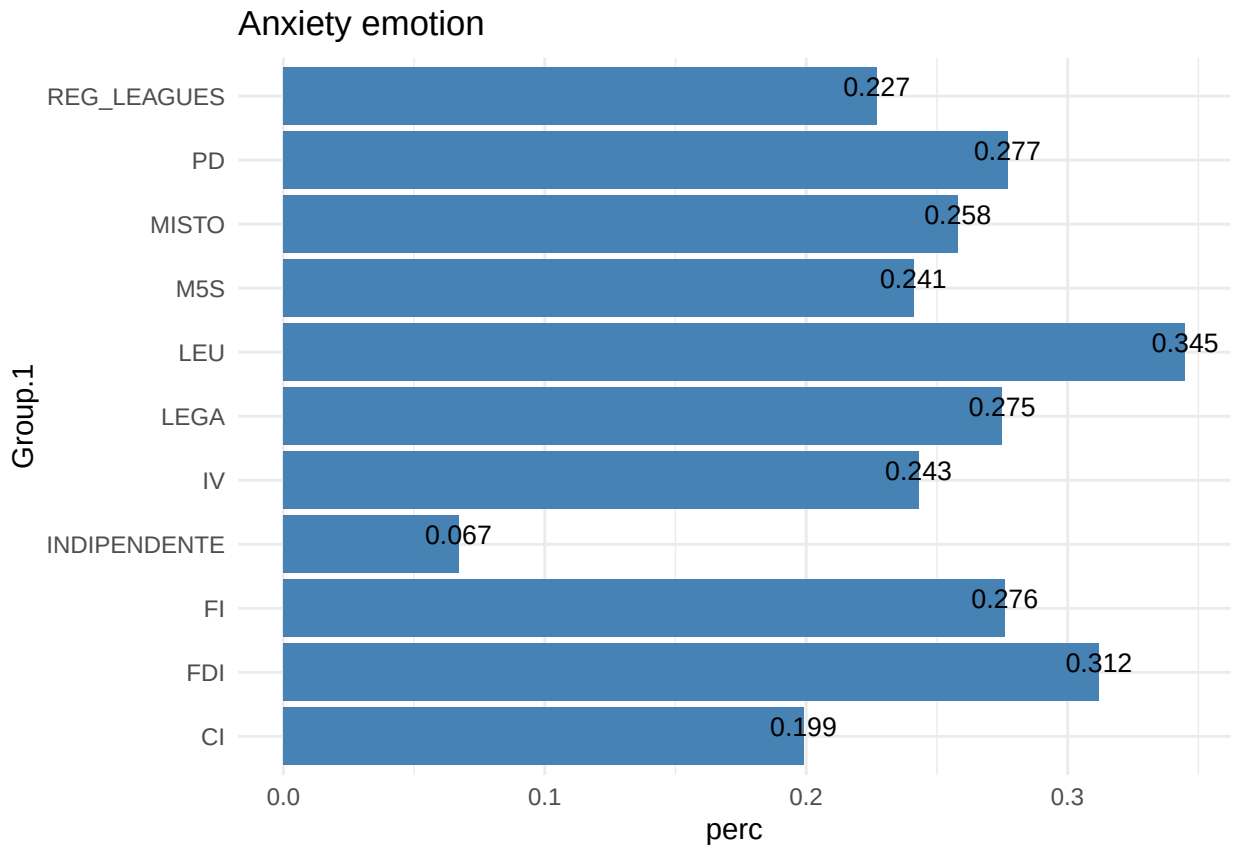


Table 4: ANGER

Group.1	perc
FDI	1.132
LEU	1.068
PD	0.891
LEGA	0.852
FI	0.805
M5S	0.801
MISTO	0.794
IV	0.793
REG_LEAGUES	0.470
CI	0.468
INDIPENDENTE	0.345

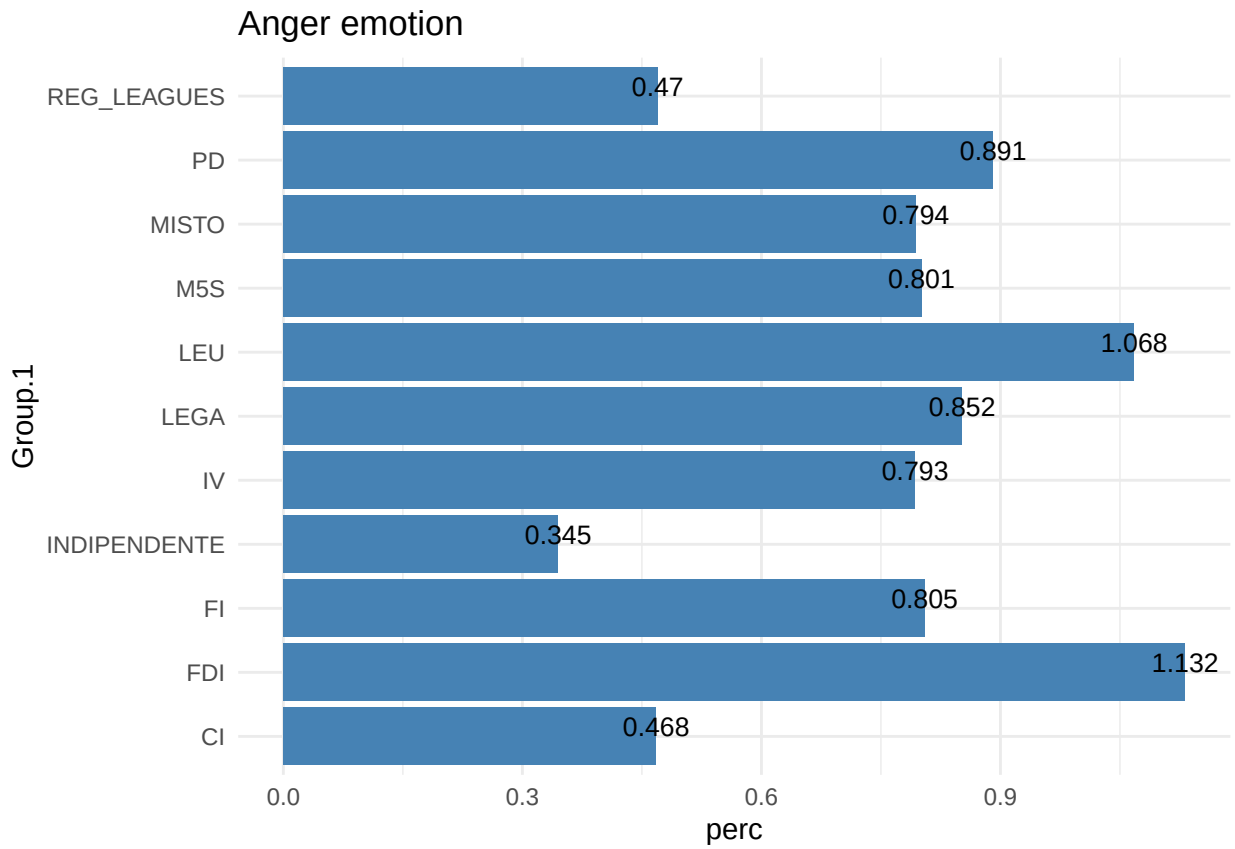


Table 5: SADNESS

Group.1	perc
FDI	0.673
FI	0.663
PD	0.638
M5S	0.591
LEU	0.587
LEGA	0.573
MISTO	0.572
IV	0.530
CI	0.523
REG_LEAGUES	0.511
INDIPENDENTE	0.414

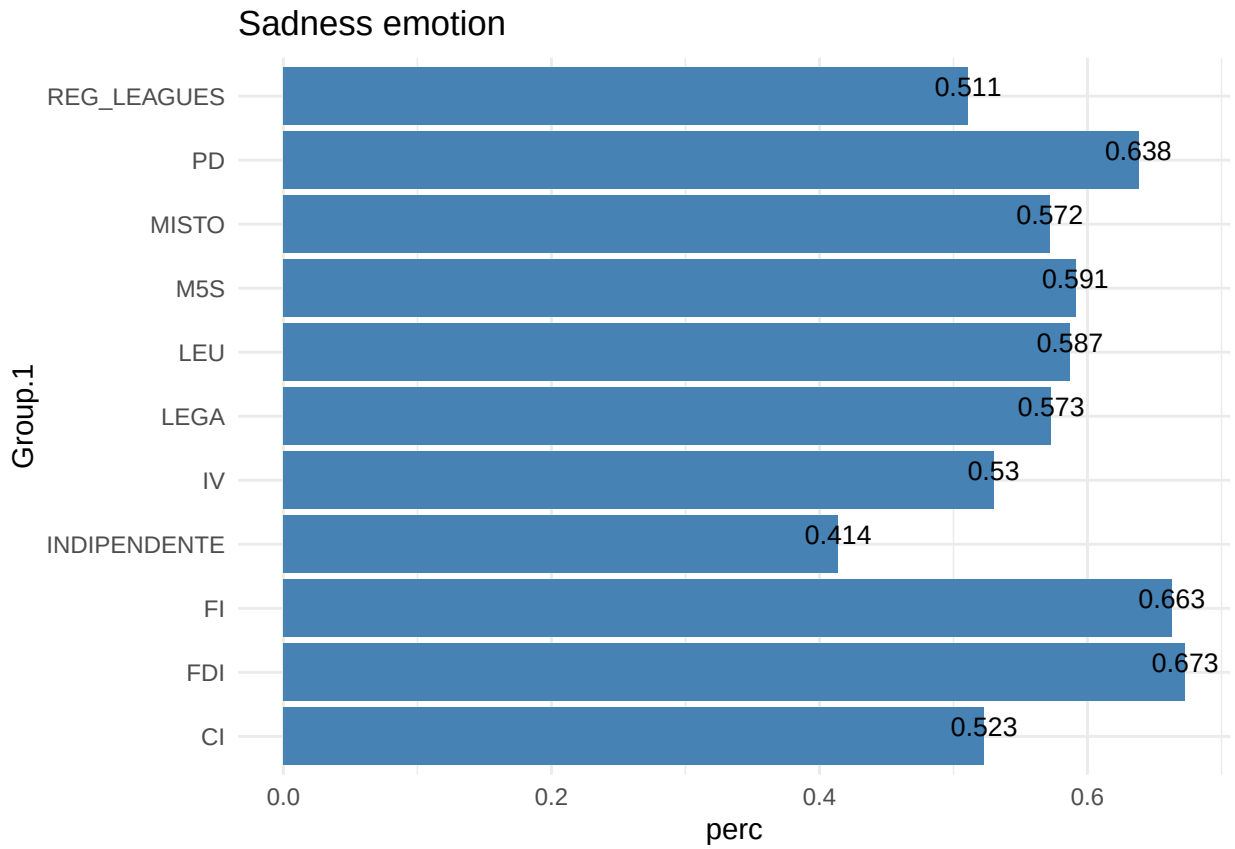
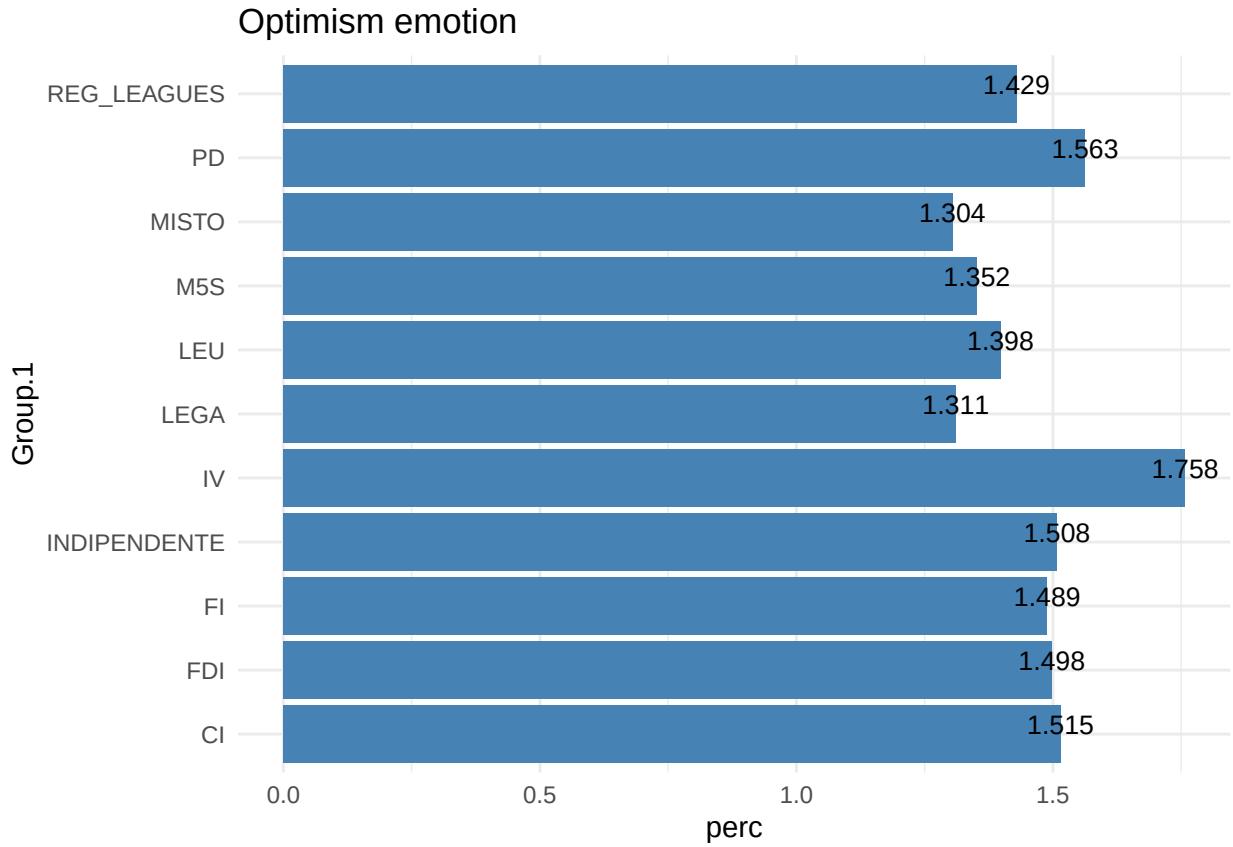


Table 6: OPTIMISM

Group.1	perc
IV	1.758
PD	1.563
CI	1.515
INDIPENDENTE	1.508
FDI	1.498
FI	1.489
REG_LEAGUES	1.429
LEU	1.398
M5S	1.352
LEGA	1.311
MISTO	1.304



Are the average values of [...] for each party statistically different from each other? The reference category is PD

```
# POSITIVE
summary(data_dict_emo$positive)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3281  0.5863  0.6542  0.6778  0.7546  1.1593
```

```r
# bivariate regression for check t-test
data_dict_emo$factor_party <- as.factor(data_dict_emo$party_id)
data_dict_emo$factor_party <- relevel(data_dict_emo$factor_party, ref = "PD")

data_dict_emo$factor_quarter <- as.factor(data_dict_emo$quarter)
data_dict_emo$factor_quarter <- relevel(data_dict_emo$factor_quarter, ref = "5")

positive_model <- lm(positive ~  factor_quarter + factor_party, data_dict_emo )

summary(positive_model)
```

```
##
## Call:
## lm(formula = positive ~ factor_quarter + factor_party, data = data_dict_emo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.26194 -0.06684  0.00093  0.04680  0.33861
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                0.710990   0.052210  13.618  < 2e-16 ***
## factor_quarter1            0.035165   0.052210   0.674  0.50234
## factor_quarter2            0.011541   0.052210   0.221  0.82556
## factor_quarter3            0.039079   0.052210   0.748  0.45611
## factor_quarter4            0.032630   0.052210   0.625  0.53358
## factor_quarter6           -0.041367   0.052210  -0.792  0.43026
## factor_quarter7            0.030252   0.052210   0.579  0.56376
## factor_quarter8            0.024362   0.052210   0.467  0.64191
## factor_quarter9            0.052797   0.052210   1.011  0.31462
## factor_quarter10           0.090541   0.052210   1.734  0.08632 .
## factor_partyCI             0.009462   0.054759   0.173  0.86321
## factor_partyFDI           -0.140003   0.054759  -2.557  0.01224 *
## factor_partyFI            -0.032835   0.054759  -0.600  0.55026
## factor_partyINDIPENDENTE -0.178239   0.054759  -3.255  0.00160 **
## factor_partyIV             0.099436   0.054759   1.816  0.07272 .
## factor_partyLEGA          -0.071907   0.054759  -1.313  0.19247
## factor_partyLEU            0.108649   0.054759   1.984  0.05029 .
## factor_partyM5S           -0.154273   0.054759  -2.817  0.00595 **
## factor_partyMISTO         -0.122489   0.054759  -2.237  0.02776 *
## factor_partyREG_LEAGUES   -0.184902   0.054759  -3.377  0.00109 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1224 on 90 degrees of freedom
## Multiple R-squared:  0.4781, Adjusted R-squared:  0.3679
## F-statistic: 4.339 on 19 and 90 DF,  p-value: 1.009e-06
```

```r
#NEGATIVE
summary(data_dict_emo$negative)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.9522  1.9364  2.3318  2.2515  2.5867  3.2025
```

```r
# bivariate regression for check t-test
data_dict_emo$factor_party <- as.factor(data_dict_emo$party_id)
data_dict_emo$factor_party <- relevel(data_dict_emo$factor_party, ref = "PD")

data_dict_emo$factor_quarter <- as.factor(data_dict_emo$quarter)
data_dict_emo$factor_quarter <- relevel(data_dict_emo$factor_quarter, ref = "5")

negative_model <- lm(negative ~  factor_quarter + factor_party, data_dict_emo )

summary(negative_model)
```

```
##
## Call:
## lm(formula = negative ~ factor_quarter + factor_party, data = data_dict_emo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.79357 -0.14849  0.00431  0.15790  0.46872
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                2.560662   0.108714  23.554  < 2e-16 ***
## factor_quarter1            0.002167   0.108714   0.020  0.98414
## factor_quarter2           -0.077716   0.108714  -0.715  0.47654
## factor_quarter3           -0.077039   0.108714  -0.709  0.48038
## factor_quarter4            0.175647   0.108714   1.616  0.10966
## factor_quarter6           -0.225225   0.108714  -2.072  0.04115 *
## factor_quarter7           -0.082757   0.108714  -0.761  0.44851
## factor_quarter8           -0.012345   0.108714  -0.114  0.90984
## factor_quarter9            0.028457   0.108714   0.262  0.79410
## factor_quarter10          -0.222362   0.108714  -2.045  0.04374 *
## factor_partyCI            -0.739253   0.114020  -6.484 4.70e-09 ***
## factor_partyFDI            0.494954   0.114020   4.341 3.71e-05 ***
## factor_partyFI            -0.056139   0.114020  -0.492  0.62366
## factor_partyINDIPENDENTE -1.173282   0.114020 -10.290  < 2e-16 ***
## factor_partyIV            -0.386425   0.114020  -3.389  0.00104 **
## factor_partyLEGA          -0.002478   0.114020  -0.022  0.98271
## factor_partyLEU            0.229343   0.114020   2.011  0.04727 *
## factor_partyM5S           -0.254663   0.114020  -2.233  0.02800 *
## factor_partyMISTO         -0.195756   0.114020  -1.717  0.08944 .
## factor_partyREG_LEAGUES   -0.777217   0.114020  -6.817 1.03e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.255 on 90 degrees of freedom
## Multiple R-squared:  0.8089, Adjusted R-squared:  0.7685
## F-statistic: 20.05 on 19 and 90 DF,  p-value: < 2.2e-16
```